CHAPTER 13

Comprehension patterns of two groups of Spanish-English bilingual codeswitchers

Rosa E. Guzzardo Tamargo and Paola E. Dussias
Universidad de Puerto Rico / The Pennsylvania State University

Current knowledge of the mechanisms underlying the production and comprehension of codeswitches derives largely from studies with bilingual speakers who do not codeswitch or who report codeswitching (CS), but do not belong to stable bilingual communities. Although these data have a place in CS research, the foundational knowledge must characterize bilinguals in communities where CS is linked to community norms. We examine this issue by comparing a group of Spanish-English bilinguals from Harlem, home to a well-defined community with regular CS, and a group of Spanish-English bilinguals who resemble participants of past lab studies. We recorded eye movements while participants read frequent and infrequent switches found in Spanish-English corpora. Despite differences between both groups, the findings revealed strikingly similar eye-movement patterns.

Keywords: intrasentential codeswitching, Spanish-English, comprehension, eye tracking, bilingual communities

1. Introduction

Many studies that focus on codeswitching (CS) have noted that this bilingual practice is strongly linked to community membership. CS is not equally used in all bilingual communities. While it has been amply documented that intrasentential CS is very common among Spanish-English bilingual communities in Miami, New York, and Texas (e.g., Carter, Deuchar, Davies, & Parafita Couto, 2011; Poplack, 1980; Trillo, 2003), highly proficient Dutch-English bilinguals in The Netherlands, for example, do not engage in intrasentential CS. Moreover, the kinds of codeswitches that bilinguals produce vary from community to community. A classic example of this is found in a study by Poplack (1987), which examined the CS patterns of two bilingual communities: one in the United States and

DOI 10.1075/ihll.11.13guz

one in Canada. The data collected in the United States were the product of a large ethnographic study in which the researchers observed and interacted with Puerto Rican Spanish-English bilinguals on a city block in East Harlem, New York (an area also known as Spanish Harlem or El Barrio). Poplack reported three main switch types in the community: (1) tag switches, (2) intersentential switches, and (3) intrasentential switches. Because each of these switch types requires different control of both languages, their occurrence was modulated by the participants' bilingual ability. Unbalanced bilingual speakers mainly produced tag switches, while the most balanced speakers generally switched intrasententially (Poplack, 1980, p. 607). Nonetheless, as a whole, a fluent and smooth transition between Spanish and English without false starts, hesitations, or lengthy pauses proved to be an integral part of the linguistic repertoire of this stable bilingual Puerto Rican community in East Harlem.

The data collected in Canada came from interviews conducted by random sampling with French-English bilinguals living in several neighborhoods in Ottawa and Hull. In this case, the participants codeswitched very differently from those in New York. The most prevalent CS occurred to fulfill four communicative functions: (1) to provide the right expression, (2) to engage in metalinguistic commentary, (3) to emphasize or bracket an English intervention, and (4) to explain, specify, or translate. The bilingual community in the United States and the one in Canada appeared to be practically opposed with respect to the types of codeswitches they produced: the Canada bilinguals produced less than 4% of tag, intersentential, and intrasentential switches, which were frequently used in New York, while the U.S. bilinguals produced less than 5% of the flag or special-purpose switches, most commonly used in Canada (Poplack, 1987, p. 67).

The central question in Poplack's (1987) study revolved around the possible factors that could account for the differences in CS patterns that were found between the two bilingual communities. Poplack explains that the linguistic or typological differences between Spanish and English, on the one hand, and French and English, on the other, could not be responsible because they are minor. However, she recognizes that part of the divergence could be due to the different data collection techniques used in each case. The ethnographic approach used in New York allowed the researchers to observe the participants more in depth and become more acquainted with them, while, with the random sampling methodology used in Canada, the researchers were never considered group members, as their social contact with the participants was shorter. Moreover, since the goal of the Ottawa-Hull study was to examine the use of French in the region, the interviewers did not overtly encourage CS by not codeswitching themselves.

Nonetheless, Poplack (1987) suggests that a more plausible explanation for the differences found in the CS patterns of these two communities has to do with

a true difference in communicative patterns based on differences between the bilingual contexts. In other words, disparate social, historical, and political factors have influenced the attitudes that each community has towards the use of English, leading to distinct CS patterns. Although French is a minority language in the Canadian context and Spanish is a minority language in the U.S. context, French has been in contact with English in Canada longer than Puerto Rican Spanish has in the United States. French enjoys more prestige in Canada – with its official national language status – than Spanish does in the United States. Bilingualism is a representative feature of New York Puerto Rican identity, but it is not associated with a particular ethnic group in Canada. These differences generated distinct attitudes towards bilingualism that resulted in dissimilar CS behaviors in both communities. The New York bilinguals consider that CS is appropriate to their dual identity and display an apparent unawareness of the particular switch types they use. They codeswitch as an overall discourse strategy to use both languages rather than to achieve any specific discursive effects; this minimizes the saliency of the switch points. Instead, the Ottawa-Hull participants show neutral affect towards their bilingual behavior and are able to correctly identify specific reasons for codeswitching. They emphasize their codeswitches "by repetition, hesitation, intonational highlighting, explicit metalinguistic commentary, etc. and use the contrast between the codes to underline the rhetorical appropriateness of their speech" (Poplack, 1987, p. 65).

The role of the speech community on CS patterns is so pervasive that it emerges even when examining groups of speakers of the same language pairs living in the same country. Poplack (1987) documented differences in CS patterns based on community attitude differences between the Ottawa and the Hull neighborhoods examined in Canada. While the Ottawa neighborhoods codeswitched more frequently with the purpose of providing the right expression, the Hull neighborhoods, particularly the upper-middle class neighborhood, preferred to codeswitch for metalinguistic commentary. Poplack related this difference to the belief in Hull that good French must exclude anglicisms (1987, p. 64).

O'Malley-Madec (2007) reports similar findings in her comparative study of two Irish-English bilingual communities in Ireland. The researcher specifically compared the community of Carroroe, within the heartland of Irish-speaking Galway, to Claregalway, a peripheral community located three miles from English-speaking Galway city. Both communities share rural identities with a similar social mix. Irish is the majority language in Carroroe, but it is the minority language in Claregalway. Despite this, both communities have similar intensity of contact with English due to the large number of Carroroe residents who commute to Galway city and work in English-speaking environments. English is attached to somewhat of a negative social identity in Carroroe; however, in Claregalway,

it is simply considered the language of the present, whereas Irish is considered to belong to the past.

