Subject pronoun realization in Spanish and English: assessing inter-linguistic functional equivalence via intra-linguistic inherent variability

Rena Torres Cacoullos & Catherine E. Travis

1 Introduction

Subject expression has been considered a paradigmatic case for grammatical convergence in studies of US Spanish, under the assumption that subject pronouns are inter-linguistically similar enough to license the prediction of boosted rates 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 just such a statement of inter-linguistic grammatical comparability, beyond overall rates of superficially similar forms, by probing the similarity of Spanish and English first person singular (1sg) pronouns.

The focus is narrowed to 1sg because first- and second-person referents are intrinsically of ‘given’ information status (Chafe 1994: 79), unlike the third person, which, additionally, may be realized as a full NP. Further, in Spanish at least, strong person-number effects are consistently found in aggregate data sets, calling for further comparisons of individual pronouns to show whether the same constraints apply across grammatical persons. Thus, the current state of our knowledge has not settled whether subject pronouns constitute a single category.

The data are approximately 1,000 clauses with 1sg subjects from the Corpus of Conversational Colombian Spanish (CCCS) (Travis 2005a) and 1,800 from the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois 2000; Du Bois et al. 2003; Du Bois and Englebretson 2004, 2005), including 151 tokens of null 1sg subjects. Both are corpora of conversation, with prosodically-based transcription using Intonation Units (henceforth, IUs) (cf. Du Bois et al. 1993: 47), each of which is represented on a distinct line (see Appendix for transcription conventions).

We begin (§2) by motivating the utility of intra-linguistic variation patterns for establishing inter-linguistic grammatical (dis)similarity. We next (§3) synthesize hypotheses about the functions of subject pronouns cross-linguistically and propose operationalizations in terms of predicted co-occurrence patterns, establishing these for Spanish yo expression and then applying them to the two candidate English equivalents, I expression (§4) and stress on I (§5).

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

2.1 The uninformativeness of overall rate comparisons

A range of rates of Spanish 1sg subject expression is depicted in Figure 1. At the two extremes, the difference in yo rate between Peninsular varieties in Madrid (32%, N = 10,185) and Puente Genil, Andalusia (50%, N = 307) is 18 percentage points, or 56%,
while in Puerto Rico, the difference between Castañer (34%, N = 1,257) and San Juan (50%, N = 949) is 16 percentage points, or 47%. Such rate differences by regional dialect or genre may be quantitatively greater than those between immigrant generations in the United States. In New York, for example, the difference in subject expression rate (all grammatical persons) between New-York born (NYBR) (38%, N = 12,077) and Newcomers (30%, N = 17,030) is 8 percentage points (Otheguy et al. 2007: 782, Table 5), or 27%. The point here is that statistically-significant rate differences need not translate into linguistically-important differences, and subject-expression frequency differences in and of themselves do not permit an inference of change in progress (as Cameron 1995: 24, n.6 cautions), as the threshold for qualifying as a ‘high(er)’ rate remains unknown.

![Figure 1](image)

**Figure 1** Comparison of overall rates of occurrence of Spanish 1sg subject pronoun *yo*.

* Madrid, Spain (Enríquez 1984) and Santiago, Chile (Cifuentes 1980-1) as reported in Silva Corvalán (1994); Castañer, Puerto Rico (Holmquist 2012); Caracas, Venezuela (Bentivoglio 1987); Cali, Colombia (Travis 2007); San Juan, Puerto Rico (Cameron 1992: 233); Puente Genil, Andalusia (Ranson 1991).

### 2.2 Variationist comparative method for establishing grammatical (dis)similarity

With the wide range of overall rates, variable subject expression in Spanish nonetheless displays uniformity in *linguistic conditioning*—the configuration of internal factors which best explains the variation in a statistically significant way—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 (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, the grammar is the same across Spanish varieties. What has never been tested, however, is whether it is similar to the grammar of English subject realization.
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, e.g., Poplack and Levey 2010). Our premise is the following:

In discourse, grammatical similarity or functional equivalence is observable in parallel favoring effects of contextual factors—which operationalize putative shared constraints or functions—contributing to the selection of a given variant over “all non-occurrences of the variant” (Labov 1982: 30).

