Alternating or mixing languages?

1   Introduction

Must language contact and change be bedfellows? Do bilinguals inevitably mix, or grammatically blend, their languages? Some notion of language mixing is implied by blended labels such as Spanglish or Türkendeutsch, and a growing number of articles, books, and conferences on bilingualism are spotlighting contact effects. To assess hypothesized contact-induced grammatical change, this chapter implements a *convergence evaluation metric* that is based on comparison of variable linguistic structures.

As a case study of the evaluation metric we take subject pronoun expression, a showpiece of convergence for non-null subject languages in contact with null-subject languages, such as Turkish spoken in the Netherlands and Spanish in the United States (e.g., Doğruöz & Backus 2007; Silva-Corvalán 1994: 145-165). Here, we travel to New Mexico in the southwestern United States to observe the most intimate contact between languages—as spoken by the same bilinguals. For subject expression as a linguistic variable, the competing variants are pronominal and unexpressed subjects, illustrated in (1) for Spanish and (2) for English.[1]

(1)
Manuel:	... *y Ø me escamé poco,*	'... and (I) got a little anxious,
	pero yo dije,	but I said,
	... Ø vine a lo que Ø vine.	it is what it is.'
		(lit. (I) came for what (I) came for.)
		(NMSEB 16.1, 24:54-24:57)

(2)
Manuel:	*.. I was able to scramble and,*
	.. find the --
	my other flashlight,
	... Ø turned it on,
	Ø worked on the one I had just broke.
	(NMSEB 16.1, 26:06-26:12)

We begin with a telling method for establishing change and diagnosing its source. The convergence evaluation metric relies on measures of grammatical similarity or difference between language varieties derived from comparisons of the structure of internal variation (Section 2). Data come from a bilingual speech corpus grounded in a well-established speech community and from comparable benchmarks

---

[1]   Examples are produced verbatim from the transcripts. Transcription conventions are presented in the Appendix. Within parentheses following examples from the NMSEB corpus is the recording number and name, and the beginning-ending time stamps of the lines reproduced; line numbers are given for the CCCS and SBCSAE examples. Speaker names are pseudonyms.

(Section 3). Since similarities between languages potentially due to shared cross-linguistic tendencies are not straightforwardly informative as to convergence, analyses center around measurable diagnostic differences. For Spanish and English subject expression, a first diagnostic difference is located in the variable context. Outside coordinate clauses, English null subjects occur only in prosodic-initial position, an environment that in Spanish disfavors null subjects, favoring, conversely, pronominal subjects (Section 4). Within the variable context in each language, a second diagnostic difference is provided by the relative magnitude of probabilistic constraints. Referent accessibility, in terms of coreferentiality with the preceding subject and clause linking, shows the same direction of effect but is a relatively stronger constraint in English than in Spanish (Section 5). On both measures, bilinguals' Spanish and English line up with their respective monolingual counterparts and, most remarkably, are different from each other.

2   Convergence evaluation metric: Comparison of variation patterns

Perhaps the kind of contact-induced change that has attracted the most attention is grammatical *convergence*, or magnified structural similarity (e.g., Aikhenvald 2002: 2). Although grammatical convergence has been widely claimed for contact sites across the globe, past and present, and is treated as a virtual concomitant of language contact, evidence has been modest at best (as noted by Poplack & Levey 2010). Even if we agree that a particular change has in fact occurred, whether we can convincingly attribute it to language contact is still a question. In this section we outline evaluation metrics for contact-induced change.

A demonstration of contact effects faces both a social and a linguistic requirement. As a social condition, while a single speaker can produce an innovation, it only qualifies as change if it is taken up by other members of the community to become a regular pattern (Hopper & Traugott 2003: 36-38; Thomason & Kaufman 1988: 94; Weinreich, Labov & Herzog 1968: 186-187). Linguistically, a *sine qua non* for any contact proposal is the requirement that external influence be disentangled from cross-linguistic tendencies of internal grammatical change. To meet these twin social and linguistic requirements, we evaluate candidate contact effects by contextualizing them both socially and linguistically, that is, with respect to the speech community and the grammatical system in which they are embedded. Thus the data come from community-based spontaneous speech corpora and the analysis centers on language-internal variation patterns.

The pessimistic view was that "the process of linguistic change has never been directly observed" (Bloomfield 1933: 347). Today, more than five decades of sociolinguistic study of changes in progress provide a model for the synchronic observation of contact effects (cf., Labov 1972: 21-23). To evaluate whether contact-induced change is underway, we compare grammars in contact to grammars not in contact. Minimally, then, the evaluation involves a non-contact benchmark with which to compare the contact variety of interest. The most direct evaluation relies on monolingual benchmarks of both languages with which to compare both contact varieties, *as spoken by the bilinguals themselves*, that is, four comparison datasets (see Figure 1 below, top panel).

What is it that can be meaningfully compared to reveal contact-induced change where it has occurred, or to verify its absence where it has not?

1. *Choice of suitable data*. Contact-induced change has been asserted on the basis of disparate data sources, such as compilations of isolated linguistic examples or elicitation in lab settings. But it might be asked whether data obtained this way can be taken as representative of usage in the community. Formal styles may distort the data due to self-monitoring or hyper-correction, in particular for bilinguals' speech in minority communities, which is often stigmatized. Even elicited traditional narratives may bring forth formulaic or fossilized forms. Systematic patterns in the everyday speech of the community are discoverable in the *vernacular*, or unmonitored production (Labov 1984: 29). Sociolinguistically constructed corpora of spontaneous speech thus remain the touchstone of bilingual data (Poplack 1993:258-271).

Implementing the convergence evaluation metric, we draw here on analogous speech data for the comparison: for the contact varieties, a bilingual corpus—capitalizing on the stretches of each language contained therein—and for the non-contact varieties, monolingual benchmark corpora of each language (see Section 3 and Table 1 below).

2. *Wariness of frequency differences.* Changes in frequency are widely appealed to for comparison of grammatical systems. Specifically, a minor variant in one language is predicted to increase modeled on a major use pattern in the contact language (Heine & Kuteva 2005: 44-75). For Spanish in contact with English, predicted is overextension of pronouns on the model of English as a non-null subject language (e.g., Otheguy & Zentella 2012: 167). But while some studies have reported the predicted increase, others have not (see overview by Carvalho, Orozco & Shin 2015: xviii-xx). The counterpart change in bilinguals' English (which has received scant if any attention) would be a lower rate of subject pronoun expression modeled on Spanish.

The convergence prediction does not fare well when it comes to overall rates in New Mexican bilinguals' Spanish, or English. We fail to find a higher subject pronoun rate in these bilinguals' Spanish, at 24% (1,353/5,571), which is close to rates reported for Madrid and Mexico City, at 28% and 25% respectively (N = 12,786, Enríquez 1984: 348; N = 1,389, Lastra & Butragueño 2015: 43). Nor do we find a higher rate of <u>un</u>expressed (null) subjects in bilinguals' English than in a corpus of monolingual English, at approximately 1.5% vs. 3% respectively (Torres Cacoullos & Travis 2018: 112, 161; see Table 1 below on datasets).

