

# Spanish-English Codeswitching in Email Communication

Rosalyn Negrón Goldberg  
Department of Anthropology  
University of Massachusetts Boston

## Abstract

While spoken codeswitching (CS) among Latinos has received significant scholarly attention, few studies have examined written CS, specifically naturally-occurring CS in email. This study contributes to an under-studied area of Latino linguistic practices by reporting the results of a study of CS in the emails of five Spanish-English bilingual Latinos. Methods are employed that are not often used in discourse analysis of email texts, namely multi-dimensional scaling and tree diagrams, to explore the contextual parameters of written Spanish-English CS systematically. Consistent with the findings of other studies of CS in CMC, English use was most associated with professional or formal contacts, and use of Spanish, the participants' native language, was linked to intimacy, informality, and group identification. Switches to Spanish functioned to personalize otherwise transactional or work-related English-dominant emails. The article also discusses novel orthographic and linguistic forms specific to the CMC context.

## Introduction

*El uso de* two or more languages in one interaction, or codeswitching, has been examined in great detail over the past thirty years. Linguists have extensively recorded, described, and interpreted spoken, usually naturally occurring, codeswitching (CS). Their work is driven by a fascination with the distinctive social actors and social spaces that arise through culture and language contact. Yet these contacts are taking place in new and unprecedented ways in the digital age. The Web and other sites for computer-mediated communication (CMC) have acquainted many people with intimate conversations with strangers and shorthand exchanges with intimates. Interactional accessibility and immediacy—via email, cell phone text messaging, online messengers, and chat rooms—are both common and expected. And CMC has become the primary mode of exchange for entire subsets of inter-personal relationships. Building on the theoretical work developed for spoken discourse, researchers are finding these to be exciting times for study of the novel and not-so-novel discursive behaviors taking place in the "writing-as-talk" realms of CMC.

Electronic discourse has been claimed to fall in the middle of the continuum between spoken and written communication (Foertsch, 1995). "Interactive written discourse" (Ferrara, Brunner, & Whittemore, 1991) such as email is spoken-like, in that it enables users to communicate with rapid feedback and informality of style (Georgakopoulou, 1997). Unlike spoken discourse, however, email interactions lack prosodic and paralinguistic cues. Users have adapted to these limitations by using emoticons and abbreviations to express unseen (and often unperformed) facial and bodily expressions. Add to this the asynchronicity of email, and it is clear that electronic communication affords the user unprecedented control over self-presentation. Because asynchronous CMC is largely planned and non-spontaneous (Hinrichs, 2006), it encourages users to be more self-aware about the subtleties of communication. Thus, computer media have acquainted users with the nuances of calculated spontaneity, edited casual-talk, and emotionally subtle aprosodic written conversation. In short, computer media afford unique constraints and

freedoms. This limits the extent to which researchers can apply findings from spoken discourse studies to "written speech" directly, and the same undoubtedly applies to research on written CS. The unique contextual and technical dimensions of CMC have inspired a growing interest in CMC spaces as sites for the written reproduction of the cultural and linguistic meaning of spoken CS (Georgakopoulou, 1997; Hinrichs, 2006; Montes-Alcalá, 2007; Paolillo, 1996; Sebba, 2002; Warschauer, El Said, & Zohry, 2002). Nonetheless, given its fairly recent beginnings as a research program, written CS and other features of multilingualism in CMC remain understudied (Danet & Herring, 2003).

Among the questions still wide open for examination are: What rules guide bilingual speakers' code choices in electronic communication? In what email contexts will a bilingual choose English over his or her native language? What are the characteristics of the interlocutors with whom a person codeswitches in email? What forms of writing emerge with CMC codeswitches? Equally important is the need to fill the considerable gap in scholarly knowledge about the online/written CS practices of Latinos. With a few exceptions (e.g., Montes-Alcalá, 2005, 2007), research on the written language practices of Latinos, the United State's fastest growing minority, has been limited to analysis of non-naturally occurring corpora (Callahan, 2004; Lipski, 1982; Pfaff & Chávez, 1986). More research is needed on bilingual Latinos' written CS practices that are situated in real-world, everyday interactional contexts. As more and more quotidian communications take place online, and especially through email, research is needed to document the trends, shifts, and innovations in online bilingual practice. In this spirit, this article reports the results of an exploratory study of CS in the routine emails of five Spanish-English bilingual Latino graduate students. Methods not often used in discourse analysis—i.e., multi-dimensional scaling and tree diagrams—are employed to explore systematically the contextual parameters of written Spanish-English CS.

## Background

Myers-Scotton (1997) has described CS as an optimization strategy employed by rational actors. A code choice, according to this formulation, reflects an actor's calculation of the expected benefits and possible costs of that choice. Such calculations necessarily involve knowledge about the rights and obligations (RO), as well as the consequences, of available code options. A key dimension for evaluating the appropriateness and potential effects of a code choice is 'markedness' (Myers-Scotton, 1983, 1998). 'Unmarked' choices are characterized by a general agreement about the corresponding RO sets, or norms of an exchange. 'Marked' choices, in contrast, flout these norms (to varying degrees) and thus draw attention to a choice and its intentions, establishing new RO sets. In the data analyzed in this study, English and Spanish are used as both unmarked and marked choices. Situational factors like email subject and email recipient determine whether Spanish or English will be the base unmarked language,<sup>1</sup> with a second marked language embedded to achieve various communicative goals. Although he attributed little agency to such switches, Gumperz (1982) described this type of CS as 'situational.' In situational switching, group membership indicators like gender, age and status, and social setting determine the appropriateness of a code choice or CS itself.

In the most thorough treatment of email CS to date, Hinrichs (2006) builds on the frameworks advanced by Myers-Scotton and Gumperz. In his analysis of Jamaican Creole (JC) and Jamaican

English (JE) in email communication, he found that JC is the marked choice. This contrasts with the trend in spoken communication in Jamaica, where JC is the unmarked choice. Hinrichs suggests that the cognitive cost of writing JC is greater than writing JE, because his study participants were more familiar with JE's writing conventions and rules. Hinrichs reports that the use of JC in a mostly JE text is usually accomplished among interlocutors of the same status (situational switching) or to provide contextual information. In this latter function, CS is itself a conversational resource and serves as a 'contextualization cue' (Auer, 1995; Bailey, 2000a; Gumperz, 2001; Li, 1994). Contextualization cues frame or highlight information needed for interpretation and provide clues to the underlying intentions of a message. Often, these are cues to what is left unsaid in a conversation. For example, Georgakopoulou (1997) found that CS from Greek to English served to lighten apologetic emails in her corpus. By switching to English for a word or phrase during an apology, writers were able to stave off a potentially face-threatening act and reaffirm the intimate nature of the exchange.

In addition to discussing Jamaican Creole's importance for presenting particular social personas and its status as a 'we-code,' Hinrichs (2006) also described the significance of CS for indexing writers' social identities. The identity functions of CS are among the most widely explored in bilingual studies (Bailey, 2000b; Bucholtz & Hall, 2005; Gafaranga, 2001; Greer, 2005; Lo, 1999; Sebba & Wooton, 1998; Williams, 2006; Zentella, 1997). Not surprisingly, studies of CS in CMC have also examined the ways that writers align themselves varyingly with social categories and groups. Paolillo (1996) argued that in a Usenet forum on Punjabi culture, Punjabi served the group-identification needs of a Punjabi online community. Similarly, Georgakopoulou (1997) emphasized the importance of CS for alignment with an in-group of Greek friends and colleagues. Drawing on Zentella (1997), Montes-Alcalá (2007) observed that CS expressed bilingual bloggers' membership in "both Hispanic and Anglo worlds" (p. 169).

