Foundations for the study of subject pronoun expression in Spanish in contact with English: Assessing inter-linguistic (dis)similarity via intra-linguistic variability

Rena Torres Cacoullos, Pennsylvania State University
Catherine E. Travis, Australian National University[*]

1. Introduction

Subject expression has been considered a paradigmatic case for grammatical convergence in studies of US Spanish, as bilinguals are thought to associate Spanish and English subject pronouns. The overwhelming preference for expressed subjects in English is thus predicted to boost the rate of expressed subjects in contact-Spanish varieties. But is Spanish subject expression an appropriate linguistic variable to ascertain convergence? A "prerequisite" to analyzing contact-induced change, as Weinreich (1968, 2) stressed, is that "the differences and similarities between the languages in contact […] be exhaustively stated". In this chapter, we offer such a statement of inter-linguistic grammatical comparability, by probing the conditioning of Spanish and English first person singular (1sg) subject expression.

A statement of Spanish-English (dis)similarity must distinguish cross-linguistically valid patterns from particular ones pertaining to the language pair. Subject expression patterns are said to be shaped cross-linguistically by cognitive factors, such as discourse cohesion and activation or accessibility (Ariel 1988, 79, Givón 1983, Levinson 1987, 384), alongside more interactional or pragmatic factors, related to the kind of action the utterance performs (Fox 1987, Oh 2005, Ono and Thompson 2003, inter alia). Quantitative analyses have confirmed accessibility effects (e.g. Nagy, Aghdasi, Denis and Motut 2011, Paredes Silva 1993), and have also revealed the mechanical factor of priming, or the perseveration of the same syntactic structure across utterances (Cameron and Flores-Ferrán 2003, 50-54, Travis 2007, 120-121), as well as a role for frequent lexically particular constructions (Travis and Torres Cacoullos 2012, 738-741; cf. Erker and Guy 2012) .

We focus this first comparison of Spanish and English subject expression on 1sg because it has yet to be empirically shown that subject pronouns constitute a single category; third person, for example, has different information status from first and second person, and first person, as a reference to the speaker, has also had attributed to it particular pragmatic functions.

We begin by motivating the value of intra-linguistic variation patterns for establishing inter-linguistic grammatical (dis)similarity. We then review the results for Spanish *yo* expression, specifically, the effect of cognition verbs (which we show is a construction effect); of subject continuity (which we reconfigure as Human Switched Reference, an accessibility effect); and of priming (which modulates subject continuity). And then we consider these same effects for English *I* expression, identifying similarities and differences which provide a basis for testing for convergence in contact varieties.

2. Linguistic conditioning as a characterization of the grammar of subject expression

2.1 The uninformativeness of overall rate comparisons

Figure 1 shows rates of 1sg subject expression in Spanish, across a range of studies, and in English. This figure confirms that, at over 95%, the rate of English *I* (vs. an unexpressed 1sg subject), in the right-most bar, is approximately two to four times greater than the rate of Spanish *yo*, in the bars to the left. The near-categorical rate of subject expression in English has been the basis of the prediction that US Spanish varieties will evince higher rates of subject expression than non-contact varieties. Yet we observe here that reported rates of Spanish 1sg subject expression also have a wide range, from about 25% to 50%. A difference in rate of use between contact and non-contact varieties of Spanish alone, then, is insufficient to claim influence from English, as the threshold for qualifying as a 'high(er)' rate remains unknown: statistically-significant rate differences need not translate into grammatically-relevant differences. Subject-expression frequency differences in and of themselves do not permit an inference of change in progress (as Cameron 1994, 24, n. 6 cautions).

Figure 1: Comparison of overall rates of occurrence of 1sg subject pronoun: Spanish *yo* and English *I*.
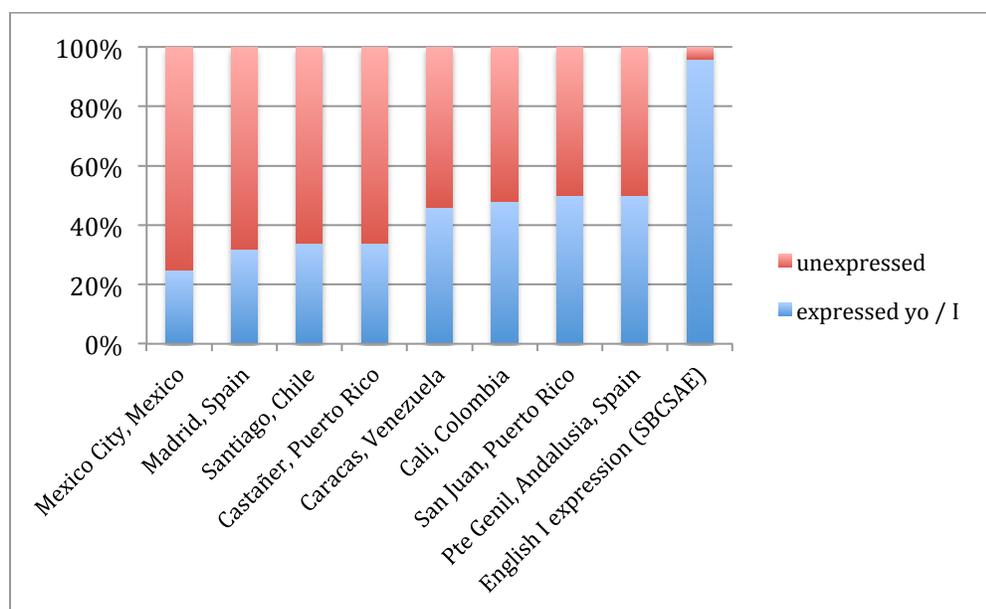
Sources: Mexico City (N=234) Lastra & Butragueño, this volume; Madrid, Spain (N=10,185) (Enríquez 1984) and Santiago, Chile (N=2,238) (Cifuentes 1980-1), as reported in Silva Corvalán (1994, 153); Castañer, Puerto Rico (N=1,527) (Holmquist 2012, 211); Caracas, Venezuela (N=721) (Bentivoglio 1987, 36); Cali, Colombia (N=878) (Travis 2007, 113); San Juan, Puerto Rico (N=949) (Cameron 1992, 233); Puente Genil, Andalusia, Spain (N=307) (Ranson 1991, 135, 138); English (Torres Cacoullos and Travis 2014b, 22).