Had we found significant differences in the predicted direction, however, we would still council wariness of frequency increases or decreases. Overall rate of use of the linguistic form of interest is equivocal as a measure of grammatical difference and therefore of change, for three principal reasons.

- Overall rate is susceptible to data distributions, since the preponderance in a dataset of some propitious context will boost the rate (Poplack & Levey 2010: 404). Overall rate may be affected by the situational factor of genre, for example, dialogic interaction vs. monologic narrative, often also as a reflection of data distributions (Torres Cacoullos & Travis 2018: Section 7.5.3; Travis 2007: 129-131). Shifting rates may be the fallout of other situational considerations, especially the impact of the interviewer and elicitation strategies (Bailey & Tillery 2004; Hernández 2009: 604-606).
- Given fluctuations in overall rate due to such extra-grammatical considerations, the threshold for a <u>linguistically significant</u> difference—even if statistically significant—is unknown. For example, within monolingual

Spanish itself, the subject pronoun rate in one variety may be double that in another (even within a single study, e.g. ranging from 26% to 50%, Cameron 1993: 306), and a notably wide range in overall pronoun rate characterizes Spanish varieties across numerous studies (Silva-Corvalán & Enrique-Arias 2017: 175-176). This means we cannot automatically interpret a rate difference as evidence of change (even one as great as 24 percentage points).

- An overall rate difference in and of itself is unrevealing as to the mechanism of change. Once a change has been substantiated, even if an internal explanation is rejected, a shift in pronoun rate is compatible with more than one kind of contact effect. Increased pronoun use could just as well be attributed to overextension following the non-null subject model of English (e.g., Heine & Kuteva 2005: 70), as to a default processing strategy for dealing with any second language (e.g., Sorace & Serratrice 2009).

3. *Appeal to conditioning of forms*. The conditioning of competing linguistic forms supplies a more discerning measure of grammatical change. Linguistic conditioning is a model of co-occurrence patterns, or of rates of use *in particular linguistic contexts*. The working hypothesis is that grammatical systems are characterized by the covariation between linguistic elements, and that in definable contexts this covariation is probabilistic (Bresnan, Dingare & Manning 2001; Cedergren & Sankoff 1974; Labov 1969). In the case of subject pronoun expression, despite overall rate differences, the linguistic conditioning is near identical across the numerous studies of Spanish varieties. For example, even with a 24 percentage point difference between San Juan and Madrid, probability values for the effect of accessibility were "startlingly similar" (Cameron 1993: 319), as they were for Caracas and Los Angeles (Silva-Corvalán & Enrique-Arias 2017: 175-176). Besides accessibility, the impact of factors such as priming, verb class and person/number on subject expression is similar across Spanish varieties (Torres Cacoullos & Travis 2018: Chapter 5). It is this shared conditioning that provides a baseline against which to compare contact varieties and determine whether change has taken place.

As a definite measure of *contact-induced* change, the convergence evaluation metric calls for identifying one or more diagnostic differences for variants that are parallel across the two languages in contact (such as pronominal vs. unexpressed subjects). The *diagnostic difference* (or "conflict site") is a difference in co-occurrence patterns, instantiated in the linguistic conditioning of internal variation (see Poplack 2018: 26; Poplack & Meechan 1998: 132). Such language-specific co-occurrence patterns enable us to diagnose convergence and rule out cross-linguistic tendencies of internal change.

The methodological desiderata for evaluating contact effects on grammar are depicted in Figure 1— comparable speech corpora of contact and non-contact varieties and, through analysis of internal variation, identification of diagnostic differences between parallel variant forms across the languages. Diagnostic differences are found in the variable context and the probabilistic constraints, considering direction and relative strength of effect.

In what follows we will illustrate how this metric allows us to go beyond typological labels such as that of null vs. non-null subject language to identify quantitative differences in the linguistic conditioning of internal variation. But first, let us situate the data in their social context.

Figure 1   Convergence evaluation metric
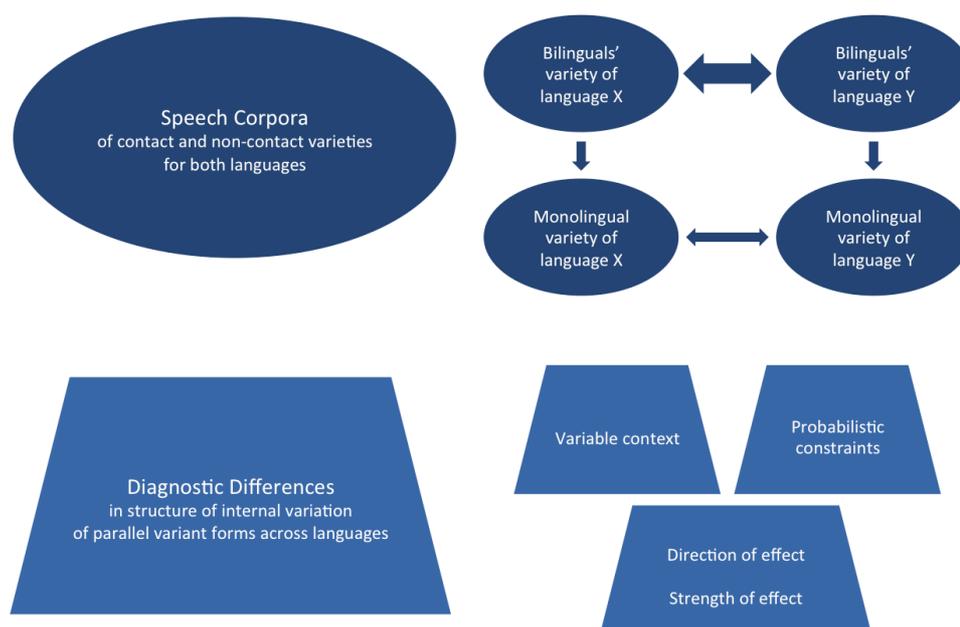
3   The bilingual speech corpus

Settled from the end of the 16th century, northern New Mexico is home to the oldest Spanish-speaking area in what is now the United States. At some 2,400 km from Mexico City, northern New Mexico was one of the most isolated areas of the Spanish-speaking world, first as a remote colony of New Spain, then on the periphery of newly independent Mexico in the early 19th century, and subsequently as part of the United States with the 1848 Treaty of Guadalupe Hidalgo (Bills & Vigil 1999: 43). Contact with English has been intense over the last 150 years: in 1850 the region was annexed, in 1878 the railroad arrived, and in 1912 New Mexico was admitted to the Union as the 47th state.

Today New Mexican Spanish is an endangered dialect, threatened by shift to English and stigmatization with respect to non-local varieties of Spanish. English was imposed in the educational system, displacing Spanish even in northern communities by the 1940s. The displacement of Spanish was not only in the official curriculum. Children were punished for speaking Spanish in school, as recounted in the excerpt in (3) (Anita, born 1941). At the turn of the 21st century, Spanish is back in the schools, but it is taught as a foreign or second language, to the detriment of the local variety (cf., Gonzales-Berry 2000). The excerpt in (4) underscores the stigmatization of bilinguals' language varieties, especially of the minority language. Here Leandro calls the local Spanish 'another kind of language sandwich'.

