

"Conversational" Codeswitching on Usenet and Internet Relay Chat

John C. Paolillo
Indiana University, Bloomington

Abstract

This article examines the use of English, Hindi, and Punjabi in four Internet communication contexts that differ on the dimensions of ethnic homogeneity and synchronicity. The linguistic theory of codeswitching predicts that code-mixing of English and Hindi or English and Punjabi primarily only occur in ethnically homogenous and "conversational" contexts. Quantitative analysis of the four contexts supports this prediction and leads to the suggestion that synchronicity is an important part of what it means for a communication mode to be "conversational."

Introduction

The Internet is commonly portrayed, both by advocates and in the popular media, as a place where people from diverse backgrounds can congregate to engage in discussion about a broad range of interests. The diversity of user backgrounds can be seen not only in areas of interest, political views and place of origin; it extends to native language as well. Thus it is not difficult to locate conversations on the Internet where Spanish, German, French, Portuguese, Chinese, and other major world languages are the primary language used. With a little more effort, one can find other conversations where less well-known languages are used, such as Punjabi and Hindi. In short, the Internet is a multilingual domain, and many users of the Internet are bilingual or multilingual (Danet & Herring, 2007; Paolillo 1996, 1999, 2007; Wright, 2004).

However, aside from the research just cited, relatively little is known about the linguistic and social reflexes of multilingualism on the Internet. On a global scale, we would like to know what the consequences of such multilingualism will be. Will linguistic diversity on the Internet increase to match the diversity of its users? Or will users gravitate toward a few widely-used languages? The answers to these questions depend on the nature of multilingualism in specific online environments, and the dynamic social processes influencing it. As of yet, however, little is known about language use in environments that are themselves multilingual, e.g., in which the participants use two or more languages. Is bilingual interaction on the Internet similar to that of ordinary bilingual face-to-face conversation? To what extent does the medium of interaction affect the way that multilingual discourse unfolds?

The answers to these questions are unlikely to be simple, because the Internet does not represent a single, homogeneous communications technology. Rather, the Internet supports a diverse set of communications technologies, with different social and interactional properties. Among text-only communications modes, an important dimension of variation is synchronicity—synchronous ("real time") communications are those in which two or more people must be online at the same time in order to communicate, and asynchronous ("non real time") communications take place when one's messages are stored for later retrieval by others (Kiesler, Siegel, & McGuire, 1984). While both modes support back-and-forth exchanges that participants call "conversations," synchronous modes permit interactional behaviors that are more like face-to-face conversation (e.g., short turns, back-channeling; see Cherny, 1999). Do these different communication modes differ in terms of how different languages are used in them? If so, can one

or another type of CMC be said to be "more like conversation" on the basis of how different languages are used?

A distinctive characteristic of bilingual face-to-face conversation is the use of two (or more) languages side-by-side, often with more than one language being used in a single sentence. This behavior is known as "codeswitching." Bilinguals themselves tend to regard codeswitching as illegitimate, since it violates standards of correctness for both languages (Gumperz, 1982; Myers-Scotton, 1993a). Perhaps because of this stigma, it is seldom found in traditional forms of writing. Nonetheless, codeswitching is pervasive in oral, face-to-face bilingual situations the world over. Gumperz (1982) suggests the term "conversational codeswitching," indicating that codeswitching is not merely characteristic of *bilingualism*, but specifically of face-to-face bilingual *conversation*.

Codeswitching has also been observed to occur in CMC, in both synchronous and asynchronous modes. However, previous studies of mixed language use online have typically focused on a single mode of CMC, e.g., personal email (Georgakopoulou, this issue), instant messaging (Lee, 2007), Internet Relay Chat (Androutsopoulos & Hinnenkamp, 2001), or Usenet newsgroups (Climent et al., 2003). The present study contributes to the understanding of Internet multilingualism through a systematic comparison of conversational codeswitching in four Internet fora.¹ Synchronous and asynchronous communication modes are investigated as used by two different, but related, online bilingual communities: the online Punjabi community (with predominantly Punjabi/English speakers) and the online Indian community (with predominantly Hindi/English speakers). The findings reveal differences in codeswitching according to both synchronicity and ethnic group, with the most productive, conversation-like codeswitching occurring in the Punjabi IRC forum. The results thus confirm a close relationship between bilinguals' online and off-line behavior, while identifying a contribution of the synchronicity of communication to its conversational qualities.

Background

Conversational Codeswitching

Sociolinguistics first adopted codeswitching as an object of study in the seminal work of Blom and Gumperz (1972) and Gumperz (1982). Since then, the study of codeswitching has largely been conducted along two parallel tracks: one investigating its grammatical constraints, the other investigating its social functions. Most research on codeswitching treats these two avenues of research as independent of one another (Tabouret-Keller, 1995), and individual studies tend to focus more on one aspect of the phenomenon than the other (e.g., Myers-Scotton, 1993a, on social factors, and Myers-Scotton, 1993b, on grammatical constraints). The two types of studies contribute two different distinctions that are important in identifying instances of conversational codeswitching as such.

A distinction made on structural grounds is that between codeswitching on the one hand, and borrowing and other linguistic convergence phenomena on the other (Gumperz, 1982). Borrowing occurs when a language adopts words or other elements from another language and incorporates them into its existing grammar, whereas codeswitching takes place when the grammatical systems of both languages (as well as the words) are used in the same exchange. For example, a great deal of the vocabulary of the academic registers of English is of Latin or French origin, but use of the academic registers of English does not involve codeswitching into Latin or French, because one need not really know any Latin or French to use those words: only

English grammar is used. Codeswitching implies use of both grammatical systems, and hence true bilingualism (knowledge of the grammar of both languages). During the Middle and Early Modern English periods (c. 1100-1600), Latin- and French-English codeswitching was widespread among educated English speakers, because at that time education required learning one or both of those languages.

Regarding social functions, an important distinction is made between *situational switching* and *metaphorical switching* (Blom & Gumperz, 1972; Gumperz, 1982). Situational switching is codeswitching that is conditioned by factors of the situation in which an interaction takes place. For example, native speakers of Spanish in Texas generally use Spanish in home settings, but switch to English in institutional settings (e.g., schools, government offices), even when others present are bilingual. The change in situation effectively bounds the interaction, so that switching takes place between interactions. In contrast, metaphorical switching is codeswitching that takes place within a single interaction. Such switching is metaphorical because it exploits associations between codes and social roles for communicative effect. For example, Myers-Scotton (1993a) describes an interaction in a bank in Kenya where a customer approaches the teller in situationally-appropriate Swahili; when the teller refuses to process a transaction which would be against the bank's rules, the customer switches to Luyia, a minority language that customer and teller happen to share. By using Luyia in asking for a personal favor, the customer covertly appeals to the teller's sense of ethnic loyalty, and obligation toward kin. These special communicative effects of metaphorical codeswitching have the status of conversational implicatures (Grice, 1976), since their interpretation is highly situationally dependent (Gumperz, 1982). Again, true conversational codeswitching implies metaphorical switching, rather than situational switching.