Indeed, despite differences in overall rate, variable subject expression in Spanish displays widely-replicated uniformity in *linguistic conditioning*, both across dialects (e.g., Cameron 1993; 1994) and genres (Travis 2007). Numerous reports on Spanish subject expression, including those in this volume, have arrived at homologous multivariate models of contextual effects on speakers' use of expressed vs. unexpressed subject pronouns in both contact and non-contact

varieties (see Silva-Corvalán (2001, 154-169) for a review and Otheguy and Zentella (2012) for a recent detailed treatment). By the criterion of linguistic conditioning, then, the grammar is the same across Spanish varieties.

What has yet to be discovered, however, is whether Spanish grammar is similar to the grammar of English in regard to subject expression. That is, it is unknown whether, in the rare instances when *I* is left unexpressed in English, the conditioning factors coincide with, or differ from, those that are operative in Spanish. It is only once this has been established that a true test of convergence can be conducted, and it can be determined whether the contact variety evinces English-specific properties, with (non)expression rates that are higher (or lower) in the same *linguistic subcontexts* as they are in English.

2.2 Variationist comparative method for establishing grammatical (dis)similarity

A thorough empirical statement of grammatical (dis)similarity is obtained from comparisons of the linguistic conditioning of variant choice (on the variationist comparative method, see Poplack and Levey 2010, Poplack and Meechan 1998, Torres Cacoullos and Aaron 2003, Torres Cacoullos and Travis 2011). Our premise is the following:

> In discourse, inter-linguistic grammatical similarity is assessed via intra-linguistic variability, by comparing the effects of contextual factors contributing to the selection of a given variant: parallel favoring effects of factors operationalizing putative shared constraints or functions indicate grammatical similarity.

Table 1 shows two independent Variable-rule analyses (Sankoff, Tagliamonte and Smith 2012) of factors contributing to choice of 1sg subject expression: Spanish expressed *yo* (vs. unexpressed) drawn from the Corpus of Conversational Colombian Spanish (CCCS) (Travis 2005); and English expressed *I* (vs. unexpressed) drawn from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois, Chafe, Myer, Thompson, Englebretson and Martey 2000-2005). Variable-rule analysis uses logistic regression to perform binomial multivariate analysis for a choice between two variants. The results for Spanish (in the left-hand pair of columns) and English (in the right-hand pair of columns) provide Probabilities (factor weights), with values closer to 1 indicating favoring effects on expressed *yo* and expressed *I*. A factor (level of a predictor variable) *favors* a variant when its frequency relative to its alternative is higher in the presence of the contextual feature represented by the factor, which itself operationalizes a hypothesis about speaker choices. We first review the linguistic conditioning shown here for Spanish (section 3), and then go on to consider that for English (section 4).

Table 1: Two independent Variable-rule analyses of factors contributing to the selection of 1sg subject pronoun: *yo* (vs. unexpressed) in conversational Colombian Spanish (CCCS) and *I* (vs. unexpressed) in conversational American English (SBCSAE)
(non-significant factor group within [ ])

	Yo		*I* (IU-initial, non-coordinated)	
N	1,020		262	
Corrected mean (Input)[1]	.48		.29	
	Prob	**N**	**Prob**	**N**
Verb class				
Cognition	.66	134/201	[.52]	53/71
Other	.46	363/819	[.49]	133/191
Priming: Realization of previous				
coreferential 1sg subject*				
Expressed *yo / I*	.58	209/372	.53	103/143
Unexpressed	.41	112/330	.35	17/32
Subject continuity:				
Human Switched Reference				
Intervening hum subj present	.56	283/519	.60	87/111
Intervening hum subj absent	.42	144/365	.42	90/137
Tense				
Imperfect**	.62	71/116	N.A.	
Present	.48	287/573	.52	71/99
Preterit / Past Tense	.48	86/186	.48	91/139

Non-significant factor groups in both analyses: Polarity, Turn Position; also included for *yo*-expression: Realization of subject of immediately preceding clause (pronouns favor).

* Realization of previous coreferential 1sg subject up to two intervening human subjects. The bulk of these are at 0 or 1 intervening human subjects (86%, 602/702 for *yo*, 98%, 171/175 for *I*).
** Includes other syncretic forms, e.g., Pluperfect, Present Subjunctive (N Imperfect=83).

3. Operationalizing (cross-linguistic) constraints or functions

The Table 1 results for Spanish correspond with those that are reported in this volume and numerous prior studies of Spanish subject expression (cf. Torres Cacoullos and Travis To appear). *Yo* is favored with cognition verbs, when the previous coreferential 1sg subject is also realized as *yo*, in (human) switched reference contexts, and with imperfective tense-aspect. Here we present a reconsideration of these amply reported contextual factors.

---

1 The corrected mean (input) indicates the overall likelihood that the variant (expressed *yo*, *I*) will occur in the variable context as defined. Comparison of input values here is not meaningful, as for *I* expression only a sample of the expressed variant was taken (see §4.1 below).

3.1 Cognition verbs: Identifying particular constructions

For Spanish 1sg subject expression, an observation beginning with, as far as we know, Bentivoglio (1987, 50-54), is that *yo* is favored by cognition (or mental activity) verbs—most frequently here *creer* 'think/believe', *saber* 'know' and *pensar* 'think'. This effect is also found in the present data, as seen in Table 1. Moreover, these verbs display not only a higher rate of expression, but also distinct linguistic conditioning, as revealed by independent multivariate analyses of cognition vs. other verbs (Travis and Torres Cacoullos 2012, 734-737). Only cognition verbs are subject to a Turn-position effect (cf. Bentivoglio 1987, 38-40): *yo* is favored in turn-initial position with cognition verbs, but no such turn taking effect is found with other verbs (as in line 3 in (1)).[2] We take this divergence in subject expression patterns as evidence for a class of cognition verbs, defined by the verb slot in a (*yo*) + COGNITION VERB$_{1SG}$ construction.

(1)   (*yo*) + COGNITION VERB$_{1SG}$

1.	Santi:	*... Están rescatando los valores [familiares],*	'they are going back to [family values].'
2.	Ángela:	*[Hm],*	'[Hm],
3.		***Yo creo*** *que es que por ahí.*	**I think** that that's where.
		Es que hay que empezar.	There's where you've got to start.'

(CCCS, Almuerzo: 1736-1739)[3]

There is growing recognition that particular expressions mold the structure of variation (e.g., Poplack, Lealess and Dion 2013, Torres Cacoullos and Walker 2009). In usage-based models (e.g., Bybee 2010, Goldberg 2006), speakers have available a number of stored constructions of differing levels of schematicity. Particular constructions may be identified in a replicable way by token and relative frequency, by unithood indices and, once again, by the linguistic conditioning of variants (Bybee and Torres Cacoullos 2009, Torres Cacoullos and Walker 2011).