In these studies, minority codes were used to affirm in-group intimacy and familiarity, contrasting with a more standard or formal lingua franca, which in each case was English. As with spoken CS, research on written CS has shown that English is associated with professional or formal contexts and the online context in general. Warschauer et al.'s (2002) study of young Egyptian professionals' online use of Arabic and English found that English was used much more frequently online and in formal emails than Egyptian Arabic or classical Arabic. Conversely, a Romanized version of Egyptian Arabic was used extensively in informal email messages and online chats. In interviews, participants reported that they switched to Egyptian Arabic when expressing highly personal content that could not be expressed optimally in English. Warschauer et al.'s research exemplifies how situational switching functions in CMC and points to the dominance of English on the Web. Similarly, Paolillo (1996) observed that the use of English-only in a Punjabi online forum frequented by Punjabi expatriates was nearly four times as great as the use of Punjabi only. Both studies show that L2 English is preferred as the language of formal or informational exchanges and is perceived as less personal than native languages.

## Methods

### Participants and Questionnaire

The data presented here were collected from five native speakers of Spanish who were also fluent in English. The five participants were recruited through an email sent to graduate student listings at a southeastern university in the United States. Two women and three men were selected for the study based on their availability for an entire week. I provided participants with a questionnaire for use as a daily journal of their emails (see Appendix A: Daily Email Tracking Questionnaire). The questionnaire asked six questions about the language(s) used in emails, relationship to email recipients, email subjects, and function of particular language choices. Additionally, I asked participants voluntarily to provide excerpts from emails they wrote. Analysis of text from personal emails, particularly those between intimates, has been limited by the restricted accessibility to private data sources. Previous research has used public mailing list emails (Davis & Brewer, 1997; Durham, 2003; Yates, 1996) or the researchers' own emails or those of their social network members (Crystal, 2001; Georgakopoulou, 1997; Montes-Alcalá, 2005; Tsiplakou, 2009; cf. Hinrichs, 2006 for a notable exception). My use of both a daily journal questionnaire and volunteered email excerpts bypassed key privacy issues, in that participants did not divulge messages they were not comfortable sharing. This strategy yielded questionnaire data on 133 emails and further thematic detail on 101 email text samples. At the same time, it points to a clear limitation of the study, in that analysis of email themes is based on a relatively limited set of "safe" emails and is not necessarily representative of the full range of email exchanges in which participants engaged.

Data on participant characteristics appear in Table 1. All five participants were graduate students in their late 20s or early 30s. With the exception of one Dominican-American woman, participants were international students from Spanish-speaking countries. The Dominican-American woman was born in the U.S. but was a native Spanish speaker who learned English as a second language. The most number of years a participant had known and spoken English was 27 years and the least was 3 ½ years. As the data will show, participants' roles as students figured into the types of emails they sent and received. In addition to their responsibilities as graduate students, all five participants worked at least part-time. Thus, the data include both academic and work-related emails.

Gender	Nationality	No. of Years Known English	No. of Emails Contributed to Study
Female	Uruguayan	3.5	46
Male	Mexican	5	23
Female	Dominican-American	27	31
Male	Venezuelan	3.5	6
Male	Puerto Rican	17	27

Table 1. Participant characteristics

Participants were given the choice either to complete the questionnaires based on seven days' worth of emails they had already written and saved in their 'sent messages' folder or to complete

the questionnaire at the end of a seven-day period, starting on the day they received the questionnaire. Three participants opted to complete the questionnaires based on previously-sent messages. To minimize self-monitoring for the two participants who opted to complete their questionnaires after a seven-day period, I was not explicit about the specific CS focus of the study. Upon return of the completed questionnaires (as email attachments), I conducted brief semi-structured interviews with each participant. The purpose of these interviews was to further explore interesting uses of language that emerged from the questionnaires and to get the participants' personal thoughts on CS in their spoken communication.

### **Multidimensional Scaling of CS Themes**

Data analysis was targeted to yield insights about the underlying factors contextualizing codeswitches, as well as the novel language uses that emerged in CS contexts. Email excerpts were coded for CS function themes using a grounded theory method. In other words, the themes emerged in the course of analyzing the email excerpts, and no *a priori* categories were used. CS function themes describe the purpose of a code choice in each of the 101 email excerpts provided by participants. To arrive at the coding categories, I examined the immediate context in which the codeswitched words or phrases were used. First I asked: What is the dominant language of the email? and assigned a category accordingly. Next I asked: Did a clear or equally efficient Spanish/English equivalent exist for a codeswitched word or phrase? This helped distinguish between CS functions that were more technical as opposed to emphatic or stylistic. For example, the word "email" was a technical term used as an alternative to the longer, less efficient Spanish equivalent *correo electrónico*. "Email" and similar technical terms are accepted and understood globally in their English forms. As such, they do not necessarily reflect a user's stylistic preferences, nor do they connote an intention to emphasize or embellish a statement. Such uses were coded as *Technical Words & Phrases*.

Other terms are also widely used or popular and non-emphatic, but they do not depend on global consensus on appropriate usage nor are they more efficient alternatives to monolingual forms. These include words like "taxes" and "research." The use of such terms in a Spanish-dominant context reflects embeddedness in particular local, bilingual contexts and functions to minimize ambiguity. For instance, the word "research" is more strongly associated with its academic connotations than the Spanish equivalent *investigación*, which can be used in academic and criminal, as well as more mundane, inquiries. The use of such terms, categorized as *Popular Words & Phrases*, reflects participants' desire for communicative efficiency and their participation in English-dominant social milieus.

In contrast to *Technical* and *Popular Words & Phrases*, some codeswitches had more stylistic or emphatic functions. These included *Novel Words, Expressions, and Personal English/Spanish*. These CS functions describe cases in which a word or phrase is switched to capture culturally-specific meanings in one language for which no ideal alternative exists in the other language. Some such switches also have humorous or friendly undertones, even if close bilingual alternatives exist. Finally, I asked: Where does the codeswitched word or phrase appear in the body of an email? For example, if Spanish was used as a greeting in a mostly English email, this choice was coded as *Spanish Greeting*.

A total of 22 CS function themes were identified. The coding produced a dataset of 101 emails along the rows and 22 themes along the columns. Each cell contained a 1 or 0, indicating the presence or absence of one of the 22 themes in a given email (101 emails x 22 themes). I then created a similarity matrix of CS themes based on the co-occurrence of these themes in each of the email text samples (22 themes x 22 themes). Each cell in the matrix held the number of times that two themes co-occurred across all respondents' emails. Multi-dimensional scaling was run on the similarity matrix.

Multi-dimensional scaling (MDS) is a multivariate and exploratory data analysis method used to uncover underlying dimensions along which themes and other data items are distributed. In an MDS analysis, themes are represented spatially according to the similarity between—or in the case of this study, the co-occurrence of—themes. Themes that co-occur in a text appear close to each other on an MDS map. Themes that do not co-occur appear further apart. An MDS map is interpreted by looking at which themes cluster together. In this study, themes that cluster together represent code choices that tend to occur in the same email contexts. Further, using an MDS map, one can look for distributions of themes along some continuum. An example of a continuum along which CS themes can be ordered is *Intimacy*. One end of an *Intimacy* continuum might include personal and highly informal language uses, and the other end might include professional and highly formal uses. Interpretation of both the underlying dimensions and clusters is based on the researcher's understanding of the data and represents the best possible guess about the factors affecting a set of relationships among themes. A key advantage of using theme co-occurrence and MDS to analyze CS in emails is that it allows the researcher to explore systematically the contexts in which certain code choices occur. MDS maps visualize complex data in a way that is easy for the human mind to consider and interpret.

