CODE-SWITCHING IN PUERTO RICAN BILINGUAL BLOGS

by

AMY HERNANDEZ

(Under the Direction of Sarah Blackwell)

ABSTRACT

Code-switching is a linguistic phenomenon that is often studied in the areas of contact linguistics, language education, and bilingualism. Most often code-switching has been studied in oral communication because most often those who code-switch use this linguistic tool in more informal settings. This paper analyzes code-switching in the context of Internet blogs. While the medium of blogging is a written one, it is a very informal setting and bloggers often use language that is more like speech than writing. For this reason, it is of interest to study code-switching in this area. In the present study, 20 Puerto Rican blogs are analyzed to examine the contact between English and Spanish in the language of the bloggers. The three hypotheses formed for this thesis are that 1) it is possible to propose a set of categories of code-switching that better describes and accounts for the data in Spanish-English Internet blogs; 2) most bloggers will use Spanish as a Matrix Language (ML) and English as an Embedded Language (EL); and 3) the most frequently occurring type of code-switch is lexical item switches. An additional goal of this thesis is to elucidate the differences between code-switching and borrowing in light of the fact that phonology can no longer play a part in the determination.

INDEX WORDS: Language Contact, Code Switching, Lexical Borrowings from English into Spanish, Blogging, Computer Mediated Communication
CODE-SWITCHING IN PUERTO RICAN BILINGUAL BLOGS

by

AMY HERNANDEZ

B.A., The University of Missouri, 1997

B.S.Ed., The University of Missouri, 1998

M.A., The University of Georgia 2004

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2009
CODE-SWITCHING IN PUERTO RICAN BILINGUAL BLOGS

by

AMY HERNANDEZ

Major Professor: Sarah Blackwell
Committee: Chad Howe
Diana Ranson

Electronic Version Approved:
Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2009
DEDICATION

I dedicate this thesis to my parents who never let me give up. Specifically, I dedicate this to my mother for taking the time to laugh with me in the rain, y a mi padre por enseñarme la definición misma de ‘buena gente’ a través de su propio ejemplo.
ACKNOWLEDGEMENTS

I would first like to thank my major professor, Dr. Sarah Blackwell, who met with me often to give me advice and suggestions. I would also like to thank her for letting me know how proud she is of me and for giving me much needed reality checks. She also helped keep me excited by my work through her interest in my project. I would also like to thank the other members of my committee, Dr. Diana Ranson and Dr. Chad Howe for giving of their time and working with me. I deeply appreciate all of their kind words and encouragement as well as suggestions to improve my work. I owe my deepest thanks to my participants for writing such interesting blogs. I am extremely grateful for my family, who never thought I was crazy for adding a second Master’s degree and who never failed to remind me how proud they are of me, even when I didn’t feel proud of myself. Lastly, I thank my friends, for letting me be crazy, for giving me an escape from my work, and for not letting me give up. Without my friends and family I never would have finished and for their support I am eternally grateful.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivations and goals</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Why study blogs</td>
<td>2</td>
</tr>
<tr>
<td>1.3 Why Puerto Rican blogs</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Definitions</td>
<td>7</td>
</tr>
<tr>
<td>1.5 Organization of project</td>
<td>9</td>
</tr>
<tr>
<td>2 LITERATURE REVIEW</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>10</td>
</tr>
<tr>
<td>2.2 Code-switching versus borrowing</td>
<td>10</td>
</tr>
<tr>
<td>2.3 Code-switching in spoken discourse</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Code-switching in written corpora</td>
<td>23</td>
</tr>
<tr>
<td>2.5 Code-switching and Internet data</td>
<td>27</td>
</tr>
<tr>
<td>2.6 Summary</td>
<td>37</td>
</tr>
<tr>
<td>3 METHODOLOGY AND DATA ANALYSIS</td>
<td>40</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Methodology</td>
<td>40</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Code-switching categories proposed by Valdés-Fallis (1976) ........................................16
Table 2: A description of code-switching categories proposed by Gumperz (1982) ....................19
Table 3: A description of code-switching categories proposed by McClure (1981) ....................21
Table 4: Typology of code-switching uses by Montes-Alcalá (2007) ........................................36
Table 5: Comparison of other researchers’ code-switching categories ...................................44
Table 6: Breakdown of posts by matrix language .................................................................66
Table 7: Percentages of CS and B as well as total number of tokens ......................................67
Table 8: Breakdown of CS categories ...................................................................................69
CHAPTER 1

INTRODUCTION

1.1 Motivations and goals

Code-switching is a linguistic phenomenon that is often studied in the areas of contact linguistics, language education, and bilingualism. This feature of language use shows a specific dexterity that allows speakers to access two codes, to switch between the two to find the right word or expression, to emphasize their utterances in just the right way, or to show linguistic solidarity. Although code-switching has often been studied in oral communication, this thesis analyzes code-switching in the context of Internet blogs. While the medium of blogging is a written one, it is a unique form of communication that is not quite like written language or spoken discourse. Blogging retains some of the features of written language such as the ability to edit a message and the lack of conversational exchange; however, bloggers use this format of expression freely and have the tendency to use language without self-editing much like they would in a conversation. For this reason, it is of interest to study code-switching in this new genre. Additionally, the study of blog language can provide insight into language change, as it is a snapshot of a speaker’s language use untainted by a filter one might use when the intended audience is known. For instance, bloggers can write posts without necessarily being concerned with the readership, and thus blogging is much like writing a diary or internal dialogue, which is not co-constructed in any way. Furthermore, blogging is distinct from conversations because there is no exchange with or influence of other speakers in the act of producing language. Nilsson (2003:30) describes the distinct qualities of ‘blogspeak’, noting that it “combines writing
and speech in a unique environment, one that supports both the written internal monologue and threads of conversation.” Therefore, the language used by bloggers is of interest to linguists because it is independently constructed.

In the present study, 20 Puerto Rican blogs are analyzed to examine the contact between English and Spanish in the language of the bloggers. The three hypotheses formed for this thesis are that 1) it is possible to propose a set of categories of code-switching that better describes and accounts for the data in Spanish-English Internet blogs; 2) most bloggers will use Spanish as a Matrix Language (ML) and English as an Embedded Language (EL); and 3) the most frequently occurring type of code-switch is lexical item switches. Modeled after a study conducted by Montes-Alcalá (2007) in which she analyzes code-switching in 15 Spanish-English bilingual blogs over the course of a year, the present study also defines categories of code-switching practices. Montes-Alcalá (2007) proposes seven categories of socio-pragmatic functions for code-switches including lexical items, triggered switches, quotes, elaboration, tags, emphatic, and free. These categories are defined in detail in Chapter 2. An additional goal of the present study is to determine the differences between code-switching and borrowing in the blogs in light of the fact that phonology cannot play a part in the determination of types of code-switches in this communication medium, even though phonology has proved useful in distinguishing between borrowing or single-word code-switches in studies of spoken language (Clyne 2003, Lipski 2005, Poplack et al. 1988).

1.2 Why study blogs?

Information age technology has made it possible for people all over the world to communicate quickly and freely. Due to the increase in access to the Internet world-wide, researchers have focused attention on language use on the Internet in recent years. However, this
is a relatively new focus for linguistic studies, and it may be years before we can fully understand the impact that the Internet has had on communication. According to Merchant:

Computer-mediated communication makes rapid combinations of sound, image and written text possible: web-pages can include images, sound files and animated written text and most e-mail systems allow these to be transferred between users as attachments. As well as this, the emergence of new linguistic conventions is very much in evidence. (Merchant 2001:295)

The very nature of e-language facilitates the use of novel forms and shortenings. Users must become proficient in this new dialect to communicate effectively using the nuances of the language of chatrooms, message boards, and blogs. This can be seen almost universally when one user asks another what a particular initialism, such as LOL (laughing out loud) or IMHO (in my humble opinion), or an acronym means. It is so much a part of on-line discourse that one need only type a particular initialism into any search engine to find a quick reference. Web sites, such as <urbandictionary.com>, have become popular resources for defining new terms, which are often computer user creations. Ross (2006:41) explains the reason that these linguistic innovations are so prevalent in on-line discourse when he states, “Speedy communication allows less time for careful, organized thought. Partly for this reason, e-mails, text messages and broadcast messages (instant messages, chat messages) are in many ways stylistically more similar to spoken language than more traditional forms.” In other words, the nature of this type of language use is less like written language and more like speech. Because of the unique nature of communication on-line, Internet language is an excellent medium for studies on code-switching.
Cheater (2006) discusses the computer user’s need to use readily-available “prefabricated” phrases that create a unique style of speech. In her study, she demonstrates how “e-English” creates new lexical items, making nouns from verbs, verbs from nouns, and initialisms, such as btw for by the way. Language use on the Internet is a readily-available way to witness linguistic innovations as they occur. Shortis (2001:3) agrees when he states, “Information and communication technology texts show language changing in a microcosm and computers can be used to track these changes in some detail.” This is of particular interest for this project on code-switching, given the increased contact with English by Internet users in general due to the popularity of the Internet. On-line communication has made it possible for speakers all over the world to contact each other easily, quickly, and inexpensively. As a result, the Internet has created new language. English has been described as the lingua franca of the Internet (Crystal 1997, Morin 2006), and user’s contact with English on-line has given rise to new contact situations that previously may not have existed.

Every social group has its own set of lexical items and structures that its members use to show in-group affinity. Shortis (2001) elaborates on this use of jargon in his text devoted exclusively to information and communication technology. There is a common stereotype of people who spend a great deal of time on-line. These people tend to be seen as anti-social or dysfunctional, overly educated “geeks”. This sub-group status easily explains why these users would wish to develop an insider language. Merchant (2001:293) conducted a study that illustrated how “the use of popular electronic communication is resulting in linguistic innovation within new, virtual social networks in a way that reflects more wide-reaching changes in the communication landscape.” This new social network can be seen as a speech community that has a desire to create a group-specific jargon. For this reason computer-mediated
communication is an excellent source for the study of informal language practices that show solidarity.

Blogging provides a unique view into language use. This new genre of computer-mediated communication can be viewed as a cross between written language and speech. Originally called *weblogs*, blogs are “frequently updated websites where content (text, pictures, sound files, etc.) is posted on a regular basis and displayed in reverse chronological order” (Schmidt 2007:1). According to Blood (2000), the term *weblog* was first used by Jorn Barger in December 1997. At the beginning of 1999, Jesse James Garrett compiled a list of websites that were weblogs. On his page he counted 23 known websites of this type. Suddenly, this form of communication grew, and by July of 1999, Pitas launched the first free tool that one could use to build a weblog. Because of the creation of this tool, more people became interested in authoring blogs, not just for specific work-related purposes, but also as a type of journal. Later that year, other software developers produced similar software. With the growing number of creators of weblogs, the name was quickly shortened to *blog* (Blood 2000). Since the inception of free blogging software, the number of blogs has become too numerous to count. Now, websites exist for the sole purpose of organizing blogs. One such website is &lt;puertoblogs.com&gt;, which is a list of blogs written by those who choose to identify with Puerto Rico. This website organizes blogs by topic, such as education, technology, politics, and personal; and, it has 1177 blogs that are broken down into 16 different categories.

Why would one choose to write and maintain a blog? Nardi et al. (2004) studied the motivations for blogging by interviewing 23 people aged 19 to 60. They discovered five major reasons people blog: documenting one’s life; providing commentary and opinions; expressing deeply-felt emotions; articulating ideas through writing; and forming and maintaining
community forums (43). They also found that because of the versatility of the medium, the content was as equally diverse as the motivations themselves. This preliminary examination into blogging is an important step in understanding the very the nature of Internet communication. For this particular project, I chose to analyze blogs from the personal section because the language used would likely be informal. Additionally, while Spanish-English code-switching has been studied formally for more than three decades from diverse linguistic perspectives (e.g. sociolinguistic, syntactic, psycholinguistic, neurolinguistic), there is much left to explore. According to Lipski (2005:1), “there are, however, some phenomena that do not fit easily into current typologies.” Thus, typologies of uses need to be revised and tested to account for more code-switching data. Furthermore, by analyzing data from blogs, we can tap into a relatively unstudied language genre.

1.3 Why Puerto Rican blogs?

Since the Spanish-American war, the English language has had an impact on the language of Puerto Rico. According to Lipski (1994), since the first United States administration in Puerto Rico, all education was mandated in English. Even though the early educational system was conducted in English, this change in language failed and lead to a collapse of the school system. Spanish was eventually recognized as the “de facto” language, but the influence of English was, and continues to be, undeniable. Because of the early influences of an American administration, English lexical items were incorporated into the lexicon. The early political control of Puerto Rico is not the only reason why English has influenced the language of Puerto Rico. “Today, the results of earlier attempts at implanting English in Puerto Rico are far outshadowed by the effects of commercial advertising, television and cinema, and Puerto Ricans with only a minimal knowledge of English use a large number of Anglicisms” (Lipski
1994:330). Additionally, the ease of travel between the mainland and Puerto Rico has kept the linguistic contact much stronger. It is the unique situation of Puerto Rico that makes this group of speakers excellent for studies on code-switching.

1.4 Definitions

Two of the major processes involved in discussions regarding language mixing are code-switching and borrowing. According to Thomason (2001), code-switching is by far the most studied result of language change resulting from contact. Code-switching is one of the ways a speaker may express the reality of being bicultural by articulating and acknowledging, both languages and cultures within a single utterance. While the distinction of what constitutes a borrowing and what is a code-switch is still openly debated in the literature, I use the criteria proposed by Myers-Scotton (1992:22), who regards relative frequency of occurrence as the most important factor in making the distinction. Poplack et al. (1988:220) also consider frequency, along with phonological integration, as the determining criterion for differentiating code-switches from borrowings. In an attempt to disambiguate the differences, Myers-Scotton (1992:36) states: “it is not that a B[orrowed] form must recur, it is that a C[ode] S[witched] form must not recur in order to be a CS form.” In other words, Myers-Scotton maintains that a code-switched form is not repeated. Therefore, a singly occurring form may be considered a code-switch, but when an embedded language form is repeated it counts as a borrowing.