Table 1 shows three independent Variable-rule analyses (Sankoff et al. 2012) of factors contributing to choice of 1sg subject form: Spanish expressed yo (vs. unexpressed), English expressed I (vs. unexpressed) and English stressed I (vs. unstressed). Variable-rule analysis uses logistic regression to perform binomial multivariate analysis for a choice between two variants. Probabilities (factor weights) with values closer to 1 indicate favoring effects on expressed yo, expressed I and stressed I, respectively, in the three sets of columns. 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 observe here similarities as well as differences across the three subject pronoun variables, comparing by the significance and direction of effect of the factors listed on the left. First, for Verb class, the favoring effect of cognitive verbs on yo reported in previous studies across Spanish varieties (e.g., Bentivoglio 1987: 60; Enríquez 1984: 240; Silva-Corvalán 1994: 162; Travis 2007: 116-117) is replicated here, while such an effect is absent for both of the English putative counterparts. We will attribute the favoring effect of cognitive verbs to particular constructions (§3.1). Second, the comparison indicates parallel effects for Subject continuity, a factor included in virtually every study of Spanish subject expression, as well as Priming, which thus far has been included in few. Considering these two factors as candidates for cross-linguistic constraints on subject realization, we establish new operationalizations as applied to Spanish yo expression in §3.2 and §3.3, and probe them for English I expression in §4.4. Third, there is divergence between the three variables in Polarity and Turn Position, which we address in §5.2.

We do not pursue TAM (tense-aspect-mood) effects, as they have yet to be satisfactorily interpreted, awaiting reports from other languages. Favoring of expressed subjects by the Imperfect and other tenses with person syncretism has been interpreted as supporting an ambiguity-resolving function, but, as Ranson (1991) demonstrated, ambiguity is not as common as constructed examples may have led us to believe. Alternatively, as Silva-Corvalán (1997, 2001) proposed, what may be relevant is the discourse function of TAM forms, such that that pronominal subjects are less likely with foregrounded verbal situations, which are coded by perfectives. Furthermore, TAM 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). Crucially, interactions

---

1 Nagy et al. (2011:140-141) report that Tense is significant in Italian, but not Cantonese, English, or Russian subject pronoun expression.
merit investigation, including the absence of a TAM effect in coreferential contexts (Cameron 1994: 33-38) or with cognitive verbs (Torres Cacoullos and Travis 2011: 253).

Table 1: Three independent Variable-rule analyses of factors contributing to the selection of 1sg subject variant: yo (vs. unexpressed) in conversational Colombian Spanish (CCCS), expressed I (vs. unexpressed) and stressed I (vs. unstressed) in conversational American English (SBCSAE) (non-significant factors within []). (adapted from Travis & Torres Cacoullos 2012; Torres Cacoullos & Travis, forthcoming; Travis & Torres Cacoullos, in preparation).

<table>
<thead>
<tr>
<th></th>
<th>Yo</th>
<th>I (IU-initial)</th>
<th>Stressed I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected mean (Input)²</td>
<td>1,020</td>
<td>177</td>
<td>1,610</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Verb class</th>
<th>Prob</th>
<th>N</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive</td>
<td>.66</td>
<td>201</td>
<td>[.45]</td>
<td>41</td>
<td>[.50] 639²</td>
</tr>
<tr>
<td>Other</td>
<td>.46</td>
<td>819</td>
<td>[.52]</td>
<td>136</td>
<td>[.50] 951</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Priming: Realization of previous coreferential 1sg subject ¹</th>
<th>Prob</th>
<th>N</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressed yo / I / stressed I</td>
<td>.58</td>
<td>372</td>
<td>.57</td>
<td>110</td>
<td>.70 86</td>
</tr>
<tr>
<td>Unexpressed / unstressed</td>
<td>.41</td>
<td>330</td>
<td>.15</td>
<td>17</td>
<td>.47 637</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subject continuity: Intervening Human Subject(s)</th>
<th>Distance³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present (&quot;switch&quot;)</td>
<td>.56</td>
</tr>
<tr>
<td>Absent (&quot;same&quot;)</td>
<td>.42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Polarity</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>[.56]</td>
<td>216</td>
<td>[.44] 27</td>
<td>.68 291</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>[.49]</td>
<td>795</td>
<td>[.51] 150</td>
<td>.46 1319</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Turn position</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn-initial Intonation Unit</td>
<td>[.49]</td>
<td>332</td>
<td>[.49] 48</td>
<td>.56 443</td>
<td></td>
</tr>
<tr>
<td>Non-turn-initial IU</td>
<td>[.51]</td>
<td>686</td>
<td>[.51] 129</td>
<td>.48 1159</td>
<td></td>
</tr>
</tbody>
</table>

Other factor groups included in analyses:
- yo expression—TAM, Realization of subject of immediately preceding clause; I expression—TAM (n.s.); stressed I—Status of verb (clausal, formulaic).