These distinctions impose methodological requirements on studies of codeswitching. With respect to the borrowing/codeswitching distinction, care must be taken not to count borrowed words in the category of codeswitches, since their communicative effects are distinct. Likewise, situational and metaphorical switching are structurally as well as socially distinct: while situational switching occurs only between sentences, metaphorical switching often occurs intra-sententially and intersententially, so that grammatical elements of both languages may be used together in the same sentence.

The type of behavior most relevant to the present study is intrasentential, metaphorical switching, since it characterizes conversational codeswitching as distinct from other bilingual behaviors (such as situational codeswitching) and language contact phenomena (such as borrowing). Nonetheless, it is important to observe all forms of other-language use for the insights they may provide about the motivations for codeswitching. If we adopt as a premise that codeswitching is a characteristic of bilingual conversation, we can then predict that bilingual discourse on the Internet should show codeswitching precisely to the extent that it is conversational. We could further predict that modes of communication that differ in terms of how conversational they are should differ comparably in terms of how much codeswitching they exhibit.

Two Modes of Computer-Mediated Communication

Computer-mediated communication (CMC) is a heterogeneous collection of technologies supporting many different types of communicative activity. Thus, one cannot talk about how CMC in general is conversational or spoken-like; one must take into account the *mode* of a given

communication (Herring, 2002; Yates, 1996a, b). Two modes of Internet communication were selected for this study: Usenet and Internet Relay Chat (IRC). Both Usenet and IRC have been said to support "dialog" or "conversation" among their participants (e.g., Baym, 1995; Hentschel, 1998; Sack, 2000; Smith, 1999; Werry, 1996), and both are widely available, decentralized services that people from a wide variety of backgrounds use to establish contact with one another for purposes ranging from computer programming to network administration to scientific research to personal advertisements to social interchange. Most of the bandwidth of both modes of communication is taken up by messages oriented toward constructing and maintaining otherwise fluid associations of users who share some common interest ("virtual communities," Rheingold, 1993). A number of online fora have formed around interests relating to nationality or national origin, and it is relatively easy to find fora that use other languages alongside English for both modes of communication.

Usenet and Internet IRC differ in their synchronicity and thus also in their resemblance to spoken communication. Usenet supports an offline, or "asynchronous," mode of communication, much like electronic mail, in which users need not be simultaneously online to communicate and interact with one another. Usenet is hierarchically organized into thousands of "newsgroups" that operate somewhat like public bulletin boards (Smith, 1999). At the time the data for this study were collected, users read messages, or "postings," with a piece of client software known as a "newsreader" that allowed a user to select one or more newsgroups from which to read messages.² Newsreaders generally organize the messages in a newsgroup into "threads," according to the subject line in the header of each message.³ Users may select a thread from which to read messages, and have the option of responding to each message as they read it. Postings on Usenet are like email messages in that they can be any length, from only the header information to many screenfuls, although the norm is around one screenful of text. Although Usenet is asynchronous, an interactive feel can develop in newsgroups. Responses to messages often come in a matter of hours (or even minutes, in the case of closely-linked servers), and lengthy threads of a dozen or more messages can develop in a short time on a busy newsgroup. Oftentimes two users will appear to take turns reading and responding to each other's messages, giving the thread the appearance of a conversation carried out between two people (Severinson-Eklundh, this issue, Part I). Most Usenet newsgroups are unmoderated, so that anyone can post a message on any topic. A small number of newsgroups are moderated, which means that messages are sent to one or more moderators who screen them for appropriateness before they are posted. Messages to one newsgroup are often "cross-posted" to other newsgroups, so that users reading either newsgroup can read the same message (Paolillo, 2000). A typical Usenet message appears in Example 1.

(1) A typical Usenet message (from soc.culture.punjab)

From: R C <xxxxxx@xxx.com>
 Subject: Re: TYPICAL INDIAN PARENTAL WARNING
 Date: Sun, 03 Mar 1996 19:15:52 -0800

SANJEEV wrote:

> I think that this is a topic that has been discussed from
 > centuries all over the world. The point here is that we dont need
 > to insult people to get our way. Can't you guys write without
 > giving judgements on someone else's brain. Put your points
 > forward...put some logic to it and wait for the other answer. If
 > we keep doing this...we will have no newsgroups or we will have
 > guy to newsgroup censoring. So keep your language in check. Some
 > people think that by abusing someone, you can get a big audience
 > on the newsgroup but that's like the wrong way to go about it.
 > Nothing personal.

Thank you, group facilitator. Maybe you should start alt.india.moderated -
 in the meantime, let us enjoy our conversations. Nothing personal.

In example 1, the current poster's contribution to the message consists entirely of the last two lines of the example. The previous paragraph, prefaced by "SANJEEV wrote:" and with each line introduced by an angle bracket (>), is an excerpt of an earlier poster's message to which the current poster is responding.

In contrast, Internet Relay Chat (IRC) is an online, synchronous communication mode, originally intended as an alternative to programs such as "talk" or "phone" (Pioch, 1993).⁴ Unlike talk and phone, however, IRC clients typically provide a window of only one or two lines for typing messages, and a participant's message becomes available to others only when the return key is pressed; backspaces and other edits are not visible to other participants. IRC is organized as a list of several thousand "channels" relayed over a distributed network of servers. Users connect to a server using a client program, which may be a telnet client, or a more specialized IRC client program. The user can then execute commands, such as listing the channels present on the server, or joining one or more channels. When a user "joins" a channel, messages typed by any of the other users on the channel start scrolling past on the user's terminal screen, as they are received by the server. Since different users are connected to different servers, there is often a propagation delay ("lag") of several seconds between the time a user types a message and the time it appears on other users' screens (Hentschel, 1998). Otherwise, the messages are displayed in "real time" in the sequence in which they are received by the server, and the only people who can send and receive messages on a particular channel are those who are "present" at any given time. Once on a channel, users can execute commands to list the nicknames or "nicks" of users present on the channel, or to send messages publicly to the channel, or privately to another user (messaging). A typical log of an interaction on IRC appears in Example 2 (line numbers have been added to the log file for reference).

