Two languages, one effect: Structural priming in spontaneous code-switching*

RENA TORRES CACOULLOS
Penn State University
CATHERINE E. TRAVIS
The Australian National University

We investigate here the contribution of code-switching and structural priming to variable expression of the Spanish first person singular subject pronoun in the New Mexican bilingual community. Comparisons with both Spanish and English benchmarks indicate no convergence of Spanish toward English grammar, including in the presence of code-switching, where the linguistic conditioning of variant selection remains unaltered. We find a language-internal and cross-language priming effect, albeit of differing strength, such that speakers' preceding coreferential (Spanish and English) subject pronouns favor subsequent pronouns, whereas unexpressed subjects tend to be followed by unexpressed subjects. Given the rarity of unexpressed subjects in English, in the presence of code-switching fewer tokens occur with unexpressed primes. Thus, code-switching has no intrinsic effect. Instead, it results in associated shifts in the distribution of contextual features relevant to priming, contrary to the convergence-via-code-switching hypothesis and in accordance with the contextual distribution-via-code-switching hypothesis, which we put forward here.

Keywords: bilingual speech, structural priming, variable subject expression, code-switching, bilingual contextual distribution hypothesis

1. Assessing code-switching as a mechanism of change

Is code-switching in and of itself a mechanism of grammatical change? At least since Gumperz and Wilson (1971), many would venture to answer "yes" (e.g. Backus, 2005, p. 334; Thomason, 2001, p. 136; Winford, 2005, p. 86). Yet there is little in the way of empirical evidence for contact-induced change-in-progress (see Poplack & Levey, 2010, for a review), and even less for the surmised role of code-switching in such change.

Since code-switching is an online phenomenon, synchronic tests of its putative role in change must be devised. To date, two quantitative community-based tests have been conducted. Both model variation patterns to arrive at a characterization of (changes in) grammatical structure, though each employs a different comparison to probe the role of code-switching. In their study of preposition stranding as a variant of preposition placement in French (e.g. *J'avais pas personne à parler avec* "I had no one to talk to"), Poplack, Zentz and Dion (2012a) classified French–English bilinguals in the national capital

region of Canada according to their propensity to code-switch. No differences were revealed in comparisons of the two speaker groups, made up of "copious" code-switchers (those with 20 or more switches per recording) and "sparse" code-switchers (those with fewer than 20 code-switches). Investigating variable subject expression in the Spanish of New Mexican bilinguals, Torres Cacoullos and Travis (2010, 2011) classified tokens according to context of occurrence, whether in the absence or presence of code-switching. No differences in subject expression patterns were found in comparisons of the two contexts. In that study, code-switching environments were delimited by the speaker's use of an English multi-word sequence in the recent discourse (specifically, the preceding ten prosodic units), raising the question of the proximity of code-switching for any cross-language effects to appear – a question which we explore here.

The variable use of Spanish subject pronouns in the United States is one of the most trumpeted loci of grammatical alteration impelled by language contact. This is considered a candidate for grammatical convergence because Spanish and English subject pronouns are thought to be strongly associated for bilinguals due to the overlap in their deictic meaning and person-number categories. The overwhelming preference for expressed subjects in English is thus predicted to boost the rate of expressed subject pronouns in contact-Spanish varieties, an idea

Address for correspondence:
Rena Torres Cacoullos, Department of Spanish, Italian, and Portuguese, Penn State University, University Park, PA 16802, USA
rena@psu.edu

put forward as early as Granda Gutiérrez (1972) and as recently as Otheguy and Zentella (2012).

In this study, we return to variable first person singular (1sg) subject (*yo* "I") expression in a unique new corpus of spontaneous code-switching, the New Mexico Spanish–English Bilingual (NMSEB) Corpus (Torres Cacoullos & Travis, 2015a), with three objectives. First, to conduct another test of contact-induced grammatical change, this time drawing on comparisons with non-contact varieties of both of the bilingual's languages; that is, not only Spanish but also English benchmarks. Second, to test whether there are grammatical differences within the same bilinguals according to a maximally proximate measure of code-switching, namely in the immediately preceding or the same clause as the instance of variable subject expression, as in (1).[1] Third, to further examine how code-switching may affect the distribution of contextual features contributing to variant choice (an expressed vs. unexpressed subject pronoun), especially those features that are relevant to structural priming.

(1)   Ivette: *for a long time Ø lo hice asina.*
       Ivette: "*for a long time* (I) did it *like this*."
                  [06 El túnico, 0:50:00-0:50:02]

2. Good data for the study of code-switching: The bilingual community and corpus

Evaluating code-switching as an impetus of language change-in-progress requires a corpus of unreflecting speech amenable to systematic quantitative analysis, drawn from a well-defined speech community of bilinguals who code-switch in their spontaneous discourse.

The first imperative, a corpus of unreflecting speech, follows from the empirical observation that the VERNACULAR, the mode of speech that is used with friends and family and in which minimum attention is paid to speech, is highly regular and thus provides the most systematic data for linguistic analysis (Labov, 1984, p. 29). While at first blush a combination of corpus- and lab-based methods might seem optimum (Gullberg, Indefrey & Muysken, 2009), experimental methods are unsuitable for the collection of New Mexican Spanish production data. Experimental procedures impose formality and self-monitoring and, additionally, are associated with schools, institutions where the native speakers and their local

varieties have long been denigrated (see examples (2) and (3) below) (see Sankoff, 1988a, pp. 145–146).

The second imperative, that data be drawn from a well-defined speech community (Labov, 2001, p. 34), rather than an assortment of bilinguals, is a consequence of the discovery that, even for typologically similar language pairs, the ways in which bilinguals combine their languages differ from community to community (see Poplack, 1998, for an illustration). A community-based approach allows us to detect coherent code-switching patterns. The present study is a departure from generalizations adduced from large numbers of participants of unknown social characteristics (including heterogeneous groups of university students), as well as from claims based on anecdotal observations or expedient (counter-)examples from a few individuals. It also differentiates itself from studies of change in communities undergoing loss or shift within three generations (as is the case for most immigrant communities, see Silva-Corvalán, 1994, *inter alia*), in examining the speech of a native, non-immigrant, community.

2.1 The contact site: Spanish and English in New Mexico

Northern New Mexico is home to "arguably the oldest continually spoken variety of Spanish anywhere in the Americas that has not been updated by more recent immigration" (Lipski, 2008, p. 193). Following settlement in 1598 from New Spain (what is, today, Mexico), Spanish speakers in the northern section of the state had minimal contact with speakers of other varieties of Spanish (Gonzales-Berry & Maciel, 2000, p. 4; Lipski, 2008, pp. 195, 202), developing over time their own distinct variety which we refer to here as Traditional New Mexican Spanish, following Bills and Vigil (2008, p. 7). Traditional New Mexican Spanish is generally Mexican grammatically but has some "independently developed" words and phonetic features (Bills & Vigil, 2008, p. 15).

New Mexico became a Territory of the United States in 1850; however, English speakers were in the minority for longer than in the surrounding region: in 1890, 70% of the population of New Mexico could not speak English, a figure which had dropped to 33% by 1910 (Fernández-Gibert, 2010, p. 48). New Mexico was awarded statehood in 1912, and English increasingly displaced Spanish in the educational system (see Gonzales-Berry, 2000). Children were punished for speaking Spanish in schools, as recounted in (2), by one of the speakers in the NMSEB Corpus.

(2)   Punishment for use of Spanish at school
       Pedro: *… usaban la jarita or you had to go*
                 out and get a load of wood.
                 [10 El timbre portátil, 0:11:16-0:11:19]

---

[1]   All examples given are from the NMSEB Corpus (Torres Cacoullos & Travis, 2015a). Examples are reproduced verbatim from the transcripts, with the addition of zeros in the original and parentheses around *I* in the translation representing an unexpressed *yo*, to ease identification, and underlining to mark the target clause(s). Transcription conventions are presented in Appendix A. Within brackets is the recording number, name and time stamp. In the translation (which appears below the original), speech that was originally produced in English appears in italics.

Pedro: "... they would use the cane *or you had to go out and get a load of wood."*

From the 1900s (with the Mexican Revolution, 1910–1920), immigration from Mexico has led to increasing contact with contemporary varieties of Mexican Spanish. While this augments the presence of Spanish overall, Traditional New Mexican Spanish is stigmatized in comparison with monolingual varieties, and thus the northward spread of Mexican Spanish also threatens the maintenance of Traditional New Mexican Spanish (Bills & Vigil, 1999, p. 56), as does the teaching of educated standard Spanish in schools as a foreign or second language (Gonzales-Berry, 2000). This disparagement of the local variety is displayed in example (3), about the speaker's granddaughter getting her homework – which the speaker had helped her with – marked as wrong. In this way, contact with English, with Mexican Spanish and with standard "school" Spanish are all said to endanger Traditional New Mexican Spanish (Bills & Vigil, 2008, p. 313; Travis & Villa, 2011).

(3) Disparagement of Traditional New Mexican Spanish at school
  Inmaculada: *.. they c- called it proper Spanish.*
  Lucy: *mhm.*
  Inmaculada: *o=r,*
    whatever,
    it was called,
    (H) but it wasn't our Spanish.
  Lucy: *hm.*
  Inmaculada: *so she got everything wrong.*
    [14 Proper Spanish, 0:26:25-0:26:32]

For students of language contact – the locus of which, as stressed by Weinreich (1963), is the bilingual speaker – the remaining speakers of Traditional New Mexican Spanish and English provide a precious window into bilingual speech phenomena, as their speech allows us to examine long-term grammatical repercussions of contact.

2.2 The community-based corpus

Speakers comprising the NMSEB Corpus are minimally third generation *Nuevomexicanos* "New Mexican Hispanos". They are bilingual in that they regularly use both languages with the same interlocutor in the same domain, "the appropriate code for the Hispano community" in New Mexico (Gonzales, 1999, p. 29). Rather than administering a battery of proficiency tests, we adopt this criterion of regular use of both languages, as observed by the fieldworkers and subsequently confirmed in the recordings (see Poplack, 1993, p. 254).

Speech samples were recorded with the goal of best approximating observation of informal, everyday vernacular speech. Eight minimally third generation *Nuevomexicano* students of the University of New Mexico, who, importantly, were all community in-group members (see Clyne, Eisikovits & Tollfree, 2001, pp. 235–236; Poplack, 1993, p. 260), conducted interviews with one or more family members or acquaintances, who make up the participants in the corpus. The fieldworkers were instructed to speak in both English and Spanish, switching between the two as they naturally would. Thus, the language switching and linguistic structures that occur in the corpus did not arise in response to direct elicitation from the interviewers.

We employed the technique of the sociolinguistic interview to elicit "narratives of personal experience" during which monitoring of speech is minimized and for which the participant, not the interviewer-researcher, is the indisputable expert (Labov, 1984, pp. 32–42). Covered were what can be considered general interest topics, such as childhood, family and work, as well as community-particular, or insider, topics, such as hunting bears, making adobe bricks and pole vaulting in the *acequia* (irrigation ditch). Often arising naturally in the recordings was information about the speaker's linguistic history and their language attitudes (such as experiences with use of Spanish and English in school, as illustrated in the excerpts in (2) and (3) above), providing insights into the social context within which the bilingual phenomena of interest arise.