Thomason (2001:132) defines code-switching as “the use of material from two (or more) languages by a single speaker in the same conversation.” Many theorists debate what segments can be included under the term code-switching. Some linguists focus only on switches that occur intersententially, while others look at sentence-internal switches. Myers-Scotton (1993b) proposes the constituent as the unit of analysis in her Matrix Language Frame model, rather than
the clause or the sentence. This structure allows for single-word switches to be included in code-switching analyses. Despite the discrepancies in what is considered code-switching, it is clear that code-switching is distinct from borrowing. Clyne (2003:71) points out that with borrowing only a single word is used from another language and morphological and/or phonological integration is likely. What is of particular interest in this study is the fact that, in the context of written blogs, phonological integration is no longer an audible factor in determining the difference between code-switching and borrowing. To make the distinction between code-switching and borrowing, I turn to definitions provided by Myers-Scotton (1992:21):

Codeswitching involves at least two languages used in the same conversation. Of these languages, one is the M[atrix] L[anguage], the language which sets the morphosyntactic frame for codeswitching utterances … One or more languages may serve as embedded languages, providing both singly-occurring lexemes in constituents otherwise in the ML, and also E[mbedded] L[anguage] islands, constituents entirely in the EL.

Borrowings differ because, although they may have arisen from code-switched forms, they are now a part of the ML mental lexicon, while code-switches remain EL material. Even though code-switches and borrowings differ, they are seen as being on a continuum, according to Myers-Scotton. The major distinction used in the present study for considering whether a single word is a code-switch or a borrowing is frequency. If a term recurred in different blogs, it was considered a borrowing. Typically recurring borrowings appeared frequently, such as blogging related terms like blog and post. In addition, I counted established borrowings (i.e. el parking and el show) and the words that were spelled using non-standard orthography in an attempt to show Spanish phonology imposed on English words, such as los estaits and los monchies. The recurring terms were ‘cultural borrowings’ related to Internet processes. I use Myers-Scotton’s
(1992:28) definition of cultural borrowing: “they stand for objects or concepts new to the ML culture.”

1.5 Organization of project

This thesis is divided into five chapters. Chapter 2 provides an in-depth look into the literature on code-switching as well as on borrowings and describes previous studies in these areas. Chapter 2 also describes studies conducted in the field of computer-mediated communication. Early work on blog research, such as the studies by Blood (2000), Gallagher (2000), Nardi et al. (2004), and Boyd (2006), investigates the nature of a blog as a medium and specifically focuses on answering the question of why one would blog. There are very few linguistic studies on blogs; however, language choice in other forms of computer-mediated communication, including simultaneous chat programs and e-mails, has been investigated. Chapter 3 describes the methodology and data analysis used in this study. It outlines a discussion of the research design and provides a description of the nature of the blogs themselves. After analyzing language choice in general in the Puerto Rican blogs, I propose changes to the socio-pragmatic categories provided by Montes-Alcalá (2007). I also compare other sets of socio-pragmatic categories, including those proposed by Valdés-Fallis (1976:58), Gumperz (1982:75-80), and McClure (1981:80-81), to justify this new taxonomy of uses for the code-switches in the blogs studied. This chapter also situates this study within the larger body of research on code-switching and language choice. Chapter 4 is a presentation and discussion of the results of the specific dataset collected. Finally, Chapter 5 concludes and summarizes the results of the present study and proposes areas of further research.
CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

This chapter outlines major contributions to research on code-switching. Section 2.2 discusses key distinctions between code-switching and borrowing that are relevant to the present study. After describing how code-switching and borrowing differ, I summarize previous research in this area of language use. Section 2.3 provides a brief overview of studies on code-switching in spoken discourse, which has been the focus of the majority of research. Section 2.4 discusses code-switching in written corpora given that Internet blogs are a written medium. Section 2.5 presents code-switching studies on Internet data, most of which have focused on simultaneous chat language or e-mail exchanges. Finally, the last section, 2.6, summarizes the nature of blog language and how this medium of language use can contribute to code-switching research.

2.2 Code-switching versus borrowing

Schatz (1989:132) lists criteria for understanding the differences among borrowings, nonce borrowings, and code-switches. With regard to the use of single word items, the distinction becomes problematic. In an effort to clarify these three processes, Schatz (1989) creates a list of features for each of them. According to Schatz (1989:132), borrowings differ from code-switches in that they require phonological, morphological, and grammatical integration and are used by monolingual speakers for lexical need. Additionally, they are treated as part of the L1 lexicon, i.e. the matrix language. The difference between code-switches and
nonce borrowings becomes less distinct. Nonce borrowings happen in the speech of bilingual speakers, phonological and morphological integration is not likely, and they are not frequently occurring. The list of features that Schatz proposes to describe these three processes (borrowings, nonce borrowings and code-switching) does not enable us to clearly differentiate between ‘nonce borrowing’ and ‘code-switching’. Poplack et al. (1988) define nonce borrowings as those that have not become integrated into the speech of monolingual speakers and therefore do not have a high frequency of use. However, according to this definition and that proposed by Schatz (1989), there is no difference between nonce borrowings and single word code-switches. Given that blog posts are written and phonological integration cannot be evaluated, these criteria are not applicable for distinguishing between code-switches and borrowings in computer-mediated communication.

Myers-Scotton (1992) makes a clearer distinction between code-switching and borrowing, which is more useful for the analysis of Internet data. Her definition of borrowing is broader in the sense that it includes much more than loanwords lexicalized into monolingual speech, and it better describes borrowings used by bilingual speakers. Myers-Scotton’s (1992) definition of code-switching is based on the identification of the Matrix Language: “Codeswitching involves at least two languages used in the same conversation. Of these languages, one is the matrix language, the language which sets the morphosyntactic frame for codeswitching utterances” (19). The distinction here is that the rules that govern code-switched utterances are provided by this Matrix Language (ML), while the other language, the Embedded Language (EL), is inserted into this morphosyntactic frame. Myers-Scotton (1992) draws further distinctions between switching and borrowing with regard to single word information from the EL. In an attempt to distinguish single word items as code-switches or borrowings, we can first
look to Poplack et al. (1988) for their discussion on ‘nonce borrowings’. We find that these are singly-occurring, or rarely-occurring forms that should not be considered code-switches. Myers-Scotton (1992:32) states that these forms show morphosyntactic integration into the ML. However, she points out that the definition of nonce borrowings as a category does not explain the difference between single word code-switches and nonce borrowings. This problem led to Myers-Scotton’s discussion as to why single lexemes may be considered code-switches and are not assumed to be borrowings or nonce borrowings. She uses the criterion of frequency as the most important characteristic to differentiate code-switching from borrowing. Frequency has been counted as a criterion for other studies regarding classification of borrowing versus code-switching (e.g. Schatz 1989, Myers-Scotton 1992, Poplack et al. 1988). Because of the problems stated above, I do not consider the category of nonce borrowings and extend the definition of borrowings to include those forms that are not necessarily lexicalized into the matrix language and are used by bilingual speakers. For the present study, I use frequency of occurrence as the major criterion for separating code-switching from borrowing. If the terms were recurrent in multiple blogs, especially in the area of Internet terminology, I counted them as borrowings. That is to say that if a single lexical form was repeated in more than one blog, I counted it as a borrowing rather than a code-switch. Morin (2006:161) conducted a study on Internet related terms in Spanish language press on-line. In her study she classified Internet terminology as borrowings, and more specifically, as loanwords. Other borrowings found in the corpus were established loanwords and those that attempt to reflect phonological integration through non-standard orthography.

Lipski (2005) studied the insertion of so into Spanish utterances to determine whether this use of an English lexeme was a code-switch or a borrowing. Lipski also provided criteria for
distinguishing code-switching from borrowing. According to his distinction, lexical borrowing is a conscious, deliberate choice that becomes lexicalized and used consistently. Because of this, the origin of the word often becomes lost and it is integrated into the speech of monolingual speakers. Code-switching, on the other hand, is produced by bilingual speakers either deliberately or subconsciously when communicating with other bilingual speakers (Lipski 2005:13). The insertion of *so* does not fit neatly into either category (code-switching or borrowing), because it is spontaneous, can occur in Spanish-only discourse with no other code-switching, and it is not found among monolingual speakers. The insertion of *so* clearly comes from English. When asked directly, the participants signaled that the use of *so* was an English word. According to the criteria for lexical borrowing proposed by Lipski (2005), the insertion of *so* does not fall under the term ‘borrowing’. However, the pronunciation of *so* was often integrated phonologically, with the use of the [o] of Spanish rather than the diphthong of English. This integration does make it more of a borrowing rather than a code-switch. Furthermore, the presence of *so* insertion in many bilingual speakers’ otherwise Spanish discourse makes this lexeme much more like a borrowing than a code-switch. However, the insertion of *so* does not appear among monolingual speakers, and it is because of this that Lipski (2005) deems it a code-switch but still considers it to be a different kind of code-switch. If we redefine borrowing to include more than lexicalized established loanwords, perhaps *so* would more easily fit into the category of borrowings. Other discourse markers are problematic in classification in the present study. According to the frequency criterion, code-switched discourse markers that were repeated in multiple blogs, such *anyway(s)* which appeared 16 times in the corpus, should be considered borrowings, but traditionally in the literature they are treated as code-switches, once again making this classification problematic.
2.3 Code-switching in spoken discourse

Code-switching is most widely studied in conversations due to the informality of naturally-occurring spontaneous speech. When bilingual speakers are communicating with other bilingual speakers, they can feel free to move from one code to the other without necessarily or consciously choosing one form over another. The medium of written language often requires authors to think about the specific structures, and, as a result, they tend to use more formal language.

Blom and Gumperz (1972) introduced the notions of ‘situational switching’ and ‘metaphorical switching’ in order to distinguish types of code-switches. Situational switching is triggered by a change in the situation, such as workplace language and home language. Metaphorical switching refers to changes in the speaker’s language choice when the situation stays the same, but the speaker wants to demonstrate a specific communicative intent. The speaker is not code-switching because the situation calls for a different code, but rather for stylistic or efficiency-related reasons. The interpretation of the speaker’s communicative intent depends upon the association between a particular language or language variety and a specific situation. For the present study, there is no change in situation as the format of blogging keeps the situation the same, and thus metaphorical switching is the focus. Because switching occurs to convey a specific communicative intent, we need to define those intentions through the definition of socio-pragmatic categories.

Auer’s Bilingual Conversation (1984) marked a turning point in the study of code-switching. He questions the way the term ‘situation’ is defined. ‘Situation’ is seen as an interactively achieved phenomenon. Using the framework of ethnomethodology and conversation analysis, Auer argues that participants in conversational interaction continuously
produced frames for subsequent activities, which in turn created new frames. He calls for a conversation-analytic approach to code-switching which would focus on “members’ procedures to arrive at local interpretations of language alternation” (Auer 1984:3). This approach depends solely on the conversational context for interpretation, in that one speaker’s turn influences the language choice of the next so much that it cannot be analyzed outside of the conversational setting. Auer (1984:11) discusses the problem of listing functions of code-switching because such a list cannot cope with the “in-principle infinite number of ways in which language alternation may be come meaningful.” That is, because of the theoretically infinite number of potential functions of code-switches, Auer argues in favor of not establishing categories of code-switching uses. However, looking at the different functions of code-switches can aid in understanding this practice as a whole. The present study describes how code-switching practices are employed in a set of Puerto Rican blogs, and how the uses are the same for the individual bloggers. The categories proposed in the present study are also used by other researchers who have studied conversational code-switching (Valdes Fallis 1976, McClure 1981, Gumperz 1982, Montes-Alcalá 2007). These categories still hold true for the instances of written code-switching found in the blogs that are analyzed in the present study.

One researcher, Valdes-Fallis (1976), conducted a study using recorded conversations to analyze code-switching patterns. She studied language alternation in Spanish-English bilinguals in the southern New Mexico area. After collecting recordings, she transcribed and coded each type of language switch. She divided language switches into 12 types. A description of her categories and definitions, as well as examples from her study, can be found in Table 1 below.
<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Situational switches</td>
<td>related to social role of speakers</td>
<td>No example given</td>
</tr>
<tr>
<td>Contextual switches</td>
<td>situation, topic, etc., are linked to the other language</td>
<td>Pues si, mira, cada vez que iba, bueno, la consulta era eight dollars (66).</td>
</tr>
<tr>
<td>Triggered switches</td>
<td>switches due to preceding or following item</td>
<td>No yo si brincaba en el trampoline when I was a senior (58).</td>
</tr>
<tr>
<td>Switching of isolated items</td>
<td>lexical need?</td>
<td>No, porque mira te subes en el trampoline y brincas en el aire (59).</td>
</tr>
<tr>
<td>Identity markers</td>
<td>stress in-group membership</td>
<td>No example given</td>
</tr>
<tr>
<td>Preformulations</td>
<td>linguistic routines</td>
<td>ese domingo, I might be there, because my friend and her husband are bringing her suegros (69).</td>
</tr>
<tr>
<td>Discourse markers</td>
<td><em>But, and, of course, etc.</em></td>
<td>you know, to go back. Pero the last time he told me I didn’t have to go back unless I flare up (67).</td>
</tr>
<tr>
<td>Metaphorical switches</td>
<td>Obvious stylistic device for emphasis or contrast</td>
<td>I started going like this. <em>Y luego decía,</em> look at the smoke coming out of my fingers…(71)</td>
</tr>
<tr>
<td>Proper nouns</td>
<td>no definition given</td>
<td><em>Aquí en State dicen que es co-</em> (82).</td>
</tr>
<tr>
<td>Quotations and paraphrases</td>
<td>may be contextual or noncontextual</td>
<td><em>Me dijo</em> that if I flare up to go back, you know to go back (66).</td>
</tr>
<tr>
<td>Sequential responses</td>
<td>speakers use language last used</td>
<td><em>Yo hasta empecé a escribir lo que soñaba.</em> Every morning I would get up …(79)</td>
</tr>
<tr>
<td>Symmetrical switches</td>
<td>blend and proportion of language alternation is made to resemble that of other speakers</td>
<td>I cut it so short and I know that if I cut it, then <em>me pudiera</em> (78).</td>
</tr>
</tbody>
</table>

Table 1: Code-switching categories proposed by Valdés-Fallis (1976:58)

After Valdes Fallis (1976) determines which of the twelve groups applied to each code-switch, she finds that, due to the regularity with which speakers code-switched, bilingual speakers have access to both systems thus giving them a “supersystem” made up of both codes (1976:84). Her early study is useful for understanding how code-switching tokens can be categorized according
to use and provides us with a point of departure for establishing an empirically based code-switching taxonomy. However, when attempting to apply her categories to the present study, I found that the definitions were not clearly explained. For example, for her category of “triggered switches” she does not describe what kinds of items serve as triggers. To further understand the categorization, I looked at the examples provided by Valdes-Fallis (1976). These examples, not clearly labeled, were also difficult to understand, and thus they did not illustrate the types of code-switching she proposed. At times, the category that was given was applied to a segment of speech that did not contain a code-switch. When presenting examples the analysis is extremely difficult to follow. For example Valdes Fallis (1976:69) presents this transcription to illustrate some of her categories:

Porque si viene, si van ustedes al bullfight ese día, ese domingo, I might be there, because my friend and her husband are bringing her suegros. Her suegros are from Tennessee y los van a traer y quieren llevarlos a una corrida de toros. So...we...she wrote to me yesterday and asked me to find out if there is a bullfight.