1. Realization of previous coreferential 1sg subject is limited here to previous mentions at a distance of two or fewer intervening human subjects. The bulk of these are at 0 or 1 intervening human subjects (86%, 602/702 for expressed yo, 96%, 122/127 for expressed I, 94% 682/723 for stressed I).
2. For stressed I, cognitive, perception and desiderative verbs are combined.
3. For stressed I, subject continuity is best configured as Distance (0-2 vs. 3 or more intervening clauses), see §5.1.

² The corrected mean (input) indicates the overall likelihood that the variant (expressed yo, expressed I, stressed I) will occur. Comparisons of input values across the three analyses are not meaningful. For I expression, only a sample of the much more frequent expressed variant was taken, yielding an artificial expression rate of 50%; for stressed I, the relationship between the Input and the overall rate of variant selection (15%) (which generally it closely reflects) appears distorted because Realization of previous coreferential 1sg subject applies to only 45% (723/1,610) of the data.
3 Operationalizing (cross-linguistic) constraints or functions

Cross-linguistically, subject expression patterns are said to be shaped by cognitive factors, such as discourse cohesion and activation or accessibility (e.g., Ariel 1988:79; Chafe 1994:74; Levinson 1987:384), alongside more interactional or pragmatic factors, related to the kind of action the utterance performs (e.g., Fox 1987; Oh 2006; Ono and Thompson 2003, inter alia). Quantitative analyses have confirmed accessibility effects (e.g. Clancy 1980; Nagy et al. 2011; Paredes Silva 1993), and have also revealed the mechanical factor of priming, or the perseverance of the same syntactic structure across utterances (Cameron and Flores-Ferrán 2003:50-54; Travis 2007:120-121), as well as a role for lexical frequency (Erker and Guy 2012) and lexically particular constructions (Travis and Torres Cacoullos 2012). We discuss these below, beginning with the latter.

3.1 Identifying prefabs and (particular) constructions

There is growing recognition that particular expressions contribute to general distribution patterns (e.g., Poplack et al. 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, including conventionalized expressions, or prefabs, which mold the structure of variation. Prefabs and particular constructions may be replicably identified by unithood indices from distributional analysis, token and relative frequency, and the linguistic conditioning of variants (Bybee and Torres Cacoullos 2009; Torres Cacoullos and Walker 2011).

For Spanish 1sg subject expression, an observation beginning with, as far as we know, Bentivoglio (1987: 50-54), is that yo is favored by cognitive (or mental activity) verbs—most frequently here creer ‘think/believe’, saber ‘know’, pensar ‘think’, acordarse ‘remember’. Beyond the higher rate, independent multivariate analyses of cognitive and other verbs reported in Travis & Torres Cacoullos (2012: 734-737) revealed distinct linguistic conditioning of yo, with a contrary direction of effect for turn position. We took this divergence in subject expression patterns as evidence for a class of cognitive verbs, defined by the verb slot in a (yo + COGNITIVE VERB\textsubscript{1sg}) construction.

A strong candidate for a lexically particular construction within the cognitive verb class is yo creo. Evidence is: (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 the contrary tendency), while priming and turn position are significant (Travis and Torres Cacoullos 2012: 739-741). Thus, particular constructions both contribute to and deviate from more general variation patterns.

The turn-position constraint that is operative with cognitive verbs is such that occurrence in a turn-initial Intonation Unit (IU), as in lines 1, 2 and 5 in (1), favors yo expression (cf. Bentivoglio 1987: 38-40).\footnote{3} We interpret turn-position as an interactional
factor, as something intrinsically tied to turn taking (and, as cross-tabulations show, independent of subject continuity). This interactional effect in the particular (yo) + COGNITIVE VERB\textsubscript{1SG} construction may have fed scholars’ perception of an abstract meaning-feature of contrast for yo in general, evidence for which, we hasten to add, is lacking (Travis and Torres Cacoullos 2012: 737).