The 23 NMSEB Corpus participants in the present report are all from northern New Mexico, were born between 1922 and 1964, comprise 14 women and 9 men and include miners, ranchers and schoolteachers; see Appendix B for particpants' background. They can be considered "early" bilinguals, with the caveats that such ways of classifying bilinguals remain disputable (see Grosjean, 1998, pp. 133–144), and, particularly relevant in the case at hand, "age of acquisition" estimates are subject to the vagaries of self-reporting for people who have grown up in bilingual communities. Participants report that their first language was Spanish (18/23), they learned English upon starting school (18/23) or (also) at home (5/23), they prefer to speak English (13/23), Spanish (6/23) or both (4/23), and they use both languages not just in the home, but also with friends and at work, corroborating the bilingual nature of the community in which these speakers live. (For an overview of the NMSEB Corpus and a sociolinguistic profile of participants, see Travis & Torres Cacoullos, 2013.)

A corpus that can be used for accountable quantitative analysis requires exhaustive transcription of the audio data. Transcriptions were done in *ELAN*, a software program that aligns the audio file to the transcription (Lausberg & Sloetjes, 2009). We use standard orthography for phonetic variants but represent lexical and morphological variants of the community, for example *muchita* "girl", seen in example (10) (vs.

muchachita), and the first person plural suffix -*nos*, seen in (18) (vs. -*mos*). The transcription adopted for the NMSEB Corpus follows the method for discourse transcription outlined in Du Bois, Schuetze-Coburn, Cummings and Paolino (1993) and is prosodically based, relying on the notion of the Intonation Unit (IU), "a stretch of speech uttered under a single coherent intonation contour" (Du Bois et al., 1993, p. 47). Each IU is represented on a distinct line in the transcription, followed by punctuation indicating its prosodic contour. The relevance of the IU, which has been described as the "most plausible basic unit of the grammar of spoken language" (Croft, 1995, p. 875), will become apparent in Section 6.1. The present data constitute approximately 18 hours of speech (60,000 IUs and 202,000 words), of which approximately 15 hours (45,500 IUs and 196,000 words) were produced by the participants.[2]

Having met the extralinguistic requirements for testing convergence impelled by code-switching – an appropriate contact setting and speaker sample – in the following section we delimit a linguistic variable that is an appropriate candidate for grammatical convergence, namely one which involves points of both overlap and discrepancy between the languages in contact.

3. Variationist comparative method to ascertain convergence

3.1 Conditioning of variation to ascertain (change in) grammatical structure

Any pronouncement of convergence that is based on departures from an idealized monolingual norm as perceived by the analyst rests shakily not only on unverifiable characterizations of the "monolingual norm", but also on the precarious equation of variation with change. But we know that, while all change involves variability, "not all variability and heterogeneity in language structure involves change" (Weinreich, Labov & Herzog, 1968, p. 188). How, then, can grammatical alteration be observed?

Linguistic variability is conditioned by contextual features, which contribute to speakers' choices among the set of VARIANTS constituting a LINGUISTIC VARIABLE (Labov, 1969) – here, speakers' choice of Spanish pronoun *yo* vs. an unexpressed 1sg subject. The prediction of the convergence-via-code-switching hypothesis (e.g. Gumperz & Wilson, 1971) has been that code-switching to English results in higher rates of expressed subjects

in Spanish. Here, however, we do not rely on overall rates of linguistic forms alone, but look to the configuration of conditioning factors (see Poplack, Zentz & Dion, 2012b, p. 250), which provides "a more penetrating characterization of grammatical structure" (Erker & Guy, 2012, p. 546). Thus, following the variationist comparative method (Poplack & Meechan, 1998), we determine the linguistic conditioning of *yo* expression in the NMSEB Corpus, and then compare this conditioning with the structure of variability in monolingual benchmarks for Spanish and English.

3.2 Data: Circumscribing the variable context

The first step in determining the linguistic conditioning of *yo* involves defining the variable context, the sum of contexts in which speakers have a choice between variants (Labov, 2005). It is only by accounting for all non-occurrences as well as occurrences of the phenomenon of interest that we can establish the factors influencing speaker choices (Labov, 1972, p. 69).

We began by extracting all tokens of finite Spanish verbs with (expressed and unexpressed) 1sg subjects produced by our participants (N = 2,324).[3] This included tokens where the pronoun and verb were separated by intervening material (N = 32), namely an adverb such as *también* "also", *nunca* "never", *quizás* "maybe", *nomás* "only, just". We also included the one token in the data of apparently emphatic *yo sí* "yes", as *yo* expression with *sí* is variable in these data. Cases of unmarked contrast have been excluded in past studies as presumed obligatory contexts (e.g. Silva-Corvalán, 2003, p. 850), but it has also been noted that there is variability in such environments (Amaral & Schwenter, 2005; Otheguy, Zentella & Livert, 2007, pp. 775–776). Further, without a clear operationalization of contrast, there is a great risk of circularity, whereby tokens are interpreted as contrastive because of the presence of *yo*, and then *yo* is described as a marker of that contrast (for operationalizations and tests of contrast, see Travis & Torres Cacoullos, 2012, pp. 738–741; Travis & Torres Cacoullos, 2014, pp. 367–371).

We then implemented the following exclusions. We excluded the notably small number of tokens (N = 9) that were contextually ambiguous, of which only two can be considered genuinely ambiguous. In four cases

---

[3] Not extracted were prosodically truncated tokens where the speaker cut off before completing, or just after, the verb and did not produce a subject pronoun (and therefore there was the potential for a truncated postverbal subject) (N = 46); tokens where there was a 1sg verb but the speech was not clear enough to definitively identify the realization of the subject (N = 39); tokens that occurred in word play or in speech produced with marked voice quality (N = 7); and unexpressed tokens which occurred in contexts of repair, where the speaker partially repeats what they had just said (N = 12).

the identity of the subject makes no difference to the interpretation of the event (what Ranson, 1991, p. 145, refers to as "person irrelevant" and Ono & Thompson, 1997, p. 488, describe as the "referent" being left "open"). One such example is seen in (4), where the subject of *tenía que meter el dedo* could mean "I had to stick in my finger", but it is equally possible that it means "one had to". There are two tokens where the ambiguity arises because the speaker doesn't complete the utterance, and another because the surrounding speech is unclear.

(4)　Inmaculada:　*(H) y ella me decía ándale lava los trastes y yo te doy lipstick.*
　　　Lucy:　　　*<@ oh= @>.*
　　　Inmaculada:　*so Ø lavaba los trastes y yo muy contenta con lipstick.*
　　　　　　　　(H) !pero ya= estaba bien gastado, Ø <u>tenía</u> que meter el dedo <@ pa' adentro pa' agarrar el lipstick @>.
　　　　　　　　[14 Proper Spanish, 0:41:02-0:41:12]
　　　Inmaculada:　"(H) and she would say to me go and wash the dishes and I'll give you lipstick."
　　　Lucy:　　　"<@ oh= @>."
　　　Inmaculada:　"so (I) would wash the dishes and I'd be all happy with lipstick. (H) but it was already really used up, (I/she/one) <u>had</u> to stick a finger <@ inside to get the lipstick @>."

We excluded the substantial number of postverbal tokens of *yo* (16% of expressed subjects in the data, 108/664), as postverbal subject pronouns are unlikely to follow the same constraints as preverbal ones (see Silva-Corvalán, 2001, p. 165). (On the patterning of first singular subject position in this corpus, see Benevento & Dietrich, published online, January 14, 2014.) Also excluded were cases where the verb appeared in a distinct IU from the pronoun (N = 30) as it is not always possible to tell in such cases whether this pronoun should be considered the subject of the verb (illustrated in (5), where *yo* precedes a subordinating conjunction and there is also truncation) (see Otheguy & Zentella, 2012, p. 236). Other cases of expressed subjects in rare configurations (N = 5) were excluded (for example *yo quizá como que tengo s-una m- --* "I maybe kind of like have a --" [06 El túnico 0:35:08]).

(5)　Bartolomé:　*...(1.2) sí pues <u>yo</u> cuando com- --*
　　　　　　　... principié la escuela <u>no sabía</u> nada en inglés.
　　　　　　　[02 La marina, 0:37:46-0:37:53]

Bartolomé　"...(1.2) yes well when <u>I</u> beg- --
　　　　　...(I) started school (I) <u>didn't know</u> anything in English."

Several non-variable contexts were also excluded. In identifying such contexts, we do not ask whether variability is theoretically possible with any given verb, but rather "we formulate ... broad definitions of clausal and lexical types where variability is low enough to disqualify them from the study" (Otheguy et al., 2007, p. 776). The non-variable contexts we identified for these data include *wh*-questions, an environment where preverbal subjects do not occur in this variety (N = 27, e.g. *qué te puedo contar?* "what can (I) tell you?" [13 La acequia, 0:01:27]) (see Silva-Corvalán, 1982, p. 103); subject relatives (N = 2, e.g. *yo era la única que Ø no sabía arrear* "I was the only one who didn't know how to drive" [06 El túnico, 0:52:02]) (see Otheguy & Zentella, 2012, p. 246; Silva-Corvalán, 1982, p. 103); and fixed expressions (N = 6, e.g. *ahí voy* "coming" [13 La acequia, 0:00:19]).

These protocols leave us with 2,220 tokens for analysis.

4. Linguistic conditioning of *yo* among bilinguals

The constraints on variable subject expression are probabilistic statements about the co-occurrence of the variants (here, *yo* and unexpressed) with elements of the linguistic context. Factors based on these contextual features operationalize hypotheses about speaker choices between variant forms. Six factor groups (independent variables, or predictors) were considered together in multivariate analysis (using Goldvarb Lion, Sankoff, Tagliamonte & Smith, 2012). Variable-rule analysis uses logistic regression to perform binomial multivariate analysis for a choice of the "1" variant (here, pronominal 1sg *yo*) vs. the "0" variant (unexpressed 1sg); the procedure determines the factor groups that together account for the largest amount of variation, in terms of stepwise increase of log likelihood, such that the addition of any of the remaining factor groups does not significantly increase the fit to the model (Sankoff, 1988b).

In Table 1 we see that proximate code-switching does not make a significant contribution; we address this in detail in Section 7. Five constraints are significant: Subject continuity (presence of intervening human subjects between coreferential mentions), Coreferential subject priming (realization of previous coreferential 1sg subject), Semantic class of verb, Reflexive clitic *me*, and Tense-aspect-mood. Probabilities, shown in the first column, are such that the closer to 1 the more the corresponding contextual feature (factor, or level of the predictor variable) favors *yo* occurrence, and the closer to 0 the greater the DISfavoring effect on *yo* (or conversely, the

Table 1. *Multivariate analysis of factors contributing to choice of subject pronoun* yo *(vs. unexpressed subject).*

N = 2,220; Input: .17[a] (Overall rate: 25%)				
	Probability	% *yo*	N	Data distribution
Subject continuity:				
Human switched reference				
Non-coreferential (intervening human subjects)[b]	.61	37%	819	43%
Coreferential (no intervening human subjects)	.42	14%	1103	57%
Coreferential subject priming:				
Realization of previous coreferential 1sg subject[c]				
Pronoun (*yo* or *I*)	.66	31%	462	35%
Unexpressed (Spanish 1sg verb)	.41	11%	865	65%
Semantic class of verb				
Cognition (*creo* "I believe", *no sé* "I don't know")	.70	47%	442	20%
Other (motion, copula, speech, perception, etc.)	.45	19%	1758	80%
Reflexive clitic *me*				
Absent	.55	27%	1906	86%
Present	.23	12%	314	14%
Tense-aspect-mood				
Imperfective (present and imperfect)	.56	31%	1423	70%
Perfective (preterit)	.37	12%	613	30%
Proximate code-switching				
Spanish only	[.52]	25%	1070	72%
Multi-word English present	[.46]	24%	424	28%

Not selected as significant: proximate code-switching (shown in [])

[a] The relationship between Input (corrected mean) and the overall rate of variant selection (which generally it closely reflects) appears distorted because the tokens coded for Priming, shown in the N column, constitute 60% (1,327/2,220) of the data.