Based on data collected by means of sociolinguistic interviews, O'Malley-Madec (2007) found clear differences in the CS behavior of the residents of these two communities. The speakers of Carroroe displayed a behavior akin to borrowing, in which they used English words and expressions – mostly nouns and discourse markers or tags – in an otherwise Irish matrix. Meanwhile, Claregalway residents consistently alternated their codes, usually to change the conversational action. Another interesting difference between both communities arose with respect to their use of English discourse markers or tags. On the one hand, speakers in Carroroe displayed a tendency to use English tags to signal informal speech style. O'Malley-Madec (2007) explains that this behavior coincides with the community's general attitude towards English: because English is negatively viewed, it only appears in casual speech. On the other hand, in Claregalway, not English, but Irish discourse markers are used for metalinguistic purposes. In this case, the researcher points out that the Irish tags are used to hide the speakers' communication needs in Irish. Once again, the differences in CS patterns found between the Ottawa and the Hull neighborhoods and between the Carroroe and the Claregalway communities are likely not due to linguistic facts given that the same two languages – French and English in the first case, and Irish and English in the second case – are involved. Instead, each group of speakers shows patterns that are attributable to the social context of the speech community in which they are produced.

The study of the mechanisms that enable speakers of bilingual communities to produce and to comprehend codeswitched language fluently and effortlessly has garnered a great deal of interest from psycholinguists (e.g., Dussias, 2003; Gullifer, Kroll, & Dussias, 2013; Kootstra, Van Hell, & Dijkstra, 2010, 2012; Litcofsky, 2013; Moreno, Federmeier, & Kutas, 2002; Proverbio, Leoni, & Zani, 2004). Unlike the sociolinguistic studies described above, this work does not consider the bilingual speech community as the unit of study (Torres Cacoullos & Travis, 2015); instead, lab experiments are carried out by randomly sampling university students who either do not engage in CS or who codeswitch but do not belong to identifiable speech communities where CS is a common practice. Although undoubtedly these data have a place in CS research, they must be tied to knowledge of how bilinguals in established speech communities process the same types of switches. This is an issue of key importance given the recent evidence from mainstream psycholinguistic studies demonstrating a close link between statistical regularities in speakers' linguistic environments and the linguistic representations that speakers recruit during language comprehension (e.g., MacDonald, 1999, 2013; MacDonald & Seidenberg, 2006). With this in mind, the goal

of the study presented here is to begin to close the gap between lab-based and corpus-based approaches to the study of CS by comparing a group of Spanish-English bilingual speakers from Harlem, the home of a well-defined speech community where CS occurs regularly, to a group of Spanish-English bilinguals who more closely resemble the participants in past lab CS studies.

Like most monolingual communication, interactions involving codeswitched language among bilinguals occur in the spoken domain. This is why most CS work has focused on production. In the work presented here, we study, instead, the comprehension of codeswitched language. Why? The few studies investigating CS from the perspective of the comprehender suggest that recognizing and integrating a linguistic code different from the one most recently encountered entails processing costs (see Van Hell, Litcofsky, & Ting, 2015, for a review). One empirical question that has not been examined, however, is whether these costs arise because processing other-language items is inherently difficult or rather because the bulk of studies that have examined the comprehension of codeswitches have used made-up sentences that are rarely actually produced by bilingual speakers. An ancillary goal of the study presented here, then, is to examine whether processing structures that have been demonstrated empirically to represent quantitatively meaningful patterns of language use in well-defined bilingual communities also give rise to processing costs.

An additional aim in studying the comprehension of CS is to add bilingual data to the discussion of how language production affects language comprehension. Psycholinguistic models of language comprehension that grant a central role to language experience (e.g., MacDonald, 1999, 2013) propose that the frequency with which certain strings are produced by speakers affects cognitive representations in profound ways: frequency of usage influences not only the way in which linguistic chunks are stored in memory and are linked to other stored items, but it also affects the ease with which linguistic items are accessed and processed. In support of this, recent psycholinguistic studies have provided evidence in favor of the claim that combinatorial distributional information associated with particular lexical items, which is derived from exposure to language, influences ease of processing during sentence comprehension (MacDonald, 2013). In these studies, sentence complexity effects observed during comprehension are said to derive from particular distributional patterns in production, which in turn create distributional regularities that shape comprehenders' interpretations (e.g., Boland, Tanenhaus, Carlson, & Garnsey, 1989; Britt, 1994; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Gennari & MacDonald, 2009; Holmes, Stowe, & Cupples, 1989; MacDonald, Pearlmutter, & Seidenberg, 1994; MacDonald & Thornton, 2009; McRae, Hare, Elman, & Ferretti, 2005; Novick, Thompson-Schill, & Trueswell, 2008; Sedivy & Spivey-Knowlton, 1994; Trueswell & Tanenhaus, 1994). However,

the vast majority of this work comes from studies with monolingual speakers. By examining how language comprehension proceeds when bilingual speakers process codeswitches that occur with different frequencies of usage, our aim is to broaden the evidential base of psycholinguistic models that propose a strong correspondence between production patterns and comprehension difficulty.

To recapitulate, the main goal of the work presented here[1] is to observe whether there are processing differences between a group of bilingual speakers who are members of a unified and well-established community of codeswitchers and a group comprised of a more varied, idiosyncratic combination of bilinguals. We achieve this by examining whether differences in production frequency of two types of codeswitches are reflected in the bilinguals' comprehension patterns.

2. Method

2.1 Participants

In this section, we present the similarities and differences between the two groups of bilinguals examined in this study. The description of the participant groups derives from recent census data, from the results that participants provided in a detailed language history questionnaire, and from the scores that they obtained in two proficiency measures – a vocabulary naming task and a grammar test – which participants completed in English and Spanish. As will be shown, although the participant groups share certain similarities, they display important differences that lead us to expect distinct processing patterns concerning the comprehension of codeswitches.