(2) A typical IRC interaction (from #punjab)

```

[339] <ashna> hi jatt
[340] *** Signoff: puja (EOF From client)
[341] <Dave-G> kally i was only joking around
[342] <Jatt> ashna: hello?
[343] <kally> dave-g it was funny
[344] <ashna> how are u jatt
[345] <LUCKMAN> ssa all
[346] <Dave-G> kally you da woman!
[347] <Jatt> ashna: do we know eachother?. I'm ok how are you
[348] *** LUCKMAN has left channel #PUNJAB
[349] *** LUCKMAN has joined channel #punjab
[350] <kally> dave-g good stuff:)
[351] <Jatt> kally: so hows school life, life in geneal, love life, family life?
[352] <ashna> jatt no we don't know each other, i fine
[353] <Jatt> ashna: where r ya from?
[354] <kally>1

```

The first user to join a channel is given special privileges, known as operator privileges or "ops." Users with operator privileges are empowered to "kick" other users off of the channel, "protect" users that they do not want kicked off of the channel, and "op" other users they wish to have similar privileges. When more than one operator exists on a given channel, the different operators can be in a hierarchical relationship, so that if one operator is protecting a user another operator is trying to kick, the operator with the higher rank wins. Popular IRC channels are often maintained 24 hours a day by a core of frequent users who have operator privileges and who guarantee ops to each other when they join. Alternatively, users with programming savvy will write programs known as "bots" (short for "robots"), whose purpose is to keep the channel open. Bots range in complexity from those that occasionally spew out canned text, to complex systems with accounts, logs of users on the channel at any given time, and commands that users with special privileges can execute to make the bot take any of the actions an operator can perform. Since bots can be left running when their owners are not present, a user with operator privileges will often start a bot on a channel, give it operator privileges, and leave it running so that the same user can assume operator privileges at a later point by sending a command to the bot.⁵

Operators have a very important function on IRC which can not be exercised in the same way on Usenet, namely that of enforcing collective norms of interaction. One important such norm is the norm for length of message. Most messages on IRC are short, about 1-2 lines in length. It is possible to prepare and send longer messages, and one often finds this done with "banners" or ascii art, but this practice is generally disliked (it is known as "flooding"), and it is frequently followed by an operator kicking the "flooder." Users may also be kicked for other reasons, such as harrassment of other users, use of taboo language, or failure to interact in a way considered appropriate by one or more operators (Herring, 1999).

Usenet and IRC are similar in that both communication modes encourage interactive, multi-participant communication. They differ, however, in a number of respects, as summarized in Table 1.

Feature	Usenet	Internet Relay Chat
Interactivity	Asynchronous, delayed	Synchronous, immediate
Length ⁶	1 to several hundred lines	1-2 lines
External organization	Hierarchy of newsgroups by subject	Flat list of channels
Internal organization	Subject-oriented "threads"	Temporally sequenced
Maintenance	By popular vote	User activity
Authorities	Usually unmoderated (no authority)	Operators
Broadcasts	Frequent cross-posting	No multi-channel messages

Table 1. Differences between Usenet and Internet Relay Chat

The important differences for the purposes of the present study are those that allow us to compare Usenet and IRC with conversation, namely interactivity and length. Regarding interactivity, the synchronous nature of IRC makes that mode of communication more conversation-like, while the asynchronous nature of Usenet more closely resembles communication by letter or memo, where there is usually a sizable time interval between the delivery of a message and its response. A similar situation holds regarding message length; the relatively shorter messages on IRC make that mode of communication more like conversation, where turns are typically short, while Usenet messages are more the length of short written letters or memos. On both counts, IRC should be more conversation-like than Usenet, and we would thus predict that IRC would have more frequent codeswitching than Usenet.

The Online Indian and Punjabi Communities

The communities selected for this study are the online Punjabi and Indian communities. These two communities correspond to two similar offline communities of Punjabi and Indian expatriates living in Canada, the United Kingdom, and the United States, and have been the subject of several previous studies (Mitra, 1997; Paolillo, 1996, 1999, 2001). As such they differ from many interest-based online communities that have no analogous presence outside of the Internet. Both communities have dedicated Usenet newsgroups (soc.culture.punjab and soc.culture.indian), and IRC channels (#punjab and #india). On both Usenet and IRC, the larger, more inclusive Indian community is more active, in total volume of messages and numbers of participants, than the smaller, more regionally-specific Punjabi community. The Punjabi online community began as an offshoot of the older Indian online community; the newsgroup soc.culture.punjab was created in September 1994 by Punjabi subscribers to the newsgroup soc.culture.indian, at a time when a number of other Indian and South Asian newsgroups were being created. The IRC channel #punjab was created about a year later, by members of the same community. Both online communities consist mostly of first and second-generation expatriates from India and Pakistan living in the US, the UK, and Canada. When the data for this study were collected in March 1996, relatively few members of these online communities read or sent their messages from sites in India or Pakistan.

These communities were selected in part because prior research indicated that some codeswitching occurs on the Usenet newsgroup soc.culture.punjab (Paolillo, 1996). The Indian community was selected for purposes of comparison because it is a more multi-ethnic and hence multilingual community that happens to share many of its participants with the Punjabi community. Past research on codeswitching suggests that greater ethnic and linguistic diversity

in a given context leads to less frequent codeswitching (e.g., Myers-Scotton, 1993a); thus we would predict that codeswitching should be more frequent in the Punjabi community than in the Indian community. With these two communities we can observe codeswitching under four distinct but comparable conditions: ethnically homogeneous, asynchronous (soc.culture.punjab) and synchronous (#punjab); and ethnically diverse, asynchronous (soc.culture.indian) and synchronous (#india), as indicated schematically in Figure 1.

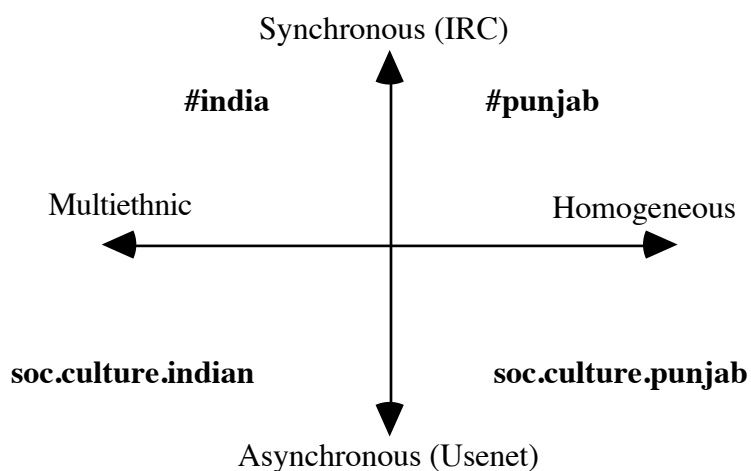


Figure 1. The four online fora

Methods

Data

Since codeswitching involves the use of elements of more than one language, it can be readily observed by examining samples of language use, classifying each linguistic element as belonging to one language or another, and counting the tokens in each category. For codeswitching to be represented accurately in the data, a representative sample of data must be collected. The complication of borrowed vocabulary must be squarely addressed, however, since counting borrowings as codeswitches would inaccurately inflate the rate of codeswitching. The data sampling and coding procedures were designed to address these two principal issues: representative sampling and the occurrence of borrowing.

Since IRC and Usenet operate somewhat differently, different data collection and preparation procedures were employed. For the Usenet groups, I downloaded all the messages available on the two newsgroups for a one-week period. Since Usenet messages headers and quotations of other messages were not relevant to my analysis, I fed the message files through a computer program which filtered out this material, and sequentially numbered each line of each message. The IRC data were collected and handled in a manner parallel to that of the Usenet data: I created "log files" of the two channels for one to two hours at a time, using the Macintosh IRC client Homer. Approximately six hours of IRC data on each channel were collected in this way. Subsequently, the log files were fed through a program that numbered each line for reference. In this way, samples of Usenet and IRC data were collected that were representative of the activity in the different fora.