(3)
Anita:	*I went into second grade,*
	and uh=,
	.. I couldn't speak --
	... Spanish .. at all.
Lucy:	*oh=.*
Anita:	*... u=m,*
	... even if I had to use the bathroom I couldn't ask in Spanish.
	... a=nd,
	if you did,
	you got punished.

(NMSEB 14, 02:34-02:47)

(4)
Leandro:	*los mexicanitos esos,*	those Mexican guys,
	they --	they --
	.. hablaban muy diferente el mexicano que nosotros,	.. spoke Spanish very differently from us,'
Ricardo:	*Yeah.*	'Yeah.'
Leandro:	*y nosotros hablábanos otra class de language sandwich*	'and we spoke another kind of *language sandwich*
	<@ que le dicen @>,	*<@ as they say @>,*
	@@@	*@@@*
	de Nuevo México.	of New Mexico.'

(NMSEB 25, 05:13-05:23)

Yet Spanish is still spoken by two-thirds of the US-born Hispanic population in northern New Mexico counties (United States Census Bureau 2014). It is important for probing contact effects that, while New Mexico has the highest "Hispanic or Latino population" as a percent of the total (47%, nearly three times the national average of 17%, United States Census Bureau 2014), it also has the lowest proportion of first-generation immigrants. Only 16% of the New Mexico Hispanic population were born outside of the US, approximately half the proportion of foreign born in other states, such as Texas (30%) and New York (39%) (United States Census Bureau 2015) (Figure 2). In the northern New Mexico counties of Mora, Rio Arriba, San Miguel and Taos, just 4% to 8% are foreign born. These demographics make this setting especially valuable for discovering contact effects since, unlike the immigrant settings in many studies, here interaction with monolingual speakers of the minority language is limited at best.

Figure 2   Proportion of Hispanic population that is foreign born in three US states (United States Census Bureau 2015)
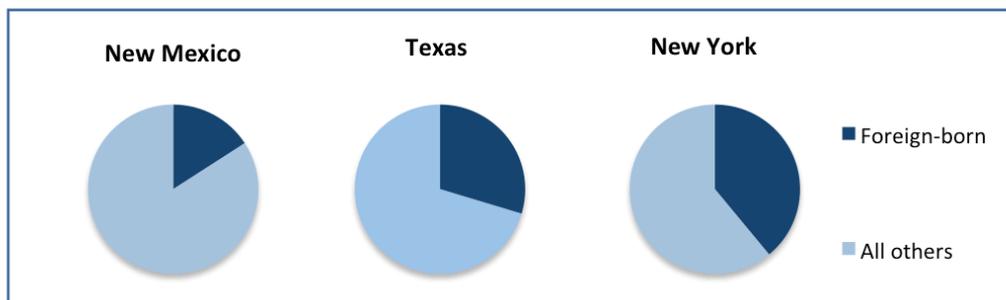
We take advantage of a bilingual corpus, the *New Mexico Spanish-English Bilingual* (NMSEB) corpus (recorded 2010-2011), that features spontaneously produced speech data, from a systematic sample of speakers, members of this speech community (cf., Torres Cacoullos & Travis 2018: Chs 2 & 3).

The bilinguals whose speech is captured in the corpus, and in the examples seen so far, are "the locus of the contact" (Weinreich 1953/1968: 1). Yet individual bilingualism constructs common in psycholinguistics, such as simultaneous vs. sequential or early vs. late bilingual, are not readily applicable to members of bilingual communities. Proficiency tests would also be unsuitable, as the testing situation is associated with the very educational institutions where the speakers and their local variety have been denigrated. Even self-reports of bilingual ability and questionnaire responses are easily misinterpreted, especially for speakers of non-standard varieties. This is because, as we can glean from Leandro's characterization of New Mexican Spanish above, self-assessments may be made under a normative lens.

Thus, the initial bilingualism criterion in selecting speakers for the NMSEB corpus was that they make regular use of both languages. A subsequent aggregate measure to validate degree of bilingualism is the distribution of clauses according language (based on the finite verb). For example, in (5), the first clause is English and the second is Spanish (in the translation to the right, roman font indicates strings originally produced in Spanish, italics represent English material). The distribution of clauses is virtually even (51% English, 49% Spanish, N = 36,000, Torres Cacoullos & Travis 2018: 67).

(5)

| Dolores: | *so I told them,* | so I told them, |
| | *que van a salir en el @Sun @.* | that they're going to come out in the @Sun @.' |

(NMSEB 22, 17:05-17:08)

The sociolinguistic interview records large amounts of speech least distorted by self-monitoring or hyper-correction to approximate speakers' vernacular (Labov 1984: 32-42). Rather than asking about particular linguistic structures, elicited are narratives of personal experience; here the authority is the speaker, not the university-affiliated interviewer. Since interviewer in-group status facilitates use of the vernacular (cf. Clyne, Eisikovits & Tollfree 2002; Poplack 1993), for the NMSEB corpus, a team of eight *Nuevomexicano* 'Hispanic New Mexican' undergraduate students were trained in the recording of sociolinguistic interviews, which they conducted with extended family members and family acquaintances. The corpus comprises 23 women and 17 men, born between 1923 and 1989. They include mineworkers, ranchers, and a variety of service employees, and most (29/40) live in rural areas.

A corpus like NMSEB can be profitably mined because it is transcribed in its entirety. It is the comprehensive transcription that allows for the counting of all clauses, the tagging of code-switching, and accountable analyses of subject expression and other structures (e.g., Aaron 2015; Balukas & Koops 2015; Benevento & Dietrich 2015; Brown 2015; Dumont & Wilson 2016; LaCasse 2018; Sankoff, Dion, Brandts, Alvo, Balasch & Adams 2015; Steuck 2018; Wilson & Dumont 2015).

An asset of the transcription for cross-language comparisons, as we will see, is that it is prosodically based. The speech stream is segmented into Intonation Units (IUs), produced with "a single, coherent intonation contour" (Du Bois, Schuetze-Coburn, Cumming & Paolino 1993: 47) (cf. Chafe 1994: 58-60). Punctuation represents "the degree of continuity that occurs at the transition point between one intonation unit and the next" (Du Bois et al. 1993: 53). In (5), the first IU projects something more, with continuing intonation indicated by the comma; the second suggests completion, with final intonation marked by the period.

A summary description of the comparison corpora in this study is given in Table 1. The bilingual nature of the NMSEB corpus means that we are able to compare the patterning of both languages, not only within the same community, but in the same speakers' speech. The monolingual benchmarks are two corpora of conversational Spanish and English, chosen because they are transcribed in IUs in the same manner as the NMSEB corpus. The analysis is circumscribed to 1sg and 3sg subjects with human specific referents (which make up the bulk of subject expression data in numerous studies, overshadowing second person and all plural subjects). A total of nearly 10,000 tokens of variable subject expression are analyzed, that is, verbs with pronominal and unexpressed subjects.

