Bilingual discourse markers in Puerto Rican Spanish

L O U R D E S   T O R R E S

Latin American & Latino Studies Program
DePaul University
802 West Belden Avenue
Chicago, IL 60614-3214
ltorres@depaul.edu

A B S T R A C T

This study examines bilingual discourse markers in a language contact situation. The focus is on how English-dominant, bilingual, and Spanish-dominant New York Puerto Ricans integrate English-language discourse markers into their Spanish-language oral narratives. The corpus comprises 60 Spanish-language oral narratives of personal experience extracted from transcripts of conversations with New York Puerto Ricans. After a review of the study of discourse markers in language contact situations, the use of English-language discourse markers is compared to the use of Spanish-language markers in the texts. The discussion considers the question of whether English-language discourse markers are more profitably identified as instances of code-switching or of borrowing. Finally, the essay explores how bilingual speakers integrate English discourse markers in their narratives with a pattern of usage and frequency that varies according to language proficiency. (Languages in contact, bilingual discourse, discourse analysis, discourse markers)*

In the past two decades, research seeking to determine the meaning and function of discourse markers has proliferated (Schiffrin 1987; Redeker 1991; Blackmore 1987, 1992; Fraser 1999; Schourup 1999). Discourse markers – words like *so*, *and*, or *y'know* – may have both grammatical and discourse meanings, and they are multifunctional. Although there are different positions on the question of what items should be designated as discourse markers, and on how they should be analyzed, at a basic level most linguists would agree that discourse markers contribute to the coherence of the discourse by signaling or marking a relationship across utterances. Theoretically speaking, as core lexical items,[1] discourse markers should not be prime targets for incorporation into bilingual speech. Linguists have posited a scale of adoptability with respect to word classes (Haugen 1950, 1956, 1969). According to this scale, nouns are the most likely word type to be borrowed, and the parts of speech that are more grammatical and less lexical are less likely to be borrowed; so, for example, interjections and discourse markers would be unlikely candidates to be borrowed. However, as Thomason and Kauf-

man 1988 point out, when languages are in contact, any linguistic feature may be borrowed from one language system to another. Current research suggests that, in contact situations, the borrowing of core lexical items such as discourse markers is quite common. In this article, I examine the use of English-language discourse markers in the Spanish oral narratives of Puerto Rican speakers with varying degrees of proficiency in Spanish and English. Although English monolingual, bilingual, and Spanish-dominant speakers all integrate English markers into their Spanish-language narratives, their language proficiency, and certain characteristics of different types of English markers condition which markers are borrowed, and how they are integrated into the Spanish-language narratives.

DISCOURSE MARKERS AND CONTACT LANGUAGES

Discourse markers may be prone to borrowing because, according to Myers-Scotton 1993, they are free forms, meaning that they stand on their own as they link propositions. According to Myers-Scotton's Matrix Language-Frame Model of Code-Switching (1993, Myers-Scotton & Jake 1995), discourse markers can also be classified as content morphemes because, like verbs, nouns, and prepositions, they assign thematic roles in discourse.[2] As content morphemes, discourse markers would be more susceptible to borrowing than are system morphemes such as quantifiers, possessives, and determiners. Myers-Scotton 1993 also argues that it is important to distinguish between "cultural" and "core" borrowed forms. Cultural borrowings are lexical items that are appropriated because they represent objects or concepts that are new to speakers and no native terms seem adequate for them. Core borrowing entails the adoption of items (such as discourse markers) that do not fill any lexical gap, and for which native equivalents are readily available. Core borrowings are like code-switches in that they occur with lower frequency than do cultural borrowings, and it is impossible to predict their recurrence. Myers-Scotton hypothesizes that core borrowed forms probably enter the language initially as code-switches and then gradually become integrated in the language as they are used more frequently alongside the native forms they often duplicate.[3]

In her study of code-switching between African languages (Swahili and Shona) and English, Myers-Scotton 1993 finds that Shona is in the process of borrowing several English-language discourse markers. In the bilingual corpus she examined, the English-language markers *because* and *but* appear 7 to 8 percent of the time, while the Shona indigenous morphemes *nokuti* and *asi*, which convey concepts similar to *because* and *but*, occur most of the time. Borrowing of the English markers occurs in the speech of speakers from all social classes and does not address a lexical gap in Shona. Many other studies document the integration of discourse markers in situations of intense language contact.

Salmons 1990 examines English discourse markers in the German speech of US-born German-American bilinguals with varying degrees of proficiency in German. He finds that the adoption of English discourse markers is accompanied by

an almost complete loss of German markers. His main interest is defining whether these discourse markers are borrowings or code-switches. Previous evidence suggests that discourse markers are a frequent locus for code-switching (Pfaff 1982, Brody 1987) and that they may serve as triggers for code-switching. Although Salmons correctly points out that the distinction between code-switching and borrowing is not categorical, the difference is important because each process indicates a different level of speaker competence in the two languages: A speaker who borrows items from another language may be monolingual, but code-switching implies a degree of bilingualism. In the German-American case, Salmons concludes that the discourse markers he studied are borrowings because they have displaced the German modal system. He argues that code-switching would be indicated if speakers had a choice between using English or German discourse markers, but in the German-American context he studied, this was no longer the case.

In a more recent study, using Myers-Scotton's Matrix Language-Frame Model of Code-switching as their theoretical basis, Goss and Salmons (forthcoming) posit a historical trajectory that leads to the incorporation of English-language discourse markers in German and to the subsequent loss of German particles in German-American dialects. Through a diachronic study of the speech in two plays focused on German-American life in the late 1800s and early 1900s, the authors suggest a process that has led to the present situation of complete loss of German discourse markers in many German-American speech communities. Initially, English-language discourse markers were introduced into German speech through emblematic code-switching. The English and German systems coexisted until the German modals began to die out. Finally, English markers totally replaced German markers, and the substitution was complete.