Yo creo 'I think' is a strong candidate for a lexically particular construction within the cognition verb class, as evidenced in (1) the high token frequency of the *yo creo* string; (2) the high proportion it comprises of all occurrences both of the lexical type *creer* and of the pronoun *yo*; and (3), in independent multivariate analysis of (*yo*) *creo*, the absence of a subject continuity effect, with a tendency toward a higher *yo* rate in Same Reference contexts, i.e. contrary to the predicted direction of effect (Travis and Torres Cacoullos 2012, 741). At the same time, the

---

2 The Turn-position effect is independent from Subject continuity; while most turn-initial tokens occur in non-coreferential contexts, within such contexts the *yo* rate remains higher with turn-initial than non-turn-initial cognition verbs. Turn-initial are tokens anywhere in the first prosodic unit (IU) of a speaker's turn (or the second IU following minimal responses such as *hm* with continuing intonation, represented with a comma, as in (1)). Cognition verbs are more likely to occur in absolute IU-initial position than 'other' verbs ((90%, 177/197 vs. 65%, 466/716 respectively), including in turn-initial IUs (94%, 65/69, vs. 74%, 160/216). The difference in *yo* rate is maintained in absolute (IU-initial) turn-initial position: 85%, 55/65 for cognition vs. 45%, 72/160 for 'other'.

3 Examples are reproduced verbatim from the transcripts (see Appendix for transcription conventions), aside from the bolding and underlining, and the addition of zeros in the original and parentheses around *I* in the corresponding translation to ease identification of unexpressed subjects. Codes in parentheses following examples refer to corpus, transcript name, and line numbers.

favoring effect of coreferential subject priming observed for other verbs remains significant for cognition verbs including *yo creo*. Particular constructions thus both contribute to and deviate from more general variation patterns and clearly merit analysts' attention.

3.2 A new subject continuity measure: Human Switched Reference

Widely replicated is the effect of 'switch reference', defined by Silva-Corvalán (1982, 104) as the context in which "the subject referent of the preceding finite verb […] is different from the referent of the […] subject" of the target finite verb. Such an effect is consistent with the notion of accessibility, whereby cross-linguistically more coding material (here, an expressed subject) is said to correspond to contexts of lesser accessibility—such as when the referent has not been recently mentioned—and less coding material (here, an unexpressed subject) to correspond to contexts of greater accessibility (Givón 1983, 18).

As a refinement of the notion of Switch Reference, Travis and Torres Cacoullos (2012, 726-729) proposed 'Human Switched Reference', which considers the presence of subjects with specific human referents intervening between coreferential subject mentions.[4] The two measures largely overlap; in the CCCS database most 1sg tokens of Switch Reference are also Human Switched Reference (87%, 519/595). The difference is illustrated in the following example; while *te digo* '(I) will tell you' (line 3) occurs in what is traditionally considered a coreferential (Same Reference) context, *lo miro* '(I) will look at it' (line 5) and *te aviso* '(I) will let you know' (line 8) would both be considered Switch Reference. But in both cases the subjects intervening between coreferential mentions are inanimate (and in a subordinate clause) and hence, by the measure of Human Switched Reference, do not count as disrupting subject continuity.

(2)
1.	Santi:	*Yo me averiguo,*	'**I**'ll find out,
2.		*y,*	and,
3.		*Y Ø te digo si hay algún apartamento,*	And (**I**)'ll tell you if there's an apartment,
4.		*Entonces,*	So,
5.		*Ø lo miro,*	(**I**)'ll look at it,
6.		*.. si tiene alcoba del servicio,*	.. if it has a service room,
7.		*.. entonces te --*	.. so --
8.		*.. Ø Te aviso.*	.. (**I**)'ll let you know.'

(CCCS, Pizza: 1332-1339)

Human Switched Reference provides a more discerning account of variable *yo* expression than does Switch Reference, as shown in Table 2: when there are no intervening human subjects between coreferential mentions (the top row), the difference between Same and Switch Reference contexts vanishes (with rates of expression of 39% and 41% respectively). On the other hand, within Switch Reference contexts (the rightmost column), there is a significantly higher rate of *yo* expression in contexts of Human Switched Reference (55% vs. 41%, *p* = .03, by Fisher's exact test).

---

[4] We develop the measure of intervening human subjects for variable 1sg expression, building on Givon's (1983, 14) Potential Referential Interference measure, which counts semantically compatible referents.

Table 2: Rate of expressed yo according to Switch Reference and Human Switched Reference (N=884) (Adapted from Travis and Torres Cacoullos 2012, 738, Table 3)

	Same Reference (0 interven. clauses)	Switch Reference (1+ interven. clauses)
Human Switched Reference		
Intervening human subject absent (0)	39% (113/289)	41% (31/76)
Intervening human subject present (1+)	No cases	55% (283/519)

3.3 Coreferential subject priming: Interaction with subject continuity

Near-ubiquitous in language variation are structural priming or perseveration effects, reported in the sociolinguistic literature as early as Poplack (1980) and Wiener and Labov (1983). *Coreferential subject priming* is seen in the tendency to repeat the form of the previous coreferential 1sg subject: a previous *yo* favors a subsequent *yo*, as in (3), and a previous unexpressed mention favors a subsequent unexpressed mention, as illustrated in lines 5 and 8 of (2) above. Coreferential subject priming is of specifiable duration, operating in the CCCS for previous mentions at distances of two or fewer intervening human subjects (Travis and Torres Cacoullos 2012, 730).

(3)
|			
Angela:	*yo ahorita no estoy trabajando.*		'Right now **I**'m not working.
	... Entonce=s,		… So,
	Es de ahorros.		it's from savings.
	.. de unos ahorritos que yo tengo,		.. from some savings that **I** have,'

<div align="right">(CCCS, Insurance: 687-690)</div>

       Priming and subject continuity interact, having non-independent effects (even though the factors themselves are distributionally independent) (on interaction, see Sankoff 1988, 986). Figure 2 depicts the rate of subject expression in contexts of Same Reference vs. Human Switched Reference, according to the realization of the previous coreferential 1sg subject.