Immediately we can identify a few problems in the analysis. The first token given is labeled as either an ‘isolated term’ switch or a ‘symmetrical switch’. Since the researcher cannot decide which type of switch this example is, it is quite difficult to apply her categories to other data. The second problem is that some code-switched information is not labeled as any type of switch while other segments are labeled as a certain type of switch where there is no language alternation. For example, friend is labeled as a triggered switch, but the information before and after is in English. While, it is useful to see how categories have been proposed to account for code-switching data, it is impossible to use Valdes Fallis’ categories because they are not well defined, nor do the examples clearly show how to apply this taxonomy to specific examples.
Gumperz (1982) also proposed a typology of code-switching functions. Drawing from research he conducted in 1976, Gumperz compared three different language pairings; Spanish-English, Hindi-English, and Slovenian-German. While the categories proposed by Gumperz (1982) share some similarities with those proposed by Valdes-Fallis (1976), there are differences. Gumperz (1982) proposes the six categories defined in Table 2 below.
<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quotations</td>
<td>direct quotation or reported speech</td>
<td>She doesn’t speak English, so, <em>dice que la reganan</em> [sic]: “Si se les va olvidar el idioma a las criaturas” (76).</td>
</tr>
</tbody>
</table>
| Addressee specification        | Switch directs messages to one of several possible addressees | Hindi-English  
A: Sometimes you get excited and then you speak in Hindi, then again you go on in English.  
B: No nonsense, it depends on your command of English.  
*Kэн hai bai?* (Who is it? Directed to a third participant) (77). |
| Interjection                   | interjections or sentence fillers               | *Andale pues. And do come again. Mm?* (77).                                                                                           |
| Reiteration                    | repetition or paraphrase                        | The three old ones spoke nothing but Spanish. Nothing but Spanish. *No hablaban ingles*[sic](78).                                    |
| Message qualification          | qualifying constructions such as sentence and verb complements or predicates following a copula | The oldest one, *la grande la de once años* (79).                                                                                     |
| Personalization vs objectivization | the distinction between talk about action and talk in action, the degree of speaker involvement in, or distance from, a message | …they tell me “How did you quit, Mary?” I don’t quit I …I just stopped. I mean it wasn’t an effort that I made *que voy a dejar de fumar por que me hace daño o* this or that uh-uh. It’s just that I used to pull butts out of the waste paper basket yeah. I used to go look in the… *se me acababan los cigarros en la noche. I’d get desperate y ahí[sic] voy al basarero [sic] a buscar, a sacar; you know* (81). |

Table 2: A description of code-switching categories proposed by Gumperz (1982:75-81)

Like Valdes-Fallis (1976), Gumperz includes the category of quotations, but he includes both code-switches for direct quotations and for reported speech. Gumperz describes the category of ‘personalization versus objectivization’ as follows:
The code contrast here seems to relate to such things as: the distinction between talk about action and talk as action, the degree of speaker involvement in, or distance from, a message, whether a statement reflects personal opinion or knowledge, whether it refers to specific instances or has the authority of generally known fact. (Gumperz 1982:80)

An example of this type of switch provided by Gumperz is the following:

A: ….they tell me “how did you quit, Mary?” I don’t quit I... I just stopped. I mean it wasn’t an effort that I made que voy a dejar de fumar por que [sic] me hace daño o this or that uh-uh. It’s just that I used to pull butts out of the waste paper basket yeah. I used to go look in the … se me acababan los cigarros en la noche. I’d get desperate y ahi voy al basarer [sic] a buscar, a sacar, you know (Gumperz 1982:81).

According to Gumperz (1982), the use of Spanish is for a more personal connection to the information expressed in that code, while English is used to show more distance from the topic. Gumperz says that English is used for talking about the problem, and Spanish is used for acting out her problem.

McClure (1981) also conducted a study in which she recorded both naturalistic and manipulated conversations; she included tasks that elicited sentence completion and sentence repetition. From her study on code-switching in Spanish-English bilingual children, McClure devised her own set of categories. Code-switching took place for quotations, addressee specification, emphasis, clarification, elaboration, focus, attention attraction or retention, parenthesis or personalization versus objectivization, and topic shift (1981:80-81). A description of the categories proposed by McClure (1981), as well as examples from her study, are listed in Table 3 below.
<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quotation</td>
<td>direction quotation</td>
<td>Then Michael told Don Pablo, ¿Sabes quién es la novia de Hector? (81)</td>
</tr>
<tr>
<td>Addressee specification</td>
<td>if a speaker switches addressees and the relevant characteristics of the new addressee are different from those of the previous addressee, a language switch is common</td>
<td>Pregúntale a Patty. Pregúntale. (To Rosa) Wasn’t I at the house? (To Patty) (81)</td>
</tr>
<tr>
<td>Emphasis</td>
<td>commands or repetitions; emphasis applies to entire sentence</td>
<td>Who is that? ¡Pégalo! (82)</td>
</tr>
<tr>
<td>Clarification</td>
<td>repetition to resolve ambiguity or clarify</td>
<td>I: ¿Dónde fuiste? R: A school. I: Huh? R: A escuela. (82)</td>
</tr>
<tr>
<td>Elaboration</td>
<td>no definition provided</td>
<td>Yo lo puedo quebrar. Yo lo también…lo pu(edo) quebrar. I can break this easy with my nose (82).</td>
</tr>
<tr>
<td>Focus</td>
<td>brings into prominence a constituent within a sentence</td>
<td>Este Ernesto, he’s cheating (83).</td>
</tr>
<tr>
<td>Attention attraction/retention</td>
<td>code-switch used as a device to attract or retain the attention of his audience.</td>
<td>Now let me do it. Put your feets down. ¡Mira! It’s Leti’s turn again (83).</td>
</tr>
<tr>
<td>Parenthesis/Personalization vs objectivization</td>
<td>rhetorical request or direct address</td>
<td>...respiran las llantas del tren, y...That’s all I could think (84).</td>
</tr>
<tr>
<td>Topic shift</td>
<td>stylistically done to mark a desired change in topic</td>
<td>T: Dile que es una casa sin techo. L: We have a pretty, uh, Christmas tree (85).</td>
</tr>
</tbody>
</table>

Table 3: A description of code-switching categories proposed by McClure (1981:81-85)

McClure’s categories are similar to those proposed by Gumperz (1982), in that both typologies include quotation, addressee specification, and personalization versus objectivization. Also, two of their categories describe the same function, but have different names. McClure’s category of ‘emphasis’ includes repetition while Gumperz includes repetition in his category of ‘message reiteration’. Additionally, both Gumperz’s ‘message qualification’ and McClure’s ‘elaboration’
are used to provide additional information. The similarities in these categories can be seen clearly in the examples provided by both researchers. Gumperz (1982:79) presents the following example to illustrate his category of ‘message qualification’: “The oldest one, la grande la de once años.” This code switch is giving additional information about the oldest child’s age. McClure (1981:82) provides an example for her category of ‘elaboration’: “Yo lo puedo quebrar. Yo lo también ... lo pu(edo) quebrar. I can break it with my nose.” Here the code-switched material gives information about how the speaker can break something. Both categories illustrate the same type of function, but have different names. These discrepancies in terminology make it difficult to apply previously proposed typologies. However, aside from the problems with the categorization of code-switches, an important finding of McClure’s study was that the categories of code-switching use were the same for both children and adults. Prior to this, many studies focused on adult speakers, and thus McClure adds a layer of understanding to code-switching research by looking at the speech patterns of bilingual children. McClure shows that children’s code-switching patterns are governed by functional and grammatical principles, just like adults’ patterns. Additionally, bilingual children follow a developmental pattern of syntactic control of code-switching. However, upon examining her typology, I found that three of her categories (focus, emphasis, attention attraction or retention) refer to the same use. Her examples for ‘emphasis’ and ‘attention attraction or retention’ do not show a clear difference in the categories. She gives the following as an example of ‘emphasis’: “Who is that? ¡Pégalo!” Then, she lists this example of ‘attention attraction or retention’ that appears to use code-switching in the same way as her category of ‘emphasis’: “Now let me do it. Put your feets down. ¡Mira! It’s Leti’s turn again” (McClure 1981:82-83). The code-switched command in each sentence does not clearly show a difference in function. There is not a clear distinction in
the categories illustrated by the examples given. When McClure discusses the distinctions among these categories, she states that ‘emphasis’ is a repetition in translation and that ‘attention attraction or retention’ is a repetition in translation that has the impact of a paraphrase. Clearly, this parameter does not enable us to distinguish clearly between these two categories.

The three studies discussed above were key investigations conducted on code-switching that outlined socio-pragmatic categories and served as the basis for the categories later proposed by Montes-Alcalá (2007) in her research on code-switching in Spanish-English blogs. For the purposes of the present project, I compared the socio-pragmatic categories proposed by all four researchers to refine the categories proposed by Montes-Alcalá (2007). While most code-switching research focuses on spoken data, blogging is a written medium, and thus it is important to examine studies on written language.

2.4 Code-switching in written corpora

As pointed out in section 2.3, the majority of the research on code-switching focuses on conversational spoken discourse due to the nature of language switching. This phenomenon occurs in more informal settings, which interactive spoken data naturally facilitates. However, some studies (e.g. Albakry et al. 2008, Martin 2005) have been conducted using written corpora to illustrate how authors use code-switching as a literary device. While Lipski (1985:73) proposes that the dichotomy between spontaneous speech and written documents is an important distinction, he goes on to state that the analysis of written texts has value in its own right: “An analysis of written code switching may be of great value in tracing the psychological variables that come into play, and promises to provide a broader perspective of the affective values of language mixing” (73). By looking at how authors choose to represent characters by using “authentic” speech, we can see how code-switching is used to achieve this. Albakry et al. (2008)
is one such study that looks at code-switching in a novel written by an Egyptian-British author. This work looks at language switching that occurs between Egyptian Arabic and English. Written in English, the novel relies on code-switching into Arabic as a literary device. Albakry et al. classify switches into the following linguistic types: ‘traditional honorific titles and terms of respect’, ‘references to customs and tradition’, ‘historical references’, ‘greetings and conversational formulas’, ‘inter-language dialectal variation’ and ‘translational transfer’ (2008:223). These terms are specific, culturally-bound items, such as terms for local food and dress. While this study does classify code-switches into classes of use, it does not give specific information about which switches are more common than others, nor does it draw larger conclusions based on the analysis.

According to Albakry et al., the use of code-switching in the novel was a way for the author to convey social and cultural elements to the reader. These researchers propose that the code-switching created by the author was a way to find a “new English” that could embrace both cultures of the author (2008:233). Martin (2005) looks at code-switching in ethnic literature in the US as a way for authors to add an extra dimension to their characters. By adding code-switched forms, authors have chosen to promote the validity of their heritage language. Martin (2005) investigates code-switching in literary texts that included the following language pairs: Spanish-English; English-Chinese; and English-Jemez, a Native American language. While Martin’s (2005) focus is on the use of literary code-switching as a way to demonstrate bicultural identity, she does talk about some code-switching strategies that are similar to those outlined by McClure (1981) and Gumperz (1982). Like these researchers, Martin (2005:406) discusses the author’s intention to emphasize information by code-switching. The juxtaposition of a language other than English in an otherwise English-language novel, by nature, makes that information
more salient. Other uses of code-switching discussed by Martin (2005:406-408) are for repetition and as a way to connect or distance one’s self from the information. Although Martin (2005) does not explicitly state that code-switching for this last purpose falls into the category of ‘personalization versus objectivization’ proposed by both McClure (1981) and Gumperz (1982), it is clearly an example of this type of code-switch. This type of code-switch occurs when speakers wish to demonstrate personal involvement through language choice. Martin (2005) discusses Anzaldúa’s (1987) use of Spanish to discuss family members and her Mexican pueblo.

*Cuando Carito, mi hermanito,* was missing in action and, later, wounded in Viet Nam [sic], *mi mamá* got on her knees and *le prometió* [sic] *a Ella* que *su hijito* volvía *vivo* she would crawl on her knees and light novenas in her honor. (Anzaldúa 1987:29-30)

Family and ethnicity are tied to Spanish for Anzaldúa, and thus she expresses information about them in Spanish because it is a more personalized language choice.

Many of the studies on literary code-switching do not focus on defining the socio-pragmatic uses of code-switches. However, Callahan (2002) uses a linguistic perspective in her analysis on literary code-switching. Callahan (2002:2) argues that there has been a “tendency to discount written code-switching as artificial.” However, the data in her study using written texts show the same syntactic patterns as oral code-switching. Departing from the idea of analyzing literary code-switching as a tool to enrich character development, Callahan (2002) focuses on the functions of code-switching in Chicano literature using the same framework that is used by researchers who describe oral code-switching. Callahan (2002) uses Myers-Scotton’s Matrix Language Frame model to investigate whether a model proposed for speech could also be applied to written data and to see whether modifications to the model were needed to account for written data. The data consisted of an analysis of 30 texts: ten short stories and nine novels,
which used English as the matrix language, and eight short stories and three novels, which were written in Spanish. Callahan (2002) finds that written code-switching is constrained by the same syntactic rules that constrain spoken code-switching. Written code-switches are used to mimic what others do in speech, so it is not strange that the rules that govern such switches be the same in written and oral discourse.