[b] Excluded are tokens involving quotation (N = 245) and where the previous coreferential subject was not identifiable (due to inaudible material) (N = 53).

[c] Tokens not coded for Coreferential subject priming are ones where the prime occurred at a distance of five or more clauses (N = 506), postverbally or in a different IU from the verb (N = 72), as well as those involving quotation or inaudible material.

favoring of unexpressed 1sg). The table also shows, in subsequent columns, for each factor, the rate (%) of *yo*, the number of tokens, and the data distribution (how much of the data in that factor group these tokens constitute).

4.1 Subject continuity

We begin with the Subject continuity effect. This is consistent with the notion of accessibility, whereby cross-linguistically less coding material (here, an unexpressed subject) is said to correspond to contexts of greater accessibility – such as when the referent has been recently mentioned – and more coding material (here, a subject pronoun) to correspond to contexts of lesser accessibility (Givón, 1983a, p. 18). We measure recency of mention in terms of intervening human subjects (Travis & Torres Cacoullos, 2012, pp. 726–729), which considers the presence of specific human subjects intervening

between the target token and the previous coreferential mention as subject produced by the same speaker.[4] This is a reconfiguration of Givon's (1983a, p. 14) Potential Referential Interference measure of accessibility, which counts semantically compatible referents in any syntactic role within the preceding three clauses.

The following examples illustrate. In (6), between the target token in line 7 and the previous coreferential mention as subject in line 1, there are three intervening clauses, and two intervening human subjects, "she" and "Bobbie". In (7), there is one intervening clause, but no intervening human subject between the target in line 5 and the previous mention in line 1; the inanimate subject of

---

[4] Coreferential plural subjects (N = 170) and coreferential second person subjects produced by the interlocutor (N = 176) do not count as previous coreferential mentions (though note that, being partially coreferential, nor do they count as "intervening human subjects").

the intervening clause does not count as disrupting subject continuity.

(6)   Intervening Human Subject present
        (non-coreferential context)
1.   Betty: *... Ø le fajé una nalgada. ((CLAPS))*
2.   Carrie: *.. @@@*
3.   Betty: *(TSK) .. she <u>was</u> starting to walk,*
4.   *... and she <u>wanted</u> to go to the fireplace,*
5.   *sit on the mantel,*
6.   *.. (TSK) ... y no la <u>dejaba</u> la Bobbie.*
7.   *<u>yo no sé</u> qué estarían haciendo allá.*
                     [13 La acequia, 0:18:16-0:18:28]
1.   Betty: "... (I) gave her a slap on the bottom.
                ((CLAPS))"
2.   Carrie: .. @@@
3.   Betty: "(TSK) .. she <u>was</u> starting to walk,
4.   ... and she <u>wanted</u> to go to the
                fireplace,
5.   sit on the mantel,
6.   .. (TSK) ... and Bobbie
                <u>wouldn't let</u> her.
7.   <u>I don't know</u> what they were doing
                there."

(7)   Intervening Human Subject absent (coreferential
        context)
1.   Miguel: *... and <u>I used to get</u> chile quite a bit.*
2.   *.. y <u>salía</u> todo.*
3.   *... carrots,*
4.   *unas carrots grandotas.*
5.   *asina= Ø <u>las sacaba</u>.*
                [04 Piedras y gallinas, 0:47:11-0:47:18]
1.   Miguel   "*... and <u>I used to get</u> chile quite a bit.*
2.   .. and everything would <u>grow</u>.
3.   ... carrots,
4.   big carrots.
5.   (I) would <u>pull them out</u> like this."

The measure of intervening human subjects, which we term Human Switched Reference, provides a more discerning account than the standard clause-based measure of Switch Reference, which considers whether the subject of the target clause is coreferential with that of the preceding clause, human or not (see Cameron, 1994). At a distance of one intervening clause between the target verb and previous coreferential mention, typically considered Switch Reference, an intervening human subject is present only 42% (110/260) of the time. This is what makes a difference: the rate of *yo* is 28% (31/110) in the presence of an intervening human subject, but 9% (14/150) in the absence of an intervening human subject (as in (7)). That is, what matters is that the "switch" *vis-à-vis* the immediately preceding clause be from a subject referring to a specific human. At distances of two or more intervening clauses (as in (6)) the difference between

the two subject continuity measures diminishes (between two and four intervening clauses, an intervening human subject is present two-thirds of the time (67%, 203/305) and the *yo* rates in the presence and absence of intervening human subjects are closer (31%, 63/203 vs. 22%, 22/102, respectively)).

As configured for the multivariate analysis in Table 1, non-coreferential contexts are those with one or more intervening human subjects within four clauses from the previous mention (N = 313), as well as those with no previous mention within the preceding four clauses (N = 506). Subject continuity makes a significant contribution, with such non-coreferential contexts favoring selection of *yo*.

4.2 Coreferential 1sg subject priming

Structural priming, the tendency to repeat the same structure or form, has been observed in both community-based studies of natural speech and in psycholinguistic experiments. The priming effect found here takes into account the realization of the speaker's most recent reference to themselves as subject, what we call coreferential 1sg subject priming. This effect is illustrated in the following two examples: in (8) the previous coreferential 1sg subject and the target token are both realized pronominally, and in (9) both are unexpressed.

(8)   Coreferential 1sg subject priming: *yo* to *yo*
        Rocío: <u>*Yo no tenía*</u> *máquina de lavar.*
                 (H) <u>*Yo lavaba*</u> *con lavadero.*
                        [05 Las tortillas, 0:02:39-0:02:42]
        Rocío: "<u>I didn't have</u> a washing machine.
                 (H) <u>I washed</u> with a washboard."

(9)   Coreferential 1sg subject priming: unexpressed to
        unexpressed
        Miguel: *ya Ø <u>no quería</u>.*
                  ... it was too hard.
                  ...(1.5) pero Ø <u>lo hice</u>.
                  [04 Piedras y gallinas, 12:54:44-12:54:46]
        Miguel: "(I) <u>didn't want</u> to any more.
                  ... it was too hard.
                  ... (1.5) but (I) <u>did it</u>."

Coreferential 1sg subject priming in a Human Switched Reference context is illustrated in (10), where there are two intervening human subjects (*Nancy* and *they*) between coreferential mentions.

(10) Coreferential 1sg subject priming with intervening
        human subjects
        Fabiola: *... las cuatro vinieron?*
                 yo [creí que la N]ancy no <u>había</u> venido.
        Jake: *[no=].*
                 ... they <u>have</u> four.

Fabiola: ... *vino también la Nancy?*
          ... *(1.3) yo creía que no más tres muchitas.*
                    [09 La salvia, 0:43:09-0:43:17]
Fabiola: "... all four came?
          I [thought that N]ancy hadn't come."
Jake: *"[no=].*
          ... they *have four.*"
Fabiola: "... *did* Nancy also come?
          ... (1.3) I thought that there were only
              three girls."

It has been found that priming is stronger "production-to-production" than "comprehension-to-production" (Gries, 2005, p. 374, and references therein), in other words, speakers are primed more by themselves than by their interlocutor. Therefore, we consider here only coreferential 1sg mentions. Coreferential mentions across quotative contexts, as in line 2 in (11), were coded separately as such, since it is not known how quotation may affect priming (N = 245).[5]

(11) Manuel: ...*(0.9) y Ø dije,*
                Ø voy a chequear ese,
                .. y luego Ø me voy a revolver .. pa'trás.
                [16 Trip to Africa pt.1, 0:24:29-0:24:34]
      Manuel   "...(0.9) and (I) said,
                (I)'m going to check on that,
                .. and then (I)'m going to go back."

How far back from the target can the prime be? Measures of distance between prime and target for other variables have included elapsed time and number of words or parsing units (modeled logarithmically in large-scale corpus studies, see Gries, 2005, p. 120; Szmrecsanyi, 2005, p. 371). Here we used clauses as a measure (following Weiner & Labov, 1983), reasoning that for a syntactic variable, such as subject expression, finite verbs "interfere" more than arbitrary material in the strength of a previous subject prime.

We operationalized the clause as any finite verb, in a main or subordinate clause, with a referential or non-referential subject, produced by any speaker. Set aside were fixed expressions that are not fully clausal, which had to be determined for both Spanish and English in this bilingual corpus.

In Spanish, not counted as clauses were fixed impersonal expressions (such as *es que* "it's that", *quién sabe* "who knows", *hay veces que* "sometimes") as well as discourse markers (such as non-literal uses of *anda/e(n)* "go ahead", *¿sabes qué?* "you know what?", *mira* "look", as in (12)) We did, however, count as clauses instances of *(yo) no sé* "I don't know" and *yo creo* "I think" (as

do Otheguy & Zentella, 2012, pp. 234–235) because, although on frequency measures they qualify as particular constructions, they are not entirely autonomous of other instances of *yo* + (cognition) verb by the measure of shared constraints, and do still behave as clauses (Travis & Torres Cacoullos, 2012, pp. 738–741) (see Torres Cacoullos & Walker, 2009).

(12) Sandra: *y mira que Ø me hallé un dólar.*
                [03 Dos comadres, 0:44:04-0:44:06]
      Sandra: "and look (I) found a dollar."

In English, non-clauses include several discourse formulae. Classified as formulaic were 1sg subject–present tense cognition verb collocations *I mean*, *I guess*, *I think*, *I remember*, *I'm sure*, *I know*, *I don't know* when they occur as parentheticals in the clause (*I guess* in (13)), or on their own in an IU, that is, as prosodically independent (*I don't know* in (14)), as opposed to when they occur introducing clausal material in the same IU (Travis & Torres Cacoullos, 2014, pp. 363–366). Also skipped over in the clause count as formulaic were expressions such as *you know* (except where it occurs with a complement clause), *you know what* (on its own in an IU)*, and *go figure*.

(13) Aurora: *when he finally .. decided to,*
                .. (TSK) I guess get a house.
                [15 Las cosas viejas, 0:05:50-0:05:53]

(14) Miguel : *se me hace que también otras cosas.*
                 .. melones,
                 I don't know.
                 ... pero mainly chile.
                 [04 Piedras y gallinas, 0:29:32-0:29:38]
      Miguel: "I think also other things.
                 .. melons.
                 I don't know.
                 ... but mainly chile."

Potential primes were counted within four clauses from the previous coreferential mention. Coreferential 1sg subject priming with this cutoff applies to approximately two-thirds of the data.