The two communities of bilinguals examined here were residents of Pennsylvania and New York. One group of speakers (n = 18; mean age 21) lived in State College, a college town in the center of Pennsylvania, where the Hispanic and Latino population nearly reaches 4% (U.S. Census Bureau, 2010a). Many of the State College participants were immigrants from Hispanic countries or had lived in U.S. cities with a large Hispanic presence, but they had moved to the State College area to pursue undergraduate or graduate studies at Penn State University (PSU). The student body at the PSU University Park campus is largely comprised by White students (71%), while the Hispanic students make up 4.7% of the

1. The research reported in this chapter was supported in part by an NSF Doctoral Dissertation Improvement Grant BCS-1123874 to Rosa Guzzardo Tamargo, Paola Dussias, and Chip Gerfen, NSF grants BCS-1535124 and OISE-0968369 and NIH Grant HD082796 to Judith Kroll and Paola Dussias and NIH Grant HDO71758 to Paola Dussias.

student population (Penn State University Budget Office, 2011). The second group of speakers (n = 42; mean age 21) lived in different areas of New York City – mainly the Bronx, Harlem, and Brooklyn – where the Hispanic and Latino population reaches 28.6% (U.S. Census Bureau, 2010b). This group of participants comprised students from The City College of New York (CCNY), located in West Harlem. In CCNY, Hispanics constitute the largest ethnic group, representing approximately 32% of the student body, whereas White students make up about 24% (The City College of The City University of New York, n.d.).

As Poplack (1980) explained in her study on the bilingual community in Harlem, these participants are part of a large, established community of bilingual codeswitchers. Meanwhile, 240 miles away from Harlem, the State College participants are more isolated from larger bilingual communities in which CS is a common practice and, thus, constitute more of a varied mix of individual bilingual codeswitchers. Even if this can only be mentioned anecdotally, it is worth noting that, during data collection, it became evident that there was a clear difference in terms of the linguistic environment in both communities: while State College was overwhelmingly English-dominant, in City College and its Harlem surroundings, one could hear interactions in both English and Spanish, coupled with constant CS. Although both participant groups constitute student populations, the census data clearly illustrate a difference in the participants' linguistic and cultural context. Unlike State College, Harlem is characterized by the strong presence of a culturally and socially cohesive group of Spanish-English bilingual speakers.

The similarities exhibited by the participant groups concern their overall language use as well as their language proficiency. All participants were either born in the United States or had immigrated there at a very early age. In the language history questionnaire, they reported speaking English and Spanish since childhood and they reported using both languages on a regular basis at the time of the study. When asked about their general language preferences, about one third of the participants (33% in State College and 36% in Harlem) preferred English, while a small percentage (6% in State College and 7% in Harlem) preferred Spanish. The remaining participants in each group (61% in State College and 57% in Harlem) reported no particular language preference and, rather, enjoyed using both languages equally. Table 1 displays the mean proficiency self-ratings that participants provided in the language history questionnaire as well as the mean scores that they obtained in the two proficiency measures used in this study.

Overall, both groups of bilinguals were proficient in English and Spanish, although their mean scores were consistently higher in English than they were in Spanish for each proficiency measure. Paired-samples *t* tests conducted within each group for each proficiency measure showed that both the State College and the Harlem participants could be categorized as English-dominant, at least with

Table 1. Mean results for participant proficiency measures

Measure	Language	Group				Independent-samples *t* tests
		State College	Paired-samples *t* tests	Harlem	Paired-samples *t* tests	
Self-ratings (/10)	English	9	$p = .041$	9	$p = .001$	$p = .579$
	Spanish	8		8		$p = .645$
Vocab test (/30)	English	20	$p = .034$	20	$p < .001$	$p = .830$
	Spanish	15		14		$p = .339$
Grammar test (/50)	English	44	$p < .001$	40	$p < .001$	$p = .005$
	Spanish	37		35		$p = .201$

respect to these particular measures. Moreover, independent-samples *t* tests conducted between both groups for each proficiency measure revealed no significant differences, with the exception of the English grammar test. In this particular measure, the State College participants obtained significantly higher scores than the Harlem participants.

Despite the similarities presented above, several differences between the participant groups arose in their responses to the language history questionnaire. Whereas the Harlem participants were mostly of Dominican descent, the State College participants comprised a more varied group, which included speakers of mainly Mexican, Puerto Rican, and Dominican descent. When asked about the cultures with which they identified, the Harlem participants displayed more of a mixed identity than the State College participants. While 76% of the Harlem participants expressed identification with both Hispanic/Latino and U.S. American cultures, 50% of the State College participants identified only with the Hispanic/Latino culture. Although both participant groups considered either Spanish or both Spanish and English as their home language, when it came to the languages used outside of their home, the State College participants reported using more English, while the Harlem participants tended to use both.

More differences between both groups became evident when they answered questions about their exposure to both languages and their CS practices. Consistent with our impressions, 83% of the State College participants stated that they were more exposed to English, while 17% said that they were equally exposed to English and Spanish. A more mixed picture arose within the Harlem group: 38% reported more overall exposure to English, 14% acknowledged more exposure to Spanish, and nearly half (48%) reported equal exposure to both languages. According to the answers provided by the participants in the language history questionnaire, it seems that the Harlem participants codeswitched more than the State College participants. In terms of CS frequency, 44% of the State College

participants indicated that they codeswitched "sometimes," while 56% said that they codeswitched "most of the time." In the Harlem group, 45% of the participants said that they codeswitched "sometimes," 33% said that they codeswitched "most of the time," and the remaining 21% reported codeswitching "always." When asked with whom they codeswitched, the State College participants mainly reported codeswitching with family members and friends. The Harlem participants, however, seemed to be less restrictive with their CS practices: besides codeswitching with relatives and friends, they reported codeswitching with bilingual acquaintances in general.