Analysis

Once the four corpora were complete, I generated keyword-in-context concordances for each using the Macintosh program *Conc*, a concordance generating program made publically available by SIL International;⁷ the concordances were exported to a text file which was then annotated using the spreadsheet program Microsoft Excel. Two types of annotation were made for each entry in the concordances: the language each entry belongs to (Hindi, Punjabi, English, etc.), and the grammatical status of the entry. Three basic types of grammatical status were considered: proper names of places and individuals, content words (referential meaning-bearing nouns, adjectives, and verbs), and "system" words (grammatical functor words, prepositions, determiners, pronouns, quantifiers etc.). For concreteness, I adopted the criteria of Myers-Scotton (1993b) in identifying "system," or grammatical function, words. I then used the sorting, counting, and statistical facilities of Excel to analyze the data further. In this way, statistical summaries of the languages used on each of the four fora were generated. Finally, the original concordances were used to investigate the functions of other language use in the corpora qualitatively.

Borrowing was taken into account in two ways. First, when the words belonging to each language were counted, content words, proper names, and grammatical function words were counted separately from one another. Proper names are "fixed" to their referents, especially in the case of individuals' proper names, even though those names may be other-language forms. Thus, such names were counted separately to avoid inflating the count of other-language forms. Furthermore, the frequency of use of the grammatical function words of a language provides an index of use of the grammar of a language, and thus, an index of the rate of codeswitching. Borrowed words are more likely to be content words than grammatical function words (cf. van Coetsem, 1988). By counting these different word types separately, it is easier to ascertain how different languages are used; i.e., whether the grammar of the other language is used, and hence codeswitching occurs, or whether primarily borrowing occurs.

The second way in which borrowing was taken into account was through a qualitative micro-analysis at points of potential codeswitch (Gumperz, 1982). This technique makes it possible both to confirm the status of individual items as codeswitches, and to verify the function of those switches. This was done using *Conc*, which allows the researcher to navigate through a text using either its concordance or its word frequency list. This meant that I could identify candidate codeswitched forms and then locate them in context, where I could identify their status and function. Summary lexical statistics were also generated with the assistance of *Conc*.

Results

In terms of overall word frequencies, non-English languages were used more often on the IRC channels than on the Usenet newsgroups. Across the two communities, in contrast, we found comparable rates of other-language use in both communications modes. In the Punjabi fora, English and Punjabi were used, while in the Indian fora, in order of prominence, English, Hindi, Bengali, and Punjabi were used, with Hindi use substantially exceeding that of Bengali and Punjabi.⁸ Counts of the different categories of words for the four fora are given in Table 2 for *soc.culture.punjab* and *#punjab*, and in Table 3 for *soc.culture.indian* and *#india*. In the tables, tallies for the Usenet groups are on the left, and those for the IRC channels are on the right. Since there was considerably more traffic on the larger Indian fora, the total word frequencies are greater for the Indian fora than for the Punjabi fora.

	soc.culture.punjab				#punjab			
	Content	System	Names	Total	Content	System	Names	Total
English	4185	4572	298	9055	3172	2952	728	6852
	46%	50%	3%	95%	46%	43%	11%	76%
Punjabi	81	14	388	483	190	106	1854	2150
	17%	3%	80%	5%	9%	5%	86%	24%
Total	9538				9002			

Table 2. Token counts for Punjabi fora by language

	soc.culture.indian				#india			
	Content	System	Names	Total	Content	System	Names	Total
English	60433	64191	445	125069	5444	6334	1484	13262
	48%	51%	0%	94%	41%	48%	11%	79%
Hindi	1449	532	6288	8269	285	479	2720	3484
	11%	4%	85%	6%	8%	14%	78%	21%
Total	133338				16746			

Table 3. Token counts for Indian fora by language

From the tables, we can see that Usenet has an average of 5% to 6% non-English use, while IRC has from 21% to 24% non-English use. This supports the prediction that the more interactive, synchronous fora would show more codeswitching. Moreover, a greater proportion of other-language system words is used on the IRC channels than on the Usenet newsgroups. This suggests that where Punjabi and Hindi are used on IRC, they are used more often together with Punjabi and Hindi grammar, i.e., as genuine code-switches. In contrast, the corresponding ratio for the Usenet newsgroups is heavily skewed in favor of content words, suggesting relatively more borrowed Punjabi and Hindi words used in grammatically English sentences. The main apparent difference between the two communities is that Punjabi content words may be slightly more common than the corresponding Hindi items on both Usenet and IRC, suggesting more borrowing in the Punjabi community than in the Indian community. These general results are discussed in more detail below.

Codeswitching on Usenet and IRC

Usenet appears to disfavor the use of the system morphemes of the alternative code, while IRC appears to favor them. Thus, IRC appears to have codeswitching, while Usenet has little. Inspection of the concordances verifies this observation. The system morphemes of Hindi found in soc.culture.indian come entirely from three messages, in which the Hindi is contained in quoted poems. Thus, for the week under study, there appears to have been no intra-sentential codeswitching on soc.culture.indian.

Quoting of poetry turns out to be one of the principal functions of other-code use on both the Indian and Punjabi newsgroups. On soc.culture.punjab, there is a similarly limited use of Punjabi, with the principal uses being quotation of poetry and song lyrics, as in example 3, and fixed, formulaic expressions such as nationalist slogans, as in example 4.⁹

(3) Quotation (song lyric)

From: xxxx@xxxxx.xxxxxx.ca (S H S)
 Subject: Re: kuldip manak in election
 Date: 2 Mar 1996 22:58:24 GMT

Rajwinder Singh (rajwi@otto) wrote:

: I was just listening to Chhinda's "ucha burj Lahore da" -- some
 : songs are good, but the lyrics are very male-chauvinistic,
 : sometimes outright misogynistic: "rannaa dee matt gichee pichhe
 : kaihndaa hai jagg saara nee."
 :
 : rs

WHAAAAAAT??!! Punjabi music misogynistic - puuuullllleeeese - it's only a song.
 Besides who defines what is chauvinistic - even western feminists are split these
 days (re: Paglia versus the rest of the pack). To hell with this analysis - just
 submit to the power of BHANGRA and enjoy.

P.s. could you translate the lyrics above into english.

(4) Fixed expression (nationalist slogan)

From: xxxxxxx@xxx.com (MJassal)
 Subject: Re: Khal S**T EATER stan??????
 Date: 27 Feb 1996 18:45:41 -0500

KHALISTAN ZINDABAD THATS ALL I HAVE TO SAY THATS ALL THAT NEEDS TO BE SAID
 [Long live Khalistan]

MANDEEP SINGH JASSAL

These uses, like the Hindi poetry, do not qualify as intra-sentential codeswitches (some intrasentential codeswitching is found on soc.culture.punjab; see example 9 below). These observations suggest that Punjabi and Hindi have only limited functions on the two Usenet newsgroups.