Overall, as shown in Table 1, previous realization as a pronoun (*yo* or *I*) most strongly favors occurrence of *yo*, which is expressed almost three times as often in this context than with a previous unexpressed 1sg subject (31% vs. 11%). When we examine this more closely, we find that the priming effect dissipates with increasing distance (see Travis, Torres Cacoullos & Kidd, to appear). At a distance of zero intervening clauses (that is, in coreferential contexts), the rate of expression in the context of a previous *yo* (seen in (8) above) vs. a previous unexpressed 1sg subject (seen in (11) above) is more than four times greater (35%, N = 148 vs. 8%, N = 551); at a distance of one or two clauses it is three times greater

---

[5] When the target and preceding occur as part of the same quote, as for the second token in (11), however, the preceding token was counted as a prime.

(40%, N = 68 vs. 13%, N = 226); and at a distance of three or four clauses, it is less than twice as great (45%, N = 38 vs. 24%, N = 86).

Note that in Table 1, previous realization as a pronoun includes both *yo* (N = 254) and *I* (N = 208). We will explore the cross-language priming effects in relation to code-switching in Section 7.3.

4.3 Cognition verbs and particular constructions

As to the remaining constraints in the multivariate analysis, for Semantic class of verb, the strong favoring effect of cognition verbs observed here replicates that reported in numerous previous studies across different varieties of Spanish (e.g. Bentivoglio, 1987, p. 60; Enríquez, 1984, p. 240; Silva-Corvalán, 1994, p. 162; Travis, 2007, pp. 116–117). Frequent among cognition verbs are the 1sg subject–present tense collocations *yo creo* "I think", with a notably high *yo* rate of 87% (94/108) (also reported to favor *yo* in other varieties (Erker & Guy, 2012, p. 536, 539; Travis & Torres Cacoullos, 2012, pp. 739–741), and *(yo) no sé* "I don't know" with a *yo* rate of 41% (56/138).

Second, *yo* expression is disfavored by reflexive *me*-marked verbs, most frequent among which is the cognition verb *acordarse* "remember" (tokens of which account for 40%, 125/314, of the reflexive verb data). Though the overall rate of expression for *acordarse* (12%, 15/125) corresponds to the average for reflexive verbs, the *yo* rate may tend to be lower in the negative polarity collocation *no me acuerdo* (at 8%, 4/49) than in affirmative *me acuerdo* (at 14%, 10/69) pointing again to lexically-particular constructions with distinctive tendencies in subject expression rates.

Finally, considering the Tense-aspect-mood constraint, *yo* is disfavored with Preterit (past perfective) verb forms. This is as predicted, and though it has been interpreted as an effect of ambiguity of person–number morphology (Preterit suffixes distinguish grammatical persons), this interpretation is not applicable for these data as the Present (which also distinguishes grammatical persons) patterns identically to the Imperfect (which is morphologically ambiguous), with relatively high *yo* rates of 32% (287/908) and 30% (152/515), respectively.

Having established the patterning for the NMSEB Corpus, we turn now to the comparisons, first with Spanish, then English, benchmarks.

5. Comparing bilingual and monolingual Spanish subject expression

5.1 The equivocality of overall rates

Spanish subject expression has been enlisted as a candidate for grammatical convergence under the
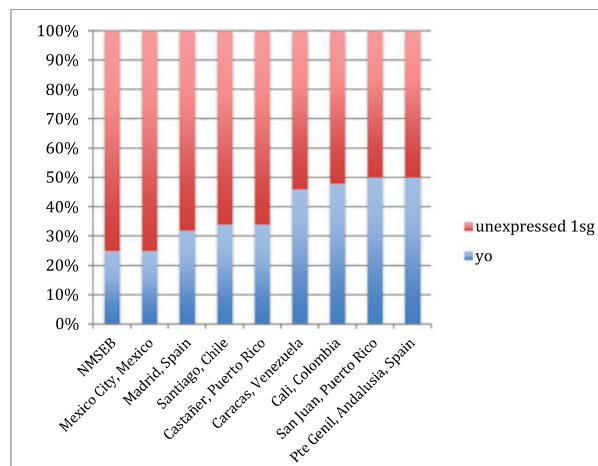

Figure 1. (Colour online) Rates of expression of Spanish 1sg subject pronoun *yo* across studies. Sources: Mexico City, Mexico (Lastra & Butragueño, 2015); Madrid, Spain (Enríquez, 1984) and Santiago, Chile (Cifuentes, 1980–1) as reported in Silva Corvalán (1994, p. 153); Castañer, Puerto Rico (Holmquist, 2012, p. 211); Caracas, Venezuela (Bentivoglio, 1987, p. 36); Cali, Colombia (Travis, 2007, p. 113); San Juan, Puerto Rico (Cameron, 1994, p. 31); Puente Genil, Andalusia (Ranson, 1991, p. 135, 138).

assumption that Spanish and English subject pronouns are similar enough that bilinguals may come to use Spanish subject pronouns at a higher rate in tandem with the overwhelming rate of subject pronoun use in English. But a fundamental problem with confining tests of convergence to comparisons of overall rates of variable occurrence is the question of the threshold for a "high(er)" rate, given the susceptibility of rates to genre, topic or other extragrammatical situational considerations, as well as disparities in overall rates between dialects.

Figure 1 depicts rates of 1sg Spanish subject expression in the NMSEB Corpus and as reported in eight other studies. Note the wide range here, where the highest, at 50%, doubles the lowest, at 25%. Also note that the NMSEB Corpus, with a rate of 25%, is among the lowest.

On the one hand the lower rate of subject expression in the NMSEB Corpus is dissonant with convergence in terms of the prediction that the overall rate of *yo* would be pulled up by virtue of bilinguals associating it with English *I*. On the other hand, it could be consonant with convergence, if convergence takes the form of bilinguals' loss of constraints favoring expressed subjects (see Silva-Corvalán, 1994, pp. 153–162). The comparison of overall rates of use across dialects and studies is, as anticipated, unrevealing as to change, contact-induced or otherwise.

Crucially, no study has identified differences in the linguistic conditioning of subject expression in Spanish. Regardless of overall rate differences, the same probabilistic constraints on subject expression in Spanish

Table 2. *Linguistic constraints on 1sg Spanish subject expression across studies.*

	NMSEB Corpus N = 2,220	Cali, Colombia N = 1,020	Caracas, Venezuela N = 892 (1sg and pl)	San Juan & Madrid N = 1,764 & N = 1,512 (all sg persons)	New York Newcomers N = 16,735 (all persons)
Subject continuity					
Non-coreferential contexts	↑	↑	↑	↑	↑
Coreferential subject priming					
Pronoun *yo*	↑	↑	*	↑[a]	*
Semantic class of verb					
Cognition (e.g. *yo creo*)	↑	↑	↑	*	↑
Reflexive					
Reflexive clitic *me*	↓	*	*	*	↓
Tense-aspect-mood					
Perfective (Preterit)	↓	↓	*	↓	*

Sources: Cali (Travis & Torres Cacoullos, 2012, p. 726); Caracas (Bentivoglio, 1987, pp. 36, 60); San Juan & Madrid (Cameron, 1994, pp. 32, 38, 40); New York Newcomers (Otheguy & Zentella, 2012, pp. 163–165).

* Not tested or not reported.

[a] For all singular grammatical persons, Cameron (1994, pp. 39–40) finds priming in coreferential contexts and a weaker effect under switch reference.

are consistently reported, including in comparisons both across dialects (e.g. Cameron, 1993, 1994) and genres (Travis, 2007). We therefore turn to the linguistic conditioning, and compare the constraints on subject expression in bilingual New Mexican Spanish and non-contact varieties, in order to corroborate grammatical similarities or differences.

5.2 Shared probabilistic constraints across varieties of Spanish

Table 2 depicts the constraints on variable Spanish subject expression in the NMSEB Corpus and as reported in four other studies. This shows that the linguistic conditioning of variable *yo* subject expression for New Mexican Spanish–English bilingual speakers (detailed above in Table 1) replicates that found in numerous studies across Spanish varieties. As indicated by the arrows, Spanish 1sg pronominal subjects in both the NMSEB Corpus and non-contact varieties are favored in non-coreferential contexts, by preceding pronouns, and with cognition verbs; they are disfavored with reflexive-marked and Preterit verb forms.

In short, the constraints on subject expression in the bilingual data of the NMSEB Corpus do not display differences with non-contact varieties of Spanish.

6. Comparing variables: Spanish subject expression vs. English subject realization

The parallel linguistic conditioning of *yo* expression in the NMSEB Corpus and non-contact varieties of Spanish, however, is insufficient to dismiss convergence.

For Spanish subject expression to serve as a diagnostic for convergence with English, cross-linguistic and language-specific tendencies in subject realization must be distinguished. As Weinreich (1963, p. 2) states at the beginning of his book *Languages in contact*, a "prerequisite" to analyzing contact-induced change is that "the differences and similarities between the languages in contact … be exhaustively stated". And yet, the patterning of subject realization in English is rarely taken into account in studies adjudicating on contact effects. Here we draw on the results for English subject realization from Torres Cacoullos and Travis (2014) and Travis and Torres Cacoullos (2014), to develop such a statement of differences and similarities (see Torres Cacoullos & Travis, 2015b). The comparison demonstrates that the grammatical structure of Spanish subject expression (as revealed in the linguistic conditioning of variability) presents clear differences from that of English *I* realization. These differences between the languages in contact, or "conflict sites" (Poplack & Meechan, 1998, p. 132), enable us to rule out cross-linguistic tendencies when making comparisons, and establish subject expression as an appropriate linguistic variable to ascertain convergence.

What cross-linguistic tendencies on subject realization may be ventured? Though the notion of null-subject, as opposed to non-null-subject, languages has enjoyed currency and finer classifications have been put forward (from both formalist (e.g. Roberts & Holmberg, 2010) and typological (Dryer, 2011) perspectives), these have yet to be substantiated by empirical study of the characteristics of the postulated language types. A survey of reported

rates across different languages plainly fails to support the proposed taxonomies. Considering just 1sg expression, rates in so-called "null-subject languages" (those in which verbs are marked for subject agreement) range from 21% in Polish (Chociej, 2011, p. 52), to 50% in Spanish (Ranson, 1991, pp. 135, 138), and in so-called "radical pro-drop languages" (in which there is no agreement marking on the verb), from 16% in Japanese (Lee & Yonezawa, 2008, p. 738) to 66% in Mandarin (Jia & Bayley, 2002, p. 110). Thus, merely classifying two languages as being of one or another of these types goes little way toward furnishing predictions about rates of expression and, more importantly, tells us nothing about the loci of (dis)similarities in the probabilistic constraints on subject expression.

Candidates for cross-linguistically valid constraints are subject continuity and priming, which we have discussed above, and coordination, which we discuss below. Each of these has been replicated in quantitative studies of variation between expressed and unexpressed subject pronouns in languages other than Spanish, subject continuity for languages as diverse as Cantonese, Italian, Russian (Nagy, Aghdasi, Denis & Motut, 2011, pp. 141–142) and Arabic (Owens, Dodsworth & Kohn, 2013, p. 268; Parkinson, 1987, p. 354); both subject continuity and priming for Auslan (McKee, Schembri, McKee & Johnston, 2011, pp. 387–389) and the Vanuatuan language Tamambo (Meyerhoff, 2009, pp. 308–309); and coordination for Finnish (Helasuvo, 2014) and Russian (Nagy et al., 2011, p. 142). Note that these probabilistic constraints cut across proposed taxonomies of languages based on the notion of null subject.