Instead of examining literary code-switching, Montes-Alcalá (2001) focuses on written code-switching found in journals. Her study shows that code-switches are not only governed by grammatical and social rules, but also are dependent upon a proficiency in both languages. Montes-Alcalá (2001) aims to capture more natural, less contrived language structures by analyzing her own journal entries for the period of a year and a half. She then categorizes the types of code-switches according to their socio-pragmatic functions whether situational and metaphoric. According to Montes-Alcalá (2001:196), situational code-switching clarifies the message when one speaker does not understand, and the other must switch to another language to facilitate comprehension. Her definition of ‘situational’ differs from that of Blom and Gumperz (1972), who define situational switching as language alternation that occurs depending on the type of situation. However, Montes-Alcalá sees ‘situational switching’ as a tool to clarify lack of comprehension on the part of the audience. Metaphoric code-switching is emphatic or contrastive; therefore, it is this type of switch that is the focus of the study by Montes-Alcalá (2001). What Montes-Alcalá (2001) contributed to the field of code switching in this study is her application of these ideas to a non-literary written corpus. Montes-Alcalá (2001) posits that the pragmatic functions of written code-switching do not exactly mirror those of oral code-switching. This may be due to the nature of written journals and the absence of interlocutors with whom to converse. For example, conversational strategies are no longer a factor, because
journal entries do not require the interaction of speakers. Text type is essential in understanding written forms of code-switching. Code-switching as a literary device is constructed by authors in an attempt to mirror natural language, but it is in no way natural.

2.5 Code-switching and Internet data

Why are Internet data of interest to linguists and relevant to code-switching? With the increase in popularity of Internet use, more researchers have become interested in communication through this medium (Crystal 2001, 2006; Cumming 1995; Herring 2001; Danet and Herring 2007). Computer-mediated communication is an area of relatively new focus for linguistic study with the rise of technology and its use in global communication. Looking at code-switching strategies in this new medium can give us insight into how code-switching is employed. Blogs, however, can be journal-like and do not involve the interaction of interlocutors, and thus the language strategies employed by a particular blogger are not directly influenced by other speakers. To better understand the nature of blogging it is critical to look at how language is used on-line and how it has been studied. Crystal (2001, 2006) has conducted extensive research in the area of language use on the Internet. He begins by explaining that the Internet, developed in the USA in the 1960s, is the world’s largest computer network with over 300 million hosts connected by the year 2005 (2006:3). This network enables users to send messages to any other registered user globally. The array of services on the Internet is growing quite rapidly and ranges from business to personal use.

When Crystal (2006) discusses the language used on the Internet, or ‘Netspeak’, he describes it as intermediate form of language that falls somewhere between written and oral styles. He finds the interaction between writing and speaking that defines Internet language the most interesting thing about this linguistic style. On-line language has features of both written
language and speech. Crystal (2006:28-30) highlights the differences between speech and written language using seven points of contrast. First, speech is necessarily temporal and involves people who are present in some way; the intended audience is clear. Written language differs from this greatly because it is permanent. The audience is not present, nor does the writer necessarily know who will read the message. Second, speech is synchronous, while writing is asynchronous. Third, extra-linguistic factors play a key role in the interpretation of speech, and deictic expressions are often included, whereas writing cannot depend on extra-linguistic cues, and temporal expressions are confusing and are therefore avoided (Crystal 2006:28). Fourth, because of the temporal nature of speech, speakers have the freedom to include invented vocabulary, obscenities, and slang. Writing, because of its permanence, tends to have more standard, non-contracted forms, and obscenities are often written using a graphic euphemism by replacing letters with symbols after the first letter of the word (29). Fifth, discussions on daily activities and the weather are usually expressed in speech, as are social relationships and opinions. Speech is especially good for these phatic uses of language because of the subtleties that are able to be articulated through prosody, tone, and facial features. In other words, speech is an excellent way to establish social bonds without focusing on conveying information. Writing is more suited to recording facts and communicating ideas given that charts and graphs can organize information and ideas. Sixth, speech allows for reformulation of ideas, but errors cannot be taken back. Writing allows for editing of errors, but a final product cannot be changed or reformulated instantaneously. Seventh, speech carries unique features that are not easily translated into the written word. Intonation, volume, rhythm, and pauses are difficult to represent faithfully in written text. Writing carries its own unique features such as capitalization,
spacing, and punctuation (30). It is easy to see how on-line language, ‘Netspeak’, falls into both categories, especially in light of the fact that it may be synchronous or asynchronous.

The informal nature of Internet language, as well as the presence of emotional symbols, makes it resemble spoken language in some ways. However, Crystal (2006:32-33) identifies several differences between face-to-face interaction and Internet language. Face-to-face interaction allows for speakers to interrupt and change the conversation and the interaction while it takes place. Internet communication requires that a participant type his or her entire message before hitting send, and as a result, the recipient receives the message as a whole. Also, there is no way for a participant to know, mid-message, whether the ideas in the message are clear to the recipient or whether some repair or further explanation is needed. To gauge the success of the message, one must wait for the response. Crystal (2006:33) points out speed as a second major factor that distinguishes Netspeak from face-to-face conversation. The Internet is dependent upon technology for the transmission of the desired message, which means that the speed of the computer and Internet connection and the speed with which the recipient reads and responds to a message are involved in how quickly the conversation progresses. There are multiple factors involved in determining the lag in a computer-mediated conversation. Participants may not receive the message immediately, as they may have navigated away from the page, left the room to complete another task, or been interrupted by another task. These distractions are not usually present in face-to-face interaction, which does not depend on technology. Also, during face-to-face interaction, interlocutors would not participate in other activities that could be viewed as rude during a conversation. However, Netspeak is not exactly like other forms of written language, in that certain conventions have been invented to make up for those features of speech that are not normally present in writing. Some of these features include the extra-linguistic
asides written parenthetically or between asterisks to show physical action or responses to the conversation, for example *sigh*. Other special conventions are the use of emoticons to show emotion, or even sarcasm to help the recipient interpret the message, such as ;), the use of all capital letters to show a rise in volume, and the spacing of letters to show speed, to name a few.

These written conventions of language use are not the only communication strategies that separate computer-mediated communication from other forms of communication. However, the differences in face-to-face interaction and Internet language can give bloggers the freedom to use language structures without considering an interlocutor. The relative anonymity of the Internet and the delay in message receiving give a speaker the option to write without a specific audience and to use structures unaltered by the presence of another speaker. Face-to-face communication is often considered more natural language use, but it is always a co-constructed event, such that the involvement of speakers is influential on the actual linguistic structures used. And, while written language is often considered more contrived because of the option to edit a message, language use on-line is distinct because Internet users often write messages without using this option.

Additionally, language use on-line is remarkable because it gives one access to a previously unavailable community of speakers. The notion of multilingual communication is particularly salient today because the language of the Internet has been, until now, primarily English, but as new coding technologies have evolved, even complicated character-based languages can easily be represented on-line. This rise of multilingualism on the Internet has sparked interest in the field of language contact, and until now has led to studies on code-switching practices among bilingual Internet users. Earlier studies on code-switching have dealt primarily with conversational data, and the few studies on code-switching and Internet language
are based on chatgroups or e-mail. One of the studies that focuses on code-switching and chatgroups is Siebenharr (2006), which analyzes data from Swiss-German Internet Relay Chat (IRC) rooms. Due to the informality of Internet language, on-line communication appears to promote the use of language varieties that had previously been absent from written forms of communication (482). As a result, on-line language use, though written, does not behave like other forms of written media. Siebenharr (2006) investigates language use in synchronous forms of IRC in a non-English speaking on-line community. Siebenharr (2006) uses Auer’s (1998) approach to code-switching, which posits that code-switching should be interpreted as verbal action. “Since chatting comprises numerous features of oral communication, code-switching terminology can be applied to the study of IRC [Internet Relay Chat]” (Sienbenharr 2006:487).

Code-switching is used by Sienbenharr to refer to communicatively and socially meaningful code alternation. Sienbenharr (2006) also points out that, while code-switching on the Internet can be analyzed using the same frameworks applied to spoken data, it does not retain the same function in IRC as it does in face-to-face communication. However, since code-switching can be seen and interpreted by all users in IRC, it still retains much of the social and communicative import that it has in oral communication. The corpus used in Sienbenharr (2006) is comprised of chat logs recorded from 2002 to 2005 run by <bulewin.ch>, a Swiss IRC network with several regional channels (489). Two types of channels were recorded: regional channels that often use the Swiss-German dialect and supraregional channels that use Standard German generally. Sienbenharr (2006) finds the greatest number of code-switches on the supraregional channel, since its speakers often switched from Standard German into Swiss-German. However, the channels that were regional and therefore used primarily Swiss-German showed fewer code-switches into Standard German. Sienbenharr (2006) posits that this is due to the younger
speakers’ desire to only use Swiss German on-line and that the Standard German they learned in school is not used in this linguistic arena. Also, the older speakers tend to switch into Swiss German from Standard German in the chatrooms, while those that more often use Swiss German as the matrix language do not code-switch as often.

Sienbenharr (2006) discusses the importance of directionality in code-switches. Code-switches from Swiss German to Standard German mark asides that are not relevant to the main discourse, and thereby mark the information in Standard German as less important. The tendency to use Swiss-German to convey important messages is supported by Siebenharr (2006), since his findings suggest that those who start out in Swiss German tend not to switch. While Sienbenharr (2006) discusses the lack of research on written forms of code-switching, it is also important to note that the language of the Internet especially that of simultaneous chat, functions much more like spoken discourse than the standard forms of written language.

Hinrichs (2006) has done similar work on code-switching and Internet language, but rather than focusing on the obvious choice of IRC, his work focuses on the language of e-mails. E-mail is not synchronous like chat, but retains the conversational element of exchange. However, the messages are written without much planning and are seen as spontaneous even though the exchange may occur over days or even weeks. Hinrichs (2006) states that, although, the language of e-mails is often spontaneous in principle, it can be thought out and planned. In addition to the time the exchange takes, the option of planning and editing makes it different from simultaneous chat. Nevertheless, the classification of e-mail language becomes problematic as the range of uses has become broader. Language use in e-mails is “less regulated by formal and pragmatic norms than in other written text types, and that it is a choice for users to make whether they wish to adhere to the norms established by paper-based writing” (Hinrichs
2006:19). It is, therefore, more likely for informal language to be used in e-mails than in other forms of written text not used on-line.

To conduct his study, Hinrichs (2006) compiled a primary corpus of e-mails written by Jamaican students, and then created a secondary corpus of postings to Jamaican discussion forums. Data were collected during February and March of 2002 and then March and April of 2003. Hinrichs (2006) found that the preferred language choice, or matrix language, for Jamaican e-mails was English. Code-switching to Jamaican Creole was the marked choice, the unexpected form, in terms of Myers-Scotton’s Markedness Model, whereas, in spoken data Jamaican Creole is the matrix language. Thus, code-switching is more expected as a conversational strategy and is less expected in writing. Hinrichs (2006:139) explains that this shift is medium related. English is the language of literacy in Jamaica, and because e-mail requires written forms, there is a cognitive cost associated with using Jamaican Creole in written language. Because of the cognitive cost, English is the preferred language for written discourse; therefore switching from English, the unmarked choice, to Creole, the marked choice, is less common in e-mails than in spoken discourse.

Hinrichs’ (2006) findings support the most recently updated version of the Markedness Model (Myers-Scotton and Bolonyai 2001), which says that speakers try to optimize the relationship between costs and communicative outcomes. While his research is the most comprehensive study on code-switching and Internet language, it is still specific to the situation of Jamaica. Studies conducted on languages that share equal status of literacy would probably yield far different results. It is for this reason that it is important to look at other language combinations and computer-mediated communication. Also, it is necessary to look at other forms of language use on-line. Studies have been conducted on the language of chat and e-mails,
but blogging is so new that very few studies have been conducted using this particular Internet language situation.

One of the only studies on code-switching and blogging comes from Montes-Alcalá (2007). Perhaps the medium of blogging has been overlooked not only because it is relatively new as a form of communication, but also because it lacks the conversational component upon which much research on code-switching depends. When Crystal (2006) was preparing the original 2001 version of his book dealing with language and the Internet, blogs were in existence, but they did not reach popularity until later and required a more detailed description, which he included in the updated version of this work on Internet language. Five years after the original publication of Crystal’s work, blogs are everywhere, and numerous terms related to the nature of blogging have appeared (238). According to Crystal (2006:239), blog was declared “word of the year” by Merriam-Webster in 2004. The style of blog language can be uniquely informal: “Here we have examples of a style of writing which has never been seen in a public, printed form, outside of literature, and even there it would take an ingenious novelist indeed to capture its innocent spontaneity and unpredictable thematic direction” (Crystal 2006:244). This is largely due to the lack of a mediator or editor to regulate the content of a personal blog. People who write blogs are free to use whatever style of language they wish.

Schmidt (2007) looked at blogging practices from an analytical framework. As Schmidt suggests, if we consider blogging as a particular framework, we must look at the practice as it fits into social interaction in general:

In order to fully explain aspects of social life, one has to connect the micro-level of individual action and the macro-level of social structures by explaining the mechanisms of how the macro-level actions are in turn (re)producing the macro-level structures.
Applied to the phenomenon of blogs, this approach leads to the idea of blogging practices, which in the most general sense consist of individual episodes in which a blogger uses specific software to attain specific communicative goals. These episodes are framed (but not solely determined) by three structural elements: rules, relations, and code. (Schmidt 2007:3)

It is this last element, the use of code, that is, language choice, which motivates the present study. This new medium is an exciting area of study specifically for code-switching practices because the user does not depend solely on shared knowledge between interlocutors. Often, a blogger writes what s/he wishes to express rather than concerning him or herself with the particular audience, especially in the area of personal blogs that have no set topic.