The borrowing of discourse markers across languages has also been amply documented in the case of non-European languages. Brody 1987, 1995, and Solomon 1995 consider the use of Spanish markers in Mayan[4] narratives, and Hill and Hill 1986 study such markers in Mexicano narratives. Brody hypothesizes that Spanish particles in Tojolabal Mayan originated as code-switching. She refers to the use of such markers as emblematic and claims that they constitute indigenous speakers' "foray into the prestige language" (1987:509). Unlike the German-American case, in Tojolabal Mayan, discourse markers do not replace native markers but often serve as doublets to the native markers; in other words, both the Mayan marker and its Spanish equivalent are produced in an utterance. Brody 1995 hypothesizes that, in the future, Spanish discourse markers may replace the Mayan markers, as in the German-American case. Solomon 1995 studies borrowed discourse markers in a Yucatec Mayan narrative and identifies a different pattern. She focuses on the Spanish discourse marker *entonces* and finds that it alternates in the narrative with the Mayan marker *ka*, with which it shares a similar semantic meaning and function. In the narrative text that Solomon studies, the Spanish and Mayan markers are found to have complementary global and local discourse functions.

Maschler 1994 studies Hebrew discourse markers in mostly English talk in order to develop a theory of "metalanguaging" in bilingual conversation. Maschler argues that switched discourse in general may function in a metalinguistic fashion, providing information about language rather than communicating other types of information in the narrational world. Discourse markers serve as boundary markers framing items such as different verbal activities, shifts in contexts, or new components in a narrative.

Thus, a review of literature concerning discourse markers in contact situations has revealed that various outcomes are possible in different contexts. Several authors posit a direct relationship between borrowing and code-switching, hypothesizing that borrowed markers initially enter the language through code-switching. In some cases, borrowed discourse markers replace native markers; they can also appear in addition to the native markers. In other situations, they can function in complementary distribution with native markers or serve as a specific metalinguistic device. As we will see, a cross-generational study of discourse markers in Puerto Rican Spanish offers insight into the actual process of the incorporation of English discourse markers into Spanish speech over time through code-switching and borrowing.

METHODOLOGY AND DATA

The data for this study are extracted from 60 narratives produced by members of three generations of Puerto Ricans from Brentwood, New York. The data were gathered as part of a large project (Torres 1997) examining language practices of this sample group, who live in a suburb of New York City that has the largest Latino community in New York state, excluding New York City itself. Through a combination of ethnographic research, questionnaires, and extensive interviews, I gathered different types of data to enable me to study a range of topics, such as language use patterns, borrowing, code-switching, verb usage, and narrative performance.[5] As a Puerto Rican who lived and worked in the community for three years, I was able to gain easy access to the community and to participate in a range of educational, social, and political activities. In the course of my research, I interviewed 30 adults ranging in age from 18 to 65. The interviews were conducted in Spanish and focused on Latino life in Long Island.

For the present study, I chose 60 Spanish narratives of personal experience (Labov & Waletzky 1967, Labov 1972) from transcripts of the interviews that I conducted with participants. I selected stories of similar length and topics. Oral narratives are a highly structured discourse type consisting of specific components, each of which is associated with specific linguistic properties (Labov & Waletzky, 1967). Narratives are automatically interpreted as having additivity, meaning that each new clause presents new information, and temporality, meaning that events are sequentially ordered (Segan, Duchan & Scott 1991). The components may include an abstract, an orientation, complicating actions,

an evaluation, a resolution, and a coda. The abstract serves as a brief summary of the narrative to be presented. An orientation section establishes the setting, the time, and the participants. The complicating actions are the portions of the narrative where the story unfolds. The evaluation sections, which can be present throughout the narrative or in a separate section, are where the speaker expresses why the narrative is interesting and worth telling. The resolution explains what finally happened, and the coda often serves as a transition from narrative time to the present.

One-third of the narratives analyzed were taken from interviews with first-generation speakers (Group 1/G1) who were born in Puerto Rico and came to New York in their teens. They have lived in Brentwood most of their lives and range in age from 40 to 60. These speakers are Spanish-dominant. Group 2 (G2) speakers, who produced the second group of narratives, came to New York before the age of five or were born there. They have spent most of their lives on Long Island and range in age from 30 to 50. They are bilingual. The third group (Group 3/ G3) comprises English-dominant persons born in Brentwood who were in their late teens or early twenties at the time of the study.

In the present study, following Schiffrin's 1987 general categorization, the discourse markers analyzed fall into the following categories: the discourse connective *and*, which has grammatical properties as a discourse coordinator; markers of cause or result, such as *because* and *so*; and the expressions *y'know* and *I mean*, whose main function is to ensure participation. Although other English language discourse markers appear in my data, I chose to focus on the markers in each of the above three categories that occur most often in my data.

And is the most frequent mode of connecting discourse in monolingual speech. It can be an option for zero, or no connector, between clauses. In this sense, it can be considered the norm in comparison to other connectors; for example, *but*, *so*, and *because* are more likely to appear before a tense switch (Schiffrin 1987). *And*'s high frequency means that there are fewer restrictions on its use than on other markers. *And* can and does occur simultaneously with or in the place of other connectors. Unlike markers that carry social or expressive meaning, *and* merely coordinates clauses, and its meaning cannot be derived independently from the ideas expressed in the clauses that surround it.