       First, the strength of the coreferential subject priming effect is moderated by subject continuity: priming is greater in Same Reference contexts (no intervening human subjects), as also observed by Cameron and Flores-Ferrán (2003, 49). This is seen in the larger gap between the lines in Same Reference (to the left) as opposed to Human Switched Reference contexts (to the right): in the former, the rate of *yo* expression is twice as high with a previous *yo* than with a

previous unexpressed mention (52%, N=192 vs. 25%, N=173), whereas in the latter the difference between 'Previous *yo*' and 'Previous Ø' narrows (61%, N=266 vs. 48%, N=252).[5]

Figure 2: Rate of expressed *yo* in Human Switched Reference vs. Same Reference (presence vs. absence of intervening human subjects) (N=883), according to the realization of the previous coreferential 1sg subject
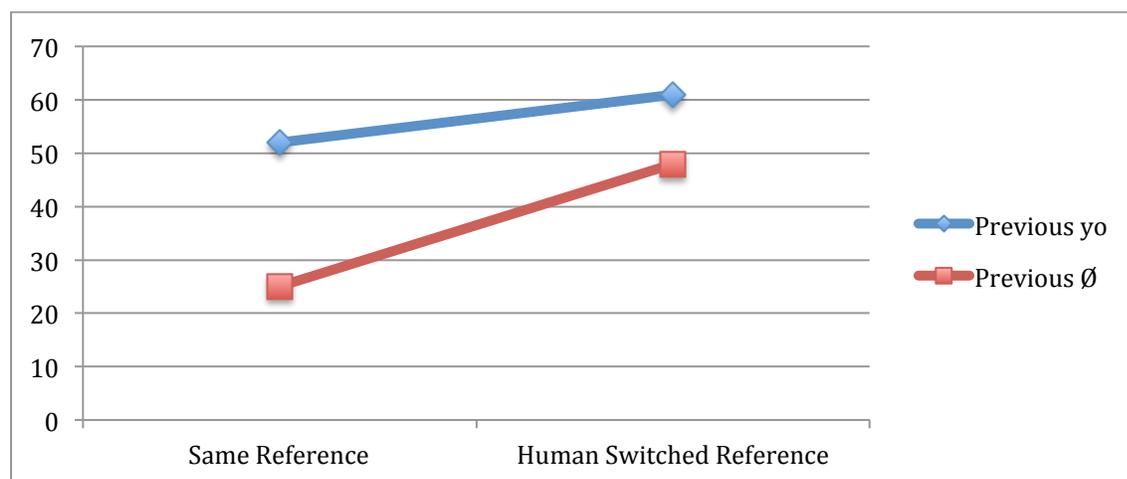
Second, the strength of subject continuity is moderated by coreferential subject priming. This is seen in the steeper slope of the Previous Ø line, which tells us that Human Switched Reference has a greater effect when the previous coreferential 1sg subject was unexpressed. In this context, the rate of *yo* goes up from 25% under Same Reference to nearly double, 48%, in the presence of one or more intervening human subjects. But when the previous mention was a *yo*, the increase in the rate of *yo* is smaller, from 52% under Same Reference to 61% under Human Switched Reference.[6]

In fact, what we have is a lower rate of *yo* with a previous unexpressed mention in a coreferential context (25%), on one side, and on the other, similarly high rates in all three other cells (48%-61%). Thus, subject continuity operates when the preceding subject is unexpressed but is neutralized by coreferential subject priming, while *yo-yo* priming operates in continuous contexts but is neutralized in non-continuous contexts (Travis and Torres Cacoullos 2012, 729-733).

3.4 Elusive tense effects

As seen in Table 1, the Imperfect favors *yo*. Although favoring by the Imperfect and other tenses with person syncretism has been interpreted as supporting an ambiguity-resolving function for subject pronouns, true ambiguity may be uncommon (Ranson 1991) and instead relevant may be

---

5 We find a similar narrowing from Same to Switch Reference contexts as traditionally conceived: the rate of *yo* expression is 52% (N=153) with a previous *yo* vs. 25% (N=137) with a previous unexpressed mention in Same Reference contexts but 59% (N=348) vs. 46% (N=321) in Switch Reference contexts.

6 Again, we find a similar difference for traditional Switch Reference, with an increase from 25% to 46% in the context of a previous unexpressed mention but only from 52% to 59% in the context of a previous *yo*.

tense(-aspect-mood) discourse function (Silva-Corvalán 1997; 2001). Furthermore, 'tense' may be more pertinent in narrative than in conversation (cf. Travis 2007, 119) and for third person than for first (Travis and Torres Cacoullos 2012, 734) and may interact with other predictors (subject continuity (Cameron 1994, 33-38) or verb class (Torres Cacoullos and Travis 2011, 253)). It is hoped that further (cross-linguistic) exploration will elucidate tense-aspect effects.

3.5 Summary: Reconsidering constraints on subject expression

In sum, the study of subject expression will benefit from a reconsideration of the widely-invoked factors influencing this variability, namely:
(1) The verb class effect as one of particular constructions.
> The (*yo*) + COGNITION VERB$_{1SG}$ construction, a frequent instance of which is *(yo) creo* 'I think', is distinguished by its linguistic conditioning, with turn-initial position favoring *yo* expression with cognition, but not other, verbs.

(2) The cognitive factor of subject continuity, as an operationalization of referent accessibility.
> The intervening human subjects-based measure (Human Switched Reference) provides a more discerning account of *yo* expression than does the clause-based measure (Switch Reference).

(3) Mechanical priming effects, in conjunction with subject continuity.
> Established here is coreferential subject priming—the realization of the previous coreferential mention as subject, which, importantly, interacts with subject continuity.

Are these candidates for cross-linguistic constraints or are they particular to Spanish? Below we test these factors for English 1sg subject expression.

4. Unexpressed ('null') subjects in English

As seen in Figure 5.2 above, while Spanish rates of 1sg expression range from approximately 25 percent to 50 percent, English rates are above 95 percent. A more telling difference, however, lies in the variable context for *I* expression, which is found to be more limited than for Spanish *yo* expression (as discussed in Torres Cacoullos and Travis 2013).

4.1 The variable context of English subject expression

A first circumscription of the variable context for English subject expression is to declarative main clauses, as in lines 2 and 3 in (4). Initial analysis revealed no cases of unexpressed *I* in interrogatives or subordinate clauses outside of instances with *and*-coordination (as in *if I go out and Ø ask for it* (17 Jim: 7)). In contrast, in Spanish, main (as opposed to subordinate) clauses have been reported either to have no significant effect (Travis 2007, 115) or to favor subject expression (Shin and Montes-Alcalá 2014, Cuadro 6).[7]

---

[7] Cameron (1992, 210-216) suggests that clause type and position of a subordinate clause in relation to the main clause may interact with subject continuity.

(4)
1.   Angela: ... (TSK) (H) and **I** put some onion powder ... in the mayonnaise,
2.   .. (H) and **Ø** put it on some .. boiled eggs.
3.   ... **Ø** Opened em up,
4.   (H) and **I** didn't stuff the eggs.
5.   (H) **I** just put that (H) mayonnaise on top.