English subject realization follows these cross-linguistic tendencies, but it also exhibits language-specific constraints. In the following two sections we look at patterns of English subject realization and test whether there has been adaptation in the NMSEB Corpus to these English-specific constraints.

6.1 *Unexpressed subjects in English vs. Spanish*

The overall rate of subject pronoun *I* in conversational English is considerably higher than that of *yo* in any Spanish variety, with UNexpressed 1sg subjects in the vicinity of just 2% (151/∼9,000) (Torres Cacoullos & Travis, 2014). But given the uninformativeness of rates discussed above, a more telling difference between the two languages lies in the variable context. In a sample of (expressed and unexpressed) 1sg subjects from the Santa Barbara Corpus of Spoken American English (SBCSAE, Du Bois, Chafe, Myer, Thompson, Englebretson & Martey, 2000–5), Torres Cacoullos and Travis (2014, pp. 22–23) find no cases of unexpressed *I* in interrogatives, relative clauses or subordinate (complement, adverbial, or *if*) clauses (other than ones involving coordination, as

in *if I go out and Ø ask for it* (SBCSAE 17, Jim, line 7)). Thus, for *I* expression, variability is found only in declarative main clauses. Such a restriction to declarative main clauses is absent from Spanish.

Within declarative main clauses, variation between expressed and unexpressed *I* occurs with both non-coordinated and coordinated verbs. We see in Table 3 that in English, coordinated verbs, defined here as verbs with coreferential subjects conjoined with *and*, present a notably lower rate than non-coordinated verbs. A similar pattern is seen in the NMSEB Corpus for Spanish, where the subject expression rate is also lowest in coordinated verbs with *y* "and", at just 5%. However, a shared pattern cannot be construed as a point of convergence in and of itself because it does not speak to whether change has taken place. This is determined through comparisons with non-contact Spanish benchmarks. A similar effect has been reported for "and" coordination across varieties of Spanish, including in Colombia (the figures for which are given in Table 3, based on the Corpus of Conversational Colombian Spanish), Mexico City (Lastra & Butragueño, 2015) and Puerto Rico (Cameron, 1992, p. 206)).

Most striking in English is the prosodic constraint that applies to non-coordinated verbs. As seen in Table 3, with non-coordinated verbs, unexpressed 1sg subjects occur virtually only in absolute initial position of the IU, as in (15), a constraint that has been intuited by a number of scholars (e.g. Napoli, 1982) as an application of a more general phonological process of (variable) "left-edge deletion" (Weir, 2012). Thus, outside "and-" coordination, the variable context must be restricted to prosodic-initial position in English.

(15)   Miles: *(H) But it was like I went [to] Bahia,*
        Jamie: *[What's] --*
        Miles: *... last Sunday,*
                  (H) ... Ø got there at eight,
                  (Hx) .. Ø left a te=n,
                  ... Ø dropped this person off at home,
                  in Foster City,
                     [SBCSAE – 02 Lambada, 848–854]

We find no such initial-position prosodic constraint for Spanish subject expression. Prosodic position thus qualifies as a "conflict site" (Poplack & Meechan, 1998, p. 132), according to which the structure of variability differs for Spanish and English 1sg subject expression, beyond patent overall rate differences.

If bilingual Spanish 1sg subject expression has been influenced by the English prosodic constraint, with non-coordinated verbs we should see higher rates of UNexpressed subjects where unexpressed subjects occur in English, namely in IU-initial position (as in line 1 in (6) above). That is, we should see lower rates of expressed subjects in IU-initial position than in non-IU-initial position. (Non-IU-initial position includes verbs

Table 3. *Rate of 1sg subject expression by "and" coordination and according to position in prosodic unit (Intonation Unit, IU), in the NMSEB Corpus, compared with non-contact English and Spanish.*

	I English[a]		*yo* NMSEB Corpus[b]		*yo* Spanish[a]	
	%	(N)	%	(N)	%	(N)
"and"-coordinated coreferential verbs	24%	(98)	5%	(73)	26%	(27)
Non-coordinated verbs	78%	(348)	27%	(2045)	52%	(883)
IU-initial	71%	(262)	34%	(1074)	55%	(645)
Non-IU-initial	99%	(86)	19%	(971)	45%	(238)

[a]English (Santa Barbara Corpus of Spoken American English) and non-contact Spanish (Corpus of Conversational Colombian Spanish) from Torres Cacoullos and Travis; in English sample *I* rate is an artificial 34%, with two tokens of expressed *I* extracted for each unexpressed.
[b]Excluded from total are tokens of an initial adverb with an unexpressed subject (*e.g. ya no me acuerdo* "I don't remember anymore" [03 La Marina, 0:37:00]) (N = 95) and tokens of *o* "or" (N = 7).

preceded by conjunctions such as *cuando* "when", *pero* "but" (as in (9) above, line 3), adverbs, fillers, or other (more substantial) material (as in (12) above).

Instead, returning to Table 3 we see that, whereas in English (SBCSAE) the highest rate of *I* expression is found in non-IU-initial position, in the NMSEB Corpus *yo* expression is highest in IU-initial position, as is also the case in the non-contact variety of Spanish.

Where we do find English parallels with Spanish is in terms of cross-linguistic tendencies for priming and subject continuity. Coreferential 1sg subject priming has an effect: though rare, when unexpressed *I* does occur, it tends to do so in clusters. The subject continuity effect in English – a higher rate of expression in Switched Human Reference contexts – is bound to unexpressed-to-unexpressed priming, which tends to occur in a coreferential context, thus raising the rate of unexpressed *I* in this context (Torres Cacoullos & Travis, 2015b).

Thus, while the NMSEB Corpus and non-contact varieties of Spanish and English all share candidate cross-linguistic patterns, on the more restricted variable context in English and on the clear "conflict site" with Spanish (that of prosodic initial position), the NMSEB Corpus data demonstrate no convergence with English.

6.2 Expressed **yo** vs. stressed **I**

It could be countered, nevertheless, that, given the supremely lopsided rates of the variants, *I* expression is not a pertinent English variable enticing bilinguals to convergence. Alternatively, English subject pronoun stress has been interpreted as equivalent to Spanish expressed pronouns. For example, Payne (1997, p. 43) claims that "Spanish pronouns correspond [functionally] to English stressed pronouns (roughly speaking)". Givón's (1983b, p. 17) topic accessibility continuum also suggests equivalency, placing unstressed pronouns in a language such as English at the same level as

person-number agreement with unexpressed subjects in a language such as Spanish (both coding more continuous or accessible participants), and placing stressed and expressed (independent) pronouns at the same level (both coding less continuous participants).

Looking again at data from the SBCSAE, Travis and Torres Cacoullos (2014, p.373) found a rate of stressed *I* of 14% (163/1,133), setting aside formulaic units (discourse marker and quotative uses of *I think* and other collocations, as noted above) which, in general, lack stress on *I*. Multivariate analysis reveals similarities and differences in the linguistic conditioning of stressed *I* and expressed *yo*. The hypothesized cross-linguistic constraint of coreferential subject priming is shared: stressed *I* to stressed *I* priming is operative, such that stressed *I* is favored when the preceding clause subject was a coreferential stressed *I*, just as we observed for *yo*-to-*yo* and unexpressed-to-unexpressed priming. However, the subject continuity constraint, though congruent with referent accessibility, is configured differently, with stressed *I* being subject to a distance effect as opposed to the local coreferentiality effect that is relevant for expressed *yo*. Stress on *I* is not favored when there is a switch in reference from the subject of the preceding clause (even in the case of Human Switched Reference) but when the previous mention is at a distance of two or more clauses. *Yo* expression in the NMSEB Corpus adheres to the Spanish, not the English, pattern, with the sharpest difference in *yo* rate being found between contexts with none vs. at least one intervening human subject.

The strong cognition verb class effect on *yo* is absent for stressed *I*, such that lexical types that might be classified as cognition verbs show disparate behavior. *I think* (the putative translation counterpart of *yo creo*) displays a stressed *I* rate close to the average (15%, 23/157), but *I don't know* (+ CLAUSE) shows a significantly higher rate of stress (20%, 12/59). Other cognition

verbs (*I guess* and *I mean*) are virtually never stressed (0/36 and 1/188 respectively). In Spanish, on the other hand, cognition verbs tend to behave as a class, which favors *yo* expression (Travis & Torres Cacoullos, 2012, pp. 734–742). The NMSEB Corpus is again aligned with monolingual Spanish, with cognition verbs favoring expressed *yo*. Besides this general verb class effect, we find among the cognition verbs a *yo* rate that is twice as high for *yo creo* "I think" (87%, 94/108) as for (*yo*) *no sé* "I don't know" (41%, 56/138), also conforming to the monolingual Spanish pattern of high rates of *yo creo* while differing from the English pattern of higher rates of stressed *I* in *I don't know*, thus demonstrating distinct lexically particular constructions in the two languages as well.

In sum, as with expression of *I*, we find no evidence for convergence with patterns of stress on *I*. The bilinguals in the NMSEB Corpus evince no alteration of Spanish constraints on 1sg subject expression and no adaptation to English language particular constraints. As characterized by the linguistic conditioning of variant choice, *yo* in the NMSEB Corpus is grammatically similar to expressed *yo* in other varieties of Spanish, and different from candidate English counterparts expressed *I* and stressed *I*.

Further, recall that, despite copious code-switching in the corpus, we found no significant effect for recent use of English in the multivariate analysis (Table 1). We now turn to consider the role of code-switching.

7. Subject expression under code-switching

In Torres Cacoullos and Travis (2010, 2011), with another dataset of New Mexican Spanish, we compared the linguistic conditioning of variants in the presence vs. absence of code-switching by "frequent" code-switchers, defined as those who produced at least 20% of their 1sg tokens in the presence of multi-word English strings within the preceding ten IUs or three clauses, whichever represented the larger discourse segment (only rarely did three clauses go over 10 IUs) (Torres Cacoullos & Travis, 2010, p. 187; Torres Cacoullos & Travis, 2011, p. 255).[6] We found no effect using this measure, but it may be that any effect for code-switching has dissipated at a distance of ten prosodic units. Here, we therefore take a maximally close measure, namely the same or the preceding clause.

7.1 Coding maximally proximate English

To register the maximally proximate presence of English with respect to instances of variable Spanish subject expression, we take into account the stretch of speech

from the immediately preceding clause to the end of the target clause (defining clauses as outlined above; Section 4.2). That is, we consider both preverbal and postverbal material. Preverbally, we include the preceding clause and material that lies between that clause and the target. For example in (16), *me acuerdo* is coded as being preceded by multi-word English produced by the same speaker (here, *all of a sudden, no sister to talk to*).

(16) Preverbal presence of English, up to and including the preceding clause
Aurora: *... y luego sin su familia,*
y ella y su familia eran muy juntos.
Ricardo: *... sh= X.*
Aurora: *y luego all of a sudden,*
... (0.9) no sister to talk to,
no?
Ricardo: *(H) yeah.*
Aurora: *(TSK) oh=,*
otra cosa,
Ø me acuerdo,
[15 Las cosas viejas, 0:34:19-0:34:29]
Aurora: "... and then without her family,
and she and her family were very close."
Ricardo: "... sh= X."
Aurora: "and then *all of a sudden,*
right?" *... (0.9) no sister to talk to,*
Ricardo: "(H) yeah."
Aurora: "(TSK) oh=,
another thing,
 (I) remember,"

Postverbally, we go to the end of the clause, defining clause completion as the point at which nothing further is "projected" (Hopper & Thompson, 2008) (see Torres Cacoullos & Travis, to appear). We count material up to the end of the IU, as in (17). We do not go beyond the IU in cases of truncation or prosodic completion (indicated by a period or question mark). We do go beyond the IU in cases of continuing intonation (indicated by a comma), but only when the completion of the minimal syntactic unit is projected by the syntax, as in (18), and not when it is not, as in (19), where *con mi flashlight* "with my flashlight" is an adjunct.