Table 1   Four datasets to probe contact-induced change: Bilingual corpus and benchmark corpora

	Bilingual corpus	Monolingual benchmarks	
Corpus name	Spanish-English New Mexico Spanish-English Bilingual corpus (NMSEB) (Torres Cacoullos & Travis 2018: Chs. 2&3)	Spanish Corpus of Conversational Colombian Spanish (CCCS) (Travis 2005: 9-25)	English Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois, Chafe, Myer, Thompson, Englebretson & Martey 2000-2005)
Speakers	40	37	207
Recordings	31	30	60
Hours transcribed	29	9	23
Words	300,000	96,000	249,000
Intonation Units	98,000	33,000	70,000
N tokens*	Spanish English		
Pronoun	1,353 196	1,143	658
Ø	4,218 98	1,659	329

*   Number of tokens of 1sg & 3sg pronominal and unexpressed (Ø) subjects within the variable context for subject expression; on the sampling of English pronouns, see 4.

Utilizing these comparable corpora, we are able to meet Weinreich's directive that "the differences and similarities between the languages in contact must be exhaustively stated" in order to assess the linguistic outcomes of contact (1953/1968: 2). The classification into null-subject and non-null subject language types implies patent differences in the frequency of pronominal subjects and consequent frequency shifts if the languages undergo convergence. As frequency shifts are an uncertain gauge of change, however, we apply the convergence evaluation metric to transform these typological labels into differences in the linguistic conditioning of pronominal vs. unexpressed subjects.

4   Diagnostic differences in the variable context

4.1   Variable context for English subject expression

In order to discover co-occurrence patterns, we first delimit a *variable context* as distinct from contexts of categorical use where a variant always or never occurs. The variable context here is the set of linguistic environments where speakers have the option of leaving a subject pronoun unexpressed. The universe of data is thus the sum total of *occurrences and non-occurrences* in a variable context as defined (Labov 2005: 7). Under this principle of accountability, "… any variable form (a member of a set of alternative ways of 'saying the same thing') should be reported with the proportion of cases in which the form did occur in the relevant environment, compared to the total number of cases in which it might have occurred" (Labov 1969: 738, n.720).

While much is known about variable subject expression in Spanish, we cannot meet Weinreich's directive unless we also delve into English (Torres Cacoullos & Travis To appear). Though rare, unexpressed subjects in English are used as alternatives of pronominal subjects. Both variants are illustrated in (6), with variation found at two loci. First, there is alternation in coordinate clauses, namely those joined to a coreferential-subject clause via a coordinating conjunction, most often *and*, as for example (10) ahead (or coordinating conjunction + ADV, e.g., *and then*, in (6), lines (g), (h) and (j)). Second, for non-coordinated verbs, thanks to the prosodically transcribed speech corpora, we can determine that the alternation occurs in prosodic-initial position only, as in (6), lines (b), (c) and (f).

(6)

a.	Rebecca:	*What did he do.*
b.	Rickie:	*Ø Just loo=ked=,*
c.		*he [had a] --*
d.	Rebecca:	* [Did he s]top walking?*
e.	Rickie:	*n=,*
f.		*Ø just kinda loo=ked,*
g.		*.. and then Ø loo=ked,*
h.		*and then .. Ø walked.*
i.	Rebecca:	*.. [Okay].*
j.	Rickie:	* [And] then he kept walking.*

(SBCSAE 08, 436-445)

The restriction of non-coordinated English unexpressed subjects to initial position is not a syntactic "root" phenomenon, as often claimed (e.g., Haegeman 2013), but is indeed prosodic. Prosodic-initial position means that the subject (or verb, with an unexpressed subject) occurs either in absolute IU-initial position, or is preceded at most by non-speech material, such as pausing, laughter, fillers *uh* or *um*. The subject of subordinate clauses and interrogatives is very rarely prosodically initial, as it tends to be preceded by a subordinating conjunction, relative pronoun, *wh*-word or a form of *do,* as in lines (a) and (d) in (6). So, it is not possible to tell whether the absence of unexpressed subjects in these contexts is due to clause type or to prosodic position. However, in main declarative clauses, while the subject most often occurs (approximately three-quarters of the time) in IU-initial position, as in lines (b), (c) and (f) in (6), it may also appear non-IU initially preceded by *well, so, of*

course, oh, illustrated in (7), for example. The deciding finding is that in these main-clause contexts, unexpressed subjects are again entirely absent.[2] Thus, it is not clause type that matters, but prosodic position.

(7)

Tom: *... Well I lost my family when I was about eleven,*
                  so I went to live with this other family,

(SBCSAE 32, 1570-1571)

In sum, coordinate clauses and prosodic-initial position constitute the English variable context for subject expression, which is thus quite restricted. In a sample of 500 pronouns (see note 2), fewer than one half occur where an unexpressed subject could have been used instead. In contrast, in Spanish, the great majority of both pronominal and unexpressed subjects are within the variable context, as eligible instances for a choice between variants.

4.2 Diagnostic difference of prosodic position

A first step, then, in identifying diagnostic differences is to determine whether the tendencies in Spanish are in the same direction as English for coordinate clauses, as well as for prosodic position even if not as a categorical restriction. To assess this, in Figure 3 we compare subject pronoun rates in coordinate clauses and, for non-coordinate clauses, in IU-initial and Non-IU-initial position, across the four datasets: monolingual Spanish and English, and bilinguals' Spanish and English. To model co-occurrence patterns within the restricted English variable context, after we extracted all tokens of English null subjects, we constructed a sample of the greatly outnumbering pronominal subjects by taking, for each null subject, the closest preceding and following eligible pronoun. This gave an artificial overall pronoun rate of 66.6%, so that the relevant comparison in Figure 3 below is that of relative heights of columns, not absolute values.

Beginning with a comparison of coordinate clauses, we can establish first that this constitutes a site of similarity across the two languages. In both monolingual Spanish and English (in the first and fourth sets of columns), there is variation between unexpressed and pronominal subjects under coordination, which, again in both, favors unexpressed subjects, seen in that the first column is the lowest within each set. As a cross-language similarity where we have a shared tendency, coordinate clauses are not readily informative as to contact-induced change, and, as expected, the tendency for the bilingual varieties (in the second and third sets of columns) is in the same direction (we return to clause linking in Section 5).

A sharp cross-language diagnostic difference lies in prosodic initial position. Focusing again first on the monolingual varieties, we now compare subject pronoun rates according to position in the IU: IU initial (e.g. *he had* and *Ø just kinda looked* in (6)) vs. Non-IU initial (e.g. *well I lost, so I went* in (7)) (i.e. the second and third column of each set). In English, in non-IU initial position, the pronoun rate is 100% because of the prosodic-initial restriction on unexpressed subjects. In Spanish, in non-IU initial position, not only is there variability but the pronoun rate is lower than in IU-initial position, the opposite direction of English. That is, in Spanish a pronominal

---

2   Based on a sample of 500 subject pronouns consisting of 10 tokens each from the 50 SBCSAE recordings that have unexpressed subjects; of these, 388 occurred in main clauses and 112 in subordinate clauses.

subject is more likely in IU initial position, as in (8), than in non-IU initial position, as in (9), where unexpressed subjects not only occur but are favored.