<div align="right">(SBCSAE, 11 This retirement bit: 759-763)</div>

The variable context for *I* expression must further leave aside discourse markers—collocations *I mean*, *I guess*, *I think*, *I (don't) know*, *I remember* and *I'm sure* when produced prosodically independently from other clausal material or when appearing as parentheticals (between the subject and verb or following the verb). Not only is *I* invariably expressed but, with the exception of *I don't know*, it is generally unstressed in these collocations when they are used as discourse markers, indicative of their formulaic status (cf. Travis and Torres Cacoullos 2014: 363-365). Likewise, *say*, *be like*, *go* and *think* when used as quotatives pattern like discourse-marker formulas in disfavoring stress on *I* and appear with unexpressed *I* only in *and*-coordinated constructions (for example, *I phoned her and Ø said* (11 Angela: 954).

Finally, also outside the variable context are contracted forms of auxiliaries *be, will, have, had,* and *would* which did not once appear in the absence of expressed *I* (i.e., there were no cases of *'m, 'll, 've, 'd*). These three sets of contexts together represent nearly half of all instances of *I* (155/320 in a random sample).

This meticulous circumscription of the variable context allows us to follow the *principle of accountability*, that occurrences and non-occurrences of a given variant be noted (Labov 1982, 30). The following analyses are based on all 151 unexpressed 1sg subjects occurring in the SBCSAE (compared with approximately 9,000 tokens of *I*) and, following the methodology applied in Leroux and Jarmasz (2005), a sample of approximately twice as many expressed *I*, constituted by the closest preceding and following coreferential token falling within the envelope of variation.[8] For example, for the unexpressed 1sg subject in line 4 in (5) below, we extract the closest following *I* produced by the speaker which is in line 9, but skip over the closest preceding in line 2 as it occurs in a subordinate clause and extract the token in line 1. In cases of sequences of two or more unexpressed, we went to the next closest previous and following eligible tokens.[9] This protocol yielded 446 tokens, in which the *I* expression rate is (an artificial) 66%, which then allows us to compare the frequency of the variants in sets of linguistic sub-contexts in order to determine the linguistic conditioning.

---

8 We obtain similar results from a sample constituted by only one preceding token of expressed *I* for each unexpressed (Torres Cacoullos and Travis 2014). We thank Matthew Callaghan for help with the extraction and coding of additional tokens.

9 In the few cases where there was no preceding or following instance (for example, when the unexpressed subject is at the start of a transcript), we looked in the other direction (i.e. to following tokens if there were none preceding); for one unexpressed token there was no expressed *I* meeting our extraction criteria in the transcript, and for a further five there was only one such token. We skipped over *I*s occurring in quoted speech (uncodable for subject continuity and priming) or in a separate Intonation Unit from the verb (as in line 2 in (5)) (inapplicable to unexpressed mentions). Because we skip *I*'s outside the variable context, the proportion of 1sg tokens occurring in coreferential contexts following this extraction method is akin to that when 1sg tokens are exhaustively extracted, at approximately 50% (comparing with a subset of this same corpus for a study of stressed *I* ((Travis and Torres Cacoullos 2014).

(5)
1.   Alan: **I** keep a diary,
2.   ... when I,
3.   ... go out of the country.
4.   ... Ø Still got em in here.
5.   .. Otherwise,
6.   (H) my problem is,
7.   ... and it's a lot easier to do it that way,
8.   but I,
9.   ... **I** keep a diary when I go out of the country,

(SBCSAE, 60 Shaggy dog story: 757-765)

Further analysis of these 446 tokens leads to additional refinement of the variable context. Table 3 depicts rates of the expressed 1sg subject pronoun in English and Spanish according to coordination and position in the prosodic unit, here, the Intonation Unit (IU) ("a stretch of speech uttered under a single coherent intonation contour" (Du Bois, Schuetze-Coburn, Cumming and Paolino 1993, 47), transcribed on a single line).

In terms of *and- / y*-coordination, English and Spanish show the same tendency: for both English and Spanish, *and- / y*-coordinated verbs (seen in line 2 and 4 in (4) for English and line 3 in (2) for Spanish) have the lowest rates of expression. In terms of position in the prosodic unit, however, the languages differ. For English, outside the context of *and*-coordination, <u>un</u>expressed 1sg subjects occur virtually only in absolute initial position, as in line 4 in (5). Though the English initial-position constraint for unexpressed subjects has been characterized in terms of an undefined notion of the sentence (cf. Napoli 1982, 99, Roberts and Holmberg 2010, 5), it is not a syntactic constraint, but a prosodic one (Torres Cacoullos and Travis 2014b, 29).

As can be seen in Table 3, in non IU-initial position expressed *I* is near categorical.[10] In contrast, in Spanish, in non-IU-initial position the rate of expressed *yo* is lower than in initial position. Here, then, we have a 'conflict site' (Poplack and Meechan 1998, 132) according to which the grammar, as instantiated in the structure of variability, is clearly different for Spanish and English 1sg subject expression, beyond patent overall rate differences.

We discuss the two loci of variability of English *I* expression in turn below: *and*-coordinated verbs and, elsewhere, IU-initial position.

---

10 Non-IU initial tokens are preceded most frequently by *so* (N=19), *(and) then* (N=11, as in line 7 in (9), *oh* (N=7), *well* (N=6) and a variety of adverbials (e.g., *first of all* (N=7). There was just one token of an unexpressed *I* in this environment (*So I sat over here, . . . before we went over to Diane's, and Ø explained the recipe to em*, 43 Alice: 486). This token is preceded by *and* in the IU and has one intervening human subject in a subordinate clause, and thus could be considered a case of coordination in a broader definition than that which we apply here (cf. Torres Cacoullos and Travis 2014: 26).

Table 3: Rate of expressed 1sg subject pronoun by coordination and according to position in prosodic unit (Intonation Unit, IU)

	English – SBCSAE	Spanish – CCCS
	% *I* (vs. Ø) N= 446 Overall sample: 34%*	**% *yo* (vs. Ø)** N=913[11] Overall sample: 49%
and/*y*-coordinated coreferential V**	24% (24/98)	26% (9/27)
Non-coordinated verbs		
Non-IU-initial V	99% (85/86)	45% (106/238)
IU-initial V	71% (186/262)	55% (353/645)

* Gross overall rate of English *I* (not considering the variable context) is > 95%.
** As defined in section 4.2.