(17) Postverbal presence of English in the same clause, same IU
Miguel: *... pues Ø le dije que I was gonna go= y,*
[04 Piedras y gallinas, 1:07:02-1:07:04]
Miguel: "... well (I) told him that *I was gonna go=* and,"

(18) Postverbal presence of English in the same clause, following IU
Monica: *.. yo creo que tendríanos,*
.. maybe twelve or thirteen years.
[11 El trabajo, 10:43:56-0:43:59]

Table 4. *Rate of* yo *according to coreferential subject priming (realization of previous coreferential 1sg subject) and code-switching (presence of multi-word English by the speaker, within same or immediately preceding clause).*

| | Realization of previous coreferential 1sg subject | | | | | | | |
| | Spanish *yo* | | English *I* | | Spanish Ø | | Overall totals[a] | |
Presence of English	%	(N)	%	(N)	%	(N)	%	(N)
Multi-word English present	38%	(26)	20%	(137)	12%	(103)	24%	(424)
Spanish only	35%	(129)	15%	(33)	13%	(466)	25%	(1070)
Overall totals[b]	38%	(254)	22%	(208)	11%	(865)		

[a]Totals are greater than the sum of corresponding columns because of tokens not coded for priming (primes at a distance of five or more clauses, involving quotation, occurring postverbally or in a different IU from the verb).
[b]Totals are greater than the sum of corresponding rows because of tokens not coded for proximate code-switching (single-word English items, N = 443; proper nouns, N = 76; interlocutor speech, N = 182).

> Monica: ".. <u>I think</u> that (we) would have been,
> .. *maybe twelve or thirteen years*."

(19) No postverbal English in the same clause
(*flashlight* in adjunct, following IU)
Manuel: ... *(H) áhi donde Ø iba mirando,*
... *con mi flashlight,*
[16 Trip to Africa pt.1, 0:24:19-0:24:22]
Manuel: "... (H) there where (I) was looking,
... with my flashlight,"

For the most stringent test, we consider here multi-word English strings by the speaker who produced the target token, as in (16), (17), and (18) above (N = 424). Single-word English-origin items not appearing in the dictionary issued by the Royal Spanish Academy, the *Diccionario de la Real Academia* (www.rae.es), were coded separately pending close analysis (N = 443), since many nouns and verbs are likely to be borrowings (Aaron, published online January 14, 2014; Torres Cacoullos & Aaron, 2003) (see discussion in Poplack & Dion, 2012; Stammers & Deuchar, 2012). Such items with at least five tokens from at least two different speakers include *daddy*, *grandma*, *lonche* "lunch", *lonchar* "to have lunch", names of years, and the conjunction *so*. Also set aside here is material produced by the interlocutor – both multi-word English strings (N = 120) and English-origin single words (N = 62). Finally, set aside for present purposes were proper nouns such as *Albuquerque*, *Navy*, *Macy's*, which have been hypothesized to trigger use of the language with which they may be culturally associated (Witteman & Van Hell, 2009) (N = 76). All other tokens were classified as occurring in the absence of proximate code-switching to English.

Following these protocols, the ratio of tokens produced in the proximate presence of multi-word English to those appearing in the absence of any English in the same or immediately preceding clause is approximately 1:2.5 (424
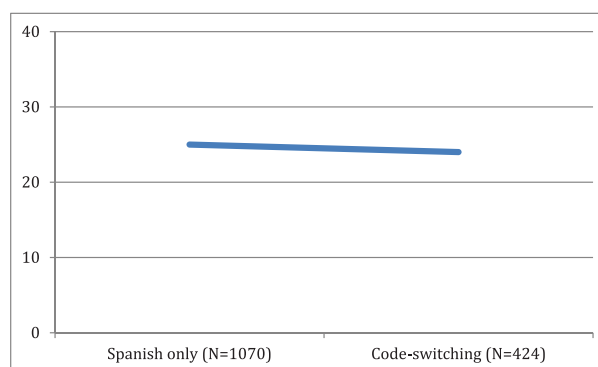

Figure 2. (Colour online) Rate of *yo* according to code-switching (presence of multi-word English by the speaker, within same or immediately preceding clause) (N = 1,494).

to 1,070), with numbers sufficiently robust to test the role of code-switching.

7.2 Proximate code-switching and yo expression

If code-switching is a mechanism of change, intrinsically promoting grammatical convergence as displayed by an elevated rate of expressed subject pronouns, we should observe a higher *yo* rate precisely in maximal proximity to an English code-switch. But, as seen in Table 4, in the right-hand Totals column, the *yo* rate when multi-word English is maximally proximate (at 24%, 103/424) is no different from that in the absence of any English (at 25%, 270/1,070) (and, not shown here, to that in the presence of a single-word English item, 24%, 106/443). This is consistent with the non-significance of proximate code-switching in the multivariate analysis seen in Table 1, and displayed graphically in Figure 2, where the horizontal line indicates the resounding lack of an effect of presence of multi-word English on the rate of subject expression.
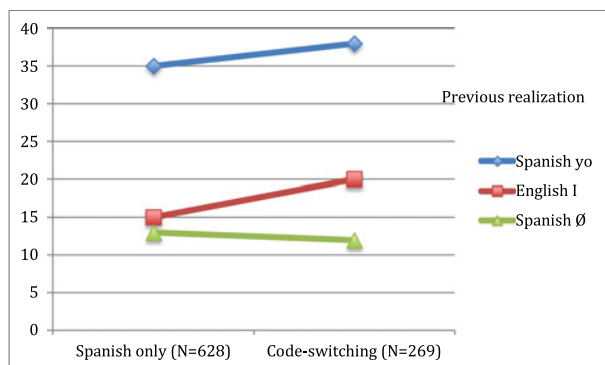
Figure 3. (Colour online) Rate of *yo* according to code-switching (presence of multi-word English by the speaker, within same or immediately preceding clause) and coreferential subject priming (previous coreferential 1sg subject realization) (N = 897).

7.3 Priming: Intra- and inter-linguistic

Despite a privileged theoretical status there is scant evidence for a contribution of code-switching to grammatical change. Ample evidence has been found, however, for structural priming in stable variation within the one language, and also between languages in psycholinguistic experiments (see e.g. Bernolet, Hartsuiker & Pickering, 2012, pp. 504–505; Fleischer, Pickering & McLean, 2012, p. 270; Hartsuiker, Pickering & Veltkamp, 2004, p. 412). Does cross-language priming apply in these natural code-switching NMSEB Corpus data?

To determine whether coreferential English *I* as well as Spanish *yo* (which were combined in the multivariate analyses in Table 1) primes *yo*, we consider the rate of *yo* according to realization of the previous coreferential mention as Spanish *yo*, English *I*, or Spanish unexpressed, shown in the first three columns in Table 4. The bottom Totals row shows a clear effect for English *I*: when the realization of the previous coreferential 1sg subject is English *I*, as in (20), the rate of *yo* is double (22%, 45/208) than that when it is an unexpressed Spanish subject (11%, 95/865) (*p* = .0001 Fisher's exact test). Stronger yet is *yo*-to-*yo* priming, with a *yo* rate nearly double again (38%, 96/254) (*p* = .0002).

(20) Cross-language coreferential 1sg subject priming: English *I* to Spanish *yo*
   Miguel: ...(1.5) I was like nineteen.
        ... years old.
        .. pero yo me acordaba.
        [04 Piedras y gallinas, 0:20:08-0:20:13]
   Miguel: "...(1.5) I was like nineteen.
        ... years old.
        .. but I remembered."

Figure 3 (based on the numbers given in Table 4) shows the rate of *yo* according to the proximate use of English

and by realization of the previous coreferential mention as Spanish *yo*, English *I* or Spanish unexpressed. As in Figure 2, the lines remain mostly horizontal, depicting the lack of an effect of proximate code-switching.[7] But the priming effect, intra- as well as inter-linguistic, is displayed by the positioning of the lines: previous Spanish *yo* results in a higher rate of *yo* expression than does previous English *I*, which itself results in a higher rate of *yo* expression than does Spanish unexpressed.

In sum, we have found that cross-language priming, reported in experimental studies, applies in natural code-switching: though indubitably weaker than intra-language (*yo*-to-*yo*) priming, there is cross-language coreferential 1sg subject (*I*-to-*yo*) priming (Travis et al., to appear). What is important here is that the higher rate of *yo* under priming from *I* seen in Figure 3 contrasts with its flat rate in proximity to any multi-word English code-switch seen in Figure 2.

8. Discussion

The question as yet untackled is whether, beneath the flat rate of subject expression, the same linguistic constraints operate in proximity to code-switching. To examine this, we conduct two further multivariate analyses to compare tokens in the presence and in the absence of proximate code-switching, testing the constraints identified in the aggregate analysis of Table 1. The answer is "yes", as shown in Table 5. The direction of effect in maximal proximity to an English string and in the absence of any English element is the same: *yo* expression is favored in non-coreferential contexts, following a 1sg pronoun and with cognition verbs; it is disfavored with reflexive-marked verbs and perfective aspect.[8] By the LINGUISTIC CONDITIONING of variation, speakers are behaving – grammatically – the same way in Spanish when they are code-switching as when they are not. Spanish constraints remain unaltered not only when the speaker's code-switch occurs within ten preceding prosodic units (Torres Cacoullos & Travis, 2010, 2011), but also within the

[7] Behind the sloping line for English *I* lies the fact that in Spanish-only contexts as compared with code-switching contexts – i.e. English in the same or immediately preceding clause – English *I* primes are not only less frequent (Figure 4 below) but – since by definition this context excludes English *I* primes in the immediately preceding clause – they have a weaker effect (recall that coreferential subject priming is strongest at a distance of zero intervening clauses; see Section 4.2; see also Travis et al., to appear).
[8] Not much should be made of the lack of significance of Semantic class under "Proximate code-switching present" in Table 5, since the total N is lower. The greater strength of Subject continuity in this analysis follows from the relative weakness of priming within rather than across languages (on the interaction between subject continuity and priming, see Cameron, 1994, pp. 39–40, Travis & Torres Cacoullos, 2012, p. 731; Travis et al., to appear).