(1) \textit{(yo) + COGNITIVE VERB\textsubscript{1SG}}

1. Ángela: \textit{Ø te acordás que yo quería comprar un [congelador]?
\textit{‘(you) remember that I wanted to buy a [freezer]?’}

2. Santi: \textit{[Pues],
\textit{‘[Well],

3. \textit{yo pienso que el congelador de pronto no,
\textit{I think that the freezer maybe not,

4. mami --
\textit{honey -- ‘

5. Ángela: \textit{yo también pienso que no.
\textit{‘I also think not.’

(2) \textit{Examples are reproduced verbatim from audio recordings (see Appendix for transcription conventions).\textsuperscript{4} Codes in parentheses following examples refer to corpus, transcript name, and line number or time stamp.\textsuperscript{5} This is a reconfiguration of Givon’s (1983a) Potential Referential Interference measure (which counts semantically compatible referents in any syntactic role within the preceding three clauses).}

3.2 A new subject continuity measure: Intervening Human Subjects

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 expressed subject of the sentence in question”. Such an effect for subject continuity is consistent with the notion of accessibility, whereby cross-linguistically less coding material is said to correspond to contexts of greater accessibility—such as when the referent has been recently mentioned—and more coding material to correspond to contexts of lesser accessibility (Givón 1983a: 18). In our case, “less” vs. “more” coding material corresponds respectively to an unexpressed vs. expressed subject, and an unstressed vs. a stressed pronoun.

As an alternative operationalization of subject continuity, Travis and Torres Cacoullos (2012: 726-729) put forward Intervening Human Subjects, which considers the presence of subjects with specific human referents intervening between coreferential mentions as subjects.\textsuperscript{5} Intervening Human Subjects and Switch Reference largely overlap; in the CCCS database most (80%, 289/365) tokens of no intervening human subjects are also “same reference”. The difference is illustrated in the following examples. In (2), only \textit{te digo ‘(I) will tell you’ (line 3) occurs in what is traditionally considered a coreferential (Same Reference) context, while \textit{lo miro ‘(I) will look at it’ (line 5) and \textit{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 hence, by the Intervening Human Subjects measure, do not count as disrupting subject continuity.

(2)

1. Santi: \textit{Yo me averiguo,
\textit{‘I’ll find out,

2. y,
\textit{and,}
Intervening Human Subjects provides a more discerning account of variable *yo* expression than does Switch Reference. In Table 2, the top row shows that when there are no intervening human subjects between coreferential mentions, 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, in the rightmost column, we find a higher rate of *yo* expression if there is a human subject in the intervening clause(s) (54% vs. 41%, *p* = .04).

<table>
<thead>
<tr>
<th>Intervening Human Subjects</th>
<th>Same Reference (0 interven. clauses)</th>
<th>Switch Reference (1+ interven. clauses)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absent (0)</td>
<td>39% (113/289)</td>
<td>41% (31/76)</td>
</tr>
<tr>
<td>Present (1+)</td>
<td>No cases</td>
<td>54% (283/519)</td>
</tr>
</tbody>
</table>

3.3 *Coreferential subject priming and the interaction of subject continuity and priming*

Ubiquitous in language variation are structural priming or perseveration effects (cf. Bock 1986; Scherre and Naro 1991; Weiner and Labov 1983, inter alia). In *coreferential subject priming*, dubbed the *yo-yo* effect by Travis (2005b), speakers have a strong 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, which is still the bulk of the data (69%, 702/1020) (Travis and Torres Cacoullos 2012: 730).

(3)

Angela: 
*yo* ahorita no estoy trabajando. ‘Right now I’m not working.
... *Entonces,*
Es de ahorros. ‘... So,
.. *de unos ahorritos que yo tengo,* it’s from savings.
... *lo miro,* .. from some savings that I have,
.. si tiene alcoba del servicio, .. (I)”ll let you know.’
.. *Te aviso.*

[CCCS, Pizza 1369]
Priming and subject expression 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 absence vs. presence of Intervening Human Subjects and in Same vs. Switch Reference, according to the realization of the previous coreferential 1sg subject.