The second category I look at includes markers that reflect a relationship of cause, result, consequence, or inference between current and previous utterances. *Because* and *so* function in discourse to mark relationships of cause and result. The semantic meaning of these markers is also relevant in their discourse function. *Because* often marks clauses that indicate subordinate ideas, while *so* often indicates main ideas. The semantic meanings of *because* and *so* are relevant to and act simultaneously with the meaning expressed in the surrounding clauses. Unlike *and*, their semantic meaning constrains their positions in the narrative.

Finally, *y'know* and *I mean* are primarily participation markers which convey a sense of speaker involvement. They may mark general consensus toward ma-

terial and convey the idea of identity, marking speaker and hearer as members of the same community. In monolingual English narratives, Schiffrin 1987 suggests that they are usually located in clauses which internally or externally evaluate a story's high point. They are sometimes used to preface the point of a story, or to shift from a specific story to a general point. They also function as hesitation markers and to hold a speaker's place.

To examine how English-language markers are used as opposed to the Spanish language markers, we must compare the use of discourse markers in Spanish and English. Although it would not be accurate to imply that the meaning, function, and distribution of markers is exactly the same in both languages, we can make the case that some discourse markers are similar both semantically and functionally, and that they can be employed in similar contexts. There are several excellent studies of English-language discourse markers in English monolingual speech[6] (Blackmore 1987, 1992; Schriffin 1987), but Spanish discourse markers are only now beginning to attract serious study (Cortes Rodríguez 1991, Pons Bordería 1998, Brizuela et al. 1999, Cepeda 1999, Poblete Bennett 1999). In a study comparing English and Spanish contrastive discourse markers, Fraser and Malumud-Makowski 1996 conclude that such markers are very similar in terms of their function in oral expression and of the type of interpretations they impose on the utterances that follow them.[7] In the present study, considering how the markers function in the oral narratives, I take the following pairs to be roughly equivalent: *and/y*, *so/entonces*, *because/por que*, *y' know/tú sabes*, and *I mean/digo*. Most of these are fairly transparent and uncontroversial translations, but *so* can have different nuances of meaning; however, the Spanish marker that most closely assumes its function in narrative is *entonces*. Probably it is most accurate to suggest that there is an overlap in meaning and function between *so* and *entonces*, but not an exact equivalence. In the oral narratives, *so* frequently reflects a causal relationship in which the first clause states the cause of the following clause; sometimes, however, it merely reflects temporality, meaning that the newly presented events or states occur later in time than those already presented. *Entonces*, by contrast, always reflects temporality and only sometimes signals a causal relationship.[8] Again, I am not suggesting that any of the markers I am considering are exact equivalents in Spanish and English, but that enough similarity or overlap of meaning, frequency, or discourse functions exists to justify their comparison in bilingual speech.[9]

RESULTS

Given these parameters, I now consider how likely it is that a particular marker will appear in Spanish, English, or both in the Spanish-language oral narratives produced by Brentwood Puerto Rican speakers. Since the use of discourse markers is optional, I study frequency and distribution of actual occurrences of the

relevant discourse markers. Although lexical items that function as discourse markers may have other functions (as conjunctions, adverbs, adjectives, etc.), the count here includes the words only when they function as discourse markers. Table 1 presents the combined distribution of the relevant English and Spanish discourse markers in my data.

Comparing the use of English and Spanish markers shown in Table 1, we see that, as might be expected given their level of contact with English, the Spanish-dominant speakers have the fewest English discourse markers in their speech, while the English-dominant group evidences the highest use of English discourse markers in Spanish-language narratives. The bilingual group more closely resembles the English-dominant in usage, but in general, they have not incorporated English discourse markers in their speech to the degree that English-dominant speakers have.[10]

For the function of linking ideas, actions, or states in narrative, all groups overwhelmingly favor Spanish-language *y* over *and*; this is not surprising, because they are producing Spanish-language narratives, and *y* is the most basic and most common connective used to move stories along in oral Spanish. Since *and* does not carry social or expressive meaning, it is not a likely source for borrowing. G1 speakers never use *and* as a connective in their narratives; G2 and G3 speakers rarely do so, and when they do, it almost always occurs either at the beginning of a code-switched clause or clauses, or as part of a series of code-switched clauses. This is shown in ex. (1), in which the speaker describes helping an immigrant get settled in Brentwood:

(1)

1	so le abrí la puerta
2	y la llevé allí
3	and the next thing I know
4	ella no tenía cama
5	ella no tiene mueble

1	'so I opened the door for her
2	and I took her there
3	and the next thing I know
4	she didn't have a bed
5	she didn't have furniture'

In general, in the narratives of all speakers, when *and* occurs, it is more accurate to refer to it as part of a code-switched sequence rather than as a borrowing. Only twice in this corpus does *and* appear as the only English word in a Spanish sequence. A case can be made that this discourse marker is entering the language as Myers-Scotton's model predicts, via code-switching in the speech of the bilingual speakers. Its low frequency and restricted use in the data make it difficult to predict whether it will be taken on as a borrowing in the future.