The questionnaire also included questions regarding the connection between the participants' CS behavior and their identity. More bilinguals in Harlem (52%) than in State College (44%) considered their CS practices an important part of their identity. The remaining participants in each group thought that CS did not have a significant effect on their identity (17% in State College; 33% in Harlem) or were not sure of a connection between the two (39% in State College; 14% in Harlem). It is possible that the participants who, despite codeswitching frequently, did not relate this bilingual behavior to their identity, were affected by the social stigma that is still often attached to CS. When asked about their reasons for codeswitching, once again both groups differed in their responses. The majority of the responses provided by the State College participants reflected lexical need ("Because sometimes when a word doesn't come out in a language quickly it comes out in the other faster," ela_f[2]) or expressive appropriateness ("Sometimes there are words in English that describe what I am trying to say and other times Spanish can describe it better," cco_f). Coinciding with Poplack's description of the participants in her 1980 study, the Harlem participants in this study seemed less aware of their specific reasons for codeswitching, other than stating that it was simply related to their bilingual identity: "I'm not sure [why I codeswitch]. It depends on the person I'm speaking with, but sometimes it's a comfort thing," (cto_f) "Usually because it feels more comfortable," (dma_m) "because I think in both languages," (mja_f) "I'm not really sure. It may be to express myself better," (mca_f) "Sometimes it just comes out that way because you know the other person understands you," (joca_f) "I code-switch because I think it sounds ok when I mix them. Other times it just happens without me noticing," (kro_f) "Because it comes natural. I don't really think about it and sometimes I don't even realize I am doing it" (jeca_f). Interestingly, one participant indirectly referenced the important role of exposure by stating that she codeswitches "from hearing others do it" (lhe_f).

2. After each participant quotation, we provide the participant code and gender (m = masculine; f = feminine).

To summarize, the participant groups share some general similarities with respect to their status (Hispanic students who have lived in the Unites States for all or most of their lives), their proficiency in both Spanish and English, their regular use of both languages, and their lack of preference for either language. Nonetheless, crucial differences between these groups surface in the linguistic and cultural environment in which they live, their identity, their exposure to both languages, and their CS practices. The Harlem participants belonged to a more unified (mostly Dominican), large, established community of codeswitchers while the State College participants comprised a more varied, reduced, and isolated group of codeswitchers. The Harlem participants identified with Hispanic/Latino and U.S. American cultures, but the State College participants identified more with the Hispanic/Latino culture only. The Harlem participants appeared to be exposed to both languages at a higher rate than the State College participants; they also codeswitched more frequently and with a more diverse group of interlocutors than the State College participants. Finally, CS seemed to come more naturally for the Harlem participants, who seemed to engage in this bilingual practice more unconsciously, simply because they considered it part of their identity. Meanwhile, the State College participants tended to pinpoint very specific reasons for codeswitching, mainly related to lexical need or expressive appropriateness. These factors could potentially have an effect on the participants' comprehension costs associated with CS.

2.2 Materials and procedure

As mentioned in the introduction, this study examined the role of the speech community in the comprehension of two types of codeswitched structures. We focused on potential costs associated with the comprehension of codeswitches, which are predicted to reflect CS distributional patterns and CS production tendencies. We focused specifically on reading comprehension. Although CS mostly occurs during oral production, there is, nevertheless, broad consensus that reading activates the system employed in auditory language processing (e.g., Perfetti, 1999). In the parsing literature, for example, Fodor (1998) proposes that syntactic processing during reading proceeds through subvocal phonological encoding of the prosody that guides auditory comprehension. Steinhauer and Friederici (2001) provided confirming evidence for this, showing that the processing of phrases in both reading and auditory comprehension is reflected in the same ERP component. We focus on reading because reading data linked with speed of performance have long been used in psycholinguistics to identify the cognitive processes associated with – and to develop theories of – language comprehension. In this way, our results are interpretable in the context of the vast general literature on reading

and comprehension processes. In addition, it is important to note that for the language pair under investigation here, Spanish-English CS is increasingly common in text and in the bilingual's reading experience as well. This is particularly true of email and chat environments, as seen in the extensive written CS corpus of Montes-Alcalá (2005a, 2005b, 2005c). In fact, CS is also long attested in literary contexts, as in the work of Latino writers, such as Alurista and Ricardo Sánchez.

The participants in this study read codeswitched sentences on a computer screen while their eye movements were recorded with an SR Research EyeLink 1000 eye tracker. After each sentence, the participants answered a comprehension question by pressing the "yes" button or the "no" button on a gamepad. All the sentences began in Spanish and codeswitched into English approximately midway into the sentence. In the experimental sentences, the switch occurred in an embedded clause and it always involved the auxiliary phrase. Two factors were considered in the creation of the experimental materials: the exact location of the switch (switch site) and the particular auxiliary involved in the switch (auxiliary type). Regarding switch site, the switch occurred either at a phrasal boundary, that is, between the subject and the predicate of the embedded clause, or between the auxiliary and the verbal participle within the predicate of the embedded clause. With respect to auxiliary type, the switch involved either the *estar* 'be' auxiliary or the *haber* 'have' auxiliary. As a result, four experimental conditions were generated, as exhibited in Table 2.

The participants read 32 experimental sentences, eight for each of the four experimental conditions. In addition to the experimental sentences, participants read five practice sentences, as well as 32 filler sentences, which were pseudorandomly interleaved among the experimental sentences. In the fillers, which also started in Spanish and codeswitched into English, the switch involved diverse structures and appeared at different points in the sentence.

Because this study focuses on the influence of CS production patterns – and exposure to them – on comprehension costs, it was important to choose codeswitched structures that differed with respect to the frequency of occur-

Table 2. Experimental conditions and sample sentences

Switch site	**Auxiliary type**	
	***estar* 'be'**	***haber* 'have'**
Switch at auxiliary	1. *El alcalde dijo que los escultores* are creating a statue for the park.	3. *El alcalde dijo que los escultores* have created a statue for the park.
Switch at participle	2. *El alcalde dijo que los escultores están* creating a statue for the park.	4. *El alcalde dijo que los escultores han* created a statue for the park.

* Translation: 'The mayor said that the sculptors are creating/have created a statue for the park.'

rence in spontaneous CS production. Of particular relevance for our purposes, several studies examining the production of Spanish-English codeswitches (e.g., Callahan, 2004; Guzzardo Tamargo, 2012; Lipski, 1978, 1985; Pfaff, 1979; Poplack, 1980) have documented a production asymmetry involving alternations within the auxiliary phrase – the locus under investigation in the current study. Specifically, the *estar*+participle switches (Condition 2 in Table 2) are more frequently found in natural bilingual corpora than the *haber*+participle switches (Condition 4). The lexical differences between the verbal phrases of Conditions 2 and 4 could possibly constitute a confound, if we were to compare them directly. Therefore, we decided to compare each of these conditions to respective baseline conditions (Conditions 1 and 3), which included a very frequent type of switch – a switch at a phrasal boundary – and whose only difference, compared to the experimental conditions, involved the language of the auxiliary.