### Tree Diagrams

Tree diagrams were used to display associations between participants' code choices and different attributes of emails they received. Using SAS statistical software,<sup>2</sup> I created frequency tables for each of the categorical variables in this study, for a total of six: 'Email exchange source,' 'Language used by original sender (not participants),' 'Language of participants' reply,' 'Language used by original sender (participants),' 'Relationship to recipient,' and 'Email subject.' To answer questions about the language(s) they used, participants had the choice of: English-only; Spanish-only; Both but English-dominant; Both but Spanish-dominant; and Both, about the same of each.

I designed three tree diagram templates. The first diagram was to display code choice frequencies according to whether the message was a reply or initiated by the participant (see Figure 2). Using the frequency tables created in the first data step, I entered into this template the number of messages that were replies and the number that were initiated by the participant ('Email exchange source'). Arrows were used to indicate categories. Boxes were used to hold frequency counts for each of the categories. Next, I focused on those messages that were written as replies. Referring once again to the frequency table with counts for the 'Language used by original sender (not participant),' I copied into the template the counts for each of the language categories used by the original sender. Then, continuing to add branches to the tree, I copied the counts for each of the code options participants used in their replies. I repeated all but the last

step for those messages that were initiated by the participant. The same process, whereby counts from frequency tables were displayed in tree diagram templates, was used to construct the remaining two tree diagrams. The tree diagrams display an aggregate decision model of sorts. They help to trace the influence of contextual factors that ultimately lead to participants' code choices.

Chi-squared tests were performed on cross-tabulated frequency tables to test for associations. These were:

- 1) Code choice x Language of email received (IV1)
- 2) Code choice x Email subject (IV2)
- 3) Code choice x Relationship to recipient (IV2)

Chi-squared ( $X^2$ ) tests were used to test the independence of the variables, namely whether code choice was independent of each of the other three variables of interest. Three pieces of information are needed to determine whether one variable (code choice) is dependent of another (IV1, IV2, IV3). One must compute the  $X^2$  test statistic. This was done using SAS statistical software. One must also select the level of statistical significance needed to reject the null hypothesis that two variables are independent of each other. In the analysis, I chose the  $p = 0.001$  level of significance in the Chi-squared distribution. This level would indicate a highly significant statistical association or dependence between variables. Then, one must compute the degrees of freedom ( $df$ ), or the parameters of a distribution. This is determined by counting the number of levels of each variable, subtracting one from each, and multiplying these values. For example, the code choice variable has five levels (English-only, Spanish-only, English-dominant, Spanish-dominant, and Both the Same). IV1 has the same five levels. Thus, the  $df$  would be  $(5-1) \times (5-1) = 16$ . Using a Chi-squared distribution chart, the  $X^2$  value needed to establish dependence of variables is 39.25, given a significance level of 0.001 and  $df = 16$ . Thus, a  $X^2$  result greater than 39.25 would indicate that code choice is dependent on the language of email received, rejecting the null hypothesis.

It is important to note that while CS is embedded in broader socio-cultural contexts, the methods used to analyze the data are designed to identify systematically the effects of immediate contextual factors and the regularities of these effects across cases. Thus, data interpretation is based on the immediate context of a switch: features like language of recipient, email subject, and position within text body. Whenever possible, I also draw on information about participants and their own explanations elicited during interviews to interpret the meaning of a switch.

## Results and Discussion

### MDS

#### *Clusters*

The MDS map shown in Figure 1 displays ties between nodes (themes) to represent co-occurrence. Three clusters stand out most distinctly. Cluster 1 consists of three themes: *Spanish Dominant Personal*, *English Greetings*, and *English Expressions*. *Spanish Dominant Personal*

was used to code emails written primarily in Spanish to family and friends about non-work or academic subjects. In these emails, Spanish was the base, and CS to English served various roles. As the cluster shows, two key English CS functions in Spanish-dominant personal contexts were *English Greetings* and *English Expressions*. *English Greetings* corresponds to any use of English to mark the start of an email message (e.g., "Hello," "Hi," and "Dear All"). Idiomatic expressions written in English in a mostly Spanish text were coded as *English Expressions*. Examples of idiomatic expressions used by participants in this study include "powers that be," "A.K.A.," and "un-God like jobs." Idiomatic expressions in English serve to capture meaning not easily translatable in Spanish.<sup>3</sup> In either language, these expressions are neat packages of cultural information. Communicatively successful use of these in a bilingual context depends on shared (bi-cultural) knowledge among correspondents. To sum up, in Cluster 1 we see that in personal or intimate emails written in Spanish, CS to English served to introduce a message or to invoke American frames of reference best captured in English. Because these three themes cluster together on the MDS map, we know that their corresponding code choices tended to occur together in the same emails.

Cluster 2 features three other co-occurring CS themes. These categories include established loan words and phrases (borrowing) not considered CS (Lawson & Sachdev, 2000): *Technical Terms*, *English Proper Names*, and *Popular Culture English Words and Phrases*. *Technical Terms* refers to switches to technical terms in English in a mostly Spanish email body (e.g., "email," "fax," and "track changes" [an MS Word feature]). These technical loan terms from English are widely used, although Spanish equivalents exist for some (e.g., *correo electrónico* ('email')). *English Proper Names* in participants' emails were usually place names or formal titles like "Professor." Finally, the *Popular Culture* theme was a broad category that included any non-technical English word or phrase that has come into common usage among bilinguals. These terms are frequently used and semantically efficient, therefore, preferred across various contexts. They include words like "Spring" and "Fall" to describe a semester, and "BBQ," "taxes," "research," "papers," and "ads." The MDS map indicates that *Technical English Terms* and *English Proper Names* were connected to both work-related and personal email contexts. Overall, Cluster 2 points to the importance of CS and loan words for efficiently and effectively communicating information not easily translated into Spanish.