Montes-Alcalá (2007) studies code switching using bilingual blogs. The first question addressed in this particular study is whether or not bloggers would code-switch freely when writing a public journal. Secondly, it attempts to explain why they might do so, and hypothesizes that bloggers’ code-switched forms would display the same social functions of code-switching that oral discourse research has shown. In her study, 15 Spanish-English bilingual blogs are analyzed over the course of a year, from January 2005 to January 2006. Due to the nature of blogging, the length and number of posts vary greatly among the bloggers. Montes-Alcalá (2007) finds that bloggers do indeed code-switch quite often and that the social stigma that code-switching carries in oral production does not seem to apply to the informal written format of the Internet. The seven socio-pragmatic functions of code-switches Montes-Alcalá (2007:166) proposes are: lexical items, trigger, quotes, elaboration, tags, emphatic, and free. Definitions of each category and examples from Montes-Alcalá (2007) are provided below in Table 4.
Lexical items include those that are single words or compound noun phrases that are switched for reasons of biculturalism rather than a lack of proficiency. She found that the category of lexical items was the most prolific and accounted for 27.6% of the tokens in her study. This type of switch is due to a lack of an exact lexical equivalent in one language or the other. There were a great number of switches for kinship and family terms. Triggered switches accounted for the fewest number of switches with 6.2% of the total. Triggered switches are those that are triggered by a switched word, such as proper nouns, borrowings or lexical item switches. The category of ‘quotes’ proposed by Montes-Alcalá (2007:167) are those switches that occur to quote someone else’s words, but she found that this category represented only 8.5% of the switches in her corpus of blogs. Elaboration was the fourth most common group with 9.5% of the switches. These
switches occurred to give more information to further explain an idea without disrupting the flow of the sentence. The category of ‘tags’ accounted for 9.3% of the changes. This group included idiomatic expressions, identity markers, linguistic crutches, and sentence fillers. The category was especially problematic because it was not only ill-defined, but it also included a set of unrelated structures that had very different uses. Because of this, it was difficult to use this category in the present study. Emphatic switches were widely found in Montes-Alcalá’s study comprising 24.3% of the total number of tokens, which is the second most frequently occurring type of code-switch. Emphasis is created by either just a code-switch or a code-switched repetition. The last category proposed, free, accounted for 14.3% of switches. This group was proposed for those code-switched forms that did not easily fit into any of the other categories, because there was no single, obvious reason for the code-switch or the switch constituted an eclectic combination of other functions. Montes-Alcalá (2007) says that these switches could be for purely stylistic reasons. However, this category was also inapplicable to the current study because it lacks specific criteria for inclusion. These code-switches did not have a commonality of function and therefore do not constitute a category of use. She further states that code-switching is a strategy employed by bilingual bloggers to acknowledge their bicultural reality.

2.6 Summary

The present study attempts to further the research by Montes-Alcalá (2007) by modifying her seven socio-pragmatic categories and then applying them to Puerto Rican blogs. Additionally, this study differentiates code-switching from borrowing in general terms to better categorize language use in this new medium of communication. Internet language provides a unique situation, different from both the face-to-face interaction of conversations and the intentional production of written discourse. The unique style of blogging can add a different
layer of analysis to previous work on code-switching research. Even though blogging is a written medium, we can expect speakers to be less careful in their language choices than in more formal settings, unless the blog is a professional one. It is this uniqueness that makes Internet data valuable. Also, studying written words presents the unique opportunity to study language contact without considering the phonological component. And, as a result, we are not dependent upon researchers’ aural interpretations of the message for analysis. Additionally, the fact that blogs are readily available, written forms of communication makes the data collection process much easier for the researcher. While blogging is mix of features attributed to speech and writing, it is neither of the two. Nilsson (2003) describes the linguistic nature of blogs though a parent/child analogy. She says that “blogspeak” is the child of a speech parent and a writing parent because this form of communication has features of both, but it is the unique combination that makes it a distinct form of communication. From speech, blog language is time bound, contains prosody, is socially interactive, is spontaneous in that it lacks editing, and is loosely structured (2003:31-32). From the writing parent, blogging language takes the features of being repeatedly revisable, elaborately structured, visually decontextualized, graphically rich, space bound, factually communicative, and contrived (32-33).

Because of the unique format of blogging that is a mix of two traditional media, this language genre provides insight into the nature of language use and code-switching. Previous typologies of code-switching uses have, to date, been difficult to apply to other studies. The focus has been medium driven. By studying a type of language that is a mix of the two, we can better understand how bilinguals use code-switching strategies. Additionally, due to varying terminology and poorly defined categories it is hard to apply these sets of categories to other
types of data. Through the proposal of this new set of categories, we can get one step closer to understanding code-switching practices.
CHAPTER 3

METHODOLOGY AND DATA ANALYSIS

3.1 Introduction

Puerto Rican discourse is a natural choice for looking at the interaction of English and Spanish because of the long history of contact between speakers of both languages, and therefore it is appropriate for a study on Spanish-English code-switching. This variety of Spanish has been influenced by English for over 100 years so it stands to reason that it would also show marked influences of this contact between Spanish and English. For this study, I chose to analyze blogs from Puerto Rican bloggers. The presence of such a blogging organizer, <puertoblogs.com>, made obtaining a corpus for the analysis of code-switching and borrowing within the medium of computer-mediated communication much easier.

3.2 Methodology

The blogging community of <puertoblogs.com> offers 1177 individual blogs as of June of 2009. The very nature of blogging keeps this number in constant flux. These blogs are written by those living on the island of Puerto Rico or by Puerto Ricans living off the island, in the US or in other countries. The blogs themselves are divided into 16 different categories ranging from education, to opinion, to sports. I analyzed the discourse of the blogs in the personal section because I expected that these bloggers would use more informal discourse styles since they are not obligated to write on specific subjects such as art or education. In a previous pilot study on bulletin board postings made by Puerto Rican Internet users, I discovered that when speakers posted messages about specific topics, such as art and education, they were less
likely to code-switch. Additionally, the personal section is the largest section of blogs, as 327 individual blogs are categorized as ‘personal’. Bloggers are given a choice to describe their blogs briefly for the index. This description includes a tag that lists the language of the blog. Bloggers identify their blogs as being in Spanish, English, or Spanglish. For the present study, I analyzed the language of those who have self-identified as users of Spanglish because I expected those bloggers to code-switch.

Of the 327 personal blogs that exist today, only 312 were posted at the time of the data collection. From August 2007 to August 2008, I analyzed the language use of 20 bloggers. I collected the entries posted during this particular year. From the 55 personal blogs that list Spanglish as the language choice, I eliminated those blogs that were removed, private, or had no postings during the year of analysis. After disregarding those blogs, I was left with 20 blogs for analysis. All blogs analyzed are free and open to the public and do not require membership into the blogging community for readership. These 20 bloggers have selected the label of Spanglish for the language of their blogs; however, six were written almost exclusively in English, 13 were written mostly in Spanish, and only one was a mix of the two, in that each entry was first posted in Spanish then translated to English. One blog showed no signs of language mixing at all, as there were no borrowings or code-switches. All 24 blog posts by this particular blogger, who self-identified his blog as being written in ‘Spanglish’, were written in standard Spanish. Thus, this blog was eliminated from analysis, leaving 19 blogs.

After collecting the entries of each blog for a year, I have a roughly 190,000 word corpus. Some blogs have as many as 113 separate entries while others have as few as two. Even though the particular number of posts greatly differed among the bloggers, I decided to keep the analysis limited to one year of postings. Blogging has its own unique style, and the topics covered range
from descriptions of daily life and work to cultural observations to travels. Each blog welcomes comments, which shows that each blogger expects an audience to some degree; however, that specific audience is unknown to the writer. The only expectation that these bloggers may have of their audience is that they understand both English and Spanish, as glosses are not expected to be given for either language and, most often, are not given.

To code the data, I first began by selecting any sequence of words that include English and Spanish. In order to isolate sequences with code-switches or borrowings the length of each sequence was determined by the blogger and not pre-determined. To extract sequences the matrix language was determined, and then I looked for uses of an embedded language. Next, I created a concordance of these sequences. This concordance included the information directly before and after the code-switch or borrowing. The next step was to label switches from one language to another as either an example of code-switching or borrowing. Using Myers-Scotton’s (1992) criteria to determine whether single word items were code-switches or borrowings, I used frequency as the most important determiner. The determination of whether to classify a lexical item as a code-switch or a borrowing was made based on frequency. That is to say if a lexical item was repeated in different blogs even once it was counted as a borrowing. However, every borrowing occurred at least three times throughout the corpus. Most borrowings were cultural borrowings related to Internet terms about blogging itself. Other well established loanwords were also counted as borrowings. After distinguishing code-switches from borrowings, I then categorized the code-switches to determine the socio-pragmatic uses of code-switched material. As I showed in Chapter 2, the seven socio-pragmatic categories proposed by Montes-Alcalá (2007) were problematic; thus, after looking specifically at the particular uses of both code-switches and borrowings, I made alterations to the categories. Because her category
‘tag’ included many types of expressions that were often unrelated, I propose that adding a category for ‘evaluation’ and one for ‘discourse markers’ better describes and accounts for the data. Also, the catch-all category of ‘free’ proposed by Montes-Alcalá was unnecessary because all instances of code-switching found in the present study fit into the revised set of categories that I proposed.

3.3 Data Analysis and Code-switching functions

Linguists have been categorizing the functions of code-switching for almost as long as code-switching itself has been an area of study. As noted in Chapter 2, Valdés-Fallis (1976) proposes 12 types of code-switches, including quotation or paraphrase, discourse marker, triggered switches, metaphorical switches for emphasis or contrast, contextual switches, identity markers, and preformulation (linguistic routines). Similarly, Gumperz (1982:75-80) introduces a typology of code-switching functions. His typology contains only six main categories. McClure (1981:80-81) also provides a taxonomy of code-switches in which she includes nine distinct categories. Four of her categories describe a use that the other three researchers group together. Montes-Alcalá (2007) proposes seven socio-pragmatic categories to describe code-switches found in her corpus of blog data. They include lexical items, triggered switches, quotes, elaboration, tags, emphatic, and free. Looking at each of these categorizations, it is easy to see how they overlap as well as how problematic terminology can be. In an attempt to understand the different terminology proposed for categories, I have compared what each of these four researchers have proposed in Table 5 below.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Situational switches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contextual switches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triggered switches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Switching of isolated items</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identity markers</td>
<td></td>
<td></td>
<td>Tags (identity markers)</td>
</tr>
<tr>
<td>Preformulations (linguistic routines)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discourse markers</td>
<td></td>
<td>Interjections</td>
<td></td>
</tr>
<tr>
<td>Metaphorical switches (emphasis or contrast)</td>
<td>Emphasis</td>
<td>Message reiteration repetition or paraphrase</td>
<td>Emphatic foregrounding or repetition</td>
</tr>
<tr>
<td>Clarification</td>
<td>repetition to resolve ambiguity or clarify</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus</td>
<td>attention</td>
<td>attraction/retention</td>
<td></td>
</tr>
<tr>
<td>Proper nouns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quotations and paraphrases</td>
<td>Quotation</td>
<td>Quotation</td>
<td>Quotes</td>
</tr>
<tr>
<td>Sequential responses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Symmetrical switches</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elaboration</td>
<td></td>
<td>message qualification</td>
<td>elaboration</td>
</tr>
<tr>
<td>personalization vs objectivization</td>
<td>personalization vs objectivization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>addressee specification</td>
<td>addressee specification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>topic shift</td>
<td></td>
<td></td>
<td>free catch-all</td>
</tr>
</tbody>
</table>

Table 5: Comparison of other researcher’s code-switching categories

We can see where some of the categories overlap, and how different linguists give different names to the same type of switch. The fact that some of these categories are poorly defined is not the only obstacle to trying to apply these typologies to other studies; it is also difficult to understand the examples provided. Originally, the present study was conducted to see whether Montes-Alcalá’s (2007) categories would explain other sets of blog data; however, upon attempting to classify the data gathered for this thesis, I encountered quite a few challenges. In
particular the ‘tag’ category grouped together seemingly unrelated functions. Because of this, I propose slightly different categories that attempt to better describe the actual function of the switched item. To address the first hypothesis of this thesis I compared the previously identified categories in order to propose my own set of code-switching uses. I propose that the following categories best account for the data from the present study: lexical item switches, triggered switches, quotations, elaboration, evaluation/emphasis, and discourse markers.

3.3.1 Lexical items

Single lexeme switches are included in the first category proposed by Montes-Alcalá (2007) under the term ‘lexical items’ in her socio-pragmatic taxonomy of code-switching. This category includes single word switches, such as nouns or adjectives, as well as those compound nouns that function as a single unit, such as *socker*[sic] *mom*. Unlike Gumperz (1982), Valdés-Fallis (1976), and McClure (1981), Montes-Alcalá sees this as an important category. What makes this category especially difficult to describe, or perhaps the reason it has been ignored by previous researchers, is the tendency to view single word switches as borrowings. Lexical switching is often confused with borrowing due to the fact that lexical items are single word switches or compound ideas that work the same way that single lexeme switches do. As discussed previously, categories such as nonce borrowings, those borrowings that are not yet established and are not used by monolingual speakers (Poplack et al. 1988), are often confused with single word code-switches. Even though this category has been left out of much of the previous research, it is a necessary category to include in this study due to the proliferation of examples of this type of switching.