Gumperz (1982) indicates that direct quotation is an environment that favors codeswitching, in that a speaker's words are typically represented by another speaker in the language in which they were originally spoken. Nonetheless, it is notable that what is quoted in the newsgroups is principally poetry. Quotation of poetry serves to establish a poster's cultural "authenticity," since it demonstrates a familiarity with either high or popular culture in the other language that a cultural outsider could not have. Fixed expressions serve a similar function, particularly when they represent nationalist sentiment; no one is more nationalist than the individual who expresses nationalism in the national tongue. However, their fixed and formulaic nature, and their occurrence typically at the beginning and the ends of messages, suggest that they should be regarded as borrowings, rather than as productive codeswitching. Both fixed expressions and quotes, since they are essentially someone else's words, are a "safe" way to use the other language with an audience that would be critical of disfluent or non-standard use of the other

tongue. They can be memorized in short, pithy chunks, or typed in verbatim from a written source. Hence they do not require users to risk creating novel expressions in a language in which they may not be fully fluent.

Hindi and Punjabi use is much more prevalent on IRC than on Usenet. Some of the IRC uses also involve quoted poetry (usually songs from contemporary films) as in example 5, and fixed, formulaic expressions such as the traditional Sikh greeting *sat sri akal* "God is truth" on #punjab, which is usually abbreviated to *ssa* (e.g., Example 2, line 345).¹⁰

(5) Badshah sings on #india

[009] *** Action: Badshah sings jaaney kyun log mohabat kya karte hain...
[Dear, why do people love...]

At the same time, other, more interactive uses of Hindi and Punjabi are found on IRC which are absent from the Usenet data. In Example 6, Jatt uses Punjabi several times to attract his interlocutor's attention. Jatt enters the scene (197) and addresses the group in English (198). When no one responds (as is typical on a busy channel when someone joins), he resorts to Punjabi (199). When again no one takes up his lead, he orients himself toward Navreen (206), who is presently engaged in another interaction (not shown here), using Punjabi. She responds in English (208), and Jatt continues the conversation in English (209, 219), until it appears that she has drifted out of the conversation, whereupon Jatt resorts to Punjabi again. Thus, three of Jatt's five turns in Punjabi (199, 206, and 228) are redundant with earlier turns taken in English. All of his Punjabi turns, even those that are not greetings *per se*, have the function of drawing his interlocutor into the interaction.

(6) Jatt initiates interaction with Navreen

[197] *** Jatt has joined channel #punjab
[198] <Jatt> YO
[199] <Jatt> Kidaan peopleee
[how's it going, people?]
[206] <Jatt> Navreen: kidaan
[Navreen, how's it going?]
[208] <Navreen> jatt: fine thanx and you
[209] <Jatt> Navreen: I am great
[219] <Jatt> Navreen: where kelly, and Neena?
[228] <Jatt> Navreen: kee hal chal hai tuhada
[Navreen, how is your health, etc.?
[231] <Navreen> jatt: excellent
[233] <Jatt> Navreen: suchin/
[Navreen, tell (me)?]
[236] <Navreen> jatt: suchee
[Jatt, I told you]
[241] <Jatt> Navreen: achaa for taan boht vadiaa
[okay, so then very good again]
[242] <Jatt> for=fir
[correction: again]
[245] <Jatt> Navreen: chah pee lee kay nahin
[Navreen, have you drunk tea or not?]

more interactive purposes than when they were used on Usenet. Some amount of quoted poetry and fixed, formulaic expressions were common to both communication modes, but in general, the IRC uses served more interactive purposes than the Usenet uses. These findings support the prediction that a more interactive communications mode (IRC) would show greater use of codeswitching.

Codeswitching in the Hindi and Punjabi Communities

While there are similarities within each of the two communications modes, important differences were also found between the two online communities. Notably, codeswitching functioned somewhat differently in the two communities. With respect to Usenet, we already noted that the only examples of Hindi use on soc.culture.indian were three instances of quoted poetry. While quoted poetry is found on soc.culture.punjab as well, we also find instances of Punjabi used for insulting; this use is illustrated in example 9 (see also Paolillo, 1996).

(9) Message by J. G. (excerpt)

...fax number, and I will send it to you. Of course you might end up
 having to read a little Punjabi, or do you not know how?
 Salla Sher Jat da bucha.
 [Brother-in-law Sher Jat's child]
 And I hear you've been trying to find out about me, through friends at
 UC Davis. Kee plan kharda Mr. Sher Jat?
 [What are you planning Mr Sher Jat?]

This use parallels the use in insults of the second person familiar pronoun in many languages (Brown & Gillman, 1972[1964]). Use of Punjabi, like the second person familiar pronoun, effectively underscores the contempt communicated by an insult by symbolizing more intimacy than the situation calls for. The insult in Punjabi carries extra punch in a context where the interlocutor's language competence is called into question, as it is in example 9, since it potentially excludes the interlocutor and undermines his or her claim to ethnic authenticity.

Regarding IRC, we have already noted that codeswitching on the channel #india has a participant-tracking function (example 6 and 7), while #punjab exhibits a participant summoning function (example 8). This summoning function is highly prevalent on #punjab; the greeting *ssa* was used 53 times, and the unabbreviated *sat sri akal* occurred 13 times. In fact, *ssa* was the most frequent Punjabi form on #punjab that was not someone's nick(name). The honorific particle *ji*, also typically used with names in direct address, was the next most frequent non-name Punjabi form, occurring 34 times total, with 22 of these occurring in the greeting formula *ssa ji* (<name> *ji*). This single formulaic use accounts for 14.9% of the non-name Punjabi tokens. A typical greeting sequence on #punjab appears in example 10.

(a) #india				(b) #punjab			
	Hindi	English	Total		Punjabi	English	Total
greet	29	420	449	greet	77	152	229
2nd turn	4	55	59	2nd turn	5	22	27
insult	3	25	28	insult	1	1	2
part	4	86	90	part	0	21	21
other	86	1303	1389	other	45	537	582
Total	126	1889	2015		128	733	861
	6.0%	94.1%			14.9%	89.1%	

Table 4. Functions of Hindi and Punjabi on #india and #punjab

Table 4 shows that there was a much higher incidence of Punjabi in greetings on #punjab, than Hindi in greetings on #india. In fact, there appears to be no relationship of language to function at all on #india; a chi-squared test for independence performed on Table 4 (a) is non-significant ($\chi^2 = 1.52$, $p = 0.823$, 4 df). In contrast, a chi-squared test for independence performed on Table 4 (b), with the cells above and below the middle line combined,¹³ is highly significant ($\chi^2 = 84.83$, $p < 0.001$, 1 df).¹⁴ Also apparent in Table 4 is a substantial difference in the use of the other language overall on the two channels: 15.9% of turns taken on #punjab contained Punjabi, whereas only 5.9% of turns on #india contained Hindi. This difference is also statistically significant ($\chi^2 = 55.68$, $p < 0.001$, 1 df).