Of the cause-and-effect markers, *entonces* and *porque* (*so* and *because*), Group 1 favors the Spanish markers by a 6-to-1 ratio, while Group 2 favors the Spanish markers by a ratio of 2 to 1, and Group 3 speakers are almost equally

TABLE 1. *Discourse markers in English and Spanish*

Types	Group 1			Group 2			Group 3		
	English	Spanish	Total	English	Spanish	Total	English	Spanish	Total
connective	0% (0)	100% (232)	232	5% (8)	95% (146)	154	7% (17)	93% (236)	253
cause/effect	6% (7)	94% (112)	119	31% (26)	69% (58)	84	49% (68)	51% (70)	138
participation	50% (13)	50% (13)	26	64% (9)	36% (5)	14	76% (75)	24% (24)	99

likely to choose an English or a Spanish marker to express relations of cause and effect. The trend suggests that as Spanish dominance decreases, so does the use of Spanish cause-and-effect markers. Of the two cause-and-effect English-language markers studied, *so* is used most frequently by all groups. The few cases of *so* in the G1 data are used exclusively in the sense of causality, while *entonces* is used to mark temporality. In ex. (2), a narrative on a speaker's educational experiences, she stated the following:

(2)

    1   y eso pasó conmigo en Puerto Rico
    2   mis papás se pasaban mudándose y eso
    3   yo estaba atrasada como, como dos años
    4   yo tenía cator-
    5   a los trece años me gradué de sexto
    6   entonces a los católicos vine a séptimo
    7   so este estaba atrasada[11]

    1   'and that happened to me in Puerto Rico
    2   my parents kept moving us and everything
    3   I was behind about, about two years
    4   I had fourte-
    5   at thirteen I graduated from sixth grade
    6   so to the Catholics I went in seventh
    7   so umm I was behind'

While *entonces* merely moves the narrative along, *so* marks a causal relationship between the statement in line 7 and what comes before it. *So* functions to mark global coherence in the resolution component of the narrative. Another example of *so* functioning as a global discourse marker occurs in ex. (3). The G1 speaker tells the story of a man who has been in New York for years and refuses to learn English. In lines 1–2, the narrator quotes this man's reasons for not learning English; in lines 3–4, he concludes that people like this man might be motivated to learn English if "English only" legislation were enacted:

(3)

    1   porque para que se burlen de mí
    2   mejor no lo hablo
    3   pero esa es una de la razones
    4   so en ese aspecto, es que yo estaría de acuerdo

    1   because if they are going to make fun of me
    2   I'd rather not speak it
    3   but that is one of the reasons also
    4   so in that respect, is that I would agree

In line 4, the marker *so* is used with a global function because all of what comes before line 4 leads to this conclusion.

    Among the third-generation speakers, the use of *so* is more generalized. There is a tendency for it to be used to reflect temporality with no indication of causality. In ex. (4), part of a long narrative about fighting back when an uncle threatens sexual abuse, a G3 woman reports:

73

(4)

    1   yo siempre he tratado de buscar alguien para–*to replace* entiende
    2   <u>so</u> mi tío estaba *right next door*
    3   <u>so</u> me sentía bien tú sabe, – *comfortable*, con él
    4   <u>so</u> I wouldn't mind y'know
    5   pero cuando él se trató de poner fresco conmigo
    6   eso me dolió tanto

    1   'I always tried to find someone to – *to replace*, understand
    2   <u>so</u> my uncle was *right next door*
    3   <u>so</u> I felt good, y'know, *comfortable*, with him
    4   <u>so</u> I wouldn't mind, y'know
    5   but when he tried to get fresh with me
    6   that hurt me so much

Line 2 is not directly causally related to line 1 or to any other issue brought up before that point. Rather, the speaker is changing topic and furthering the narrative story. Lines 3 and 4 could be interpreted as causally related to line 2; therefore, the use of *so* here would match the use of this marker in the G1 narratives. In ex. (5), a G3 narrative on fights in the neighborhood, an English-dominant speaker reports:

(5)

    1   y este, yo estaba sentada en el porch mío
    2   para coger un poco de aire <u>y'know</u>
    3   y para ver que nadie estuviera pasando cerca de mi casa
    4   porque si ven drogas de, <u>y'know</u>, cerca de tu casa
    5   pues piensan que
    6   uno también está en, <u>y'know</u> en to'
    7   <u>so</u> allá yo estoy sentada en el porch
    8   cuando veo …

    1   'and umm, I was sitting on my porch
    2   getting a bit of air y'know
    3   and to see that no one would be passing close to my house
    4   because if they see drugs in, y'know, near your house
    5   well they think that
    6   one is in, y'know, in everything
    7   <u>so</u> there I am sitting on my porch
    8   when I see …

Again, the speaker uses *so* in the orientation of a narrative even though there is no causal relation indicated in the text. Rather, the speaker is getting back to her point and moving the narrative along sequentially.

So usually appears in the resolution component of G1 narratives, but its distribution is more generalized across the various narrative components in G2 and G3 narratives. It may be that when this English language discourse marker is first integrated into a Spanish contact variety, it functions in a more restricted role, occurring at the global level of discourse (see Stoltz & Stoltz 1995, quoted in Solomon 1995:290). Bilingual and English-dominant speakers, in contrast, use *so* also to shift from one component of a narrative to the next, to return to the main

point of a story, to begin evaluations, or simply to move the narrative along. These are all uses of *so* that occur in monolingual English narratives (Schriffrin 1987).

The high frequency of *so* in the speech of the English-dominant speakers suggests that it first entered the community's Spanish in the speech of this group, perhaps through code-switching. In 25% of the cases where it appears in the narratives of G3, it is part of a code-switched sequence. However, it is now fairly well established as a borrowing for this group as well as for G2, the bilingual group, where it never appears as part of a code-switched sequence. This is also true of the few occurrences of *so* that appear in monolingual G1 narratives.