(8)
  Patricia: *él se fue para Estados Unidos,*   'he went to the United States,'
                                                                                            (CCCS 09, 387)

(9)
  Rocío: *.. Entonces Ø se fue.*   'So (he) went.'
                                                                                            (CCCS 20, 1450)

Figure 3   Cross-language similarities and diagnostic differences in the variable context:
             Subject pronoun rate in coordinate clauses and according to prosodic position, in
             four datasets
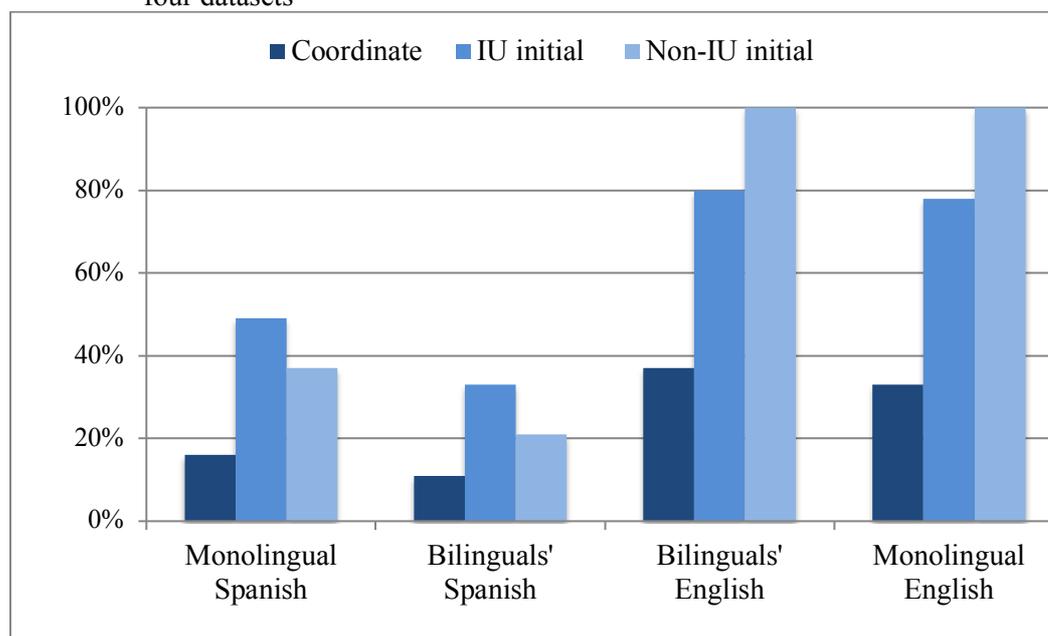
* As English is based on samples of two pronominal subjects extracted for each null subject, the valid comparison is that of relative heights of columns, not absolute values.[3]

4.3   Bilinguals' Spanish and English according to prosodic position

    Let us now consider bilinguals' Spanish and English. We first compare bilinguals' Spanish to monolingual Spanish patterns (seen in the first and second sets of columns), using the diagnostic difference of prosodic position to determine if there is change. Remember that pertinent are the relative heights of the columns, since there

---

[3]   Diagnostic <u>difference across languages</u>: in IU initial position the subject pronoun rate is 1.3 times as high that in non-IU initial position for Monolingual Spanish (49% vs. 37%) but .78 (78 vs. 100) for Monolingual English. The difference is mirrored in the bilingual data: 1.6 times as high for Bilinguals' Spanish (33 vs. 21) but .80 for Bilinguals' English (80 vs. 100). Compare the <u>similarity within languages</u>: in Spanish the rate in coordinate contexts is .33 as high that in IU initial contexts for both Monolingual (16 vs. 49) and Bilingual data (11 vs. 33); in English it is again virtually identical between Monolingual, at .42 (33 vs. 78), and Bilingual data, at .46 (37 vs. 80).

are differences in overall subject pronoun rate within monolingual Spanish which owe nothing to contact (see Section 2).[4] As reflected in the relative heights of the columns within each set, pronominal subjects are favored in IU-initial over non-IU-initial position, the opposite direction to what we just saw for English. Thus, New Mexico bilinguals' Spanish subject expression displays the same conditioning as monolingual Spanish and differs from monolingual English.

One further question remains: might the bilingual speakers' use of English have been influenced by the Spanish pattern, per a scenario of shift-induced "substratum interference" (Thomason & Kaufman 1988: 38ff)? The bilingual nature of the corpus allows us to make the comparison needed to address this question. The answer is seen in the juxtaposition of bilingual speakers' English with monolingual English patterns (the third and fourth sets of columns), revealing that the prosodic-initial restriction holds in both varieties. Thus, just as bilinguals' Spanish lines up with monolingual Spanish, their English lines up with monolingual English.

In summary, we have used a diagnostic difference in the variable context to evaluate whether bilinguals' varieties display more Spanish-like or more English-like grammatical behavior. Application of the convergence evaluation metric allows us to determine that, against the hypothesis of contact-induced structural change, the languages in contact—as spoken by the same bilinguals—are as grammatically distinct from each other as the monolingual benchmarks are.

5   Diagnostic differences in the probabilistic constraints

5.1   Accessibility as a cross-linguistic probabilistic constraint

Having located a diagnostic difference between the languages in their variable contexts, we now turn to probabilistic constraints, looking to compare direction and relative strength of effect.

While overall rates of pronominal subjects are disparate across null-subject languages (e.g., for 1sg, ranging from 21% in Polish to 49% in European Portuguese (N = 536, Chociej 2011: 52; N = 704, Posio 2013: 269), the probabilistic constraints are largely the same. Indeed, transcending typological labels, cross-linguistic effects of accessibility, priming, coordination or clause type, and person/number on subject expression have been reported for languages as diverse as Arabic and Mandarin, and even English (Jia & Bayley 2002; Leroux & Jarmasz 2005; Owens, Dodsworth & Kohn 2013).

We focus here on accessibility, the effect that has been most replicated but also the one that, as a discourse-pragmatic constraint, has been singled out as vulnerable to contact-induced change (e.g., Heine & Kuteva 2005: 70; Sorace 2004: 143). A cross-linguistic generalization is that less "coding material" (or less phonetic bulk—unexpressed subjects) corresponds to more accessible referents, and more coding material (pronominal subjects) to less accessible referents (e.g., Givón 1983). For subject expression, accessibility is usually conceived in terms of coreferentiality of the target subject with that of the preceding clause. The prosodic transcription allows

---

4   Here, the lower rates in all contexts in Bilinguals' Spanish compared with Monolingual Spanish happen to be driven by genre differences in data distribution (Travis & Torres Cacoullos 2018: Section 5.3).

for a refined measure of accessibility, by adding the structural criterion of syntactic and/or prosodic linking between clauses.