A similar analysis was conducted to ascertain whether the relative status of the interlocutors was relevant in the context of use of the other language. In the dynamics of IRC, it matters a great deal who is operator on a given channel, since operators are empowered to prevent others from being part of the discourse on a channel. On both #india and #punjab, operator status is maintained by bots, and by round-the-clock monitoring.¹⁵ Thus, operator privileges are the most readily discernable social characteristic that relates to an individual's membership in a core network of channel participants (Paolillo, 1999, 2001). As with other social networks (cf. Chambers, 1995; Milroy, 1992), we might expect the linguistic behavior of core members (i.e., operators) to be different from that of non-core members (non-operators), whose presence on the channel may be more transient.

For this analysis, four different categories of interaction were considered: operators addressing other operators, operators addressing non-operators, non-operators addressing operators, and non-operators addressing non-operators. Each turn was coded for whether it employed other-language morphemes or not, and all turns involving commands, pure vocatives, expressives, smileys, and ascii art were excluded from the analysis. The results appear in Table 5.

		#india			#punjab		
spkr	addr	Hindi	English	Total	Punjabi	English	Total
non-op	non-op	4	25	29	72	310	382
non-op	operator	3	43	46	14	84	98
operator	non-op	37	578	615	31	241	272
operator	operator	82	1243	1325	11	98	109
	Total	126	1889	2015	128	733	861

Table 5. Language use by operator status on #india and #punjab

Again, the two channels turn out to be different in a number of ways. First, a greater proportion of the participants on #india are operators, hence the "bottom-heavy" numbers in the first half of Table 5. Nonetheless, the distribution of language use according to operator status of participants appears to be completely random on #india ($\chi^2 = 2.89$, $p = 0.409$, 3 df). On #punjab, however, the distribution is significantly non-random ($\chi^2 = 9.362$, $p < 0.05$, 3 df). Figure 1 represents the break-down of this latter chi-squared figure by each row in Table 5, using a pie-chart. More than 50% of the total chi-squared value comes from non-operators addressing non-operators, where a greater-than-expected use of Punjabi is encountered. The remainder of the chi-squared value comes principally from the fact that operators, in their interactions with other operators and with non-operators, use less Punjabi than expected.

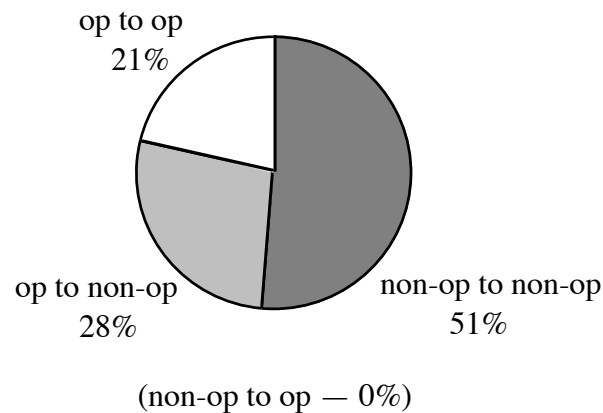


Figure 2. Relative contribution of each interaction type to total χ^2 in table 5 (#punjab)

At first sight, the greater-than-expected use of Punjabi by non-operators is surprising, since one might expect use of Punjabi to function as an index of membership in the core participant network. However, the situation represented by the distribution in Table 5 and Figure 2 is consistent with an interpretation of the (majority of) the members of the non-operator group being "aspirers" to the core membership of the network (Chambers, 1995). Aspirers overshoot the linguistic behaviors of the core members in the hope of attaining core member status, or of simply obtaining the approval of core members. Further research would be necessary to confirm this interpretation.

The reason why a similar pattern is not found in the #india corpus may have to do with the larger core network of operators on #india, and the more fluid and contested nature of operator status there. Much of the traffic on #india consists of the result of contests between operators who vie for control of the channel by attempting to "de-op" and kick other operators. This is accomplished by cliques of operators forming alliances, sponsoring and protecting one another, and with the instrumentality of bots that are "secured" from tampering by other operators. Clever operators (and IRC administrators known as "IRCCops" who dislike bots because of the bandwidth they consume) can sometimes manage to "hack" (modify) or de-op and kick another operator's secured bot, and thereby take control of the channel. All of this activity accounts for a great deal of the messages displayed on the channel—at times, interactionally adjacent messages of two participants on #india may be separated by 50 lines or more of system messages. In other words, core network membership on #india is a hotly contested status. In the absence of an uncontested core, there is no group of "insiders" to establish local norms of code use.

In summary, although the overall frequencies of codeswitching in both communities are roughly comparable, there are qualitative differences in the codeswitching of the two communities both on IRC and on Usenet. Codeswitching is used for insults on soc.culture.punjab, but not on soc.culture.indian; on #punjab, summoning and greeting uses predominate, whereas on #india, participant-tracking uses are found. These results are summarized in Table 6 below.

Usenet	Asynchronous Intra-sentential codeswitching infrequent Fixed and quoted uses of non-English language
Internet Relay Chat	Synchronous Intra-sentential codeswitching, alongside fixed and quoted uses of non-English language Codeswitching used as an interactional strategy
Indian sub-community	Ethnically heterogeneous Codeswitching less frequent
Punjabi sub-community	Ethnically homogeneous Codeswitching more frequent
#india	Codeswitching used for participant tracking No association between codeswitching and operator status
#punjab	Codeswitching used for greeting and attention-getting Non-operators addressing operators use more Punjabi

Table 6. Summary of findings

Discussion

The results of this study support the first of the two hypotheses stated earlier: more codeswitching was found in synchronous IRC than in asynchronous Usenet. The second hypothesis, that codeswitching would occur more often in the less ethnically-diverse of the two South Asian communities, was not supported in terms of codeswitching frequency: rates of other-language use were comparable across the two communities. Although the Punjabi and the Indian groups use codeswitching for different communicative functions, these differences do not seem to indicate that codeswitching is more productive in one community than in the other.

In addition to the results related to the hypotheses, it is possible to make two larger observations for the corpus as a whole. First, other language uses tend to be fixed and formulaic in all four fora. Second, use of South Asian languages such as Hindi and Punjabi, even on Internet fora dedicated to South Asian cultural interests, is rather limited. The predominant language for most purposes in all four groups is English.