With regard to participation markers, *sabes* and *digo* (*y'know* and *I mean*), all groups use English markers at least 50% of the time. Group 2 uses English-language participation markers by a 2-to-1 ratio, and Group 3 prefers them at 3 to 1 (see Table 1). For Groups 1 and 3, they are the most frequently used English language discourse markers. Interestingly, however, they function differently in the narratives of each group.

Fifty-six percent of all the English-language discourse markers produced by Spanish-dominant Group 1 fall into the participation category (*y'know* and *I mean*). It seems likely that, of all discourse markers, the Spanish-dominant group would first incorporate this type into their speech, since it has little semantic content yet is frequently heard in the speech of English-dominant and bilingual speakers.[12] As a highly salient feature of the speech of this community, it seems like a good candidate to be taken on as a metalinguistic feature. In Group 1 narratives, this marker is used in order to signal that the narrator is changing languages, if only for a brief time. For example, in ex. (6), the Group 1 narrator twice signals a code-mixed sequence by using *y'know*:

(6)

1	pero en ese hearing van ellos tienen un equipo muy bueno
2	y entonces no pudimos tener ese hearing van
3	hasta ... cuando fue ... mayo, y'know of 1989 or March
4	y ya y'know it's almost the end
5	casi es el final del, del año

1	'but in that hearing van they have good equipment
2	and then we were unable to have that hearing van
3	until ... when was it ... May, y'know of 1989 or March
4	and already y'know it's almost the end
6	it's almost the end of, of the year'

Many Spanish-dominant speakers used English participation markers this way. For G1, Spanish-language participation markers were mostly found in evaluation clauses in the narratives. If English markers are not used to frame an English code-switch, which Maschler 1994 calls a metalanguaging function, they are found in evaluation sections of the narratives. Sometimes these conditions overlap. In ex. (7), a narrative on a problem at work, a G1 speaker reports:

(7)

> 1   en el trabajo, por ejemplo, cuando hablo con los jefes
> 2   hay veces, hay veces y'know, que I get mad
> 3   ¿sabes porque?
>
> 1   at work, for example, when I talk with the bosses
> 2   there are times, there are times y'know, that I get mad
> 3   do y'know why?

Here, the speaker uses an English-language marker before a switch to English, and the marker is concurrently part of an evaluation of the story he goes on to tell. Like the other discourse markers studied, *y'know* probably enters the language initially as part of a code-switched sequence.

G2 speakers have few participation markers in their narratives, and when they appear, they are most frequent in the evaluation section of the narrative. These speakers' use of the English-language marker *y'know* mirrors their use of Spanish *tú sabes*, which also appears only in evaluation sequences.

For English-dominant speakers, these markers involve transition from Spanish to English in about 30% of the cases. Most of the time, however, they occur surrounded by Spanish discourse. G3 shows more widespread use of English-language participation markers in different components: They appear in the orientation (38%) and evaluation sections (40%), with less use in complicating action clauses (22%). At first glance, it seems that this marker mainly serves an interactional function to ensure the attention and participation of the listener. It may be that G3 speakers rely on this participation marker because they are telling stories in their less dominant language and want to check frequently and make sure that the listener is following the narrative. However, a closer examination of the function of *y'know* casts doubt on this conclusion. In a challenge to Robin Lakoff's assertion that the use of fillers like *y'know* are characteristic of women's insecurity in language, Holmes 1986 argues that *y'know* has numerous functions, ranging from uses that express speaker certainty and confidence to those that signal uncertainty and lack of confidence. Using Holmes's taxomony, I found that over 60% of the uses of *y'know* in the G3 data fit the certainty criteria established by her. For example, there were fewer instances of the type seen in exx. (8)–(11), expressing uncertainty, and more examples like those seen in exx. (12)–(15), expressing certainty.

(8)

> 1   pero yo no estuve seria con él hasta los quince años
> 2   so tuve con él, dos años, y' know, on and off
>
> 1   but I wasn't serious with him
> 2   so I was with him, two years, y' know, on and off

(9)

> 1   porque eran bien, bien, antipáticas
> 2   y son bien come … y' know
>
> 1   'because they were real, real pains in the ass
> 2   and they are real full of … .y'know'

(10)

    1   y me llevó a small claims court
    2   que es donde uno va
    3   cuando es discriminado . . . y'know.

    1   and she took me to small claims court
    2   which is where you go
    3   when you are discriminated against . . . y'know'

(11)

    1   y dejó la nena botada por allá
    2   bien, bien . . . y'know son bien careless

    1   and she left the girl alone over there
    2   real, real . . . y'know they are real careless

In ex. (8), the speaker qualifies her meaning; in ex. (9), she seeks assurance that I am following her, because she leaves out half the word; in ex. (10), the speaker seeks confirmation that I know the type of place she is thinking of; and in ex. (11), she can't think of the word she wants, so she hedges and code-switches to English. All of these qualify as uncertain uses of the participation marker, and most are signaled by pauses and hesitations. More frequently, however, *y'know* is used in cases like exx. (12)–(15), which express certainty:

(12)

    1   porque como ahora están los VCRs y eso
    2   y'know, uno renta la película y popcorn y all of that
    3   y se entretiene uno en la casa, y'know

    1   'because now there are the VCRs and everything
    2   y'know, you rent a movie and popcorn and all of that
    3   and you entertain yourself at home, y'know

(13)

    1   y este, yo estaba sentada en el porch mío
    2   para coger un poco de aire, y'know

    1   and um, I was sitting on my porch
    2   to get a bit of air, y'know

(14)

    1   porque ella ya sabía
    2   que yo era tofe también y'know

    1   because she already knew
    2   that I was tough y'know

(15)

    1   so y'know when it comes to protecting my family
    2   honey, yo voy a hacerlo

    1   so y'know when it comes to protecting my family
    2   honey, I am going to do it

Ex. (12) and (13) are "attributive" uses of *y'know*; the speaker is saying, "I know you understand the type of thing I am talking about." Exx. (14) and (15) are from a narrative in which a speaker explains why her neighbors call her "Mike Tyson."