4.2 And-coordinating constructions: debunking VP conjunction

As seen in Table 3, *and*-coordination very strongly favors u̲n̲e̲x̲p̲r̲e̲s̲s̲e̲d̲ *I* in the second conjunct, where the rate of expression is 24% (24/98), compared with 78% (271/348) elsewhere. We define *and*-coordinated clauses narrowly as clauses with coreferential subjects (that is, no intervening human subjects) conjoined with *and,*, as in lines 1-2 and 3-4 in (4) above, and in (6) below.[12] This is the pertinent delimitation of *and*-coordination, because the rate of expression of *I* in non-coreferential contexts in clauses introduced with *and* is near the overall rate (19/20). Furthermore, though other conjunctions have been included in this set in past treatments of English null subjects (e.g., Quirk, Greenbaum, Leech and Svartvik 1985, 910), not only are conjunctions other than *and* much less frequent with 1sg clauses (e.g., *but* occurs just 11 times and *or* does not occur at all), they show no indication of having the same tendency as *and* of favoring unexpressed *I* (5 of 6 *but*-clauses in coreferential contexts have expressed *I*).

(6)

      Alina: *.. I̲ w̲e̲n̲t̲ and **Ø saw** their house the other night*.

                                          (SBCASE, 06 Cuz: 1474)

      Besides the lower rate of *I* expression, coreferential clauses conjoined with *and* are distinguished in tending to occur with positive polarity (95%, 93/98 of *and*-coordinated vs. 83%, 218/262 elsewhere), and in the Past rather than Present tense (75%, 70/93, vs. 58%, 139/238). Furthermore, *I* expression here is not subject to priming (as it is in other contexts, see §4.3).

---

[11] Not included in the count for IU position are cases (N=51) with an unexpressed subject in which the verb is preceded in the IU only by an adverbial which the expressed subject variably precedes or follows (most frequently *ya* (N=30) and *ahor(it)a*) (N=12)). Also excluded are interrogatives (N=19), and where *yo* occurs in a different IU from the verb (N=37).

[12] We include *and* with minimal other material, such as *and um*, but not with more substantial material, such as *and then*.

Thus, we conclude that coreferential clauses conjoined with *and* represent a distinct schematic construction, depicted as [I VERB$_{1SGi}$ *and* Ø VERB$_{1SGi}$].

It has been proposed that instances without a subject pronoun in the second conjunct of verbs conjoined with *and* (as in (6)) are best analyzed as VP-conjunction, involving a single clause with two predications, rather than two clauses with an unexpressed subject in the second (e.g., Huddleston 2002, 238 (§233.231), Quirk et al. 1985, 942 (§913.944)). Reliable criteria for identifying VP-conjunction, however, have been lacking.

Here we consider not only the absence of an expressed subject, but take into account morphosyntactic, semantic and prosodic criteria, considering to be more tightly linked clauses that have the same tense, refer to the one event, and occur on the same Intonation Unit, as material occurring on one IU tends to have a tighter syntactic relationship than material occurring across IUs (Chafe 1994, Ch. 9, Croft 1995, Torres Cacoullos and Travis 2014a)). This prosodic criterion holds true for the [I VERB$_{1SGi}$ *and* Ø VERB$_{1SGi}$] construction: when *and* coordination occurs across IUs the rate of <u>non</u>-expression for the second conjunct is lower than when both conjuncts appear on the same IU (42/66 across IUs, 32/32 on the same IU in our sample, though see (9) below for an illustration of an expressed subject in this context).

In (6), the morphosyntactic, semantic and prosodic criteria all converge with the unexpressed subject in the second conjunct: both verbs occur in the same tense, the acts of *going* and *seeing* could be considered part of the one event, and they appear in the one IU. However, we also find very similar examples spread across IUs, and even with intervening material, as in lines 4-6 in (7), where *look* and *make sure* have the same tense and occur simultaneously. And in (8), the two parts of the conjunct refer to distinct events (the first making mayonnaise, the second putting that mayonnaise on the eggs), but here also we retain the unexpressed subject. Finally, we note that even when these features coincide, we do occasionally find expressed subjects, illustrated in line 7 in (9) (as compared with line 1, of which it is a near repetition, but where the subject is left unexpressed). These examples counter the notion of a discrete category of 'VP-conjunction'. Rather, we seem to have a continuum, from two verbs that are very tightly associated (as in (6)), which favor unexpressed subjects more, to those that are less so (as in (8)), which favor unexpressed subjects less.

(7)
1. Fred: ... (H) I look at my bank sta- .. bank statements.
2.   And,
3. Wess: Mhm.
4. Fred: ... Ø <u>look</u> through my checks.
5.   When they come in.
6.   .. And Ø **make** sure that it's fine.

(SBCSAE, 59 You baked: 754-756)

(8)
Angela: ... (TSK) (H) and <u>I put</u> some onion powder ... in the mayonnaise,
            .. (H) and Ø **put** it on some .. boiled eggs.

(SBCSAE, 11 This retirement bit: 759-760)

(9)
1. Walt: then I go and **Ø talk**,
2.   <VOX oh he's gonna say this and,
3.   (H) he's gonna say that and,
4.   (H) he's gonna be in my face and,
5.   da-na-da-na-da-na-da-na= VOX>.
6.   ... And I'm like,
7.   ... and then I go and **I talk** to him.

(SBCSAE, 21 Fear: 1092-1098)

Within the same-IU coreferential coordinating construction [I VERB$_{1SGi}$ *and* Ø VERB$_{1SGi}$], we find two lexically specific constructions, which together make up 88% (28/32) of its instances. These are with a quotative as the second verb, [*I* VERB$_{1SGi}$ *and* Ø VERB-OF-SPEECH$_{1SGi}$], for example, *I phoned her and Ø said* (11, Angela: 954), and with a motion verb as the first verb, [*I* GO/VERB-OF-MOTION$_{1SGi}$ *and* Ø VERB$_{1SGi}$], for example, *That's why I went out and Ø bought the coffee.* (49, John: 1176) (see Torres Cacoullos and Travis 2014b, 31, for more discussion of these constructions).

To return, then, to the question of cross-linguistic equivalencies, we note that it is false to assert that English differs from so called null-subject languages in a syntactic structure of VP-coordination. Rather, English, like Spanish, demonstrates a disfavoring of expressed subjects in coreferential *and*-coordinated constructions. This may be a candidate cross-linguistic constraint, being also reported for Finnish (Helasuvo Forthcoming) and Russian (Nagy et al. 2011, 142).

4.3 Prosodic-initial position: identifying conflict sites

As was seen in Table 3, variability in *I* expression outside *and*-coordinated constructions is restricted to absolute IU-initial position. The multivariate analysis for this context is shown on the right of Table 1 above. This indicates that, unlike Spanish, there is no cognition-verb effect in English. However, as for Spanish *yo*, significant factor groups are Subject continuity (Human Switched Reference) and coreferential subject Priming.