Based on findings from previous studies on natural Spanish-English production, we predicted differences between the bilinguals' comprehension of the *estar*+participle switches and the *haber*+participle switches. Specifically, we expected the *haber*+participle switches to produce more processing difficulty than the *estar*+participle switches, when each of them was compared to its baseline condition. These processing differences between *haber*+participle switches and *estar*+participle switches would reflect the CS patterns found in natural corpora, in which the *estar*+participle switches appear more frequently than the *haber*+participle switches. Based on the differences between the State College community and the Harlem community discussed in the previous section, we expected the Harlem participants to be more sensitive to the differences between the *estar*+participle switches and the *haber*+participle switches, given that (1) they potentially have more CS experience than the State College participants and (2) their CS experience has been more homogenous than that of the State College participants. In other words, we expected the Harlem participants to show greater processing differences between the two types of codeswitches than the State College participants.

3. Results

To examine the processing difficulties of the *estar*+participle switches and the *haber*+participle switches, we extracted reading measures of the participle in each of the experimental conditions. We chose the participle as the critical region because it is the moment during reading in which the participant has processed both the structures and the codeswitches under examination. We extracted two reading measures of the critical region for analysis: gaze duration and total time.

Table 3. Percent of correct responses to the comprehension questions by condition and participant group

Condition	Percent of correct responses	
	State College participants	Harlem participants
1. *estar* – switch at auxiliary	93%	90%
2. *estar* – switch at participle	91%	91%
3. *haber* – switch at auxiliary	91%	91%
4. *haber* – switch at participle	90%	92%

Gaze duration is a measure of early processing and refers to the sum of all fixation durations in the critical region from first entering it until leaving it (Rayner, 1998, p. 377). Total time is a measure of later processing and represents the sum of all fixation durations in the critical region, including all regressive fixation durations to it (Rayner & Duffy, 1986, p. 196). By analyzing both of these reading measures, we were able to determine different levels of sensitivity to the types of codeswitches under examination. Before discussing the eye-tracking results for each participant group, it is important to mention their accuracy results on the comprehension task. Table 3 displays the percent of correct responses to the comprehension questions by condition for each group of participants. For all four experimental conditions, the participants mean accuracy is equal to or greater than 90%, revealing that both groups of participants were paying attention to the task and that they understood the content of the sentences that they read.

In order to analyze the eye-tracking data, we conducted separate analyses for each participant group. Two-way repeated measures analyses of variance (ANOVA) evaluated the effect of auxiliary type and switch site on the two reading measures (gaze duration and total time). Auxiliary type (*estar* versus *haber*) and switch site (switch at auxiliary versus switch at participle) were the within-subjects factors. First, we present the results for the Harlem participants, followed by those for the State College participants. For gaze duration, the results for the Harlem group yielded a main effect of switch site, $F(1,41) = 18.04$, $p < .001$, but not of auxiliary type, $F(1,41) = .26$, $p = .616$. There was a significant by-participant interaction of auxiliary type and switch site, $F(1,41) = 5.30$, $p = .026$. Paired-samples t tests were conducted to follow up the significant interaction. They indicated significant mean differences between Conditions 3 and 4, $t(41) = 4.67$, $p < .001$. However, no significant mean differences were found between Conditions 1 and 2, $t(41) = 1.57$, $p = .123$. For total time, the results bore a main effect of auxiliary type, $F(1,41) = 7.41$, $p = .009$, a main effect of switch site, $F(1,41) = 27.87$, $p < .001$, and a significant by-participant interaction of auxiliary type and switch site, $F(1,41) = 9.49$, $p = .004$. Subsequent pairwise contrasts revealed significant

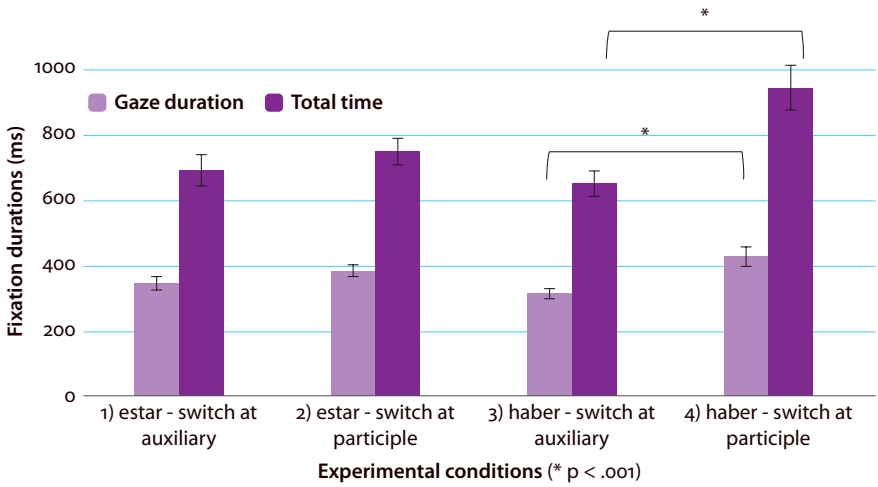

Figure 1. Mean fixation durations in the critical region by condition for the Harlem participants

mean differences between Conditions 3 and 4, $t(41) = 5.22$, $p < .001$, but not between Conditions 1 and 2, $t(41) = 1.32$, $p = .194$. Therefore, during the measures of both early and later processing, the Harlem participants read *estar*+participle switches and their corresponding phrasal boundary switches similarly. However, they read *haber*+participle switches significantly more slowly than their analogous phrasal boundary switches. The eye-tracking results for the Harlem participants are displayed in Figure 1.