These instances are emphatic uses of *y'know* to intensify the point the speaker is making. Rather than expressing uncertainty, these uses of *y'know* "serve to positively reassure the addressee of the validity of the proposition" (Holmes 1986:7) that is being uttered. Interestingly, G3's less frequent use of the Spanish-language equivalent *tú sabes* mirrors their use of *y'know*; it also appears widely in different components and functions to signal both certain and uncertain speech.

DISCUSSION AND CONCLUSIONS

First, we must return to the question of the status of discourse markers as borrowing or code-switching. In this article and in most of the literature on discourse markers, they are treated as borrowings, but this is open to question. If we apply to my data the usual criteria used to distinguish between these two phenomena, we get a mixed picture. Myers-Scotton 1993 posits that frequency of occurrence is the best criterion to distinguish single-item borrowings from code-switches. Whereas borrowed forms appear rather frequently in a large corpus, code-switches may appear only once. Poplack and Sankoff 1984 propose that not only frequency of use but also native language displacement, morphophonemic and syntactic integration, and community acceptability should be considered in identifying single words as code-switches or borrowings. They suggest that, as more of these criteria are met, we can be more confident that we are dealing with loanword adoption rather than some other phenomenon. My data have a high frequency of use of discourse markers, which suggests borrowing. On the other hand, the corpus does not display native-language replacement of discourse markers, which may suggest code-switching. In this situation, the criterion of morpho-syntactic replacement is unhelpful. Finally, speakers are aware that equivalents are available in Spanish, which again suggests code-switching. In the present study, the discourse markers studied can be analyzed variously as both code-switches and borrowings, at different stages in their integration process. We are capturing a change in progress in a speech community with speakers of varying proficiency; thus, complete substitution of borrowed discourse markers would not be expected. These data support the position that it is best to consider code-switching and borrowing as phenomena on a continuum rather than as entirely different processes.

A review of the literature on discourse markers in language contact situations offers a range of outcomes, from marker doubling to the disappearance of native markers. In the case of Brentwood Puerto Rican Spanish, all speakers, regardless of language dominance, use English markers in their Spanish speech production. Whereas Spanish-dominant speakers use English-language discourse markers in a restricted function, in the oral narratives of the bilingual and English-dominant speakers there is often overlap in the function and distribution of Spanish and English discourse markers. English discourse markers are used variably, and Spanish markers are not replaced outright. Unlike the German and Mayan situations

discussed above, the Spanish system of discourse marking is maintained to some degree by all speakers. The participants in the German bilingual study were fourth-to sixth-generation speakers of a variety that has little vitality in the US. Clearly this is not the case for Spanish in the US context; even though individual bilingualism may be in flux, in many communities (such as Brentwood) a stable societal bilingualism is the norm. In the Mayan case, Brody hypothesizes that Spanish markers may eventually replace Mayan markers, but she supplies no evidence to support this claim, and such a claim would lack support also in the case of Spanish marker use in Brentwood.

Although English and Spanish markers are not in complementary distribution, language proficiency and type of marker do affect the level of integration. We have seen that no one borrows the discourse marker *and*. The few cases of *and* with Spanish that do occur are found in the G3 data and are primarily in code-switched sequences. Of all the discourse markers considered, if *and* enters Spanish language discourse at all, it is exclusively via code-switching. As a marker that is semantically empty and that serves no expressive or metalinguistic functions in the narratives, it is not a good candidate for either borrowing or code-switching.

At the other extreme, *y'know* is a frequently used English-language discourse marker that seems to be firmly established as a borrowing. In most of the narratives, it is often preceded and followed by Spanish-language discourse. Of the markers examined, it is the most frequently employed English-language marker for Spanish-dominant and English-dominant speakers. Participation markers seem to enter the language more easily than other markers; this may be because they have the least effect on the propositions set forth in the narratives, yet they are a highly salient language feature. They serve as a kind of external indicator of involvement or attitude toward material discussed in the stories. G2 and G3 speakers prefer the English-language participation marker, and G1 speakers are as likely to choose it as to choose a Spanish participation marker. G2 speakers generally used English-language participation markers as part of an external evaluation, and they used the Spanish discourse markers in the same way. In the cases of the Spanish-dominant and English-dominant speakers, these markers function primarily as participation markers or conversational markers; they create a cohesive link between the speaker and his or her audience. In the narratives, they can express the degree of certainty or uncertainty that speakers have with respect to the content they are sharing, or with the language in which the message is rendered. As Valentine 1991 points out, *y'know* is also a marker that is much more frequent in the speech of young people. This may be why the G1 participants, who are in their teens and twenties, evidence such a high usage of this particular marker in their narratives.