Table 5. *Separate multivariate analyses of factors contributing to choice of subject pronoun* yo *(vs. unexpressed subject).*

	Proximate code-switching N = 424; Input .18 (overall rate: 24%)				Spanish only N = 1070; Input .16 (overall rate: 25%)			
	Probability	% *yo*	N	Data distribution	Probability	% *yo*	N	Data distribution
Subject continuity:								
Human switched reference								
Non-coreferential	.68	39%	157	42%	.58	35%	368	41%
Coreferential	.37	14%	219	58%	.45	15%	534	59%
Coreferential subject priming:								
Realization of previous coreferential 1sg subject								
Pronoun (*yo* or *I*)	.59	23%	163	61%	.65	31%	162	26%
Unexpressed (Spanish 1sg)	.37	12%	103	39%	.45	13%	466	74%
Semantic class of verb								
Cognition	[.58]	37%	100	24%	.75	50%	207	20%
Other	[.47]	20%	322	76%	.43	19%	853	81%
Reflexive clitic me								
Absent	.54	27%	349	82%	.55	28%	922	86%
Present	.33	13%	75	18%	.21	11%	148	14%
Tense-aspect-mood								
Imperfective	.61	31%	280	70%	.54	30%	694	72%
Perfective	.26	9%	122	30%	.41	14%	275	28%

See notes to Table 1. Not selected as significant with proximate code-switching: Semantic class of verb (shown in []).

immediately preceding or the same structural unit, the clause.

Thus code-switching to English affects here neither the overall rate nor the linguistic conditioning of Spanish subject expression. It does however affect the distribution of the data. Realization of previous coreferential 1sg subject displays a reversal of the relative frequencies of the factors, or contextual features, shown in the fourth column (labeled data distribution) for each analysis. Under proximate code-switching, close to two thirds of the data (61%) occur in the context of a previous pronoun, while when English is absent, just one quarter (26%) of the data occur in this context.

Therefore, the difference code-switching makes is in the data distribution, that is, the relative frequencies of preceding pronouns and preceding unexpressed subjects. As seen in Figure 4 and Table 5, with a maximally proximate code-switch to English (in the same or immediately preceding clause) ("Code-switching"), the ratio of instances of the variable whose previous coreferential 1sg subject was realized as unexpressed, with respect to tokens whose previous realization was a pronoun, is approximately 1:1.6 (39% to 61%). In the absence of any English within the same or preceding
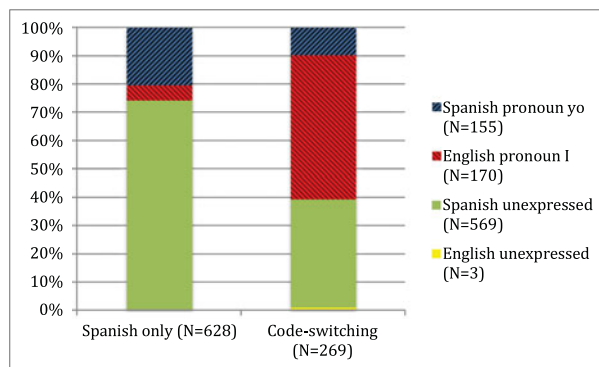

Figure 4. (Colour online) Contextual distribution of primes: previous coreferential 1sg subject realization according to code-switching (presence of multi-word English by the speaker, within same or immediately preceding clause) (N = 897).

clause ("Spanish only"), the ratio is reversed, 2.8:1 (74% to 26%)!

Given the rarity of unexpressed subjects in English (here, 3/211 English previous coreferential mentions),

when speakers are code-switching, fewer instances of variable Spanish subject expression occur in an environment favorable to unexpressed subjects – that of a previous unexpressed subject – than is the case in the absence of code-switching. In this sense, the presence of proximate code-switching to English "interferes" with intra-linguistic priming, resulting in fewer opportunities for unexpressed-to-unexpressed priming.

We venture the hypothesis that rather than code-switching intrinsically inducing grammatical alteration, what is at work is associated shifts in the frequency of contextual features contributing to variant choice; in particular, those relevant to both intra-language priming of the same structure and to cross-language priming of a parallel structure. Contrary to the convergence-via-code-switching hypothesis, the present data support this CONTEXTUAL DISTRIBUTION VIA-CODE-SWITCHING hypothesis.

9. Conclusion

Our approach to the theme of this special issue of *BLC* on cross-language effects in bilingual production and comprehension has been a community-based search to pinpoint such effects, appealing to both rates and constraints in variable Spanish first person singular subject expression. Adopting the variationist comparative method, we asked the question: when New Mexican bilinguals use *yo* as opposed to an unexpressed 1sg subject, do they adhere to Spanish constraints or has there been some adaptation to English patterns of 1sg subject realization? As revealed in multivariate analyses,

the linguistic conditioning of variant choice in these bilinguals is the same as in monolingual Spanish varieties but divergent from patterns of 1sg subject realization in English, including in the presence of maximally proximate code-switching. This is evidence that patterns of 1sg expression in Spanish do not converge with those in English for speakers in this bilingual community, even when code-switching.

If code-switching shapes bilingual grammar(s), this should have been discernable in this study given the contact setting (one of prolonged contact) and the variable under consideration (1sg subject expression, which we established is an appropriate linguistic variable by which to measure convergence). The abundance of code-switching in the NMSEB Corpus permitted a stringent test of the much-surmised role of code-switching in promoting convergence; to wit, whether there is an elevated rate of pronominal subjects in Spanish when there is a switch to English in the same or immediately preceding clause. Not only is no such global effect of code-switching substantiated, the same structural coreferential 1sg subject priming effect is observed both within and across languages.

The direct (cross-language) effect in operation thus appears to be one of priming of parallel structures. Indirectly, code-switching is accompanied by altered distributions of contexts of occurrence. Rather than intrinsically inducing grammatical alteration, as measured by comparisons of the linguistic conditioning of variant selection in bilingual and monolingual communities, code-switching makes a difference to the relative frequencies of relevant contextual features contributing to variant choice, as per the bilingual contextual distribution hypothesis.

Appendix A. *Transcription conventions (Du Bois et al., 1993)*

Line break	new Intonation Unit	[ ]	overlapped speech
.	final intonation contour	X	unclear syllable
,	continuing intonation contour	@	one syllable of laughter
?	appeal intonation contour	<@ @>	speech produced while laughing
--	truncated intonation contour	(H)	in breath
-	truncated word	(Hx)	out breath
..	short pause (0.5 sec)	(())	researcher's comment
...	medium pause (0.5–0.7 sec)	!	booster; word uttered with force
... ( )	timed pause (over 0.7 sec)	(TSK)	click
=	lengthened syllable		

Appendix B. Characteristics of the NMSEB Corpus speakers in this study*

Transcript	Speaker	Year born	Sex	Education level	Current residence	Birthplace	Occupation
1	Susan	1934	F	high	Albuquerque	Albuquerque	Stay at home mom
2	Bartolomé	1928	M	middle	Albuquerque	Sthn Colorado	Retired fire-fighter
3	Sandra	1943	F	college	Española	Española	Retired
4	Miguel	1944	M	middle	Valencia	Valencia	Laborer
5	Rocío	1945	F	high	Santa Fe	Santa Fe	Retired school teacher aid
6	Ivette	1946	F	high	Albuquerque	Valencia	Factory worker
7	Samuel	1922	M	college	Taos	Taos	School coach
8	Inmaculada	1952	F	college	Albuquerque	San Miguel	Social worker
9	Fabiola	1954	F	college	Taos	Taos	Secretary
9	Molly	1939	F	middle	Taos	Taos	Retired school cook
7, 10	Pedro	1953	M	college	Taos	Rio Arriba	School administrator
11	Mónica	1941	F	high	Albuquerque	Taos	Factory worker/school custodian
12	Marta	1964	F	college	Río Arriba	Albuquerque	Guest services manager
12	Victoria	1959	F	college	Río Arriba	Española	Retired school teacher/counselor
13	Betty	1925	F	high	Sandoval	Rio Arriba	Retired
14	Anita	1941	F	high	Albuquerque	San Miguel	Executive director
15	Aurora	1962	F	college	Sandoval	Española	Teacher
12, 16	Manuel	1954	M	middle	Río Arriba	Rio Arriba	Electrician/rancher
17	Javier	1936	M	high	Taos	Taos	Rancher/janitor
18	Francisco	1963	M	high	Río Arriba	Rio Arriba	Miner
24	Diana	1941	F	high	Taos	Taos	Dry cleaner
24	Marco	1941	M	middle	Taos	Taos	Miner
25	Leandro	1931	M	middle	Río Arriba	Taos	Miner

* See Footnote 2 in the main text.

References

Aaron, J. E. Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*. doi:10.1177/1367006913516021 Published online by Sage, January 14, 2014.

Amaral, P. M., & Schwenter, S. A. (2005). Contrast and the (non-)occurrence of subject pronouns. In D. Eddington (ed.), *Selected proceedings of the 7th Hispanic Linguistics Symposium*, pp. 116–127. Somerville, MA: Cascadilla Press.

Backus, A. (2005). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism, 9,* 307–340.

Benevento, N., & Dietrich, A. I think, therefore *digo yo*: Variable position of the 1sg subject pronoun in New Mexican Spanish–English code-switching. *International Journal of Bilingualism*. doi:10.1177/1367006913516038. Published online by Sage, January 14, 2014.

Bentivoglio, P. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2012). Effects of phonological feedback on the selection of syntax: Evidence from between-language syntactic priming. *Bilingualism: Language and Cognition, 15,* 503–516.

Bills, G. D., & Vigil, N. A. (1999). Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology, 53,* 43–66.

Bills, G. D., & Vigil, N. A. (2008). *The Spanish language of New Mexico and southern Colorado: A linguistic atlas*. Albuquerque, NM: University of New Mexico Press.

Cameron, R. (1992). Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation. Ms., University of Pennsylvania.

Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change, 5,* 305–334.

Cameron, R. (1994). Switch reference, verb class and priming in a variable syntax. *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on Variation in Linguistic Theory, 30,* 27–45.

Carvalho, A. M., Orozco, R., & Shin, N. L. (eds.) (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, DC: Georgetown University Press.

Chociej, J. (2011). Polish null subjects: English influence on heritage Polish in Toronto. Ms., University of Toronto.

Cifuentes, H. (1980–1). Presencia y ausencia del pronombre personal sujeto en el habla culta de Santiago de Chile. *Homenaje a Ambrosio Rabanales: Boletín de Filología de la Universidad de Chile, 31,* 743–752.

Clyne, M., Eisikovits, E., & Tollfree, L. (2001). Ethnic varieties of Australian English. In D. Blair & P. Collins (eds.), *English in Australia*, pp. 223–238. Amsterdam & Philadelphia, PA: John Benjamins.

Croft, W. (1995). Intonation units and grammatical structure. *Linguistics, 33,* 839–882.

Dryer, M. S. (2011). Expression of pronominal subjects. In M. S. Dryer & M. Haspelmath (eds.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. [http://wals.info/chapter/101, retrieved July 20, 2013].

Du Bois, J. W., Chafe, W. L., Myer, C., Thompson, S. A., Englebretson, R., & Martey, N. (2000–5). *Santa Barbara Corpus of Spoken American English, Parts 1–4.* Philadelphia, PA: Linguistic Data Consortium.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (eds.), *Talking data: Transcription and coding in discourse*, pp. 45–89. Hillsdale, NJ: Lawrence Erlbaum.

Enríquez, E. V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: ConsejoSuperior de Investigaciones Científicas, Instituto Miguel de Cervantes.

Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variation subject personal pronoun expression in Spanish. *Language, 88,* 526–557.

Fernández-Gibert, A. (2010). From voice to print: Language and social change in New Mexico, 1880–1912. In S. Rivera-Mills & D. J. Villa (eds.), *Spanish of the U.S. Southwest: A language in transition*, pp. 45–62. Madrid: Iberoamericana.