The same analyses were conducted for the group of State College participants. For gaze duration, the results showed a main effect of auxiliary type, $F(1,17) = 5.21$, $p = .036$, a main effect of switch site, $F(1,17) = 39.74$, $p < .001$, and a significant by-participant interaction of auxiliary type and switch site, $F(1,17) = 4.76$, $p = .043$. Paired-samples *t* tests were conducted to follow up the significant interaction. They indicated significant mean differences between Conditions 3 and 4, $t(17) = 5.31$, $p < .001$, but not between Conditions 1 and 2, $t(17) = 1.21$, $p = .244$. For total time, the results of the repeated measures ANOVA yielded a main effect of auxiliary type, $F(1,17) = 8.29$, $p = .01$, and of switch site, $F(1,17) = 15.19$, $p = .001$. Moreover, the results displayed a significant by-participant interaction of auxiliary type and switch site, $F(1,17) = 13.49$, $p = .002$. Once more, the follow-up paired-samples *t* tests revealed significant mean differences between Conditions 3 and 4, $t(17) = 4.97$, $p < .001$, but not between Conditions 1 and 2, $t(17) = 1.11$, $p = .282$. For the State College group, *haber*+participle switches were read significantly more slowly than when the switches occurred at the phrasal boundary but *estar*+switches did not show significant reading differences. This pattern arose

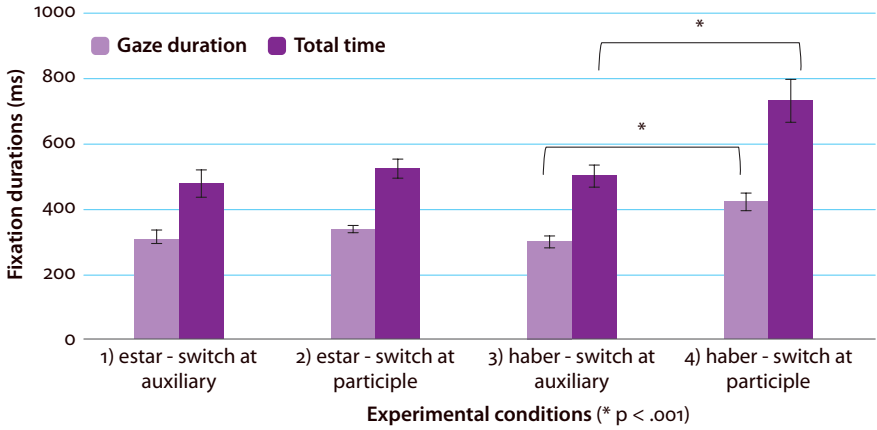

Figure 2. Mean fixation durations in the critical region by condition for the State College participants

with the measures of early and later processing alike. Figure 2 exhibits the eye-tracking results for the group of State College bilinguals.

4. Discussion

The results presented above reveal the way in which two groups of participants process two types of codeswitches. Although both types of switches involve a Spanish auxiliary and an English participle, *estar*+participle switches are more frequent in bilingual codeswitched production than *haber*+participle switches (Callahan, 2004; Guzzardo Tamargo, 2012; Lipski, 1978, 1985; Pfaff, 1979; Poplack, 1980). Models of monolingual comprehension (e.g., Gennari & MacDonald, 2009; MacDonald, 2013; MacDonald & Thornton, 2009), predict that for bilinguals who belong to communities in which Spanish-English CS is frequently present, differences in the distributional patterns of diverse types of codeswitches should be reflected in the bilinguals' comprehension costs. The two groups of participants included in this study differed in terms of their membership in a community of codeswitchers. While the Harlem participants belonged to an established CS community, the State College participants constituted more of a heterogeneous group of bilingual codeswitchers with different Hispanic backgrounds and from different areas in the United States. Given the differences between both groups of participants, we expected the distinct distributional patterns of *estar*+participle and *haber*+participle switches to be more clearly reflected in the results of the Harlem participants than in those of the State College participants.

The results for the Harlem participants matched our predictions. These speakers did not display any additional processing costs when they read the *estar*+participle switches, compared to switches that occurred at a phrasal boundary. However, they showed significant processing differences between the *haber*+participle switches and the corresponding switches at the phrasal boundary (the baseline condition). In this case, reading *haber*+participle switches entailed more processing costs than reading switches at the phrasal boundary. These processing costs reflect the CS patterns found in natural Spanish-English production corpora. By means of their comprehension, these participants proved to be sensitive, from very early on during processing, to the frequency differences between the two types of switches. The results of the State College participants were surprisingly similar to those of the Harlem participants. That is, the State College bilinguals displayed greater processing costs during the comprehension of *haber*+participle switches than during the comprehension of *estar*+participle switches, when each of these conditions was compared to its baseline condition. And, like the Harlem participants, the State College participants exhibited these processing costs differences both in the early and later measures of processing. In other words, these speakers were just as sensitive as the Harlem participants to the frequency differences of *estar*+participle and *haber*+participle switches found in spontaneous production.

These results lead us to consider two possible explanations. First, the increased frequency of *estar*+participle switches over *haber*+participle switches may be structural in nature. In Spanish, the auxiliary *estar* has been considered to be less bounded to its participle than *haber* (e.g., Torres Cacoullos, 1999). For instance, *estar* is not always accompanied by a past participle; it can occur with other expressions such as adverbial phrases (e.g., "están en casa" / '[they] are at home') and adjectival phrases (e.g., "están enfermos" / '[they] are sick'). Moreover, when *estar* is accompanied by a participle, it does not have to appear adjacent to the participle: intervening material between *estar* and the participle is permissible (e.g., "están siempre trabajando" / '[they] are always working'). In addition, *estar* and the present participle can exchange positions (e.g., "trabajando están" / 'working [they] are'). The situation with *haber* is quite different. When *haber* occurs, it is always accompanied by a past participle; the only exceptions occur with the use of the impersonal *haber* (e.g., "había café para todos" / 'there was coffee for everyone') and certain expressions that are not productive (e.g., "he aquí la respuesta" / 'have here [here is] the answer'). In addition, *haber* and its past participle cannot be separated by any intervening material and their positions in a sentence are fixed. Related to all this is the fact that *estar* still retains part of its original lexical locative meaning (Torres Cacoullos, 1999) while *haber* seems to have grammaticalized to the point that it has lost all of its original meaning of possession. The

lexical freedom of *estar* may be associated to the possibility of a switch between *estar* and its following word or phrase – in this case, the participle. The auxiliary *haber*, however, does not have the same lexical freedom as *estar*, and is, instead, overwhelmingly followed by its participle in Spanish. This degree of boundedness could be responsible for the scarce occurrence of switches between *haber* and its following participle. If this were, indeed, the case, then the CS patterns and the processing costs associated with *estar*+participle switches and *haber*+participle switches would be common to all Spanish-English bilingual communities, as long as they have similar levels of Spanish proficiency.