After *y'know*, *so* is the next most frequently used English-language discourse marker. *So* seems to be the marker that is most fully integrated as a borrowing. Traces of its use as a code-switched element are found only in the data of the

English-dominant speakers. When it appears in the speech of monolingual Spanish and bilingual speakers, it is never preceded or followed by English text. In English-dominant speech, only 25% of the occurrences of *so* appear in code-switched sequences. In most cases, and across generations, immediate prior use of English is not triggering the form, and the form is not triggering subsequent English use. Similarly to the case of *but* and *because* in Shona – which also appear without environmental triggers of their use (Myers-Scotton 1993) – *so* appears to be well on the way to becoming an established borrowing in Brentwood Puerto Rican Spanish. Of the markers studied, *so* is also the English-language marker most frequently employed by the bilingual speakers; this finding parallels the French Canadian case presented by Mougeon and Beniak 1985 concerning the incorporation of *so* in bilingual speech. Mougeon and Beniak 1985 argue that borrowing of core lexical items such as discourse markers is associated with situations of intense language contact. They analyze the use of *so* in the discourse of Canadian bilingual adolescents, whom they categorize as either low, mid, or high users of French. They discover that mid-level users of French use *so* more frequently than others. The authors deduce from this that balanced bilingualism is probably a prerequisite for the borrowing of core lexical items; as such, a high frequency of *so* usage could be taken as an indicator of balanced bilingualism. In my data, G2 speakers most resemble Mougeon and Beniak's mid language users because they are the most balanced bilinguals among all speakers in my sample. It is also interesting to note that the G1 participant who most uses *so* in her speech is the Spanish-dominant person who has the best grasp of English; this fact supports the findings of the aforementioned study.

In a study of the use of French discourse markers in the French speech of Anglophones in Montreal, Sankoff et al. 1997 also find that the most frequently used English language discourse markers are *y'know* and *so*.[13] The authors do not find that the use of these English markers is correlated with those speakers who were less fluent in French, nor were speakers substituting these high frequency English markers for any of the French markers they studied.

In Brentwood Puerto Rican Spanish, all speakers use the same range of English and Spanish discourse markers in their speech, yet each group follows its own patterns, and sometimes the same discourse markers function quite differently in different oral narratives. With each successive generation, there is a loosening of restrictions on English marker use, in that the English-dominant speakers show a more generalized use of English discourse markers across all components of their oral narratives. As Mougeon and Beniak 1991 and Myers-Scotton 1993 posit, all three discourse markers probably enter the language as code-switches, and as they become more frequently used, they acquire the status of borrowings. At the present moment in the community studied, *and* seems to be in the initial stage of entering Spanish discourse via code-switching; *y'know* and *so* are further along in the process of being integrated as core borrowings. However, all the English-language discourse markers studied coexist with Spanish-language mark-

ers that encode similar information and functions. As Mougeon and Beniak propose, citing Weinreich and Haugen, the use of core borrowing in a situation of intense language contact may suggest the degree to which all bilingual speakers are experiencing acculturation, regardless of their level of proficiency.

Future studies of discourse markers in bilingual speech can offer more insight into how speakers organize discourse and direct attention to various elements of their talk when they have two systems of markers at their disposal. Diachronic studies of the importation of discourse markers in bilingual speech will reveal if two systems of discourse marking can coexist indefinitely in bilingual speech, or if eventually one set of markers will be eliminated. Thus, the study of bilingual discourse markers promises to make contributions to theories of bilingualism, contact linguistics, and language change.

NOTES

* I thank Ann Russo for her invaluable criticism of this work. I also thank the anonymous *Language in Society* reviewers for helpful suggestions and critical comments on an earlier version.

[1] In language contact studies, linguists (cf. Weinreich 1974 [1953], Haugen 1969, Myers-Scotton & Okeju 1973 and Mougeon & Beniak 1991) identify words from the common vocabulary of a language – discourse markers, interjections, interaction markers, etc. – as core lexical items.

[2] Myers-Scotton 1993 proposes that content and system morphemes can be distinguished according to three properties. Content morphemes are [−Quantification] and either [+Thematic Role-Assigner] or [+Thematic Role-Receiver]. The last two features have to do with relating the thematic role of morphemes to the predicate-argument of a constituent.

[3] Mougeon and Beniak 1991 also hypothesize that the borrowing of core vocabulary like connectors and other discourse markers may originate as code-switches because discourse markers tend to occur at prime switch points (clause-initially or finally).

[4] See Brody 1987, 1992 for a discussion of the function of Spanish particles (conjunctions, discourse markers, etc.) in a variety of Mayan languages.

[5] See Torres 1997 for a more extensive discussion of the community and methodology.

[6] See Fraser 1999 and Schourup 1999 for overviews on the rapidly expanding work on English-language discourse markers.

[7] See Brizuela et al. 1999 for a discussion of the acquisition of discourse markers as indicators of register in the Spanish speech of bilingual children in the US.

[8] See Pons Bordería 1998 and Cortes Rodríguez 1991 for an extensive discussion of the functions of *entonces* in Peninsular Spanish varieties.

[9] Future studies that thoroughly analyze the use of specific discourse markers in Spanish may find that some of the markers have different functions and frequency patterns when compared with similar English language discourse markers. For the markers studied so far, however, this has not been found (cf. Fraser & Malamud-Makowski 1996). Also, it is probably true that different dialects of a language may favor particular markers and patterns of usage.

[10] G2 speakers have the lowest number of discourse markers over all (G1=398, G2=252, G3=490); this is probably due to the fact the G2 narratives tend to be somewhat shorter than the narratives told by G1 and G3 speakers.

[11] All Puerto Rican phonetic dialect features have been regularized in the transcripts.

[12] Urciuoli 1996 reports that Puerto Ricans use the discourse marker *you know* as an index of race, class, and locale. Speakers who used *you know* in their speech were judged to be Puerto Rican or Black, from the lower class, and poorly educated. I did not collect information on speakers' attitudes toward the discourse markers used in this study.

[13] Sankoff et al. 1997 conclude that, in the case of Anglophones in Montreal, as fluency in French grammar increases, so does the use of French discourse markers. High use of French discourse markers correlates with a high degree of integration into the local French speech community.