To better understand the differences between this type of switching and borrowing it is critical to look at some of the motivations for such a switch as well as the problematic
terminology other researchers use to describe this category. Clyne (2003:111) sites lexical renewal as one of the reasons that lexical switching is such a common practice. Other reasons a speaker may choose to use a lexical item from another language range from the need for the specificity of a particular term to the preference for a more general term. One language may have a specialized meaning attached to a certain term that can only be expressed appropriately in the original language. Clyne (2003:111-114) describes the process of the use of single lexeme material as ‘lexical transference’ rather than ‘lexical switching’ and gives six ways in which ‘lexical transfers’ are used in his corpora. His blanket term of ‘lexical transference’ conflates borrowings and code-switches in that he includes both phenomena in this category. The unifying feature of this group is that it describes a single term. The uses, however, include both borrowings and code-switches. The first use of ‘lexical switching’ is to express items that do not have real equivalents in the other language. This would include any new referent introduced to a culture that takes its name from the language of the new concept or object. The word coyote in English is an example of this type of borrowing. A second use of lexical switching deals with the interpenetration of domains where the language of the switched item is the dominant language for the domain. This use is clearly demonstrated in the present study because the Internet is dominated by the use of English, and thus Internet terminology is often expressed in English regardless of the language of communication. In the present study, Internet terms, such as blog, post, and site, were almost exclusively expressed in English, even if the rest of the sequence was Spanish. A third use of ‘lexical transference’ was to express concepts in cases where the term in the original language has been lost. In other words, the term has been replaced with the foreign term and the original is no longer in use. These types of ‘lexical transference’,
following Clyne’s terminology, were included as examples of borrowings in the present study. These first three uses of ‘lexical transference’ are clearly borrowings.

Continuing with Clyne’s (2003:111-114) description of ‘lexical transference’ we see that the last three uses describe code-switches. The fourth use of ‘lexical transference’ is for reasons of facility or economy. Perhaps a shorter form or a more general term from the other language is preferred over the translation equivalent in the host language. For example, a Spanish speaker may prefer the shorter form *popcorn* from English to the longer *palomitas de maíz* of Spanish. The fifth type of ‘lexical transference’ occurs for the preference of a verb that takes fewer arguments and is, therefore, simplified. For example, Clyne discusses the preference of the English *remember* over the German *sich erinnern* because the German verb requires a reflexive pronoun. The sixth type, according to Clyne (2003), is quotation, whereby one speaker chooses to repeat or quote a switched lexical item used by the other speaker. For example: “How is your *prima*?” “*My prima* is doing well.” However, these types of ‘lexical transference’ illustrate the inconsistencies in the definitions and use of terms. The last three uses of Clyne’s so-called ‘lexical transference’ are clear examples of code-switching. Clyne has chosen not to categorize all single word switches as borrowings, but has grouped them as a separate category which includes both borrowings and code-switches. With these uses, the line between borrowings and code-switches becomes blurred. Clyne’s (2003) lexical transference conflates borrowings with lexical code-switching. The fact that some researchers propose that single word switches are cases of borrowing (Poplack et al. 1988) or ‘lexical transference’ (Clyne 2003) has contributed to the absence of the category of ‘lexical items’ as a type of code-switch.

Montes-Alcalá (2007) describes lexical switches as those traditionally labeled ‘lexical need’. However, this category is problematic because the analyst cannot always determine what
constitutes lexical need for a language user. According to Montes-Alcalá, this type of switch is not used by a speaker due to a lack of proficiency in one language or the other, as the use of a code-switched form due to ‘lexical need’ does not fill a lexical gap. Instead, the speaker has chosen the language of the term for reasons of subtlety in meaning, or because of higher frequency of exposure to the term in one language over the other. Many of these switches are culturally-bound words such as family member terms or specific concepts. According to Montes-Alcalá (2007), this category was the most prolific one found in the blog study she conducted. Indeed, lexical switches are a productive category of code-switches due to the relative ease with which a term from one language can be inserted into a larger string of discourse expressed in another language.

One of the problematic categories proposed by Montes-Alcalá (2007:168) was labeled ‘tag’ and included an assortment of unrelated forms. After reanalyzing her cited instances of ‘tag’ switches, I found that one of them included a code-switch that is really a lexical item. This example fits better in the category of ‘lexical item’: “nada que hacer más que ejercitarlo para que vuelva a estar en tip top shape” (2007:168). The expression tip top shape has been lexicalized, and is working as a single idea, and therefore is considered a single lexical item. During an informal survey of ten native English speakers, all participants completed a fill-in the blank task with shape after having been given the prompt, tip top. By contrast, when asked to provide the next term after in good, a range of different answers were given. This shows that tip top shape has been lexicalized and now works as a single unit.

### 3.3.2 Triggered switches

A second category proposed by Montes-Alcalá (2007) is ‘triggered switches’, which are those switches that seem to be caused (i.e. “triggered”) by a single word switch. These triggers
tend to be proper nouns, borrowings, or single lexical item switches that are uttered, or written, in one language and cause the speaker to continue in that language. Triggered switches occurred, albeit rather infrequently, in the present study. Schatz (1989:152) discusses triggered switches in a broader sense and discusses three types of triggers: discourse triggered switches, psycholinguistically triggered switches, and linguistically triggered switches. Discourse triggered switches are those that occur because of a specific speech situation, such as quotation and conversational asides; however, in the present study I do not view these as triggered switches as it is problematic to consider the discourse type as a trigger. Psycholinguistic triggers, too, are problematic if we classify them as triggers. Schatz (1989:154) considers emotional outbursts and word-search problems to be triggered by psycholinguistic factors. In the present study, I have considered only triggered switches that are linguistically triggered by specific items such as proper nouns, borrowings, and other single lexical item switches, such as *le dio high five*. She won the game. In this example the lexical switch, *high five*, serves as the trigger, and the blogger continued in English. Clyne (1967) drew attention to the process of switches being triggered by certain lexical items. These triggers, which he refers to as ‘transversion’, are said to facilitate a switch.

Clyne (2003:162) provides a list of trigger-words involved in switches that can be applied to the present study.

Such words are phonologically unintegrated lexical transfers (or ones with a low level of integration) [which in the present study would be considered lexical item code-switches], proper nouns, bilingual homophones (words that sound the same or nearly the same in two or more languages or in the speaker’s idiolects in the two or more languages).
While phonological integration is not a factor in the blogging data, examples of triggered switches can still be found in the data from the present study as well as that conducted by Montes-Alcalá (2007). Montes-Alcalá found these switches to be the least frequently occurring type, but triggered switches occurred in both directions, from Spanish to English and from English to Spanish.

3.3.3 Quotations

For the purpose of this study the category of quotation includes segments of discourse that the blogger attributes to another speaker, whether real or hypothetical. This category includes those switches made by the blogger to show that another speaker is involved, and to show that it is a deviation from the regular discourse. It is impossible to know whether these quotations are exactly what the speaker said, but the bloggers have chosen to report the words of others by switching language, thus highlighting the quoted speech.

Montes-Alcalá defined this category to encompass switches that occurred to quote someone else’s words. However, only 8.5% of the switches in her data were of this type. Perhaps conversations lend themselves more toward quoting the words of others more so than blogs, and thus researchers who work on conversational code-switching have the tendency to include this category. However, this category is problematic in the fact that switching languages to quote someone else’s utterance is colored by interpretation. In other words, what a speaker puts in quotation marks may not necessarily be the exact words of others. Written quotation marks can be used by bloggers to signal any number of things, from emphasis to language switch to quotation. According to Tannen (2007:112), the repetition of speech in another context fundamentally changes it, and thus creates constructed dialogue rather than reported speech. However, the distinction of whether the code-switched utterance is constructed dialogue,
reported speech or a direct quotation is unimportant for the present analysis. I include this category in my set because there are instances when a blogger is using a language switch to attribute the utterance to another speaker whether real or hypothetical.

3.3.4 Elaboration

Montes-Alcalá (2007:168) proposes the category ‘elaboration’ and defines it as code-switches that further explain an idea or add information to the context, which are, at times, parenthetical comments. This category includes switches that add information about the discourse, but they do not interrupt the normal flow. McClure (1981:82) also includes elaboration as a category of code-switching function and cites examples from her own work on the discourse of bilingual children. An example of this type of code-switch from her study is as follows: “Yo lo puedo quebrar. Yo lo tambien… lo pu(edo) quebrar. I can break this easy with my nose.” Gumperz (1982:79) categorizes switches used to elaborate or give additional information as ‘message qualification’. Although Gumperz used a different name, the function is the same. The following example from Gumperz (1982:79) illustrates this use: “We’ve got all… all these kids here right now. Los que estan ya criados aqui, no los que estan recien venidos de México. They all understood English.” Valdés-Fallis (1976) names this category ‘paraphrase’. Although given different labels, these categories describe the same discursive function of giving more information. This additional information can be marked parenthetically and is therefore often easily identifiable.

3.3.5 Evaluation/Emphasis

The category of ‘evaluation/emphasis’ accounts for linguistic forms that express speaker/blogger opinion. Often, by using idiomatic expressions or swear words, the blogger
expresses his or her opinion about an utterance while at the same time highlighting the segment through a code-switch. Thompson and Hunston (2000:5) define ‘evaluation’ as:

the broad cover term for the expression of the speaker’s or writer’s attitude or stance towards, viewpoint on, or feelings about the entities or propositions that he or she is talking about. That attitude may relate to certainty or obligation or desirability or any number of other sets of values.

Some of the types of evaluation found in the blogging data are emphasis, disdain, disbelief, and humor. The more general function of this category is the expression of speaker opinion, which can happen through intensifiers, such as the insertion of a swear word adverbially to intensify the adjective. This can be seen in example (33) in section 4.2.5 of Chapter 4, where the blogger writes *mil fucking millas de esa mierda*. Another type of emphatic switch involves the repetition of the same or similar information in both languages. Also, foregrounding can be achieved by switching languages. This can occur within clauses or between clauses. This type of emphatic use conveys the speaker’s evaluation of an utterance. The speaker evaluates the message as important enough to highlight the information with a change in language. When a speaker/blogger makes evaluative comments on a string of discourse, the evaluation does not have to be in terms of ‘good’ or ‘bad’; an evaluation can contain emphatic statements or comments as well.

Another type of evaluation is emphasis. Bednarek (2006:3) includes importance as an evaluative parameter. For example, excerpt (29) in section 4.2.5 of Chapter 4, *Porque? Why in the fucking hell?*, contains an emphatic question in English inserted into the otherwise Spanish discourse which highlights the speaker’s emphatic evaluation of the situation. In the present study, this is the second largest group of code-switches and it comprised 18.79% of the instances.
in the corpus. Valdés-Fallis (1976), McClure (1981:82), and Montes-Alcalá (2007:168) all discuss ‘emphasis’ as a separate category, including code-switching that serves to foreground or repeat information. Valdés-Fallis (1976) considers ‘emphasis’ and ‘repetition’ to be separate categories, but others do not. Gumperz (1982) includes only the category of ‘repetition’, while Montes-Alcalá (2007) and McClure (1981) discuss ‘emphasis’. While repeating information can be emphatic, the fact that a code-switch occurs at a particular point, by nature, makes the information more salient for the recipients, and the result is that they focus more attention on the message. The emphatic nature of this type of code-switching allows it to be viewed as a subset of ‘evaluation’. The speaker/blogger deems this utterance important, and thus places emphasis on it by code-switching. The other types of evaluation often include swear words or idiomatic expressions. At times, the use of a swear word acts as an intensifier, while at others it conveys anger or disdain. Idiomatic phrases, such as *WTF*, are also used in an evaluative sense.

Evaluation is also conveyed though ubiquitous Internet terminology used to account for the lack of suprasegmental features of speech. Initialisms such as *LOL* and *WTF* are seen throughout monolingual blogging data in English. These expressions are as easily transferred as tag questions and discourse markers, but they add an evaluative component to the utterance. Expressions such as *LOL* are common in Internet language regardless of the language of the blog, post, e-mail, etc.

The category ‘tag’ proposed by Montes-Alcalá (2007) is the most problematic of all the categories proposed. In this singular category, she includes: idiomatic expressions, swear words, discourse markers, and sentence fillers. These particular categories carry out very different discursive functions. Therefore, instead of grouping these types of expressions together, I propose the categories of ‘evaluation/emphasis’ and ‘discourse markers’ in order to better
describe and account for the instances of code-switching in question. Of the five examples provided by Montes-Alcalá (2007) grouped under ‘tags’, two belong in my new category of ‘evaluation/emphasis’ and two function as ‘discourse markers’. The following are examples from the study by Montes-Alcalá. These use idiomatic expressions in Spanish to express speaker opinion and are evaluative by nature, because they use emphatic expressions. In one case, a blogger writes, “Hijole mano, remember he borrowed that cd too del Flaco Jimenez?” The second example from the study also expresses the opinion of the speaker/blogger: “Chingado, I don’t have a job right now and I’ve been traveling across the state like a madman” (Montes-Alcalá 2007:168).

3.3.6 Discourse Markers

Discourse markers give information about how we are to interpret a sequence of utterances. For example, they mark deviations from previous segments (e.g. by the way) or a return to topic (anyway), or they set up oppositions between segments (yet). Fraser (1999:931) defines discourse markers as:

- a class of lexical expressions drawn primarily from the syntactic classes of conjunctions, adverbs, and prepositional phrases. With certain exceptions, they signal a relationship between the interpretation of the segment they introduce, S2, and the prior segment, S1.
- They have a core meaning, which is procedural, not conceptual, and their more specific interpretation is ‘negotiated’ by the context, both linguistic and conceptual.

It is precisely this ‘core meaning’ that makes discourse markers easy candidates for code-switches. English discourse markers can convey a subtle difference in meaning not conveyed by their Spanish counterpart and vice versa. In addition, the ease of separating a discourse marker from the rest of the context facilitates code-switches. Some linguists have avoided the study of
discourse markers because their exact classification has been debated. What counts as a
discourse marker for some does not for others. According to Fraser (1999:932), the term
discourse marker means different things for different researchers. Various terms are used to
refer to this group of words, including ‘cue phrases’ (Knott and Dale 1994), discourse
connectives (Blakemore 1987 and 1992), and discourse particles (Schourup 1985), to name a
few. The notion that this category of words and phrases is difficult to define is not a new one.
However, it is clear that this group shares some important characteristics and clearly has
discursive features that aid in the interpretation of an utterance. Some of the major criteria for
the definition of discourse markers listed by Schourup (1999) deal with connectivity and
optionality. Many discourse markers are clearly separated from the rest of the discourse both
syntactically and prosodically. While prosody is not a factor in blogging data, we can look at
code-switching as just another way to separate the discourse marker from the rest of the phrase.
This category, which is absent from the taxonomy proposed by Montes-Alcalá (2007), accounts
for the following two examples she listed as belonging to the ‘tags’ category: “So… aquí de
nuevo en el Disney Learning Center ...” and “However, estoy de acuerdo con eso (Montes-Alcalá
2007:168).”
CHAPTER 4
RESULTS AND DISCUSSION

4.1 Introduction

This chapter addresses the following three hypotheses for this thesis: 1) it is possible to propose a set of categories of code-switching that better describes and accounts for the data in Spanish-English Internet blogs; 2) most bloggers will use Spanish as a Matrix Language (ML) and English as an Embedded Language (EL); and 3) the most frequently occurring type of code-switch is lexical item switches. Section 4.2 provides examples for each of the socio-pragmatic categories that describe the functions of the code-switches found in the present study. Section 4.3 addresses the second hypothesis regarding the directionality of switches. Finally, section 4.4 presents a general discussion of the results found in the study.