A second possibility is that the codeswitching practices of the State College participants are more similar to those of the Harlem participants – both qualitatively and quantitatively – than we initially anticipated. Although to a lesser degree than the Harlem community, the State College participants may codeswitch enough and may be sufficiently exposed to CS to display processing differences between frequent (*estar*+participle) and infrequent (*haber*+participle) codeswitches. In the case of the Harlem group, most of the participants had spent most of their lives in Harlem, occasionally combined with other CS communities, such as the Bronx. In State College, however, the situation was different. These participants had usually spent part of their lives in a Spanish-speaking country or in other parts of the United States, but had moved to State College, more of a monolingual English community, to complete their undergraduate and graduate degrees. This suggests that the participants in State College constituted less of a unified community and more of a heterogeneous combination of bilinguals. Nonetheless, it is important to note that most of these Spanish-speaking participants would join university groups and associations that brought the Hispanic or Latino students together. We can confidently state that many, if not all, of the State College participants belonged to these associations because one of the recruitment strategies for this study included contacting these particular groups. Therefore, these participants may have developed more of a cohesive community than one would have thought, in which Spanish-English CS is very present. This possibility suggests that, sometimes, the more varied groups of bilinguals that are typically recruited for lab studies can display CS practices, both in production and in comprehension, that are similar to those of more established communities.

To determine which of these two possibilities is more likely, additional data is needed. In the first case, it would be necessary to collect data from a group of highly proficient bilinguals who belong to a community of non-codeswitchers. If this group were to display similar reading patterns, then it would suggest that structure-specific reasons related to the two auxiliaries in Spanish account for how these two types of switches are processed by bilingual speakers. In the second case, a more detailed ethnographic study of the bilingual community in State

College would provide helpful samples of their bilingual behavior, and, specifically, their CS practices. If their CS production patterns clearly resemble the patterns that have been reported for the Harlem participants, then the results in this study would no longer seem so surprising. For the time being, we believe that the second possibility is more plausible, given the results of previous psycholinguistic studies that have reported clear correspondences between production patterns and comprehension costs. The Spanish-English bilinguals in State College may be part of a fluent and coherent community of codeswitchers. It may well be that precisely because of State College's isolation from other larger, more established bilingual communities, these bilinguals have created a close-knit community of codeswitchers. This picture may become clearer if we hypothesize that codeswitches fall on different points of a continuum. At one end are switches that are constrained by the interactions of the two grammatical systems that come in contact; at the other are switches that are emergent and derive from community-driven practices. We propose that while the codeswitches on the first end of the continuum would not present major processing challenges for any bilingual speaker, those on the other end of the continuum would only be easily processed by speakers of the particular communities in which those switches arise and become community norms. We are unable to examine this idea with our current data; it does, however, merit further investigation.

5. Conclusion

This study set out to answer the following question: do speakers who belong to a stable bilingual community in which CS is a common practice process codeswitches differently from bilinguals whose membership to a CS community is more idiosyncratic? To address this question, the eye-movement patterns of two groups of speakers were compared: a group of Spanish-English bilingual speakers from Harlem, the home of a well-defined speech community where CS occurs regularly, and a group of Spanish-English bilinguals who more closely resemble the participants in past lab studies. The two types of switches examined included frequent *estar*+participle switches and less frequent *haber*+participle switches. The participants completed a language history questionnaire and a series of proficiency measures in order to gather information about their language skills, acquisition, use, and exposure. This information, coupled with our impressions during data collection, helped us to grasp the differences between the two communities. The bilinguals of both communities were proficient in both languages and used Spanish and English on a regular basis. However, CS seemed to be more present in the life of the Harlem participants than in that of the State College participants.

The State College community appeared to be English-dominant, while the Harlem neighborhood could be identified as a truly balanced bilingual community. In her 1987 study, Poplack found three to four times more switching in Ottawa than in Hull because the Ottawa community had French and English influence, while the Hull community had more French and less English influence in the environment (p. 60). Based on these findings, it was reasonable to expect differences in the processing patterns of the Harlem and the State College participants, due to differences in their use of and exposure to CS. The results of the eye tracking study displayed unexpected similarities between the processing patterns of both participant groups. The Harlem and the State College participants processed the more common *estar*+participle switch with more ease than the less frequent *haber*+participle switch. These findings do not suggest that studying bilinguals who belong to CS communities is not necessary. Instead, they point towards the development of a community of codeswitchers in State College whose CS practices resemble, at least to a certain degree, those of a stable community of codeswitchers, such as the Harlem community.

References

Boland, J. E., Tanenhaus, M. K., Carlson, G., & Garnsey, S. M. (1989). Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, 18(6), 563–576. doi:10.1007/BF01067160

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33(2), 251–283. doi:10.1006/jmla.1994.1013

Callahan, L. (2004). *Spanish/English codeswitching in a written corpus*. Amsterdam: John Benjamins. doi:10.1075/sibil.27

Carter, D., Deuchar, M., Davies, P., & Parafita Couto, M. C. (2011). A systematic comparison of factors affecting the choice of matrix language in three bilingual communities. *Journal of Language Contact*, 4, 153–183. doi:10.1163/187740911X592808

Dussias, P. E. (2003). Spanish-English code mixing at the auxiliary phrase: evidence from eye-movement data. *Revista Internacional de Lingüística Iberoamericana*, 1(2), 7–34.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2), 285–319. doi:10.1023/A:1023258301588

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58–93. doi:10.1006/jmla.1997.2512

Gennari, S. P., & MacDonald, M. C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111, 1–23. doi:10.1016/j.cognition.2008.12.006

Gullifer, J., Kroll, J. F., & Dussias, P. E. (2013). When language switching has no apparent cost: lexical access in sentence context. *Frontiers in Psychology*, 4, 1–13. doi:10.3389/fpsyg.2013.00278

Guzzardo Tamargo, R. E. (2012). *Linking comprehension costs to production patterns during the processing of mixed language*. (Unpublished PhD dissertation). The Pennsylvania State University, University Park, PA.