R E F E R E N C E S

Blackmore, Diane (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
———— (1992). *Understanding utterances*. Oxford: Blackwell.
Brizuela, Maquela; Andersen, Elaine; and Stallings, Lynne (1999). Discourse markers as indicators of register. *Hispania* 82:128–41.
Brody, Jill (1987). Particles borrowed from Spanish as discourse markers in Mayan language. *Anthropological Linguistics* 29:507–21.
———— (1995). Lending the 'unborrowable': Spanish discourse markers in indigenous American languages. In Carmen Silva-Corvalán (ed.), *Spanish in four continents: Studies in language contact and bilingualism*, 132–47. Washington, DC: Georgetown University Press.
Cepeda, Gladys (1999). La variación pragmático-discursiva, entonacional y sociolingüística de los conectores conjuntivos en el habla de Valdivia, Chile. In José Serrano (ed.), *Estudios de variación sintáctica*, 103–20. Madrid: Iberoamericana.
Cortes Rodríguez, Luis (1991). *Sobre conectores, expletivos y muletillas en el español hablado*. Malaga: Editorial Librería Agora.
Fraser, Bruce (1999). What are discourse markers? *Journal of Pragmatics* 31:931–52.
Fraser, Bruce, & Malamud-Makowski, Monica (1996). English and Spanish contrastive discourse markers. *Language Sciences* 18:863–81.
Goss, Emily, & Salmons, Joseph (forthcoming). The evolution of a bilingual discourse marking system: Model particles and English markers in German-American dialects. *International Journal of Bilingualism*.
Haugen, Einar (1950). The analysis of linguistic borrowing. *Language* 26:210–31.
———— (1956). *Bilingualism in the Americas: A bibliography and research guide*. Publications of the American Dialect Society, 26. Tuscaloosa: University of Alabama.
———— (1969). *The Norwegian language in America: A study in bilingual behavior*. Bloomington: Indiana University Press.
Hill, Jane, & Hill, Kenneth (1986). *Speaking Mexicano: Dynamics of syncretic language in Central Mexico*. Tucson: University of Arizona Press.
Holmes, Janet (1986). Functions of *y'know* in women's and men's speech. *Language in Society* 15:1–22.
Labov, William (1972). The transformation of experience in narrative syntax. In his *Language in the inner city*, 354–96. Philadelphia: University of Pennsylvania Press.
————, & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In June Helm (ed.), *Essays on the verbal and visual arts*, 1214–44. Seattle: University of Washington Press.
Maschler, Yael (1994). Meta-languaging and discourse markers in bilingual conversation. *Language and Society* 23:325–66.
Malamud-Makowski, Monica (1987). Discourse markers in Spanish. Ph.D. dissertation, Boston University.
Mougeon, Raymond, & Beniak, Edouard (1986). *Linguistic consequences of language contact and restriction*. London: Oxford University Press.
Myers-Scotton, Carol (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.
————, & Jake, Janice (1995). Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential code-switching. *Linguistics* 33:981–1024.
———— & Okeju, John (1973). Neighbors and lexical borrowings. *Language* 49:871–89.
Pfaff, Carol (1982). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. In Jon Amastae and Lucía Elías-Olivares (eds.), *Spanish in the United States*, 264–97. Cambridge: Cambridge University Press.
Poblete Bennett, María Teresa (1999). La variación en la contrucción de textos orales ligada al uso de marcadores discursivos-conversacionales. In José Serrano (ed.), *Estudios de variación sintáctica*, 89–102. Madrid: Iberoamericana.
Pons Bordería, Salvador (1998). *Conexion y conectores*. Valencia: Universidad de Valencia.
Poplack, Shana, & Sankoff, David (1984). Borrowing: The synchrony of integration. *Linguistics* 22:99–135.
Redeker, Gisela (1991). Review article: Linguistic markers of discourse structure. *Linguistics* 29:1139–72.

Segal, Erwin; Duchan, Judith; & Scott, Paula (1991). The role of interclausal connectives in narrative structuring: Evidence from adults' interpretations of simple stories. *Discourse Processes* 14:27–34.

Solomon, Julie (1995). Local and global functions of a borrowed native pair of discourse markers in a Yucatan Mayan narrative. In Jocelyn Ahlers et al. (eds.), *Proceedings of the 21st annual meeting of the Berkeley Linguistic Society, Feb. 17–20: General session and parasession on historical issues in sociolinguistics*, 287–98. Berkeley: University of California Press.

Salmons, Joe (1990) . The bilingual discourse markings: Code-switching, borrowing, and convergence. *Linguistics* 28:453–80.

Sankoff, Gillian; Thibault, Pierrette; Nagy, Naomi; Blondeau, Helene; Fonollosa, Marie-Odile; & Gagnon, Lucie (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change* 9:191–217.

Schiffrin, Deborah (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schourup, Lawrence (1999). Discourse markers. *Lingua* 107:227–65.

Thomason, Sarah Grey, & Kaufman, Terrence (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.

Torres, Lourdes (1997). *Puerto Rican discourse: A sociolinguistic study of a New York suburb*. Mahwah, NJ: Lawrence Erlbaum.

Urciuoli, Bonnie (1996). *Exposing prejudice: Puerto Rican experiences of language, race, and class*. Boulder, CO: Westview.

Valentine, Tamara (1991). Getting the message across: Discourse markers in Indian English. *World Englishes* 10:325–34.

Weinreich, Uriel (1974[1953]). *Languages in Contact*. The Hague: Monton.

(Received 10 February 2000; Final revision received 9 January 2001)