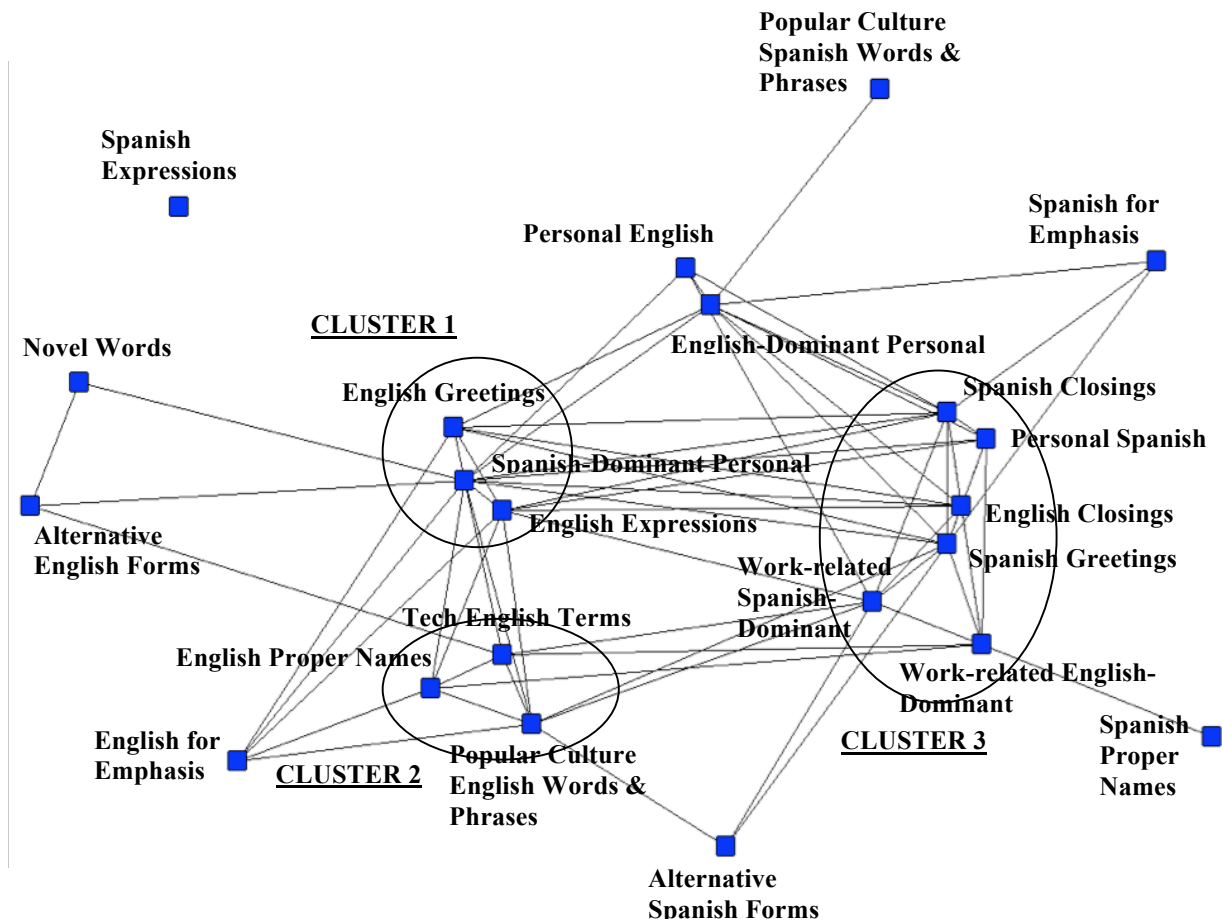


Figure 1. Multi-dimensional scaling of email function themes

Cluster 3 is a more loosely connected set of themes, which includes *Spanish Closings*, *Personal Spanish*, *English Closings*, *Spanish Greetings*, and both *Spanish Dominant* and *English Dominant Work-Related* emails. Here, work-related categories include both job and academic emails. *Spanish* and *English Closings* refer to code switches occurring at the end of emails to conclude a message (e.g., "See you later," *Saludos* ('greetings'), "Best," and *Hasta Pronto* ('see you soon')). *Personal Spanish* is the use of Spanish words or phrases with a friendly, intimate, or humorous tone in an otherwise English-dominant text. Such uses included *una copa de vino* ('a cup of wine'), *borracho* ('drunk'), and *casa* ('house'). CS scholars point out that Spanish-speaking bilinguals often briefly switch to words and phrases like *bueno* (denotes acceptance or agreement), *lo que sea* ('whatever'), *pos* ('well'), *ándale pues* ('OK then' or 'let's go!'), to mark their Latino ethnicity (Jacobson, 1982; Toribio, 2002). This type of CS is economical and particularly compatible with the often time-constrained informality of CMC.

Of the six themes appearing in Cluster 3, four refer to English-dominant email contexts. The close proximity of both English and Spanish work-related themes suggests that this cluster refers mostly to emails with work-related functions. Thus, the cluster suggests that switches to Spanish functioned to personalize otherwise transactional or work-related English-dominant emails.

Note that co-occurrence also reflects usage frequency. For example, *Spanish Expressions*, the use of Spanish idiomatic phrases in English-dominant contexts, is unconnected to other themes on the MDS map. This is because none of the participants made these types of switches. A simple count of themes that co-occur at least three times with other themes (as depicted by the connecting lines on the MDS map) yields ten themes that refer to Spanish-dominant contexts in which participants switched to English for various purposes. In contrast, six themes refer to English-dominant contexts with Spanish codeswitches. It appears, then, that the participants in this study more often switched to English in Spanish-dominant emails. English was more often used to supplement or enhance Spanish than the other way around.

I also referred back to the raw data for straight counts of the CS themes, specifically those related to greetings and closings. Counts based on the raw data indicate the number of times that CS functions occurred at all (as opposed to frequency counts based on co-occurrence). According to this simple count, there were three *English Greetings* and four *English Closings* in Spanish-dominant emails. Conversely, there were six *Spanish Greetings* and eight *Spanish Closings* in English-dominant emails. Thus, among this study's participants, embedded Spanish closings and greetings occurred twice as much as those in English. Email greetings and closings as a CS phenomenon have received little attention in the literature. Yet, from both the sender and the recipient's point of view, these brief switches may be among the most obvious acts of CS as a way to lessen distance and/or identify oneself with one's interlocutors. Greetings and closings tend to appear separated from the text body and are thus more likely to be considered separately from code-switched pieces embedded in longer stretches of writing. This separation creates a higher contrast between codes and renders the short greeting/closing an efficient sociolinguistic cue. While not generalizable, this modest finding suggests that brief switches to Spanish in English-dominant messages are a preferred way to personalize emails or lessen the distance from email recipients. The positioning of both *Spanish Greetings* and *Spanish Closings* in Cluster 3, which seems to reflect work-related emails, lends further credence to this observation.

Finally, it is instructive to consider un-clustered themes shown at the peripheries of the MDS map. Two themes in particular, *Novel Words* and *Deviant English*, represent participants' innovative use of both languages in the email context. As Sebba (2003) has argued, email represents an "unregulated space" where standard orthographic norms can be suspended. Here, participants join elements of both codes for efficiency or to signal ethno-linguistic identity membership. Thus, in switching from Spanish to English, some wrote English words according to Spanish orthography. For example, one participant wrote to her friend: *¿quieres que te de un raid?* ('would you like me to give you a ride'). Rather than write the English word "ride," the participant benefited from the writing environment to create an orthographically seamless switch. *Raid* does not interrupt the sentence's base Spanish orthographic rules. The chosen spelling suggests that in speech, the word is phonologically integrated into its Spanish base. I should note that the word *raid* is commonly used in Mexico as an alternative to *aventon* ('a lift'). However, the participant who used the word is from Uruguay. As a student in a Latin American Studies graduate program, however, she had frequent and often close interactions with fellow students from many Latin American countries, including Mexico. Her lexicon had apparently been influenced by these interactions.

Another participant wrote in an email to her friend, *el estaba berguachin* ('he was birdwatching'). In her interview, the participant who wrote this line shared that *berguachin* originated in her verbal communication with a bird-watching group. The word came about as she and a group of Spanish-speaking friends tried to teach Spanish to an English-speaking friend. In simplifying Spanish translations for the English speaker, they applied Spanish phonology to "bird watching" and encouraged him to think of it and use it as a Spanish word.<sup>4</sup> Thus, *berguachin* emerged orally, and in its migration to CMC developed matching orthography. For the bird-watching group, knowledge and usage of *berguachin* identifies its members. As it was used in the participant's email to her fellow bird-watcher, this inventive CS thus served both linguistic and group identification functions.

As indicated by the theme *Deviant Spanish*, the computer-mediated environment also allowed the use of slightly deviant spellings. They are deviant in the sense of diverging from standard Spanish orthography, but in many ways they are quite ordinary, insofar as substituting characters or letter clusters for letters not found on a keyboard is standard practice for Spanish-English bilinguals writing in English. Still, given that most word processing programs include hundreds of characters not found on keyboards, the use of some deviant spellings suggests a need for speed and efficiency when writing emails. In one case, for example, the absence of Spanish letters on the keyboard led participants to compensate by using available letters to represent the necessary phonemes (e.g., *manyana*: [ny] replaces Spanish [ñ] in *mañana* ('tomorrow')).

In another example, deviant spellings suggested a principled efficiency. Characters available on the keyboard facilitated the creation of a slightly modified Spanish word *todos / todas* ('all'). *Todos* is the masculine variant of the word and standard as a greeting to a mixed gender group. *Todas*, on the other hand, is only used to address an all-female group. One participant, who preferred a gender-neutral address that did not impose the masculine variant on a group of male and female email recipients, preferred to write *tod@s* (see Zentella, 2003 for an example of this strategy in academic writing). He explained that his use of the word was political rather than functional. What is important for the current discussion is that this gender-inclusive term would not be possible to express orally using a single Spanish word. It is a relatively new and creative construction that has diffused through electronic communication. Arguably, the use of this politically-nuanced greeting in my data also functioned to signal the writer's socio-political awareness and feminist leanings to an audience of similarly-minded women and men. Neither of the above examples of non-standard Spanish orthography is unique to bilingual individuals. I include them in the present analysis because the forms were used in bilingual email texts.