Fleischer, Z., Pickering, M. J., & McLean, J. F. (2012). Shared information structure: Evidence from crosslinguistic priming. *Bilingualism: Language and Cognition, 15,* 568–579.

Givón, T. (1983a). Topic continuity in discourse: An introduction. In Givón (ed.), pp. 1–41.

Givón, T. (1983b). Topic continuity in spoken English. In Givón (ed.), pp. 343–363.

Givón, T. (ed.) (1983c). *Topic continuity in discourse: A quantitative cross-linguistic study*. Amsterdam & Philadelphia, PA: John Benjamins.

Gonzales, M. D. (1999). Crossing social and cultural borders: The road to language hybridity. In L. Galindo & M. D. Gonzales (eds.), *Speaking Chicana: Voice, power and identity*, pp. 13–38. Tucson, AZ: University of Arizona Press.

Gonzales-Berry, E., & Maciel, D. R. (eds.) (2000). *The contested homeland: A Chicano history of New Mexico*. Albuquerque, NM: University of New Mexico Press.

Gonzales-Berry, E. (2000). Which language will our children speak? The Spanish language and public education policy in New Mexico, 1890–1930. In Gonzales-Berry & Maciel (eds.), pp. 169–189.

Granda Gutiérrez, G. de. (1972). *Transculturación e interferencia linguistica en el Puerto Rico contemporáneo (1898–1968)*. Puerto Rico: Editorial Edil.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34,* 365–399.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1,* 131–149.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. E. Bullock & A. J. Toribio (eds.), *The Cambridge handbook on linguistic code-switching*, pp. 21–39. Cambridge: Cambridge University Press.

Gumperz, J., & Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In D. Hymes (ed.), *Pidginization and creolization of languages*, pp. 151–167. Cambridge: Cambridge University Press.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages. *Psychological Science, 15,* 409–414.

Helasuvo, M.-L. (2014). Searching for motivations for grammatical patternings. *Pragmatics, 24,* 453–476.

Holmquist, J. (2012). Frequency rates and constraints on subject personal pronoun expression: Findings from the Puerto Rican highlands. *Language Variation and Change, 24,* 203–220.

Hopper, P. J., & Thompson, S. A. (2008). Projectability and clause combining in interaction. In R. Laury (ed.), *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions* (TSL 80), pp. 99–123. Amsterdam & Philadelphia, PA: John Benjamins.

Jia, L., & Bayley, R. (2002). Null pronoun variation in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics, 8,* 103–116.

Labov, W. (1969). Contraction, deletion, and inherent variability of the English copula. *Language, 45,* 715–762.

Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Basil Blackwell.

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (eds.), *Language in use: Readings in sociolinguistics*, pp. 28–53. Englewood Cliffs, NJ: Prentice-Hall.

Labov, W. (2001). *Principles of linguistic change: Social factors* (vol. 2). Oxford: Blackwell.

Labov, W. (2005). Quantitative reasoning in linguistics. In U. Ammon, N. Dittmar, K. J. Mattheier & P. Trudgill (eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society* (vol. 1), pp. 6–22. Berlin: Mouton de Gruyter.

Lastra, Y., & Butragueño, P. M. (2015). Subject pronoun expression in oral Mexican Spanish. In Carvalho et al. (eds.), pp. 41–60.

Lausberg, H., & Sloetjes, H. (2009). Coding gestural behavior with the NEUROGES-ELAN system. *Behavior Research Methods, Instruments, & Computers, 41,* 841–849.

Lee, D.-Y., & Yonezawa, Y. (2008). The role of the overt expression of first and second person subject in Japanese. *Journal of Pragmatics, 40,* 733–767.

Lipski, J. M. (2008). *Varieties of Spanish in the United States*. Washington, DC: Georgetown University Press.

McKee, R., Schembri, A., McKee, D., & Johnston, T. (2011). Variable "subject" presence in Australian Sign Language and New Zealand Sign Language. *Language Variation and Change, 23,* 375–398.

Meyerhoff, M. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change, 21,* 297–317.

Nagy, N. G., Aghdasi, N., Denis, D., & Motut, A. (2011). Null subjects in heritage languages: Contact effects in a cross-linguistic context. *University of Pennsylvania Working Papers in Linguistics, 17,* Article 16. Available online at http://repository.upenn.edu/cgi/viewcontent.cgi?article=1202&context=pwpl.

Napoli, D. J. (1982). Initial material deletion in English. *Glossa, 16,* 85–111.

Ono, T., & Thompson, S. A. (1997). Deconstructing 'zero anaphora' in Japanese. *Berkeley Linguistics Society, 23,* 481–491.

Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.

Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish of New York: Toward the formation of a speech community. *Language, 83,* 770–802.

Owens, J., Dodsworth, R., & Kohn, M. (2013). Subject expression and discourse embeddedness in Emirati Arabic. *Language Variation and Change, 25,* 255–285.

Parkinson, D. B. (1987). Constraints on the presence/absence of 'optional' subject pronouns in Egyptian Arabic. *15th Annual Conference on New Ways of Analyzing Variation*, pp. 348–360.

Payne, T. E. (1997). *Describing morphosyntax: A guide to field linguists*. Cambridge: Cambridge University Press.

Poplack, S. (1993). Variation theory and language contact: Concepts, methods and data. In D. R. Preston (ed.), *American dialect research*, pp. 251–286. Amsterdam & Philadelphia, PA: John Benjamins.

Poplack, S. (1998). Contrasting patterns of code-switching in two communities. In P. Trudgill & J. Cheshire (eds.), *The sociolinguistics reader: Multilingualism and variation* (vol. 1), pp. 44–65. London: Edward Arnold.

Poplack, S., & Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change, 24,* 279–315.

Poplack, S., & Levey, S. (2010). Contact-induced grammatical change: A cautionary tale. In P. Auer & J. E. Schmidt (eds.), *Language and space: An international handbook of linguistic variation* (vol. 1), pp. 391–419. Berlin: Mouton de Gruyter.

Poplack, S., & Meechan, M. (1998). Introduction: How languages fit together in codemixing. *International Journal of Bilingualism, 2,* 127–138.

Poplack, S., Zentz, L., & Dion, N. (2012a). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition, 15,* 203–225.

Poplack, S., Zentz, L., & Dion, N. (2012b). What counts as (contact-induced) change. *Bilingualism: Language and Cognition, 15,* 247–254.

Ranson, D. L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change, 3,* 133–152.

Roberts, I., & Holmberg, A. (2010). Introduction: Parameters in minimalist theory. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan (eds.), *Parametric variation: Null subjects in minimalist theory*, pp. 1–57. Cambridge & New York: Cambridge University Press.

Sankoff, D. (1988a). Sociolinguistics and syntactic variation. In F. J. Newmeyer (ed.), *Linguistics: The Cambridge survey* (vol. 4), pp. 140–161. Cambridge: Cambridge University Press.

Sankoff, D. (1988b). Variable rules. In U. Ammon, N. Dittmar & K. J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society* (vol. 2), pp. 984–997. Berlin & New York: Walter de Gruyter.

Sankoff, D., Tagliamonte, S., & Smith, E. (2012). Goldvarb LION: A variable rule application for Macintosh. University of Toronto. [http://individual.utoronto.ca/tagliamonte/goldvarb.htm]

Silva-Corvalán, C. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elías Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, pp. 93–120. New York: Cambridge University Press.

Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.

Silva-Corvalán, C. (2001). *Sociolingüística y pragmática del español* (Georgetown Studies in Spanish Linguistics). Washington, DC: Georgetown University Press.

Silva-Corvalán, C. (2003). Otra mirada a la expresión del sujeto como variable sintáctica. In F. Moreno Fernández, F. G. Menéndez, J. A. Samper, M. L. Gutiérrez Araua, M. Vaquero & C. Hernández (eds.), *Lengua, Variación y contexto: Estudios dedicados a Humberto López Morales* (vol. 2), pp. 849–860. Madrid: Arco Libros.

Stammers, J. R., & Deuchar, M. (2012). Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Bilingualism: Language and Cognition, 15,* 630–643.

Szmrecsanyi, B. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory, 1,* 113–149.

Thomason, S.G. (2001). *Language contact: An introduction*. Washington DC: Georgetown University Press.

Torres Cacoullos, R., & Aaron, J. E. (2003). Bare English-origin nouns in Spanish: Rates, constraints and discourse functions. *Language Variation and Change, 15,* 289–328.

Torres Cacoullos, R., & Travis, C. E. (2010). Variable *yo* expression in New Mexico: English influence? In S. Rivera-Mills & D. J. Villa (eds.), *Spanish of the U.S. Southwest: A language in transition*, pp. 185–206. Madrid: Iberoamericana.

Torres Cacoullos, R., & Travis, C. E. (2011). Using structural variability to evaluate convergence via code-switching. *International Journal of Bilingualism, 15,* 241–267.

Torres Cacoullos, R., & Travis, C. E. (2014). Prosody, priming and particular constructions: The patterning of English first-person singular subject expression in conversation. *Journal of Pragmatics, 63,* 19–34.

Torres Cacoullos, R., & Travis, C. E. (2015a). New Mexico Spanish–English Bilingual (NMSEB) Corpus, National Science Foundation 1019112/1019122. [http://nmcode-switching.la.psu.edu/]

Torres Cacoullos, R., & Travis, C. E. (2015b). Assessing inter-linguistic (dis)similarity via intra-linguistic variability for subject expression. In Carvalho et al. (eds.), pp. 83–102.

Torres Cacoullos, R., & Travis, C. E. (to appear). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism* 19.

Torres Cacoullos, R., & Walker, J. A. (2009). On the persistence of grammar in discourse formulas: A variationist study of *that. Linguistics, 47,* 1–43.

Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change, 19,* 101–135.

Travis, C. E., & Torres Cacoullos, R. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics, 23,* 711–748.

Travis, C. E., & Torres Cacoullos, R. (2013). Making voices count: Corpus compilation in bilingual communities. *Australian Journal of Linguistics, 33,* 170–194.

Travis, C. E., & Torres Cacoullos, R. (2014). Stress on *I*: Debunking unitary contrast accounts. *Studies in Language, 38,* 360–392.

Travis, C. E., Torres Cacoullos, R., & Kidd, E. (to appear). Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition*.

Travis, C. E., & Villa, D. J. (2011). Language policy and language contact in New Mexico: The case of Spanish. In C. Norrby & J. Hajek (eds.), *Uniformity and diversity in language policy: Global perspectives*, pp. 126–140. Bristol: Multilingual Matters.

Weiner, E. J., & Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics, 19,* 29–58.

Weinreich, U. (1963). *Languages in contact: Findings and problems*. The Hague: Mouton.

Weinreich, U., Labov, W., & Herzog, M. I. (1968). Empirical foundations for a theory of language change. In W. P. Lehmann & Y. Malkiel (eds.), *Directions for historical linguistics: A symposium*, pp. 95–188. Austin, TX: University of Texas Press.

Weir, A. (2012). Left-edge deletion in English and subject omission in diaries. *English Language and Linguistics, 16,* 105–129.

Winford, D. (2005). Contact-induced change: Classification and processes. *Diachronica, 22,* 373–427.

Witteman, M. J., & Van Hell, J. G. (2009). Triggered codeswitching: An experimental test of the triggering hypothesis. Presented at the 7th International Symposium on Bilingualism, Utrecht, The Netherlands, July 8–11, 2009.