Why do the IRC samples have more codeswitching than the Usenet samples? According to Gumperz (1982, p. 95), the existence of shared context is a requirement for conversational codeswitching, since it is only through reference to that shared context, a context that is subcultural and local, that the interpretations of codeswitched utterances can be unraveled. If this is the case, it suggests that among the bilingual participants of different CMC fora, codeswitching will be employed to the extent that participants share enough context to evaluate

individual instances of other-code use successfully. IRC creates shared context to a greater degree than Usenet. Connecting to an IRC server and joining a chat channel entails necessarily sharing all public messages with all other participants—the comings and goings of participants, kickings and bannings, etc. This ebb and flow of messages in real time is a context that all participants on an IRC channel share. In contrast, because Usenet is asynchronous and because newsreaders allow users to choose which messages will appear on their screens, Usenet participants can easily regulate the rate and volume of what they read, by skipping threads and being selective in reading messages. To the extent that participants are selective, the context that is shared among Usenet participants is reduced. These properties are a consequence of the architecture of the two CMC systems, and especially their synchronicity, which determines whether or not users have time to be selective.

The tendency for utterances in Hindi and Punjabi to be fixed and formulaic in nature may reflect the fact that fixed utterances are readily interpreted, even with a minimum of shared context. A common general motivation for code choice can be identified in all four fora: group membership, both in the global sense of ethnic authenticity, and in the local sense of membership in a core network, favors other-language use. Fixed and formulaic uses reflect more transparently the macro-societal "we" and "they" meanings of the different codes (Gumperz, 1982, p. 73). Thus a Punjabi IRC participant who uses *ssa ji* is suggesting to others "Hey, I'm a Punjabi too!," while a Hindi speaker who quotes poetry or film songs suggests "I'm an authentic Indian because I have knowledge of these authentic cultural symbols." Their content renders these macro-societal "we" and "they" meanings salient, by packaging ethnic, cultural, or nation-centered messages in the other tongue. Thus when the shared context is relatively impoverished, as it is in CMC in general, we can expect a greater incidence of codeswitched formulae and quotations.

Finally, the predominance of English on all four fora can be explained by a convergence of various sociolinguistic factors relating to the South Asian expatriate context and the larger context of the Internet itself at the time (Paolillo, 1996). First among these is the high prestige accorded to English in India itself, which favors English in formal and educational contexts. Those who participate on the Internet are principally English-educated, whether in India or abroad. Second, the audience of messages on IRC and Usenet is often an international one, due to the global reach of the Internet. In any given interaction, the exact audience of a message may be unknown, and the "real life" social identities of participants may be difficult to verify. Last, because computer network technology and use continue to be dominated by English-speaking countries, Internet communication encourages the use of English both socially, in terms of the prestige associated with being a knowledgeable Internet user, and technologically, for example in terms of the ASCII standard code that makes it easier to type in English than in other languages (Danet & Herring, 2007; Yates, 1996b). These circumstances favor the use of English as an international language of wider communication on the Internet.

Conclusions

Certain general inferences may be drawn from these findings about the nature of code choice on the Internet. First, the tendency for synchronous modes to favor (and asynchronous modes to disfavor) codeswitching, if supported by further research, has two general consequences. On the one hand, it supports the view that the degree to which a CMC mode is "conversational" depends at least in part on its synchronicity. On the other hand, it suggests a new interpretation of what it means for a mode of communication to be "conversational," rooted in an understanding of how shared context arises, both macro- and micro-sociolinguistically.

The consequences of this perspective need to be examined for other communication media. On the basis of the present study, one might make the following predictions about where codeswitching is more likely to be found. Communication media that support simultaneous communication of any sort, whether they use voice or text, would be expected to foster the creation of shared context, and hence favor codeswitching. This is the principal characteristic that face-to-face and telephone conversations share with IRC. Communication media that impose a substantial delay between sending and receiving messages, whether voice or text, or which otherwise impose limits on the ways that shared context is created, would be expected to disfavor codeswitching. Common communication media that share this characteristic with Usenet are voice-mail and telephone answering machine messages. These and other media would make interesting contexts in which to test the claims about interactivity and codeswitching advanced here.

Finally, methodological consequences for the study of codeswitching can be noted. Codeswitching researchers have generally agreed that the places where interesting codeswitching phenomena are found are in conversation. However, since codeswitching is often a stigmatized behavior for those who engage in it, it has traditionally been difficult to collect relevant conversational data without the observation process interfering with the codeswitching behaviors of the people being observed. Elaborate procedures have sometimes been arranged to insure that high-quality data with sufficient instances of codeswitching are collected (e.g., Poplack, 1993). Because of this, codeswitching research has been a somewhat specialized enterprise, relying on a restricted characterization of "conversation" for candidate sources of data. What CMC modes such as IRC, Usenet, and others offer are new sources of codeswitching data, where high-quality participant observations can be collected unobtrusively and with less methodological overhead. Equally importantly, the investigation of codeswitching in CMC promises to refine our understanding of the circumstances that condition bilingual language behavior by drawing attention to synchronicity and shared context in place of an otherwise under-analyzed notion of "conversation."

Notes

1. This article reports on research originally conducted in 1996 and circulated in a pre-publication version for 15 years. Subsequent studies, including some cited here, have cited ideas and examples from the pre-publication version.
2. Google Groups now provides a Web-based front end for accessing Usenet newsgroups.
3. Newsreaders construct threads by matching a reference header line across different messages. If a poster has edited the subject line in responding to a message, it is normally still understood as belonging to the thread of the message it responds to, but it may be displayed by some newsreaders under a different thread heading.
4. See Herring (2002) on the origins of Usenet and IRC.
5. This practice is explicitly frowned upon in most IRC codes of etiquette (e.g., Pioch, 1993).
6. Messages with zero content (e.g., because the writers intentionally leave them blank) occasionally appear in both modes.
7. Conc is available at <http://www.sil.org/>. The most recent version as of this writing is Conc 1.80b3.

8. Because of the extent to which Hindi, Punjabi, and Bengali (especially the first two) are structurally congruent, there are a number of words that cannot be assigned conclusively to one language or the other. Such forms were counted as Punjabi in Table 2, and Hindi in Table 3.
9. The postscript in example 3 indicates one of the shortcomings of collecting data from anonymous sources—it is not always possible to know the language proficiency levels of the people being studied.
10. A number of high-frequency English formulae are similarly abbreviated, as is the case on other IRC channels (Hentschel, 1998; Werry, 1996) and in other interactive modes of CMC such as MUDs and MOOS (Cherny, 1999) and multiplayer online games (e.g., Herring, Kutz, Paolillo, & Zelenkauskaite, 2009). Examples are *brb* "be right back," *re(-hi)* "hello again," *back* "I'm back now," etc.
11. The nick "chirag" may itself be intended as a variation of Chira G (ji); this was not counted as a greeting.
12. Vocatives are sometimes identifiable as Hindi or Punjabi because they occur with the honorific morpheme *ji*, e.g., *ji Haydays*. However, this is complicated by the fact that some people's nicknames contain this morpheme as an integral part, e.g., *Sharmaji* and *Dave-G*, to name two. See also note 7.
13. It is necessary to combine the cells in this table to meet the criterion for performing the chi-squared test of having a minimum expected value of 5 in each cell of the table. Combining the cells as indicated treats both types of greeting moves together as opposed to any other type of turn.
14. Excluding the abbreviated greeting formulae *ssa* and *ssa ji* from the #punjab counts does not affect the overall pattern ($\chi^2 = 7.886$, $p < 0.01$, 1 df). 1
15. At any given time, operators from the different time zones of Australia, the U.K., the U.S., and Canada could be found, all of whom know each other from participation on IRC, and who cooperatively support each others' continued operator status by giving each other super-user privileges on each other's bots.