### *Dimensions*

As stated above, MDS helps to uncover underlying dimensions that explain the clustering and distribution of themes. Dimensions can be gleaned horizontally, vertically, or diagonally. The MDS map in Figure 1 suggests two main dimensions that elucidate why certain sets of themes co-occurred more than others. The first dimension is evident horizontally, and suggests a distribution of themes from English CS in Spanish-dominant texts (on the left) to Spanish CS in English-dominant texts (on the right). This dimension is more obvious than revealing, but confirms that the data are coherent and reliable. The second dimension is detectable vertically. Themes related to personal content and personalized language strategies appear at the top, and

transition to work-related themes appears towards the bottom. This suggests that there is some differentiation in written CS, depending on whether the email is intended for personal or work-related communication.

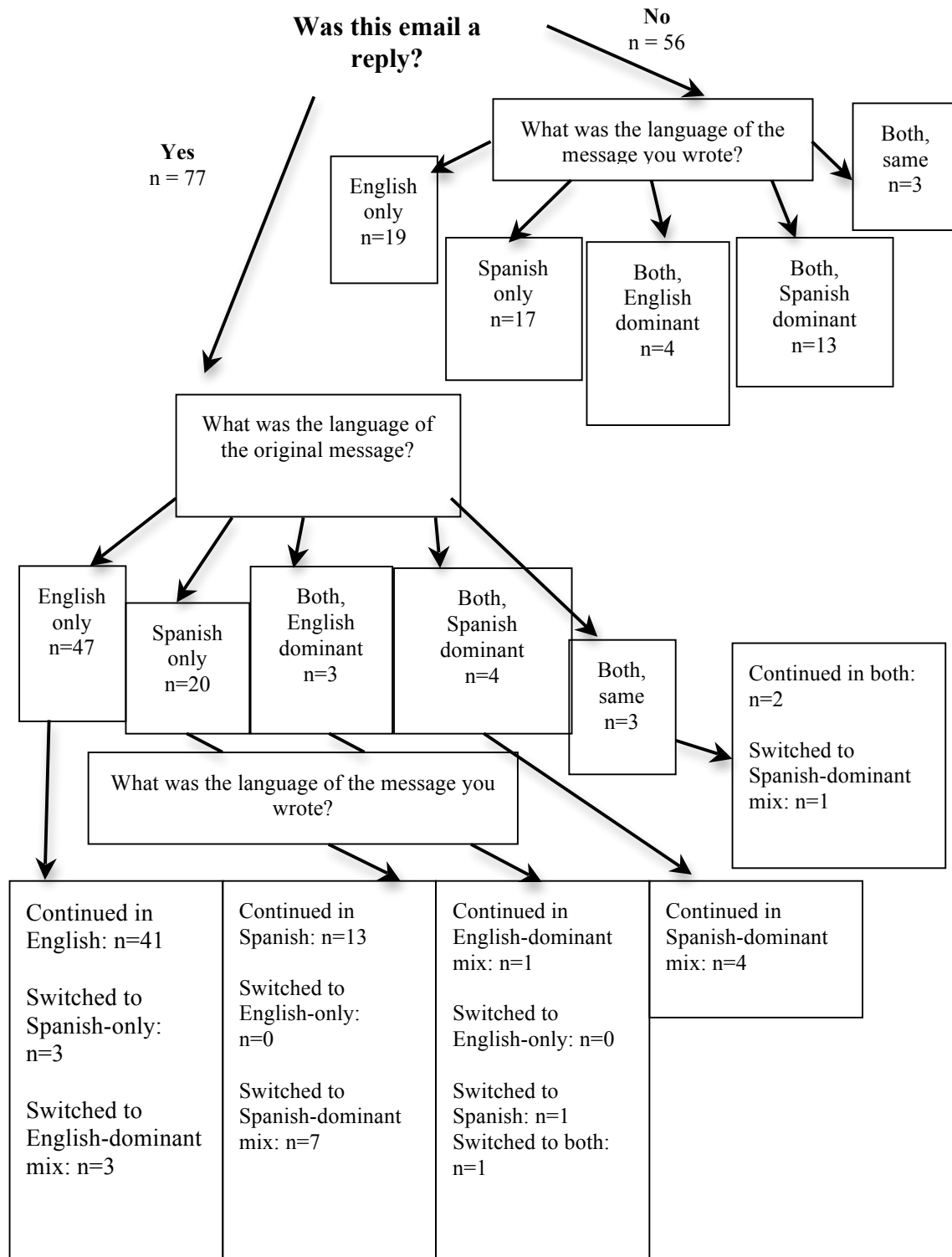
### Tree Diagrams

Figures 2, 3, and 4 present the results of the analysis of the email journal questionnaires. Figure 2 shows the relationship between the language(s) of emails received and the language(s) of participants' replies. The Chi-square test of association between the recipient email language and language of reply was highly statistically significant ( $\chi^2 = 106.5$ ,  $p < .001$ ). Recall from the methods section that a  $\chi^2$  value of 39.25 is needed to reject the null hypothesis that participants' code choices and language of emails received are independent. The results of the analysis show a  $\chi^2$  value well above that needed.

Three general patterns emerged. First, emails originating from others were almost two times more likely to be in English than in Spanish. This contrasts with participants' code choice when they were the initiators of email exchanges; initiating messages were written in Spanish slightly more often. Despite the apparent preference for using Spanish in their emails (or rather, the apparent preference for sending emails to their Spanish-speaking contacts), the participants overwhelmingly replied in English to emails written in English. However, there were some cases in which participants replied only in Spanish to English-initiated emails. This differed from what was observed for replies to emails written in Spanish, where there was never a switch to English-only.

Interviews with participants point to at least two reasons for this finding. In talking about his CS behavior in spoken communication, one participant said that if a close Spanish-speaking friend were to approach him in English, he would say "Why are you talking to me in English?" and redirect the conversation to Spanish. The assumption is that among intimates who both speak Spanish, Spanish is the preferred and expected language. In other words, English breaks with the interpersonal conventions between two native Spanish speakers. Another explanation taken from interviews with participants has to do with competence in English, the L2 for all five participants in this study. Participants noted that in both spoken and written communication, Spanish is preferred when clarity is a priority. Thus, (cross-email) switches to Spanish-only in replies to English-initiated emails may be due to participants' wishes to be able to express themselves more clearly.