First, the priming effect seen here is that unexpressed subjects are favored when they are preceded by another unexpressed coreferential subject (with a Probability of .35). That is, unexpressed *I* is rare, but when it does occur, it tends to do so in clusters, as in (10). The example in (11) (from the same conversation) illustrates a similarly structured narrative excerpt, this time all with expressed *I*.

(10)
1. Tom3:   (TSK) (H) **Ø** Graduated from there,
2.   ... u=m,
3.   ... b- **Ø** got rea=l homesick for this part of the country.
4.   (H) ... **Ø** ... Came back out here,
5.   **Ø** went through law school at the University of New Mexico,

(SBCSAE, 32 Handshakes all around: 1654-1658)

(11)
1. Tom2   **I** stayed in the merchant navy,
2.   u=m,
3.   .. (H) my first trip was around the world out to the Middle East,
4.   uh=,
5. Tom1:   (H)
6. Tom2:   **I** came back,
7.   **I** went to school for a few months,

(SBCSAE, 32 Handshakes all around: 320-326)

Second, Table 1 indicates that English shares with Spanish the effect of Subject continuity, an obvious candidate for a cross-linguistic constraint on subject expression. Though there are cases of unexpressed 1sg in non-coreferential contexts (as in line 9 in (12)), it is disfavored, and Human Switched Reference is an environment strongly propitious to expressed *I*.

(12)
1. Alan: .. Had a pouch,
2.   like,
3.   sorta like you've got.
4.   I carried it around.
5.   Well any rate,
6.   ... (H) my wife had fallen in lo=ve,
7.   ... with a,
8.   ... Mexican artist by the name of ... Nierman.
9.   ... Ø Forgot his first name.

(SBCSAE, 60 Shaggy dog story: 104-112)

However, the Subject continuity effect on *I* expression is at least in part dependent on the priming effect, as can be seen by the cross-tabulations in Table 4. This shows the rate of expressed *I* by previous realization (coreferential subject priming) and Human Switched Reference (subject continuity).

Table 4: Rate of expressed *I* in Human Switched Reference vs. Same Reference (presence vs. absence of an intervening human subject) according to the realization of the previous coreferential 1sg subject, for IU-initial (non-*and*-coordinated) verbs (N=173)[13]

	Realization of previous coreferential 1sg subject					
	Unexpressed (N=31)		Expressed (*I*) (N=142)		Total	
Subject continuity						
Same Reference	40%	10/25	71%	80/112	66%	90/137
Human Switched Ref		6/6	73%	22/30	77%	27/35
TOTAL	52%	16/31	72%	102/142		

Though the data are sparse, in the aggregate (seen in the Totals column), the tendency for subject continuity is the predicted one, with a lower rate of expressed *I* in Same Reference, or coreferential contexts (66%), than in Human Switched Reference, or non-coreferential (77%) contexts. But within Same Reference contexts (on the first row), 1sg is more likely to be expressed if the previous realization was also expressed (71%) than if it were unexpressed (40%) (*p* = 0.0046 by Fisher's exact test). Further, most cases (25/31) of a previous unexpressed subject occur in coreferential contexts. On the other hand, when the previous realization was expressed *I* (in the second column), subject continuity makes no difference, with a rate of expression of 71-73% in both Human Switched Reference and Same Reference contexts. Thus, the subject continuity effect in English is bound to unexpressed-to-unexpressed priming, which tends to occur with coreferential subjects, thus lowering the rate of expressed 1sg subjects in this context. Recall that in Spanish, similarly, subject continuity (Human Switched Reference) is moderated by priming, with a weakened effect under *yo*-to-*yo* priming.

In summary, the two loci of variable *I* expression are coreferential *and*-coordination and prosodic-initial position. A parallel favoring of non-expression in *y*-coordinated verbs applies to Spanish *yo*, while prosodic-initial position qualifies as a conflict site between the two languages. Within the variable context of prosodic-initial position, the favoring effect of cognition verbs replicated here for *yo* is absent for English *I* expression. However the comparison indicates parallel effects in Spanish and English for Subject continuity (again, replicating an effect found in virtually every study of Spanish subject expression) and for Priming (which thus far has received little attention).

5. Conclusion

Despite disparate rates of expression, 1sg subject pronouns in Spanish and English are conditioned similarly by the candidate cross-linguistic constraints of *and*-coordination, subject continuity and priming. For both Spanish *yo* and English *I*, the subject continuity effect is one of Human Switched Reference (in terms of intervening human subjects from the previous coreferential subject). As concerns priming, we find an effect for coreferential subject priming

---

13 Excluded are 2 tokens for which, due to unclear speech, it is not possible to determine if there is 0 or 1 intervening subject.

(that is, the realization of the previous coreferential subject), which as demonstrated modulates the subject continuity effect. Differences are also observed. The well known favoring effect of cognition verbs on *yo* is confirmed here for Spanish (and is interpreted as an effect of particular constructions), but is absent for English *I* expression. And, crucially, a prosodic constraint exists for English only, such that outside of *and*-coordinating constructions, unexpressed 1sg subjects virtually only occur in prosodic initial position. This prosodic constraint thus provides an ideal test for convergence.

These findings allow us to answer the question of whether Spanish subject expression in Spanish-English bilingual speech is an appropriate linguistic variable to ascertain convergence. Having distinguished candidate cross-linguistic patterns from language-specific ones, we can answer 'yes', since the linguistic conditioning of *I* expression presents identifiable differences from that of *yo* expression, especially the prosodic-initial position constraint.

The analysis of **intra-**linguistic variability has allowed us to distinguish superficial from structural **inter-**linguistic similarity, and calls into question assumptions of blanket equivalence (or incommensurability) of English and Spanish subject expression. Beyond 'higher' or 'lower' overall rates, the conflict sites in the linguistic conditioning of variant choice identified via systematic quantitative analysis will enable analysts to gauge grammatical (dis)similarity (Torres Cacoullos and Travis to appear).

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

.	final intonation contour	-	truncated word
,	continuing intonation contour	=	lengthening
?	appeal intonation contour	[ ]	speech overlap
--	truncated intonation contour	(H)	in-breath
..	short pause (about 0.5 secs)	(TSK)	click
...	medium pause (> 0.7 secs)	<VOX VOX>	marked voice quality

References

Ariel, Mira. 1988. "Referring and accessibility." *Journal of Linguistics* 24(1), 65-87.

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

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan and Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In *Formulaic language, Vol. 1: Distribution and historical change*, edited by Roberta L. Corrigan; Edith Moravcsik; Hamid Ouali; and Kathleen Wheatley, 187-217. Amsterdam: John Benjamins.