The main goal of this study was to categorize the socio-pragmatic functions of code-switches. Using Montes-Alcalá (2007) as a starting point, I refined the categories proposed. Prior to categorizing the types of code-switches, I had to distinguish between single word code-switches and borrowings. Surprisingly, of the 665 total examples of code-switches or borrowings, 154 were borrowings, which make up 23% of the total number of tokens. One might expect to find more borrowings than code-switches in Puerto Rican blog data for two reasons: 1) Puerto Rican Spanish includes a number of established borrowings; and 2) the language of the Internet is so tied to English that terms related to this area are expressed in English. The majority of borrowings were meta-blogging language terms. The largest number of borrowings came from the terms blog or blogs: 28.57% of the borrowings involved these
terms. The next largest percentage comes from post or posts, which occurred 14 times in the corpus. Other well-established borrowings such as parking, web, and show occurred very few times. Only 26 tokens came from established borrowings, while the other 128 borrowings were Internet-related terms. In other words, 80% of the borrowings were Internet-related terms. A breakdown of the total number of posts by each blogger, the language that dominated posts, as well as the total numbers of borrowings and code-switches can be found in Table 7 in section 4.3. First, I counted the number of posts for each blogger, and subsequently I separated those written in English or Spanish to get the percentage by which each language dominated the discourse of an individual blogger. Then, each language switch was identified as either a borrowing or a code-switch. After separating the instances of code-switching, I categorized each socio-pragmatic function of the individual code-switches. For a breakdown of this categorization see Table 8 in section 4.5.

4.2 Categories of code-switches

4.2.1 Lexical items

In the present study, this type of code-switching was the most common, accounting for 61.06% of the total number of tokens of code-switches. The following examples of lexical item code-switches are found in the Puerto Rican blog data. The spelling and punctuation were preserved from the original post of the user.

(1) por ejemplo que pase algo en el background check y que no me den
(2) si es una socker [sic] mom ten más cuidado porque esas cabronas son …
(3) Clubbing Somewhere Around San Juan With The Primas [title for a photo]
(4) me convenci a mi mismo que Danielle murio la ultima vez que tuvo su relapse y esta persona ya no era la mujer que amaba
(5) giving me a long discussion about our youth and their sinverguencerias - did I
mention this was at 7:45 AM

(6) I got some dirty looks from the viejitas (old ladies) but who cares!!?!

(7) pero perdió respeto y el "street cred" con el resto de las bandas.

These examples show the range of code-switches of this type. While most of the switches occurred from Spanish to English, there are examples in which a Spanish lexical item was inserted into an English phrase. For example, with the primas, and their sinverguencerias, and from the viejitas all demonstrate the insertion of a Spanish lexical item in an otherwise English phrase. Examples (1) el background check, (2) una socker mom, and (7) el street cred are compound nouns that function as one idea, and thus are one lexical item. We also can see that in all of the above examples, the words are inserted into the matrix language and take the determiner from that language. El background check and el street cred take the determiners from Spanish, as they are terms inserted into otherwise Spanish utterances. The viejitas and their sinverguencerias are used with English determiners because English is the matrix language and thus provides the determiner.

4.2.2 Triggered switches

Triggered switches were also the most infrequently-occurring type in the blogs, only accounting for 2.94% of the data. These switches only occurred 15 times throughout the corpus. The following are examples of triggered switches from the present study. The trigger is in bold, but otherwise the spelling and punctuation from the original post has been preserved.

(8) I've been studying for the MCAT like crazy. Wish me good luck... Oh & le cambie el nombre al blog. I like this one more.

(9) Mas tarde mientras mi papa intentaba darle de comer a uno de ellos, el chico le dijo: abuelito bello I love you. God knows where he picked that up, they don't speak spanish.

(10) Tampoco que tengo una hermana mayor que vive en Nova Scotia, Canada. Well
she's in town for a couple of days. This is going to be fun!

(11) Una escala en Chicago on our way to Cleveland. Fue una excelente idea quedarnos por unos días.

(12) y le dio high five. She won the game.

(13) No, no fué el naked guy in my room...)

(14) Y estoy pintando esta pared de “bright yellow”. ¡Good morning, Sunshine! Yo me siento feliz.

A number of things serve as triggers for code-switches. In example (8) a borrowing, blog, serves as a trigger for a switch into English. Quotations can also serve as triggers. For instance, in example (9) the blogger was quoting the words of a boy who used both Spanish and English in the same sentence, abuelito bello I love you. Since the last thing the boy said was in English, the blogger continued writing in English. Proper nouns are common triggers. Examples of this type of trigger can be found in (10), Nova Scotia, Canada. Well..., and (11), en Chicago on our way to Cleveland. Lexical items serving as triggers can be seen in examples (12) dio high five. She won the game, (13) el naked guy in the room, and (14) pared de bright yellow. ¡Good morning, Sunshine!. Of the 15 tokens of this type of switch in the data, there were eight instances in which a lexical item served as the trigger.

4.2.3 Quotations

The data collected for the present study also have very few examples of this type of code-switch. Only 3.13 % of the total number of code-switches were quotations. Nonetheless, this category of code-switching is common to all of the previously mentioned typologies, including Gumperz (1982), McClure (1981), Valdés-Fallis (1976), and Montes-Alcalá (2007). The following examples of quotations from the blogs analyzed preserve the original spelling and punctuation and illustrate the different types of speakers that are quoted.
Y no estoy jodiendo ni es un cliché, esas preguntas me las hicieron varias veces. Pero a la misma vez prepárate para responder preguntas como “Wow! from Puerto Rico!, I love that place, It’s my favorite vacation spot, Why in the world would you leave that to come here??!!??! Do you miss the mountains???”

You don’t need to know Spanish to go there, however it can come handy if you know at least the basic. Things like “Buenos días,” “Buenas tardes,” and “Gracias” can work wonders. If you don’t know any Spanish but can say the following: “No hablo español, ¿Habla usted inglés?” can help you a LOT.

I couldn't help to smile at the sticker on the back of the wheelchair... "que Dios te multiplique lo que a mi me deseas" - I thought about the power of wanting something...

So he called me back. No le respondi pero le envíe un mensaje de texto que decía: "Sorry but I can't see you anymore, I thought you were an adult"

Cuando entramos por la puerta el recepcionista dijo "ha look who's here" yo pensaba que

la hija de mi amiga me dice: "This area is very pretty... So different. It's clean..."

Y de momento, toma: “Nydia, you’re pregnant”.

This category includes the quoted words of hypothetical speakers as in examples (15), (16), and (17). However, these three hypothetical speakers are very different. In example (15) preguntas como “Wow! from Puerto Rico!... Why in the world would you leave?, there is no attachment to a particular source; it is something someone might say, or an approximation of comments that the blogger has heard before, but these are not the exact words of another. Example (16), Things like “Buenos días” and “Buenas tardes”, also supposes a hypothetical speaker, but this quotation is metalinguistic commentary about what one could say in Spanish in a given situation. Example (17), que Dios te multiplique lo que a mi me deseas, is quoting a written sign and has no specific speaker. And, while all three of these examples use hypothetical speakers, they are not the same kind of speaker. The last four examples are those in which the speaker is known; the blogger is code-switching to highlight reported speech. In example (18), un mensaje de texto que decía:
Sorry but I can’t…, the speaker is known by the blogger. This could be a direct quotation, as it is in the form of a text message, and the blogger could have looked at her phone while writing this particular post. However, we cannot say for sure that this was the case. Example (19), *el recepcionista dijo “ha look who’s here”*, is a traditional example of this type of code-switch. It is posed as a direct quotation, but it is more likely that it is constructed dialogue. While some of the examples are attempts at reported speech, which could also be constructed dialogue on the part of the blogger, not every quotation has to have a specific, identifiable speaker, as is the case in the first three examples. Examples (18), (19), (20), and (21) are attributable to another speaker, and the code-switch makes this change in voice more noticeable. These last four examples are instances of the traditional notion of how code-switching is used for quotations, whereby an identifiable speaker is involved.

### 4.2.4 Elaboration

The present study includes 31 tokens of this type, which constitute 6.07% of the total. The following examples show how code-switching is used for elaboration in the context of blogging. The original spellings and punctuation have been preserved from the original posts.

(22) y cuando digo todos **I mean mom, dad, serocha, adrianna, Sarah, Jonathan (my sisters husband), the boys and me** – a ver el show musical de Lion King.

(23) ahora que se fueron la casa se siente sola. **I think this is all for now.**

(24) I'm only 24. When did I miss the "you have to get married before 25" memo?? **Todo paso en mi trabajo.**

(25) Esta bastante frio la temperatura, **spring should hurry up a little bit.**

(26) Esta es la primera vez en año y medio que trato de escribir un cuento. **I’m really really really rusty.** Pero pues tenia la idea y quería hacer algo con ella, **if only to get back on the bicycle.**
These switches add information to the discourse, but do not interrupt the flow of the message. The switch in example (22), *cuando digo todos I mean mom, dad…*, clarifies who *todos* means, but does not disrupt the flow of the phrase. Likewise, in example (23), ..*se siente sola. I think that is all for now*, the switch provides additional information, and simultaneously functions as a type of closing. The switch in example (24), *todo paso en mi trabajo*, signals the beginning of a story. Examples (25) and (26) also serve the purpose of adding additional information about the discourse as a secondary comment.

### 4.2.5 Evaluation/Emphasis

Example (31) below illustrates that *LOL* can be found in otherwise Spanish discourse. The speaker/blogger adds an element of evaluation to aid the reader in interpreting the message appropriately. The blogger wishes to convey that he finds this humorous. Example (32) also involves this type of evaluation, as the blogger conveys his or her attitude toward the message by inserting the word *funny* in the middle of the message. These expressions are also syntactically separate from the discourse, as indicated not only by the punctuation used to separate the phrases, but also by a language switch. The following examples are taken from the Puerto Rican blog data.

(27)  *esta chica me envio varios mensajes de texto* insulting me out of nowhere.

(28)  *allow me to see la grandeza de tu misericordia*, invite to share the joy of love

(29)  *Porque? Why in the fucking hell?* Porque me interesa sus vidas?

(30)  *la cámara instantánea es otro mambo, pero it’s equally worth it.*

(31)  *… pero me ronca los cojones… LOL!!!*

(32)  *e incluso les encanta Puerto Rico y nos admiran. Funny.* Me gustaria saber donde estan esos “rasitas”

(33)  *mil fucking* millas de esa mierda!
(34)  Ay, a veces no puedo conmigo mismo *sigh*

(35)  un enlace de entrada desde Primerahora. **WTF?** No entiendo, quien carajo escribio eso???

(36)  los vigilantes. **awesome.** tocan de nuevo estos dias.

(37)  Ah, ¿vive alli? **Nice,**… Nunca he entrado a esos apartamentos

In all of these examples, the speaker/blogger chooses to highlight his/her evaluation of the situation by switching languages. For instance in example (32), *les encanta Puerto Rico y nos admiramos. Funny. Me gustaria saber...*, the use of *Funny* is meant to convey sarcasm on the part of the blogger. S/he is letting the reader know that s/he does not actually think the statements are humorous, which seems to be further implied by the separation of the word from the rest of the utterance through punctuation as well as a language switch.

Interestingly, some of the evaluations are not just emphatic uses of code-switches, but are aids for interpreting the message. The example in (32) is one such message. The blogger has evaluated the situation and adds code-switched commentary to show how he/she feels about the message. This can be seen with ubiquitous Internet expressions such as *LOL* and *WTF*, as in examples (31) and (35). Bloggers also choose to use evaluative words such as *Funny* in example (32), *nice* in example (37), and *awesome* in example (36). And, while *Funny* shows sarcasm, *nice* and *awesome* are uses of positive comments. Evaluations can be expressed by linguistic representations of physical actions, such as example (34), in which the expression *sigh* shows the blogger’s disappointment.

4.2.6 Discourse Markers

In the present study, 41 examples, 8.02% of code-switched discourse markers, were found, and in each of the cases an insertion of English discourse markers was made into an otherwise Spanish utterance. The switch was unidirectional from Spanish to English, as no
English blog posts inserted Spanish discourse markers. The most common discourse marker found in the data was *anyway*(s). There were 16 instances found in the discourse of at least four different bloggers. The next most frequently used discourse marker was *so* with 11 instances, followed by *actually* and *BTW* or *by the way*. Some other discourse markers found in the data were *needless to say*, *yet*, and *you know*. Most of the tokens of code-switched discourse markers in the present study have one of two functions: marking continuance of a previous idea, or marking a deviation. *By the way*, or *BTW*, is used to mark a deviation from prior discourse, while *anyway*(s) and *so* mark a return to topic or continuation through showing a cause and effect relationship. The following examples contain discourse markers found in the Internet blogs analyzed. The discourse markers appear in bold. The spelling and punctuation have been preserved from the original post.

(38) Es que he estado haciendo tantas mierdas estos días que no me sobra tiempo ni para rascarme el culo pelú. *Anyway*, el viaje que hicimos de Florida a Texas

(39) ¿Alguien sabe las contestaciones a esto? *BTW*, el banco que tiene mi carro

(40) Que *by the way*, si con algo yo me siento bastante cómoda
g

(41) por eso escribo estas lineas…*So*, por que nos sentimos solos??