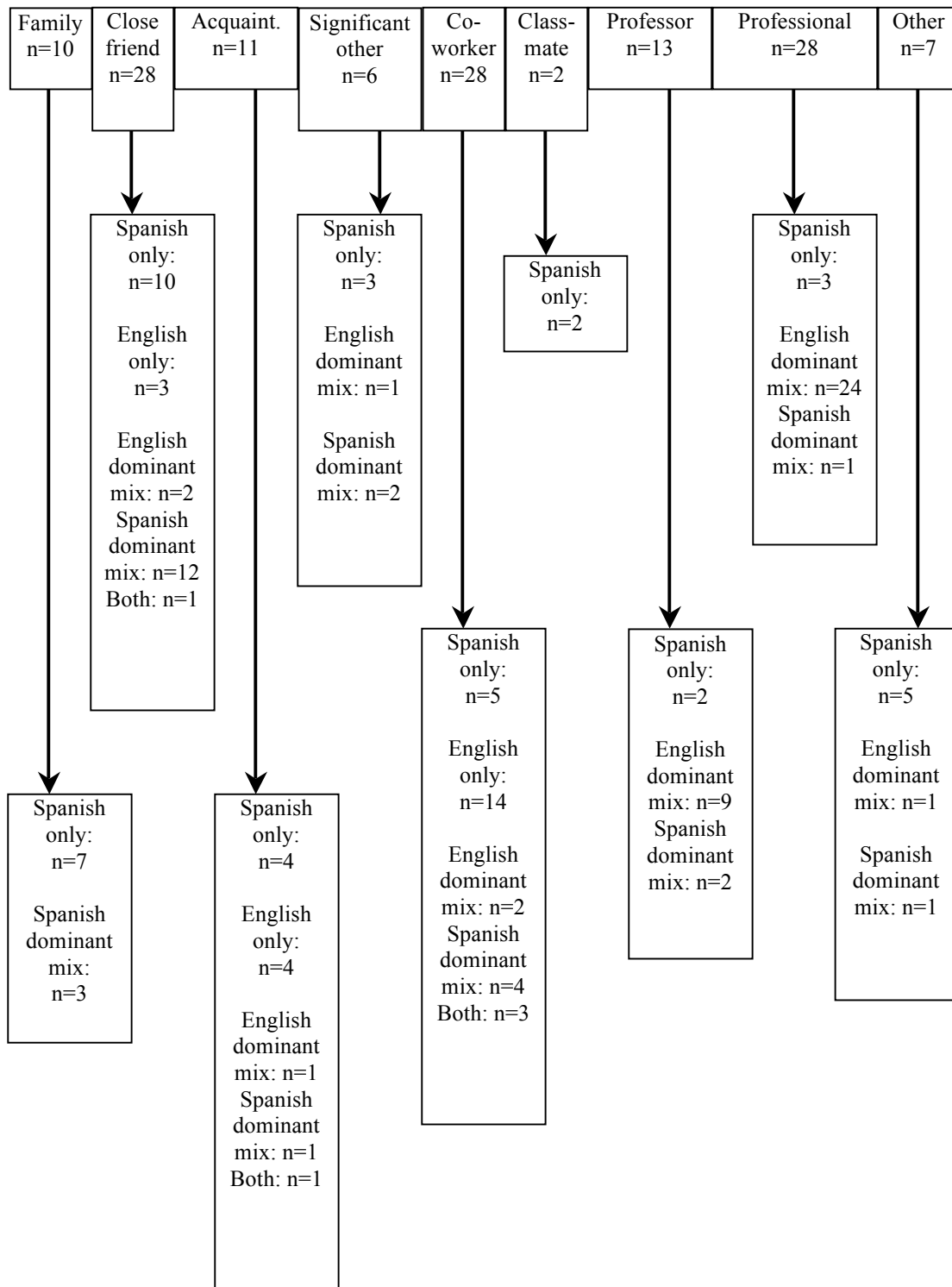
Next I examined the link between code choice and the participants' relationship to email recipients (Figure 3). Spanish was the preferred language when communicating with family and friends. English, in contrast, was most often the language of professional and academic communication, even in cases where participants' professional or academic contacts were bilingual. The association between relationship of recipient and code choice was also statistically significant ( $\chi^2 = 72.5$ ,  $p < .001$ ). The  $\chi^2$  was greater than the approximately 60 needed, at the 0.001 significance level with  $df = 32$ , to reject the null hypothesis that participants' email code choices are independent of their relationship to email recipients.



( $\chi^2 = 106.5, p < .001$ )

Figure 2. Language choice: Language of original email

### What is your relationship to the recipient?



( $\chi^2 = 72.5, p < .001$ )

Figure 3. Language choice: Relationship to email recipient

### What is the subject of the email?

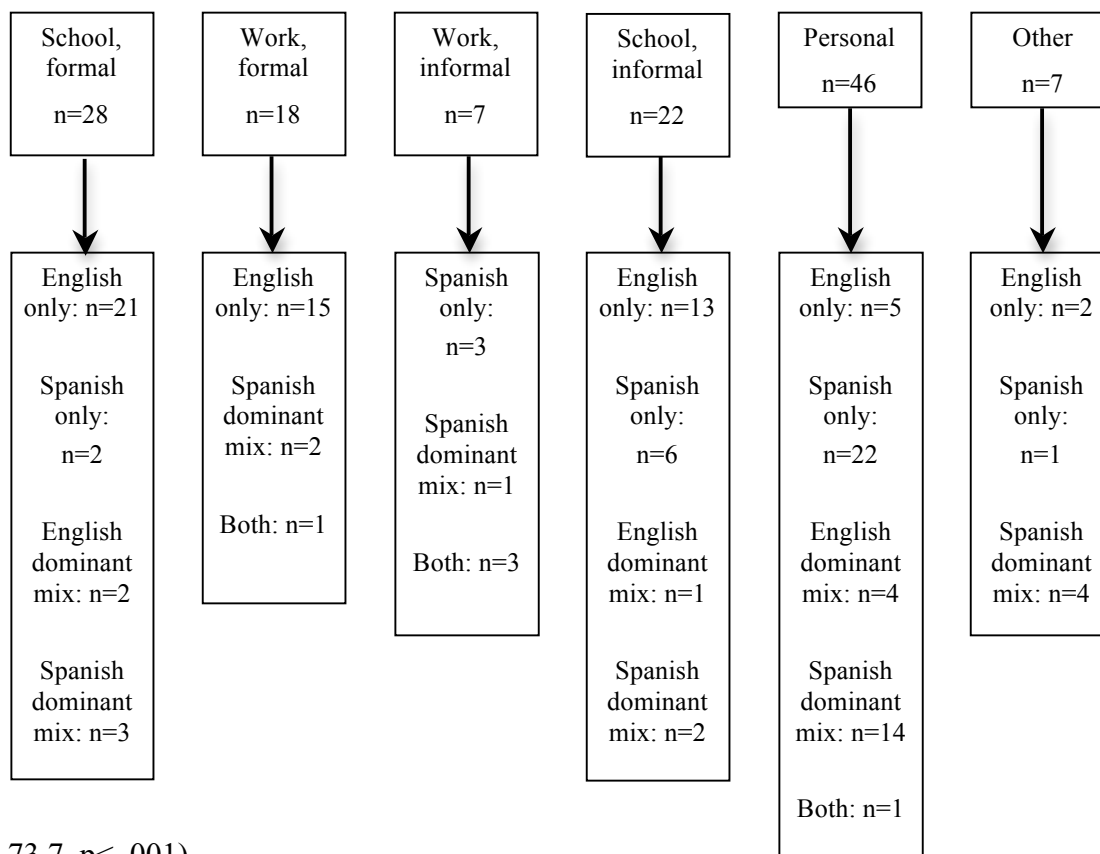


Figure 4. Language choice: Subject of email

Finally, a similar pattern was observed with code choice and email subject (Figure 4). In both formal (directed to professors) and informal (directed to classmates) academic emails, English and mixed English-dominant writing were preferred. However, a few emails regarding work-related subjects of an informal nature (e.g., sending funny anecdotes to co-workers) were written in Spanish or both, but not exclusively in English. As for emails of a personal nature, Spanish-only or mixed Spanish-dominant were the preferred choices. The association between code choice and email subject was statistically significant ( $\chi^2 = 73.7, p < .001$ ). The  $\chi^2$  was greater than the 45.32 needed, at the 0.001 significance level with  $df = 20$ , to reject the null hypothesis that email code choice is independent of the email subject.

### Conclusion

This study sought to add to the limited data available about the bilingual email practices of Latinos. While the results from this small study of five Latino email users cannot be used to generalize about Spanish-English CS in email communication, several findings are consistent with, and thus support, past research on CS in email communication. Further, this study

represents a novel approach to CMC CS, highlighting the value of drawing on a varied range of methodological approaches.