References

- Androutsopoulos, J., & Hinnenkamp, V. (2001). Code-Switching in der bilingualen Chat-Kommunikation: Ein explorativer Blick auf #hellas und #turks. In M. Beißwenger (Ed.), *Chat-Kommunikation* (pp. 367–402). Stuttgart: Ibidem.
- Baym, N. K. (1995). The emergence of community in computer-mediated communication. In: S. G. Jones (Ed.), *Cybersociety* (pp. 138–163). Thousand Oaks, CA: Sage.
- Brown, R., & Gilman, A. (1972). The pronouns of power and solidarity. In: P. Giglioli (Ed.), *Language and social context* (pp. 252–282). Hammondsworth: Penguin. Reprinted from T. A. Sebeok (Ed.), *Style in language* (pp. 253–276). Cambridge, MA: MIT Press.
- Chambers, J. (1995). *Sociolinguistic theory*. Oxford: Blackwell.
- Cherny, L. (1999). *Conversation and community: Chat in a virtual world*. Stanford, CA: CSLI Publications.

- Climent, S., Moré, J., Oliver, A., Salvatierra, M., Sànchez, I., Taulé, M., & Vallmanya, L. (2003). Bilingual newsgroups in Catalonia: A challenge for machine translation. *Journal of Computer-Mediated Communication*, 9(1). Retrieved June 24, 2006 from <http://jcmc.indiana.edu/vol9/issue1/climent.html>
- Danet, B., & Herring, S. C., Eds. (2007). *The multilingual Internet: Language, culture, and communication online*. New York: Oxford University Press.
- Giles, H., Taylor, D. M., & Bourhis, R. V. (1973). Towards a theory of interpersonal accomodation through speech: Some Canadian data. *Language in Society*, 2(2), 177–192.
- Grice, H. P. (1975). Logic and conversation. In: P. Cole & J. L Morgan (Eds.), *Syntax and semantics*. Vol 3: *Speech acts*. New York: Academic Press.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.
- Hentschel, E. (1998). Communication on IRC. *Linguistik online* 1/98. Retrieved June 24, 2006 from <http://www.linguistik-online.de/irc.htm>
- Herring, S. C. (1999). The rhetorical dynamics of gender harassment on-line. *The Information Society*, 15(3), 151–167.
- Herring, S. C. (2002). Computer-mediated communication on the Internet. *Annual Review of Information Science and Technology*, 36, 109–168.
- Herring, S. C., Kutz, D. O., Paolillo, J. C., & Zelenkauskaite, A. (2009). Fast talking, fast shooting: Text chat in an online first-person game. *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences*. Los Alamitos, CA: IEEE Press. <http://ella.slis.indiana.edu/~herring/hicss.bzflag.pdf>
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10), 1123–1134.
- Lee, C. K. M. (2007). Linguistic features of email and ICQ Instant Messaging in Hong Kong. In B. Danet & S. C. Herring (Eds.), pp. 184–208.
- Milroy, J. (1992). *Linguistic variation and change*. Oxford, UK: Blackwell.
- Mitra, A. (1997). Virtual commonality: Searching for India on the Internet. In: S. G. Jones (Ed.), *Virtual culture: Identity and communication in cybersociety* (pp. 55–79). Thousand Oaks, CA: Sage Publications.
- Myers-Scotton, C. (1993a). *Social motivations for codeswitching*. Oxford, UK: Oxford University Press.
- Myers-Scotton, C. (1993b). *Duelling languages*. Oxford, UK: Oxford University Press.
- Paolillo, J. C. (1996). Language choice on soc.culture.punjab. *Electronic Journal of Communication*, 6(3). <http://www.cios.org/EJCPUBLIC/006/3/006312.HTML>
- Paolillo, J. C. (1999). The virtual speech community: Social network and language variation on IRC. *Journal of Computer-Mediated Communication*, 4(2). Retrieved June 24, 2006 from <http://jcmc.indiana.edu/vol4/issue4/paolillo.html>
- Paolillo, J. C. (2000). Visualizing Usenet: A factor-analytic approach. *Proceedings of the Thirty-Third Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Press.

- Paolillo, J. C. (2001). Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics*, 5(2), 180–213.
- Paolillo, J. C. (2007). How much multilingualism? Language diversity on the internet. In B. Danet & S. C. Herring (Eds.), pp. 408–430.
- Pioch, N. (1993). *A short IRC primer*. Manuscript. <ftp://cs.bu.edu/irc/>
- Poplack, S. (1993). Variation theory and language contact: Concept, methods and data. In D. Preston (Ed.), *American dialect research* (pp. 251–286). Amsterdam/Philadelphia: John Benjamins.
- Poplack, S., & Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages* (pp. 199–232). Cambridge, UK: Cambridge University Press.
- Rheingold, H. (1993). *The virtual community: Homesteading on the electronic frontier*. Reading MA: Addison-Wesley.
- Romaine, S. (1989). *Bilingualism*. Oxford, UK: Blackwell.
- Sack, W. (2000). Conversation Map: A content-based Usenet newsgroup browser. *Proceedings of the International Conference on Intelligent User Interfaces*. New Orleans, LA: Association for Computing Machinery.
- Smith, M. A. (1999). Invisible crowds in cyberspace: Measuring and mapping the social structure of USENET. In: M. A. Smith & P. Kollock (Eds.), *Communities in cyberspace* (pp. 195–219). London: Routledge.
- Taboret-Keller, A. (1995). Conclusion: Code switching research as a theoretical challenge. In: L. Milroy & P. Muysken (Eds.), *One speaker, two languages* (pp. 344–355). Cambridge, UK: Cambridge University Press
- van Coetsem, F. (1988). *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris.
- Werry, C. C. (1996). Linguistic and interactional features of Internet Relay Chat. In: S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 47–64). Amsterdam/Philadelphia: John Benjamins.
- Wright, S., Ed. (2004). Multilingualism on the Internet. Special issue, *International Journal on Multicultural Societies*, 6(1).
- Yates, S. J. (1996a). Oral and written linguistic aspects of computer conferencing. In: S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 29–46). Amsterdam/Philadelphia: John Benjamins.
- Yates, S. J. (1996b). English in cyberspace. In S. Goodman & D. Graddol (Eds.), *Redesigning English: New texts, new identities* (pp. 106–140). London: Routledge.

Biographical Note

John C. Paolillo is Associate Professor of Informatics and Computing at Indiana University. His research focuses on the analysis of social behavior in online environments and the application of quantitative methods to linguistic analysis.