First, the priming effect is greater in ‘coreferential’ (no Intervening Human Subjects, or Same Reference) contexts. This is seen in the larger gap between the lines in the absence of intervening human subjects, where the rate of *yo* expression is 52% (N = 192) with a previous *yo* but 25% (N = 173) with a previous unexpressed mention (the same in terms of Same Reference: 52%, N = 153 vs. 25%, N = 137). The difference narrows in the presence of intervening human subjects, to 61%, N = 266 vs. 48%, N = 252 (in terms of Switch Reference: 59%, N = 348 vs. 46%, N = 321).

![Figure 2](image_url)

Figure 2  Rate of subject expression in presence vs. absence of Intervening Human Subjects (left, N = 883) and in Switch vs. Same Reference (right, N = 959), according to the realization of the previous coreferential 1sg subject

Second, the strength of subject continuity is moderated by priming. This is seen in the steeper slope of the Previous Ø line, which tells us that subject continuity has a greater effect when the previous coreferential 1sg subject was unexpressed. In this context, the rate of *yo* goes up from 25% in the absence of an intervening human subject to 48% in the presence of one or more (in terms of Switch Reference, from 25% to 46%). But when the previous mention was a *yo*, the increase in the rate of *yo* is smaller, from 52% in the absence of an intervening human subject to 61% in the presence of one or more (in terms of Switch Reference, from 52% to 59%).

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 (52%-61%). Thus, subject continuity (whether in terms of Intervening Human Subjects or Switch Reference) operates when the preceding subject is unexpressed but is neutralized by (*yo-yo*) priming, while priming operates in coreferential contexts but is neutralized in non-coreferential (Intervening Human Subjects or Switch Reference) contexts (Travis and Torres Cacoullos 2012: 729-733).
In summary, candidate cross-linguistic constraints on subject pronoun realization are: (1) Operationalizations of interactional-contrastive functions of subject pronouns. Turn position is one such operationalization that applies to the Spanish (yo) + COGNITIVE VERB₁SG construction, a frequent instance of which is the even more lexically particular construction (yo) creo. (2) As an operationalization of referent accessibility, the cognitive factor of subject continuity. The human-subjects-based measure (Intervening Human Subjects) provides a more discerning account of yo expression than does the clause-based measure (Switch Reference). (3) Mechanical priming effects, established here as coreferential subject priming—the realization of the previous coreferential mention as subject. There is interaction between subject continuity and priming.

4 Unexpressed (“null”) subjects in English

4.1 The variable context of English subject expression

A gross token count in the SBCSAE yields an overall rate of unexpressed 1sg subjects in the vicinity of 2% (151/9,000). A more telling difference with Spanish lies in the variable context for I expression. We find no cases of unexpressed I in interrogatives, relative clauses or subordinate (complement, adverbial, or if) clauses, other than ones involving coordination. That is, outside of instances of second conjuncts in and-coordination, as in lines 3 and 5 in (4), unexpressed 1sg is found only in declarative main clauses, as in lines 2 and 4 in (4) (Torres Cacoullos and Travis Forthcoming). 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 Under review, Cuadro 6).

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

(SBCSAE, 11 This retirement bit: 759-763)

The variable context for I expression must further leave aside discourse markers—collocations such as I mean, I guess, I think, I (don’t) know, when produced prosodically independently from other clausal material or 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. This lack of variation supports their formulaic status (see §5.3). Likewise, quotatives say, be like, go and think appear with unexpressed I only in and-coordinated constructions (for example, I phoned her and Ø said (11 Angela: 954) and, as we will see, pattern like discourse-marker formulas in disfavoring stress on I. Finally, also outside the variable context are contracted forms of auxiliaries be, will, have, had, would which did not once appear in the absence of expressed I (i.e., there were no cases of ’m, ’ll, ’ve, ’d) (Torres Cacoullos and Travis Forthcoming).
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 and a sample of expressed I constituted by the closest eligible (within the envelope of variation) preceding coreferential I for each unexpressed. The two contexts in which variation between expressed and unexpressed I occurs are and-coordinated verbs and absolute prosodic-initial position.