As students in an American university and actors embedded in English-dominant networks, participants wrote more English-only emails than Spanish-only ones. This was in part because most messages were replies to English emails. However, participants' preference for Spanish was evident in participant-initiated emails in which Spanish was more often used. This makes sense, given that four out of five of the participants had been in the U.S. less than 15 years (three participants for less than five years), and given that English was the L2 for all participants. In Hinrichs (2006)' study of email CS among Jamaicans, Jamaican Creole (JC) was the marked code in written communication, while Jamaican English (JE) was the unmarked code. Hinrichs suggests that the cognitive cost of writing JC is greater than writing JE, because his study participants were more familiar with JE's writing conventions and rules. Along similar lines, participants in the present study were more familiar with the conventions and rules of written Spanish. Thus, Spanish was more often used with other Spanish-English bilinguals as the unmarked, base code onto which English was added for clarity or emphasis. That is, English was more often used to supplement Spanish than the other way around. The supplemental use of English in Spanish-dominant emails served as a contextualization cue (Bailey, 2000a; Gumperz, 2001) to signal distinctly American frames of reference. This was important as a way to draw on shared cultural information for efficient expression and interpretation. English-dominant text, on the other hand, was the preferred choice in professional or academic email exchanges.

As with past research that has documented the role of English as a lingua franca in multilingual online contexts (Kelly Holmes, 2004; Wright, 2004) and as the language of business and technology (Paolillo, 1996; Warschauer et al., 2002), in this study use of technical terms and common English words and phrases communicated concepts not easily translatable into Spanish. In contrast to this practical function of English, an important function of switches to Spanish was to personalize otherwise work-related emails. Consistent with other trends in this study, this suggests that Spanish was associated with intimate and informal communication. In interviews, participants confirmed that in their face-to-face communication with close Spanish-speaking friends and relatives, the expectation is that Spanish be spoken. Therefore, while in some instances participants replied in Spanish-only to English-only emails, they never replied in English-only to Spanish-only emails. Accordingly, Spanish was preferred with family and friends and English with professors and co-workers. In emails where work- or school-related themes intersected with a more personal relationship and informal tone, Spanish or a mix of both English and Spanish was preferred, but not English-only. It seems, then, that participants' relationship to email recipients was a key determinant of email language choice. Finally, as has been observed about the CMC environment in general (Danet & Herring, 2007), the freedoms and limitations of the medium encouraged novel spellings and morphological constructions. In some cases, these novel forms had group identification functions.

Paolillo (1996) argues that in the Usenet forum dedicated to Punjabi culture that he studied, Punjabi served the group identification needs of the Punjabi online community. The present study shows that Spanish, especially, was important as a group identification tool—it was the language of intimate, informal communication and a way to affirm group membership. Given this, the benefits of bilingualism in the email context are intriguing and merit further investigation. Email can lead to unintended meanings, miscommunication, and misunderstanding

with negative consequences, due to lack of prosodic cues. However, it seems that bilingualism affords the email writer greater flexibility and control over a message. This can happen in two ways evinced in this study's data: 1) either English or Spanish can be used to express concepts not easily translated into one or the other language; 2) Spanish in its role as an intimate / friendly code can help to soften or humorize a message and connect two email interlocutors (see also Georgakopoulou, 1997). Brief switches to Spanish in an English-dominant text, such as in greetings or closings, are efficient ways to express an *intention* to be viewed favorably by an email recipient. Spanish greetings and closings in English work-related emails, for example, may stand out for the reader from the rest of the text, thus highlighting the switch itself, the writer's decision to switch, and the pragmatics of that choice.

The methodological approach taken in this study departs from approaches used in previous studies of CS in CMC. MDS and tree diagrams shed useful light on the contextual parameters of CS in email communication and exploit some methodologically convenient aspects of email. By the time they reach the researcher, email data are already partly analyzed (by the email program and the research participant herself). Features in email programs enable writers to organize emails sent and received into various analytically-relevant categories (e.g., "Work," "Family," "Friends"), and allow easy counts of message frequencies by sender/recipient or subject. Subject lines, when available, cue email recipients (and researchers) to the focal theme of a given message. And, of course, email comes transcribed for ready text analysis, faithful to the communicative idiosyncrasies of the writer. Tried and tested discourse and conversation analytic methods can only benefit from thinking in fresh ways about this relatively new type of data. This study adds to the growing body of work on bilingualism in CMC and advocates expanding the methodological toolkits we bring to the task.

## Notes

- <sup>1</sup> The base language is the structurally dominant language in which a second language is embedded.
- <sup>2</sup> The data analysis for this paper was generated using SAS software Version 8 of the SAS System for Windows 2004, SAS Institute Inc., Cary, NC, USA.
- <sup>3</sup> With regard to loan translations of idiomatic expressions (calques) and other instances of borrowing, I agree with Myers-Scotton's (1992) assertion that CS and borrowing are not two distinct processes. Therefore, I do not make too fine a point of distinguishing between codeswitches and borrowings in the data. As a general rule, I take frequency of use as the main criterion for classifying an utterance as borrowing. However, except for *English Technical Terms* and *Popular Culture English Words and Phrases*, which were used more frequently by participants, there were not enough email excerpts to speculate about the use frequency of certain words or phrases that might qualify as borrowings (e.g., *raid*/"ride" in the discussion on *Deviant Spellings*).
- <sup>4</sup> Possible translations of the gerund "birdwatching" include *ornitología* ('ornithology'), which has scientific connotations less fitting to the hobby group, and *mirar pájaros* ('to watch birds'), which does not imply a hobby.

## Appendix A. Daily E-mail Tracking Questionnaire

### *Date of E-mail:*

Please provide information on each of the e-mails that you sent.

#### E-Mail #1

- Was this e-mail a reply to an e-mail received from someone else?
  - yes (if yes, please answer the next question)
  - no
- Was the e-mail that you replied to written in:
  - English only
  - Spanish only
  - Both English & Spanish, but mostly English
  - Both English & Spanish, but mostly Spanish
  - Both English & Spanish, about the same each
- Choose one response that best describes your relationship to the recipient of your e-mail:
  - family member
  - close friend
  - casual friend (acquaintance)
  - significant other
  - co-worker / academic colleague
  - classmate
  - professor / instructor
  - professional / academic contact
  - other, please specify: \_\_\_\_\_
- What language was your e-mail written in?
  - English only
  - Spanish only
  - Both English & Spanish, but mostly English
  - Both English & Spanish, but mostly Spanish
  - Both English & Spanish, about the same each
- If you used both English & Spanish, please briefly describe what kinds of things you wrote in English and what kinds of things you wrote in Spanish. *(For example, some people write their greetings and closings in Spanish, and the rest of their e-mail in English.)*
- Choose the response that best describes the subject of the e-mail:
  - school related, formal
  - school related, informal
  - work related, formal
  - work related, informal
  - personal
  - other, please specify \_\_\_\_\_

**Are there any excerpts from this e-mail that you would be willing to provide as a sample?** If yes, please cut and paste here.