Holmes, V. M., Stowe, L., & Cupples, L. (1989). Lexical expectations in parsing complement-verb sentences. *Journal of Memory and Language*, 28(6), 668–689. doi:10.1016/0749-596X(89)90003-X

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2010). Syntactic alignment and shared word order in code-switched sentence production: Evidence from bilingual monologue and dialogue. *Journal of Memory and Language*, 63, 210–231. doi:10.1016/j.jml.2010.03.006

Kootstra, G. J., Van Hell, J. G., & Dijkstra, T. (2012). Priming of code-switches in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, 15(4), 797–819. doi:10.1017/S136672891100068X

Litcofsky, K. A. (2013). *Sentential code-switching and lexical triggering: A neurocognitive study*. (Unpublished MA thesis). The Pennsylvania State University, University Park, PA.

Lipski, J. M. (1978). Code-switching and the problem of bilingual competence. In M. Paradis (Ed.), *Aspects of bilingualism* (pp. 250–264). Columbia, SC: Hornbeam Press.

Lipski, J. M. (1985). *Linguistic aspects of Spanish-English language Switching*. Tempe, AZ: Center for Latin American Studies, Arizona State University.

MacDonald, M. C. (1999). Distributional information in language comprehension, production and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Lawrence Erlbaum Associates.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226. doi:10.3389/fpsyq.2013.00226

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703. doi:10.1037/0033-295X.101.4.676

MacDonald, M. C. & Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 581–611). Amsterdam: Elsevier/Academic Press. doi:10.1016/B978-012369374-7/50016-X

MacDonald, M. C., & Thornton, R. (2009). When language comprehension reflects production constraints: Resolving ambiguities with the help of past experience. *Memory & Cognition*, 37, 1177–1186. doi:10.3758/MC.37.8.1177

McRae, K., Hare, M., Elman, J. L., & Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7), 1174–1184. doi:10.3758/BF03193221

Montes-Alcalá, C. (2005a). ¡Mándame un e-mail! Cambio de códigos español-inglés online. In L. A. Ortiz López & M. Lacorte (Eds.), *Contactos y contextos lingüísticos: El español en los Estados Unidos y en contacto con otras lenguas* (pp. 173–185). Madrid, Spain: Lingüística Iberoamericana.

Montes-Alcalá, C. (2005b). When Spanish meets English: An overview of bilingualism in the United States. In R. Ahrens & K. Stierstorfer (Eds.), *Symbolism: An international annual of critical aesthetics*. Vol. 4. New York, NY: AMS Press.

Montes-Alcalá, C. (2005c). "Dear amigo": Exploring code-switching in personal letters. In L. Sayahi & M. Westmoreland (Eds.), *Selected Proceedings of the Second Workshop on Spanish Sociolinguistics* (pp. 102–108). Somerville, MA: Cascadilla Proceedings Project.

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching *palabras* (words): An electrophysiological study of code switching. *Brain and Language*, 80, 188–207. doi:10.1006/brln.2001.2588

Novick, J. M., Thompson-Schill, S. L., & Trueswell, J. C. (2008). Putting lexical constraints in context into the visual-world paradigm. *Cognition*, 107(3), 850–903. doi:10.1016/j.cognition.2007.12.011

O'Malley-Madec, M. (2007). How one word borrows another: The process of language-contact in two Irish-speaking communities. *International Journal of Bilingual Education and Bilingualism*, 10(4), 494–509. doi:10.2167/beb456.0

Penn State University Budget Office. (2011). Enrollment by Race/Ethnicity. Retrieved from <http://www.budget.psu.edu/FactBook/StudentDynamic/MinorityEnrolbyEthnicity.aspx?YearCode=2011&FBPlusIndc=N>

Perfetti, C. A. (1999). Comprehending written language: a blueprint of the reader. In Peter Hagoort & Colin M. Brown (Eds.), *The neurocognition of language processing* (pp. 167–208). Oxford: Oxford University Press.

Pfaff, C. W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55(2), 291–318. doi:10.2307/412586

Polack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7–8), 581–618.

Poplack, S. (1987). Contrasting patterns of code-switching in two communities. In E. Wande, J. Anward, B. Nordberg, L. Steensland, & M. Thelander (Eds.), *Aspects of multilingualism: Proceedings from the Fourth Nordic Symposium on Bilingualism, 1984* (pp. 51–77). Motala, Sweden: Borgströms.

Proverbio, A. M., Leoni, G., & Zani, A. (2004). Language switching mechanisms in simultaneous interpreters: An ERP study. *Neuropsychologia*, 42, 1636–1656. doi:10.1016/j.neuropsychologia.2004.04.013

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201. doi:10.3758/BF03197692

Sedivy, J., & Spivey-Knowlton, M. (1994). The use of structural, lexical, and pragmatic information in parsing attachment ambiguities. In C. Clifton, L. Frazier, & K. Rayner, *Perspectives on sentence processing* (pp. 389–413). Hillsdale, NJ: Lawrence Erlbaum Associates.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain response: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, 30(3), 267–295. doi:10.1023/A:1010443001646

The City College of The City University of New York. (n.d.). City College Facts. Retrieved from <http://www1.ccny.cuny.edu/advancement/news/factsheet.cfm>

Torres Cacoullos, R. (1999). Variation and grammaticalization in progressives: Spanish –ndo constructions. *Studies in Language*, 23(1), 25–59. doi:10.1075/sl.23.1.03tor

Torres Cacoullos, R., & Travis, C. E. (2015). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism*, 19(4), 365–386. doi:10.1177/1367006913516046

Trillo, M. E. (2003). The code-switching patterns of the Rio Linda community of El Chamizal in El Paso, Texas: An emic perspective of syntactic constraints. *Dissertation Abstracts International: Section A. The Humanities and Social Sciences*, 2524-A.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner, *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Lawrence Erlbaum Associates.

U.S. Census Bureau. (2010a). *State & County QuickFacts*: New York City, NY. Retrieved from <http://quickfacts.census.gov/qfd/states/36/3651000.html>

U.S. Census Bureau. (2010b). *State & County QuickFacts*: State College, PA. Retrieved from <http://quickfacts.census.gov/qfd/states/42/4273808.html>

Van Hell, J., Litcofsky, K. A., & Ting, C. Y. (2015). Intra-sentential code-switching: cognitive and neural approaches. In J. W. Schwieter (Ed.), *The Cambridge handbook of bilingual language processing* (pp. 459–482). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107447257.020