Cameron, Richard. 1992. Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation. Unpublished Unpublished PhD thesis, University of Pennsylvania, Philadelphia, PA.

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.

Cameron, Richard. 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(2), 27-45.

Cameron, Richard and Nydia Flores-Ferrán. 2003. "Perseveration of subject expression across regional dialects of Spanish." *Spanish in Context* 1(1), 41-65.

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

Croft, William. 1995. "Intonation units and grammatical structure." *Linguistics* 33, 839-882.

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

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

Erker, Daniel and Gregory R. Guy. 2012. "The role of lexical frequency in syntactic variability: Variation subject personal pronoun expression in Spanish." *Language* 88(3), 526-557.

Fox, Barbara. 1987. *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press.

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

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Helasuvo, Marja-Liisa. Forthcoming. "Searching for motivations for grammatical patternings." *Pragmatics* Special Issue on "Approaches to Grammar for Interactional Linguistics", guest editors Elizabeth Couper-Kuhlen, Ritva Laury and Marja Etelämäki.

Holmquist, Jonathan. 2012. "Frequency rates and constraints on subject personal pronoun expression: Findings from the Puerto Rican highlands." *Language Variation and Change* 24(2), 203-220.

Huddleston, Rodney. 2002. The clause: Complements. In *The Cambridge Grammar of the English Language*, edited by Rodney Huddleston and Geoffrey K. Pullum, 213-322. Cambridge: Cambridge University Press.

Labov, William. 1982. Building on empirical foundations. In *Perspectives on Historical Linguistics*, edited by Winfred P. Lehmann and Yakov Malkiel, 17-92. Amsterdam: John Benjamins.

Leroux, Martine and Lidia-Gabriela Jarmasz. 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.

Levinson, Stephen C. 1987. "Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena." *Journal of Linguistics* 23(2), 379-434.

Nagy, Naomi G., Nina Aghdasi, Derek Denis, and Alexandra Motut. 2011. "Null subjects in heritage languages: Contact effects in a cross-linguistic context." *University of Pennsylvania Working Papers in Linguistics* 17(2), Article 16.

Napoli, Donna Jo. 1982. "Initial material deletion in English." *Glossa* 16, 85-111.

Oh, Sun-Young. 2005. "English zero anaphora as an interactional resource." *Research on Language and Social Interaction* 38(3), 267-302.

Ono, Tsuyoshi and Sandra A. Thompson. 2003. "Japanese *(w)atashi/ore/boku* 'I': They're not just pronouns." *Cognitive Linguistics* 14(4), 321-347.

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

Paredes Silva, Vera Lucia. 1993. "Subject omission and functional compensation: Evidence from written Brazilian Portuguese." *Language Variation and Change* 5(1), 35-49.

Poplack, Shana. 1980. The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In *Locating language in time and space*, edited by William Labov, 55-67. New York: Academic Press.

Poplack, Shana, Allison Lealess, and Nathalie Dion. 2013. "The evolving grammar of the subjunctive." *Probus* 25(1), 139-195.

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

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

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.

Ranson, Diana L. 1991. "Person marking in the wake of /s/ deletion in Andalusian Spanish." *Language Variation and Change* 3(2), 133-152.

Roberts, Ian and Anders Holmberg. 2010. Introduction: Parameters in minimalist theory. In *Parametric variation: Null subjects in minimalist theory*, edited by Theresa Biberauer; Anders Holmberg; Ian Roberts; and Michelle Sheehan, 1-57. Cambridge / New York: Cambridge University Press.

Sankoff, David. 1988. Variable rules. In *Sociolinguistics: An international handbook of the science of language and society*, vol. 2, edited by Ulrich Ammon; Norbert Dittmar; and Klaus J. Mattheier, 984-997. Berlin / New York: Walter de Gruyter.

Sankoff, David, Sali Tagliamonte, and Eric Smith. 2012. Goldvarb LION: A variable rule application for Macintosh. University of Toronto. URL http://individual.utoronto.ca/tagliamonte/goldvarb.htm.

Shin, Naomi Lapidus and Cecilia Montes-Alcalá. 2014. "El uso contextual del pronombre sujeto como factor predictivo de la influencia del inglés en el español en Nueva York." *Sociolinguistic Studies* 8(1): 85–110.

Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. In *Spanish in the United States: Sociolinguistic aspects*, edited by Jon Amastae and Lucía Elías Olivares, 93-120. New York: Cambridge University Press.

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

Silva-Corvalán, Carmen. 1997. "Avances en el estudio de la variación sintáctica: La expresión del sujeto." *Cuaderno del Sur: Letras, Homenaje a Beatriz Fontanella de Weinberg* 27, 35-49.

Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.

Torres Cacoullos, Rena and Jessi Elana Aaron. 2003. "Bare English-origin nouns in Spanish: Rates, constraints and discourse functions." *Language Variation and Change* 15(3), 289-328.

Torres Cacoullos, Rena and Catherine E. Travis. 2011. "Using structural variability to evaluate convergence via code-switching." *International Journal of Bilingualism* 15(3), 241-267.

Torres Cacoullos, Rena and Catherine E. Travis. 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, Rena and Catherine E. Travis. 2015. "Gauging convergence on the ground: Code-switching in the community (Special issue, Gauging convergence on the ground: code-switching in the community, edited by Catherine E. Travis and Rena Torres Cacoullos)." *International Journal of Bilingualism*.

Torres Cacoullos, Rena and Catherine E. Travis. To appear. "Two languages, one effect: Structural priming in code-switching." *Bilingualism: Language and Cognition* (Special issue edited by Margaret Deuchar).

Torres Cacoullos, Rena and James A. Walker. 2009. "The present of the English future: Grammatical variation and collocations in discourse." *Language* 85(2), 321-354.

Torres Cacoullos, Rena and James A. Walker. 2011. Chapter 18: Collocations in grammaticalization and variation. In *Handbook of Grammaticalization*, edited by Bernd Heine and Heiko Narrog, 225-238. Oxford: Oxford University Press.

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. and Rena Torres Cacoullos. 2012. "What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation." *Cognitive Linguistics* 23(4), 711-748.

Travis, Catherine E. and Rena Torres Cacoullos. 2014. "Stress on I: Debunking unitary contrast accounts ". *Studies in Language* 38.2: 372-404.

Weiner, E. Judith and William Labov. 1983. "Constraints on the agentless passive." *Journal of Linguistics* 19(1), 29-58.

Weinreich, Uriel. 1968. *Languages in contact*. The Hague: Mouton.