## References

- Auer, P. (1995). The pragmatics of code-switching: A sequential approach. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 115-135). Cambridge, UK: Cambridge University Press.
- Bailey, B. (2000a). Switching. *Journal of Linguistic Anthropology*, 9(1/2), 241-243.
- Bailey, B. (2000b). The language of multiple identities among Dominican Americans. *Journal of Linguistic Anthropology*, 10(2), 190-223.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4/5), 584-614.
- Callahan, L. (2004). *Spanish/English codeswitching in a written corpus*. Amsterdam / Philadelphia: John Benjamins.
- Crystal, D. (2001). *Language and the Internet*. Cambridge, UK: Cambridge University Press.
- Danet, B., & Herring, S. C. (2007). Introduction. In *The Multilingual Internet* (pp. 3-39). New York: Oxford University Press.
- Danet, B., & Herring, S. C. (2003). Introduction: The multilingual Internet. *Journal of Computer-Mediated Communication*, 9(1). Retrieved November 20, 2009, from <http://jcmc.indiana.edu/vol9/issue1/intro.html>
- Davis, B. H., & Brewer, J. P. (1997). *Electronic discourse: Linguistic individuals in virtual space*. Albany, NY: SUNY Press.
- Durham, M. (2003). Language choice on a Swiss mailing list. *Journal of Computer-Mediated Communication*, 9(1). Retrieved November 20, 2009 from <http://jcmc.indiana.edu/vol9/issue1/durham.html>
- Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication*, 8(1), 8-34.
- Foertsch, J. (1995). The impact of electronic networks on scholarly communication: Avenues to research. *Discourse Processes*, 19(2), 301-328.
- Gafaranga, J. (2001). Linguistic identities in talk-in-interaction: Order in bilingual conversation. *Journal of Pragmatics*, 33(12), 1901-1925.
- Georgakopoulou, A. (1997). Self-presentation and interactional alliances in e-mail discourse: The style- and code-switches of Greek messages. *International Journal of Applied Linguistics*, 7(2), 141-164.
- Greer, T. (2005). Co-constructing identity: The use of 'haafu' by a group of multi-ethnic Japanese teenagers. In J. Cohen, K. T. McAlister, K. Rolstad, & J. McSwan (Eds.), *ISB4: Proceedings of the 4<sup>th</sup> International Symposium on Bilingualism*. Somerville, MA: Cascadilla Press.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge, UK: Cambridge University Press.
- Gumperz, J. (2001). Interactional sociolinguistics: A personal perspective. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 215-228). Malden, MA: Blackwell.
- Hinrichs, L. (2006). *Codeswitching on the web: English and Jamaican Creole in e-mail communication*. Amsterdam / Philadelphia: John Benjamins.

- Jacobson, R. (1982). The social implications of intra-sentential code-switching. In J. Amastae & L. Elías-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 182-208). Cambridge, UK: Cambridge University Press.
- Kelly Holmes, H. (2004). An analysis of the language repertoires of students in higher education and their language choices on the Internet (Ukraine, Poland, Macedonia, Italy, France, Tanzania, Oman and Indonesia). *International Journal of Multicultural Societies*, 6(1), 29-52.
- Lawson, S., & Sachdev, I. (2000). Codeswitching in Tunisia: Attitudinal and behavioural dimensions. *Journal of Pragmatics*, 32(9), 1343-1361.
- Li, W. (1994). *Three generations, two languages, one family: Language choice and language shift in a Chinese community in Britain*. Clevedon, UK: Multilingual Matters.
- Lipski, J. (1982). Spanish-English code-switching in speech and literature: Theories and models. *The Bilingual Review*, 9(3), 191-212.
- Lo, A. (1999). Codeswitching, speech community membership, and the construction of ethnic identity. *Journal of Sociolinguistics*, 3(4), 461-479.
- Montes-Alcalá, C. (2005). ¡Mándame un e-mail! Cambio de códigos español-inglés online. In L. Ortiz & M. Lacorte (Eds.), *Contacto y contextos lingüísticos: El español en los Estados Unidos y en contacto con otras lenguas* (pp. 173-185). Madrid / Frankfurt: Iberoamerican / Vervuert.
- Montes-Alcalá, C. (2007). Blogging in two languages: Code-switching in bilingual blogs. In J. Holmquist, A. Lorenzino, & L. Sayahi (Eds.), *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics* (pp. 162-170). Somerville, MA: Cascadilla Proceedings Project.
- Myers-Scotton, C. (1983). Comment: Markedness and code choice. *International Journal of the Sociology of Language*, 44, 115-136.
- Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual and Multicultural Development*, 13(1-2), 19-39.
- Myers-Scotton, C. (1997). Rational actor models and social discourse analysis. In E. R. Pedro (Ed.), *Discourse Analysis, Proceedings of the First International Conference on Social Discourse Analysis* (pp. 177-99). Lisbon: Edicoes Colibri/Associacao Portuguesa de Linguistica.
- Myers-Scotton, C. (1998). A theoretical introduction to the markedness model. In C. Myers-Scotton (Ed.), *Codes and consequences: Choosing linguistic varieties* (pp. 18-38). New York: Oxford University Press.
- Paolillo, J. C. (1996). Language choice on soc.culture.punjab. *Electronic Journal of Communication*, 6(3). Retrieved November 20, 2009 from <http://www.cios.org/ejcpublish/006/3/006312.html>
- Pfaff, C., & Chávez, L. (1986). Spanish/ English codeswitching: Literary reflections of natural discourse. In R. von Bardeleben & D. Briesemeister (Eds.), *Missions in conflict: Essays on US-Mexican relations and Chicano culture* (pp. 229-254). Tübingen, Germany: Narr.
- Sebba, M. (2002, October). Regulated spaces: Language alternation in writing. Paper presented at the *Colloquium on Code-Switching, Class and Ideology, 2nd International Symposium on Bilingualism, SIB2002*, Vigo. Retrieved November 20, 2009 from <http://www.ling.lancs.ac.uk/staff/mark/vigo/regspace>

- Sebba, M. (2003). Spelling rebellion. In J. Androutsopoulos & A. Georgakopoulou (Eds.), *Discourse constructions of youth identities* (pp.151-172). Amsterdam / Philadelphia: John Benjamins.
- Sebba, M., & Wooton, T. (1998). We, they and identity. In *Code-switching in conversation: Language, interaction and identity* (pp.262-289). London: Routledge.
- Toribio, A. J. (2002). Spanish-English code-switching among US Latinos. *International Journal for the Sociology of Language*, 158, 89-119.
- Tsiplakou, S. (2009). Doing (bi)lingualism: Language alternation as performative construction of online identities. *Pragmatics*, Special issue on *Language, discourse and identities. Snapshots from the Greek context*, 19(3).
- Warschauer, M., El Said, G. R., & Zohry, R. (2002). Language choice online: Globalization and identity in Egypt. *Journal of Computer-Mediated Communication*, 7(4). Retrieved November 20, 2009 from <http://jcmc.indiana.edu/vol7/issue4/warschauer.html>
- Williams, A. M. (2006). *Bilingualism and the Construction of Ethnic Identity among Chinese Americans in the San Francisco Bay Area*. Ph.D. Dissertation, Department of Linguistics, University of Michigan.
- Wright, S. (Ed.) (2004). Multilingualism on the Internet. Special issue, *International Journal of Multicultural Societies*, 6(1), 5-13.
- Yates, S. J. (1996). Oral and written aspects of computer conferencing: A corpus based study. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social, and cross-cultural perspectives* (pp. 30-46). Amsterdam / Philadelphia: John Benjamins.
- Zentella, A. C. (1997). *Growing up bilingual*. Oxford, UK: Blackwell Publishers.
- Zentella, A. C. (2003). 'José, can you see?': Latin@ responses to racist discourse. In D. Sommer (Ed.), *Bilingual games. Some literary investigations* (pp. 51-66). New York: Palgrave Macmillan.

### Biographical Note

Rosalyn Negrón Goldberg [rosalyn.negron@umb.edu] is an Assistant Professor in the Department of Anthropology at the University of Massachusetts Boston. She specializes in the study of language and ethnic flexibility among U.S. Latinos.

### License

Any party may pass on this Work by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing License. The text of the license may be accessed and retrieved via Internet at [http://www.dipp.nrw.de/lizenzen/dppl/dppl/DPPL\\_v2\\_en\\_06-2004.html](http://www.dipp.nrw.de/lizenzen/dppl/dppl/DPPL_v2_en_06-2004.html)