4.2 And-coordinated constructions, general and particular

First, and-coordination very strongly favors unexpressed I in the second conjunct. Coordinated clauses are narrowly defined as adjacent clauses with a coreferential subject, as in (5) (or lines 2-3 in (4)), or clauses separated at most by a single subordinate clause. Most (78/83) coordinated instances are adjacent, as may be represented in the schema [VERB1SGI and VERB1SGI].

(5) [VERB1SGI and VERB1SGI]
   Angela: (H) I just put that (H) mayonnaise on top.
   .. (H) and Ø put it on some .. boiled eggs.
   (SBCSAE, 11 This retirement bit: 759-763)

This [VERB1SGI and VERB1SGI] coreferential coordinated construction occurs across IUs, as in (5), more than half (46/78) the time. The rate of unexpression for the second conjunct across IUs is 83% (38/46), whereas on the same IU unexpression is near-categorical (32/32, in our sample). Within this same-IU [VERB1SGI and VERB1SGI] coreferential coordinated construction 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 VERB1SGI and Ø VERB-OF-SPEECH1SGI], for example, I phoned her and Ø said (11, Angela: 954), and with a motion verb as the first verb, [I GO/VERB-OF-MOTION1SGI and Ø VERB1SGI], for example, That's why I went out and Ø bought the coffee. (49, John: 1176).

4.3 Conflict site: the English prosodic-initial constraint for non-coordinated verbs

Table 3 indicates a parallel coordinated-verb constraint for Spanish, where the rate of expressed yo is also lowest in coreferential y-coordinated constructions, as in (2), lines 1-3. However, in English, outside the context of coreferential and-coordinated verbs, unexpressed 1sg subjects occur virtually only in absolute initial position of the prosodic unit, as in line 2 of (4) and lines 4-6 of (6). In non IU-initial position—preceded by so (N = 11), (and) then (N = 7), but (then, it was like), or okay, no, well, cause—I is expressed (95%). There were just two tokens of an unexpressed I in this environment, and both of these are preceded by and in the IU and could broadly be considered

---

6 For example, .. (SNIFF) So I sat over here, before we went over to Diane’s, and Ø explained the recipe to em, (SBCSAE, 43 Try a couple of spoonfuls: 484-486)
instances of coreferential and-coordination, in that the clause separating the conjuncts, though main, could be interpreted as a parenthetical insert.

In contrast, in Spanish the expression rate for coordinated constructions is close to that for non-IU-initial verbs (here, not statistically different, \( p = 0.2332 \)) — including verbs preceded by conjunctions such as *cuando* ‘when’ and *pero* ‘but’, fillers, or other (more substantial) material. Furthermore, the highest rate of *yo* is found precisely in absolute IU-initial position.

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

<table>
<thead>
<tr>
<th></th>
<th>English – SBCSAY</th>
<th>Spanish – CCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Ø (vs. <em>I</em>)</td>
<td>% Ø (vs. <em>yo</em>)</td>
</tr>
<tr>
<td>Coordinated V (and/y)*</td>
<td>12% (10/83)</td>
<td>30% (9/30)</td>
</tr>
<tr>
<td>Non-IU-initial V**</td>
<td>95% (40/42)</td>
<td>45% (106/238)</td>
</tr>
<tr>
<td>IU-initial V</td>
<td>57% (101/177)</td>
<td>55% (353/645)</td>
</tr>
</tbody>
</table>

* Coordinated Vs are those in a coreferential clause conjoined by *and*/y occurring either adjacent to the first conjunct or with a first subordinate clause intervening.

** Both unexpressed English tokens in Non-IU-initial position are preceded by *and* on the IU. Non-IU-initial V in Spanish includes *y* initially with non-coordinated verbs (\( N = 15 \)).

The English initial-position constraint has been viewed as a case of a more general phonological process of (variable) “left-edge deletion” (e.g., Weir 2012). The circumscription of the variable context to declarative main clauses (§4.1) follows from this constraint, which is prosodic in nature.  

### 4.4 Priming and subject continuity

The analysis of variable *I* expression in absolute IU-initial position is shown in the middle columns of Table 1. The one significant factor group is coreferential subject priming, such that unexpressed subjects are favored when they are preceded by another unexpressed coreferential subject. That is, unexpressed *I* is rare, but when it does occur, it tends to do so in clusters, as in (6). Priming has strong parallel effects for and-coordinated verbs as well (Torres Cacoullos and Travis Forthcoming).