(43) Tal vez he puesto más en juego mi escencia de lo que se me había ocurrido. *Actually,…* puse en juego mi escencia en manos de varios otros

These examples show that it is only the discourse marker that is switched in the discourse. The use of the discourse marker *anyway* in (38) shows a return to topic, while examples (39) and (40) mark a deviation from the topic. *BTW* and the unabbreviated form *by the way* mark asides from the original message of the post. *So*, in example (41), works like *anyway*, in that it is a return to
the original thought of the blogger. In example (42), *yet* is serving to mark a contrast in contradictory phrases between what the speaker expects and what has actually happened. The blogger expects that things had changed so much that it would never be the same, and the expectation then is that ‘she’ would not be there. The use of the discourse marker *yet* to introduce the unexpected information shows the contrast between expected and unexpected information. In example (43), the discourse marker *actually* serves to refine the message to make it a stronger assertion. These phrases have clear discursive functions that relate propositions in the message. Each of these examples has discourse markers expressed in English in otherwise Spanish utterances.

4.3 Matrix Languages

It was expected that Puerto Rican bloggers would use Spanish as the matrix language and English as the embedded language. However, I expected this from Puerto Ricans living in Puerto Rico. Since this demographic information was not always given, it was impossible to use this as part of the analysis. While more bloggers, in general, used Spanish as a matrix language the actual number of posts in English outnumbered those written in Spanish. Twelve bloggers posted more often in Spanish, six posted in English, and one blogger posted in both Spanish and English. For the total number of posts in Spanish and English as broken down by the individual blogger see Table 6 below.
<table>
<thead>
<tr>
<th>Blog</th>
<th>ML Spanish posts</th>
<th>ML English posts</th>
<th>mixed language posts</th>
<th>Total</th>
<th>Language dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>46</td>
<td>29</td>
<td>0</td>
<td>75</td>
<td>Spanish</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Spanish</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>22</td>
<td>0</td>
<td>22</td>
<td>English</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>15</td>
<td>7</td>
<td>25</td>
<td>English</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>109</td>
<td>1</td>
<td>113</td>
<td>English</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>Mixed</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>52</td>
<td>English</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>11</td>
<td>0</td>
<td>17</td>
<td>English</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>Spanish</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>73</td>
<td>0</td>
<td>77</td>
<td>English</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>Spanish</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>Spanish</td>
</tr>
<tr>
<td>13</td>
<td>9</td>
<td>8</td>
<td>0</td>
<td>17</td>
<td>Spanish</td>
</tr>
<tr>
<td>14</td>
<td>68</td>
<td>8</td>
<td>0</td>
<td>76</td>
<td>Spanish</td>
</tr>
<tr>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>Spanish</td>
</tr>
<tr>
<td>16</td>
<td>69</td>
<td>1</td>
<td>1</td>
<td>71</td>
<td>Spanish</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>Spanish</td>
</tr>
<tr>
<td>18</td>
<td>20</td>
<td>5</td>
<td>0</td>
<td>25</td>
<td>Spanish</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Spanish</td>
</tr>
<tr>
<td>Total</td>
<td>287</td>
<td>334</td>
<td>14</td>
<td>635</td>
<td></td>
</tr>
</tbody>
</table>

Table 6: Breakdown of posts by matrix language

For those who used Spanish as the matrix language, code-switching was a more common practice. Those who posted in English primarily also code-switched, but much less frequently.

Additionally, I expected that those posts that used English as a matrix language would include borrowings that might be used by the bloggers as an identity marker. This was not the case. The posts that were in English included very few borrowings in general. For a breakdown of the number of posts made by each blogger and the number of code-switches and borrowings found in those blogs see Table 7 below.
Table 7: Percentages of B and CS as well as total number of tokens

<table>
<thead>
<tr>
<th>Blog</th>
<th># of posts</th>
<th>Language of posts</th>
<th># of borrowings</th>
<th># of CS</th>
<th>Total number of tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75</td>
<td>Span 61%</td>
<td>30</td>
<td>42</td>
<td>72</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>Span 100%</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>Eng 100%</td>
<td>0</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>25</td>
<td>Eng 60%</td>
<td>7</td>
<td>52</td>
<td>59</td>
</tr>
<tr>
<td>5</td>
<td>113</td>
<td>Eng 96%</td>
<td>5</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>Mixed 83 %</td>
<td>1</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>52</td>
<td>Eng 100%</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>Eng 64%</td>
<td>2</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>Span 100%</td>
<td>3</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>77</td>
<td>Eng 94%</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>7</td>
<td>Span 100%</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>14</td>
<td>Span 100%</td>
<td>26</td>
<td>12</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>17</td>
<td>Span 52%</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>76</td>
<td>Span 89%</td>
<td>12</td>
<td>97</td>
<td>109</td>
</tr>
<tr>
<td>15</td>
<td>18</td>
<td>Span 100%</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>71</td>
<td>Span 97%</td>
<td>29</td>
<td>173</td>
<td>202</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>Span 100%</td>
<td>0</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>18</td>
<td>25</td>
<td>Span 80%</td>
<td>38</td>
<td>13</td>
<td>51</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>Span 100%</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>635</td>
<td></td>
<td>154</td>
<td>511</td>
<td>665</td>
</tr>
</tbody>
</table>

To calculate the dominant language of the blog, I first determined the number of the posts with English as the matrix language and with Spanish as the matrix language. The percentages above reflect the dominant language. For example, in blog 1 of the 75 posts, 46 used Spanish as the matrix language while 29 used English. For the mixed blog, blog 6, five of the six posts were first written in Spanish and then translated into English, such that an individual post changed matrix languages in the middle to repeat the same message. The sixth post was written entirely in English.

### 4.4 Discussion

As expected, the language of these particular bilingual bloggers shows a great deal of influence from English on blogs written largely in Spanish. By contrast, most of the blog
messages written primarily in English showed only few instances of Spanish influence. One might hypothesize that these blogs were written by Puerto Ricans living in the US who have very little knowledge of Spanish. This information was not included in the demographic information provided by the bloggers themselves, so it was not possible to test this hypothesis with the current data. Often, code-switching happened from Spanish to English. The more common language holistically for the blog posts was Spanish, and in the primarily Spanish blogs, code-switching was far more common than borrowing. The seven categories proposed by Montes-Alcalá (2007) were not as descriptive as they needed to be. I conflated her ‘emphatic’ with my ‘evaluation’ category because I view ‘emphasis’ as a subset of ‘evaluation’. Code-switching for emphatic reasons is a way to draw attention to the information. Signaling that a segment is important by emphasizing it constitutes a type of evaluation. In addition, I eliminated the problematic ‘tag’ and ‘free’ categories because they were ill-defined at best. Montes-Alcalá (2007) proposes the ‘free’ category as a catch-all for those switches that are not used for any obvious reason or for those that are an eclectic combination of other functions. While this category is the third most occurring in the research by Montes-Alcalá (2007), I find that in the present study it is an unnecessary category due to the revisions I have made to the categories. I also added the category of discourse markers to better describe the data taken from these 19 bloggers who have written a total of 635 posts during the course of one year. For a specific breakdown of the number of types of code-switches see Table 8 below.
It was expected that lexical items would be the most frequently occurring type due to the ease with which single lexical items can be inserted into discourse much the same way borrowings are easily integrated into monolingual speech. And, in fact, in the present study the most common type of code-switch is lexical item switches.

When attempting to refine previous categories, I considered the nature of discourse markers carefully. Clearly, they are distinct from the other categories. Discourse markers rely on functional meaning rather than semantic meaning, and this defining feature already distinguishes them from the other types of code-switches. Referring back to Lipski (2005) and his study on the insertion of so from English into Spanish discourse, we remember that he categorized this discourse marker as a special kind of code-switch. He found that so was
phonologically integrated into Spanish. This is often the major criterion for distinguishing a borrowing from a code-switch, in spoken data. However, he maintained that it was still a code-switch even with this integration. Other discourse markers found in the present study were treated as code-switches, although again, they are distinct from other code-switches. Using frequency as the determining factor to distinguish code-switching from borrowing, I counted words that were repeated in two or more blogs as borrowings, except in the case of discourse markers. English discourse markers were unidirectionally inserted into Spanish discourse and were repeated sometimes as often as 16 times. Thus, according to my criteria I should have perhaps considered these discourse markers borrowings. However, in light of Lipski’s (2005) findings, I count them as code-switches, but with the acknowledgment that they behave differently from the rest. Perhaps with further research I will, in fact, find that they do indeed behave more like borrowings than code-switches.

Similarly, I must also acknowledge that perhaps I should treat set Internet expressions, such as LOL and btw as borrowings. With further research on these fixed expressions it is possible that I will find that they are used prolifically among monolingual Internet users as borrowings. However, for this analysis I have considered them as code-switches and classified them according to their functions. With further research on these Internet initialisms, I will better understand the nature of these terms and will be able to refine my categorization. Truly, the difficulty of separating written single word code-switches from borrowings has been challenging during this investigation. It is only through further research that we can better understand how to classify these forms.
CHAPTER 5

CONCLUSION

Code-switching has been studied using various media; however, the newness of the Internet has sparked interest in the study of language on-line. Computer-mediated communication, in itself, is a newer area of investigation, with little existing research on languages other than English. Because of this, language choice on-line has become a topic of interest for linguistic analysis. Other studies on code choice in computer-mediated communication have tended to focus on e-mail exchanges or synchronous chat programs. To date, blogging as a genre has not been studied extensively in the area of language contact. Montes-Alcalá (2007) analyzes code-switching in Spanish-English bilingual blogs. It is due to this study that I recognized a need for more research in this area.

The journal-like nature of blogging makes this type of communication perfect for studies on language use. While the conversational context is absent, there is still a strong desire on the part of bloggers to write for some unknown reader. Blogs are not a series of notes only intended for the author to read; they are opinions and discussions posted for the world to see. Bloggers write in such a way that the whole world, and yet no one in particular, is their audience. It is this participation in an on-line community that makes it easy for bloggers to use informal language. The blogs posted to <puertoblogs.com> are all produced by bloggers who claim to be Puerto Rican or live in Puerto Rico, and thus they share cultural ties and common ground. It is the organization of blogs that makes a study of code-switching in them possible. All bloggers are
familiar with both Spanish and English and can freely post in either language, or in a combination of the two, depending on what they feel comfortable using.

For this study, I address the following hypotheses: 1) it is possible to propose a set of categories of code-switching that better describes and accounts for the data in Spanish-English Internet blogs; 2) most bloggers will use Spanish as a Matrix Language (ML) and English as an Embedded Language (EL); and 3) the most frequently occurring type of code-switch is lexical item switches. Also, I differentiate borrowings from code-switches and propose a categorization of uses for particular code-switches. While the distinction between whether a single word counted as a borrowing or a code-switch was problematic due the nature of written language, I used frequency as a criterion. I determined that if an item was repeated, except in the case of discourse markers, it was a borrowing. Over the course of one year, from August 2007 to August 2008, I read the entries of 19 different bloggers who wrote as many as 113 individual blog posts or as few as two during the year. All of the bloggers stated that their blogs were written in Spanglish, but only one participant showed a true mix of the two languages such that one language did not dominate. Of the 19 blogs, most of the authors wrote predominantly in Spanish with instances of code-switching to English. Six bloggers wrote in English. In the English-dominated blogs, there were far fewer instances of switches to Spanish, perhaps due to a lack of knowledge of Spanish. Of the 665 tokens of code-switches and borrowings, 154 were borrowings, while 511 were code-switches. The socio-pragmatic uses proposed by Montes-Alcalá (2007) were problematic in that her ‘tag’ category included a mix of expressions, such as swear words, discourse markers, and idiomatic phrases that were not necessarily grouped by function and were certainly not all ‘tags’. For this reason I proposed a revised set of categories. I removed the ‘tag’ category as well as her ‘free’ category and added ‘discourse marker’ and
‘evaluation’, which included her ‘emphatic’ category, as well as those code-switches that reflected the speaker’s attitude or emotions toward the message. These included expressions such as *LOL*, communicating that the blogger found the message funny, or *WTF*, showing that the blogger was irritated. My proposed categories are: lexical items, triggered switches, quotations, elaboration, evaluation, and discourse marker. I found that these six categories could account for the data without the need for a catch-all ‘free’ category. The largest number of code-switches found was of the ‘lexical item’ type. Over half of the code-switches (61.06%) were single-word or compound-word switches. The next largest category was ‘evaluation’, which made up 18.79% of the tokens, followed by ‘discourse marker’, which had 41 tokens (8.02%) of the 511 instances of code-switching. Elaborated code-switches were 6.07% of the total number of tokens. The categories with the fewest instances were ‘quotation’ with 3.13% and ‘triggered’ with 2.94% of the total number of code-switches. Of the three proposed hypotheses, I find that it is indeed possible to propose a set of categories to describe code-switching uses. The second hypothesis was a bit more complicated to answer than anticipated. The data in the present study supported the claim that more bloggers use Spanish as a matrix language; however the limited number of participants means that more data is needed. The third hypothesis was also confirmed because, indeed, single lexical item switches were by far the most common.

The linguistic processes of code-switching and borrowing are obviously not only found in conversational data, but also abundantly in blog posts of bilingual authors. For further research, I would like to compare the language of the Puerto Ricans living on the island to those living in the United States. I propose that the language of those living on the mainland would include more English and more borrowings than code-switches. However, the demographic data given
by the bloggers in the present study lacked the necessary information to include this distinction in the current analysis. Additionally, more information such as length of residency and language proficiency are important for addressing this particular issue. To complete this study, a survey of bloggers that includes demographic information is required. Further, I plan to compare Puerto Rican blogs with those of Spain to address the influence of Internet English. Puerto Ricans have a long history of contact with English, and it is important to compare this group to others who do not have this history of contact. By comparing other blogs from other groups of people, I can also better address the nature of Internet expressions such as *LOL* and *btw*. I can determine if these ubiquitous Internet expressions are present in monolingual Spanish language blogs, thus making them borrowings rather than code-switches. Through the collection of more data I can better understand the differences between code-switches and borrowings. The more instances I find of single word switches the easier it will be to refine my originally proposed categories.
References