We look to the subject of the preceding clause for three degrees of accessibility, as illustrated in the following examples. In (10) the target subject in *Ø was going* is coreferential with that of the preceding clause, and it is linked, syntactically via the conjunction *and* as well as prosodically via continuing intonation (marked by the comma). (Clauses may also be linked by prosody alone, as in the last two lines in (2) above, and (c) in (6).) In (11), the target subject in *Ø went* is coreferential only, because the preceding IU by the same speaker has final intonation (marked by the period), and there is also a backchannel from the interlocutor intervening. In (12), the target subject in *Ø never had money* occurs in a non-coreferential context, because the preceding clause subject is not a coreferential 1sg but *it*. As we proceed from contexts of greatest to least accessibility, pronominal subjects are most disfavored in contexts of structural linking, less so in coreferential contexts in the absence of linking, and are most favored in non-coreferential contexts.

(10)
Alfredo: *...(0.6) so <u>he</u> ... got on the treadmill,*
          and <u>Ø</u> was going,

                                                                (NMSEB 31, 49:52-49:54)

(11)
Leandro: *he just left.*
Ricardo: *.. yeah.*
Leandro: *.. Ø went in the forest service and,*

                                                                (NMSEB 25, 59:20-59:22)

(12)
Pedro: *...(2.0) so it was pretty good.*
          .. it was a pretty good time.
          ...(1.3) Ø never had money,
          Ø never had to worry about money.

                                                                (NMSEB 10, 18:46-18:52)

5.2   Diagnostic difference of relative strength of accessibility

Figure 4 shows probability values for pronominal subjects in each of the three levels of accessibility, from four independent multivariate analyses, one for each dataset.[5] Direction of effect is seen in the order of probability values. The closer to 1, the more a pronominal subject is favored in the corresponding context (level); the closer to 0, the stronger the disfavoring effect. An indication of the relative strength of the effect is seen in the range of the highest and lowest probability values, as ranked vis-à-vis other predictors in the analysis (included here besides accessibility were priming, verb class, tense-aspect; see note on Figure 4).

---

[5] The probability values are from variable rule analysis using the Goldvarb Lion program (Sankoff, Tagliamonte & Smith 2015) and indicate the strength of constraints relative to the "input probability" (which in turn reflects the overall rate of subject pronoun expression in each dataset) (Sankoff 1988).

Beginning with the direction of effect, we see that it is as predicted in both Spanish and English, with the order of the three columns by height the same in each set: pronominal subjects are favored most in non-coreferential contexts and least in structurally linked coreferential contexts, which, conversely, are the most favorable to unexpressed subjects. This cross-language similarity is consistent with the generalization mentioned above that more phonetic material codes less accessible referents. As a cross-linguistic tendency it is unrevealing as to other-language influence.

But we can also pinpoint a difference in the accessibility constraint, in relative magnitude of effect. This is conveyed by the divergence between the tallest and shortest columns (the most and least favoring contexts) in each set in Figure 4, as compared with other predictors in each analysis. Accessibility is far more important for English than Spanish subject expression, being ranked higher than any other constraint within the English regression analyses, with a range of probability values that is approximately twice as large as the next largest. In the Spanish analyses, on the other hand, it is of the same or smaller size (for further discussion, see Torres Cacoullos & Travis 2018: Chapter 6 and 7).[6]

In this way, co-occurrence patterns within each variable context have brought to light a second pair of cross-language similarities and diagnostic differences. English shares with Spanish (and other languages) a graded accessibility effect on subject expression. But the relative strength of this effect is an undeniable quantitative difference between the languages, yielding a refined criterion for testing convergence that, like the delimitation of the variable context, goes beyond the uncertain criterion of overall rate differences.

---

[6] On quantifying relative strength for comparisons, see Grafmiller and Szmrecsanyi (To Appear).

Figure 4   Cross-language similarities and diagnostic differences in probabilistic constraints: Subject pronoun probability values according to degree of accessibility, in four datasets
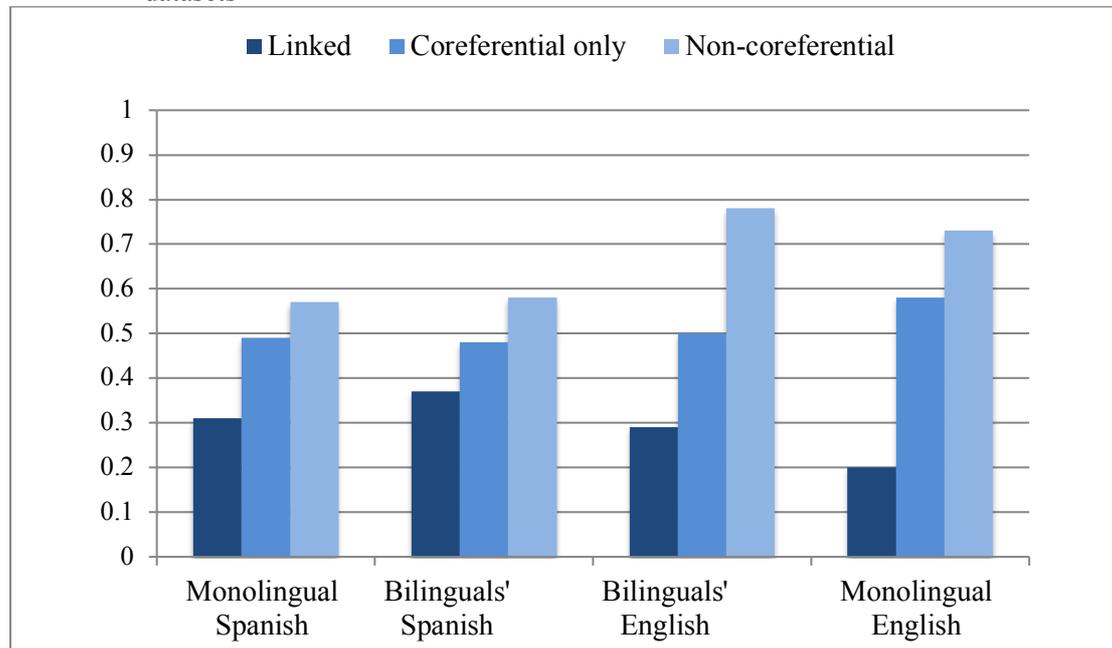
*   Also included in the logistic regression analyses are Priming, Verb class, Tense-aspect, Person (for the Spanish datasets), and Presence of English (for Bilinguals' Spanish). (For full results and corpus information, see Torres Cacoullos & Travis 2018: 165, 213-214.)

5.3   Bilinguals' Spanish and English according to strength of accessibility

What of bilinguals' Spanish and English? On the diagnostic difference of the importance of accessibility, bilinguals' Spanish patently aligns more with monolingual Spanish and their English more with monolingual English than do the bilingual varieties with each other—accessibility is no stronger in bilinguals' than in monolinguals' Spanish.

Bear in mind that when bilinguals are engaging in code-switching (alternating multiword strings), the previous mention of the target subject may be in the other language. For example, in (13), the previous mention of the target subject in *Ø venía* 'I used to come' is the subject in *I used to hoe,* so we have a Spanish subject with a coreferential and structurally linked English subject. Table 2 shows that the Spanish subject pronoun rate with a previous English subject is virtually identical to that with a previous Spanish subject, at each degree of accessibility. Accessibility impacts the choice between a Spanish pronominal or unexpressed subject no more strongly when the previous mention by the speaker is in English than when it is in Spanish.

(13)

Miguel: *... I used to hoe over there,*   *... I used to hoe over there,*
            y luego Ø venía y Ø escarbaba   and then (I) used to come and dig here.'
            aquí.

                                                     (NMSEB 04, 47:37-47:40)

Table 2   Spanish subject pronoun rate in bilinguals' Spanish according to degree of accessibility and the language of the previous mention*

	Linked	Coreferential only	Non-coreferential
Language of previous mention			
English	10% (13/127)	18% (22/125)	25% (74/295)
Spanish	9% (96/1077)	22% (141/628)	25% (269/1,059)

* N=3,305, previous mention only as either a pronominal or unexpressed variant within the preceding 10 clauses by same speaker (previous English Ø, N = 6). No difference between languages in pairwise comparisons of degrees accessibility is significant (Fisher's exact test).

To summarize this section, by the criterion of the strength of accessibility effects on subject expression, bilinguals' Spanish does not evidence even subtle adoption of grammatical properties of English but nor does bilinguals' English evidence properties of Spanish. The languages in contact—as spoken by the same speakers—are grammatically distinct.

6   Conclusion

Languages come into close contact through the bilinguals who use them. The language contact can be no closer than when the same speakers use the languages within the same discourse event, as is captured in a bilingual corpus (seen here in the speech of Manuel in (1) and (2), among other examples). The convergence hypothesis assumes that such language use results in the two languages somehow becoming mixed or blended.

To demonstrate convergence—magnified structural similarity—between languages, we must identify differences by which to compare varieties in contact with varieties not in contact. While a structure categorically present in one language and absent in another would be a qualitative difference, an "ungrammatical" structure in one language that is disfavored in the other would instead represent grammatical similarity, or a shared tendency (Givón 1979: 22-43) (cf. Bresnan et al. 2001: 29). Thus, we appeal to comparison of the structure of internal variation as an evaluation metric for convergence. By relying on diagnostic differences—dissimilarities in the internal variation patterns of forms that are parallel across languages—we can determine whether change has occurred, and if so, if it is in the direction of the contact language. For Spanish and English subject expression, as well as several shared cross-linguistic tendencies, there are also differences. For example, the qualitative (categorical) prosodic-initial position restriction of the English variable context is not even a quantitative tendency in Spanish, thus supplying a diagnostic difference to test contact induced change. A second cross-language difference lies in the greater importance of accessibility in English, in which it has a far stronger effect relative to other probabilistic constraints than it does in Spanish. The diagnostic differences become the basis for comparisons across appropriate speech corpora.

The *coup de grâce* for grammatical convergence is delivered here by juxtaposing bilinguals' English and Spanish with comparable benchmarks. On the criterion of the diagnostic differences identified, there is no hint of change in either their Spanish or their English, which remain no less different from each other than their monolingual counterparts are. Applications of the convergence evaluation metric

have arrived independently at continuity rather than change for structures as diverse as subject expression, word order, and mood choice in different language pairs (e.g., Carvalho & Bessett 2015; LaCasse 2018; Nagy et al. 2018; Poplack & Levey 2010: 404-406). Morphosyntactic change is thus not an inexorable outcome of contact. In bilingual communities, speakers can independently bring to bear language-specific grammatical principles, as they alternate between languages.

Acknowledgments

Appendix: Transcription Conventions (Du Bois et al. 1993)

Examples are reproduced verbatim from the transcripts, with the exception of the removal of in-breaths (H), for readability purposes.

Carriage return	new Intonation Unit	..	short pause (0.5 secs)
.	final intonation contour	...	medium pause (0.5-0.7 secs)
,	continuing intonation contour	...(N)	timed pause (> 0.7 secs)
?	appeal intonation contour	=	lengthened syllable
--	truncated intonation contour	@	one syllable of laughter
[ ]	overlapped speech	<@ @>	speech produced while laughing

References

Aaron, Jessi Elana. 2015. Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, 19(4), 459-480.

Aikhenvald, Alexandra Y. 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.

Bailey, Guy, & Tillery, Jan. 2004. Some sources of divergent data in sociolinguistics. In: Carmen Fought (ed.), *Sociolinguistic variation: Critical reflections* (pp. 11-30). Oxford: Oxford University Press.

Balukas, Colleen, & Koops, Christian. 2015. Spanish-English bilingual VOT in spontaneous code-switching. *International Journal of Bilingualism*, 19(4), 423-443.

Benevento, Nicole, & Dietrich, Amelia. 2015. I think, therefore *digo yo*: Variable position of the 1sg subject pronoun in New Mexican Spanish-English code-switching. *International Journal of Bilingualism*, 19(4), 407-422.

Bills, Garland D., & Vigil, Neddy A. 1999. Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology*, 53(1), 43-66.

Bloomfield, Leonard. 1933. *Language*. Chicago/London: University of Chicago Press.

Bresnan, Joan, Dingare, Shipra, & Manning, Christopher D. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: Miriam Butt & Tracy Hollaway (eds.), *Proceedings of the LFG01 Conference* (pp. 13-31). Stanford: CSLI Publications.

Brown, Esther L. 2015. The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, 19(4), 387-406.

Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico, and Madrid, Spain. *Language Variation and Change*, 5(3), 305-334.

Carvalho, Ana M., & Bessett, Ryan M. 2015. Subject Pronoun Expression in Spanish in Contact with Portuguese. In: Ana M. Carvalho, Rafael Orozco, & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 143-166). Georgetown: Georgetown University Press.

Carvalho, Ana M., Orozco, Rafael, & Shin, Naomi Lapidus. 2015. Introduction. In: Ana M. Carvalho, Rafael Orozco, & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. xiii-xxvi). Georgetown: Georgetown University Press.

Cedergren, Henrietta, & Sankoff, David. 1974. Variable rules: Performance as a statistical reflection of competence. *Language*, 50, 333-355.

Chafe, Wallace. 1994. *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chociej, Joanna. 2011. Polish null subjects: English influence on heritage Polish in Toronto. Department of Linguistics, University of Toronto. Ms.

Clyne, Michael, Eisikovits, Edina, & Tollfree, Laura. 2002. Ethnolects as in-group varieties. In: Anna Duszak (ed.), *Us and others: Social identities across languages, discourses and cultures* (pp. 133-157). Philadelphia, PA: John Benjamins.

Doğruöz, A. Seza, & Backus, Ad. 2007. Postverbal elements in immigrant Turkish: Evidence of change? *International Journal of Bilingualism*, 11(2), 185-220.

Du Bois, John W., Chafe, Wallace L., Myer, Charles, Thompson, Sandra A., Englebretson, Robert, & Martey, Nii. 2000-2005. Santa Barbara corpus of spoken American English, Parts 1-4. Philadelphia: Linguistic Data Consortium.

Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, & Paolino, Danae. 1993. Outline of discourse transcription. In: Jane Edwards & Martin Lampert (eds.), *Talking data: Transcription and coding in discourse* (pp. 45-89). Hillsdale: Lawrence Erlbaum Associates.

Dumont, Jenny, & Wilson, Damián Vergara. 2016. Using the variationist comparative method to examine the role of language contact in synthetic and periphrastic verbs in Spanish. *Spanish in Context*, 13(3), 394–419.

Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto Miguel de Cervantes.

Givón, T. 1979. *On understanding grammar*. New York: Academic Press.

Givón, T. 1983. Topic continuity in discourse: An introduction. In: T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-linguistic study* (pp. 1-41). Amsterdam: John Benjamins.

Gonzales-Berry, Erlinda. 2000. Which language will our children speak? The Spanish language and public education policy in New Mexico, 1890-1930. In: Erlinda

Gonzales Berry & David R. Maciel (eds.), *The contested homeland: A Chicano history of New Mexico* (pp. 169-189). Albuquerque: University of New Mexico Press.

Grafmiller, Jason, & Szmrecsanyi, Benedikt. To Appear. Mapping out particle placement in Englishes around the world: A study in comparative sociolinguistic analysis. *Language Variation and Change*, 30.

Haegeman, Liliane. 2013. The syntax of registers: Diary subject omission and the privilege of the root. *Lingua*, 130, 88-110.

Heine, Bernd, & Kuteva, Tania. 2005. *Language contact and grammatical change*. Cambridge: Cambridge University Press.

Hernández, José Esteban. 2009. Measuring rates of word-final nasal velarization: The effect of dialect contact on in-group and out-group exchanges. *Journal of Sociolinguistics*, 13(5), 583-612.

Hopper, Paul J., & Traugott, Elizabeth Closs. 2003. *Grammaticalization*, 2nd ed. Cambridge: Cambridge University Press.

Jia, Li, & Bayley, Robert. 2002. Null pronoun variation in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics*, 8(3), 103-116.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language*, 45(4), 715-762.

Labov, William. 1972. *Sociolinguistic patterns*. Oxford: Basil Blackwell.

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

Labov, William. 2005. Quantitative reasoning in linguistics. In: Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier, & Peter Trudgill (eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society* (pp. 6-22). vol. 1. Berlin: Mouton de Gruyter.

LaCasse, Dora. 2018. The subjunctive in New Mexican Spanish: Maintenance in the face of language contact. Doctoral dissertation, Pennsylvania State University.

Lastra, Yolanda, & Butragueño, Pedro Martín. 2015. Subject pronoun expression in oral Mexican Spanish. In: Ana M. Carvalho, Rafael Orozco, & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 39-57). Georgetown: Georgetown University Press.

Leroux, Martine, & Jarmasz, Lidia-Gabriela. 2005. A study about nothing: Null subjects as a diagnostic of the convergence between English and French. *University of Pennsylvania Working Papers in Linguistics*, 12(2), 1-14.

Nagy, Naomi, Iannozzi, & Heap, David. 2018. Faetar null subjects: a variationist study of a heritage language in contact. *International Journal of the Sociology of Language* 249:31-47.

Otheguy, Ricardo, & Zentella, Ana Cecilia. 2012. *Spanish in New York: Language contact, dialect levelling, and structural continuity*. Oxford: Oxford University Press.

Owens, Jonathan, Dodsworth, Robin, & Kohn, Mary. 2013. Subject expression and discourse embeddedness in Emirati Arabic. *Language Variation and Change*, 25(2), 255-285.

Poplack, Shana. 1993. Variation theory and language contact: Concepts, methods and data. In: Dennis R. Preston (ed.), *American dialect research* (pp. 251-286). Amsterdam: John Benjamins.

Poplack, Shana. 2018. *Borrowing: Loanwords in the speech community and in the grammar* Oxford: Oxford University Press.

Poplack, Shana, & Levey, Stephen. 2010. Contact-induced grammatical change: A cautionary tale. In: Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: An international handbook of linguistic variation* (pp. 391-419). vol. 1: Theories and methods. Berlin: Mouton de Gruyter.

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

Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica*, 47(1), 253-291.

Sankoff, David. 1988. Variable rules. In: Ulrich Ammon, Norbert Dittmar, & Klaus J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society* (pp. 984-997). vol. 2. Berlin: Walter de Gruyter.

Sankoff, David, Dion, Nathalie, Brandts, Alex, Alvo, Mayer, Balasch, Sonia, & Adams, Jackie. 2015. Comparing variables in different corpora with context-based model-free variant probabilities. In: Rena Torres Cacoullos, Nathalie Dion, & André Lapierre (eds.), *Linguistic variation: Confronting fact and theory* (pp. 335-346). New York: Routledge.

Sankoff, David, Tagliamonte, Sali, & Smith, Eric. 2015. Goldvarb Yosemite: A variable rule application for Macintosh. University of Toronto. URL http://individual.utoronto.ca/tagliamonte/goldvarb.htm.

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

Silva-Corvalán, Carmen, & Enrique-Arias, Andrés. 2017. *Sociolingüística y pragmática del español*, 2nd ed. Washington, DC: Georgetown University Press.

Sorace, Antonella. 2004. Native language attrition and developmental instability at the syntax–discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(2), 143-145.

Sorace, Antonella, & Serratrice, Ludovica. 2009. Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), 195-210.

Steuck, Jonathan. 2018. The prosodic-syntactic structure of intra-sentential multi-word code-switching in the New Mexico Spanish-English bilingual community. Doctoral dissertation, Penn State University.

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

Torres Cacoullos, Rena, & Travis, Catherine E. 2018. *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.

Torres Cacoullos, Rena, & Travis, Catherine E. To appear. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics*.

Travis, Catherine E. 2005. *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin / New York: Mouton de Gruyter.

Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19(2), 101-135.

Travis, Catherine E., & Torres Cacoullos, Rena. 2018. Discovering structure: Person and accessibility. In: Naomi Lapidus Shin & Daniel Erker (eds.), *Questioning theoretical primitives in linguistic inquiry (Papers in honor of Ricardo Otheguy)* (pp. 67-90). Amsterdam/Philadelphia: John Benjamins.

United States Census Bureau. 2014. 2010-2014 5-Year American Community Survey, http://factfinder2.census.gov, *Last accessed: 25 March 2016*.

United States Census Bureau. 2015. 2011-2015 5-Year American Community Survey, http://factfinder2.census.gov, *Last accessed: 13 July 2017*.

Weinreich, Uriel. 1953/1968. *Languages in contact: Findings and problems*. The Hague: Mouton.

Weinreich, Uriel, Labov, William, & Herzog, Marvin I. 1968. Empirical foundations for a theory of language change. In: Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics: A symposium* (pp. 95-188). Austin, TX: University of Texas Press.

Wilson, Damián Vergara, & Dumont, Jenny. 2015. The emergent grammar of bilinguals: The Spanish verb *hacer* 'do' with a bare English infinitive. *International Journal of Bilingualism*, 19(4), 444-458.