(6)
1. Miles: *(H)* But it was like *I* went [to] Bahia,
2. Jamie: [What's] --

---

1 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 (\( N = 19 \)), tokens where *yo* occurs in a different IU from the verb (\( N = 37 \)), and one rare token with a clitic pronoun in a different IU from the verb (*Por eso es que no me -- registro en las clases con ustedes, oyó*).

8 In an (admittedly small) sample of 20 relative and 25 subordinate clauses, the *I* is 100% IU-internal.
In summary, the two loci of variable I expression are coreferential and-coordinated constructions and prosodic-initial position. A mechanical priming constraint applies across the board, but subject continuity is not an independent effect, being bound to coreferential and-coordinating constructions and to unexpressed-to-unexpressed coreferential subject priming for IU-initial verbs. Thus, prosodic position, together with particular constructions, qualifies as a “conflict site” (Poplack and Meechan 1998b: 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.
5 Stressed subject pronouns in English

In a typological perspective, the topic accessibility continuum (Givón 1983b: 17) places stressed pronouns in a language such as English and independent pronouns in a language such as Spanish at the same level, both coding discontinuous or less accessible participants. This presumed equivalence is made explicit in the statement that “[…] Spanish pronouns correspond [functionally] to English stressed pronouns (roughly speaking)” (Payne 1997: 43). In data from the SBCSAE, the rate of (perceived) stressed I is 13% (346/1,1861) (Travis and Torres Cacoullos In Preparation), which could be taken as supporting the comparability of Spanish pronouns with English stressed, rather than merely expressed, pronouns. Does the test of parallel linguistic conditioning uphold this claim?

5.1 Priming and subject continuity: distance vs. coreferentiality

The linguistic conditioning of stressed I (rightmost columns in Table 1) includes a coreferential subject priming effect, as with expressed yo and I. And also as before, this priming effect is stronger in coreferential contexts: in the absence of Intervening Human Subjects, the rate of stress with a previous stressed I is 21% (N = 63) but 8% (N = 464) with a previous unstressed mention, whereas in the presence of one (or two) Intervening Human Subjects the corresponding rates are 27% (N = 22) vs. 13% (N = 166).

But subject continuity has a different configuration from yo (and I) expression, being an effect of distance rather than one of local coreferentiality (whether in terms of the presence of an Intervening Human Subject or Switch Reference). The rate of stressed I in the context of one Intervening Human Subject (13%, N = 239) is not significantly higher than in the absence of Intervening Human Subjects (11%, N = 776) (nor in Switch (9%, N = 184) vs. Same Reference contexts (11%, N = 564)). Instead, the best split, indicated in Table 1, is between occurrences at distances of three or more clauses from the previous coreferential 1sg subject and occurrences at lesser distances. This distance effect on stressed I operationalizes the proposed function of speakers “bring[ing] the idea of themselves back into the active consciousness of the listeners” (Chafe 1994: 87).

5.2 Contrastive functions of stressed I

Stressed I is further distinguished by Turn position and Polarity, measures of interactional and semantic contrastive functions, respectively. First, occurrence in a turn-initial IU (or an IU following a minimal response, as in (7) (where I* indicates stress on I)), favors stressed I. The tendency toward higher stress in turn-initial position is maintained at distances of two or fewer clauses, indicating the independence, at least in part, of the turn-position effect from that of distance (subject continuity). It may therefore be interpreted as having to do with turn-taking and thus as operationalizing a contrastive function in an interactional sense (as with Spanish cognitive verbs, §3.1).

(7)
1. Lenore: .. So you have your own equipment,
2. but,
3. Lynne: .. (TSK) (H) No.
4. I* don't have my own equipment at all.

(01 Actual Blacksmithing 50-54)
Stressed $I$ is also more likely in denials, as in (7), than in affirmations. Negative polarity would be a measure of contrast as a semantic rather than interactional notion, in that negation, considered a “contrastive semantic element” (Sun and Givón 1985: 346) (cf. Myhill and Xing 1996: 342), has more to do with the content of what is said in a fairly narrow context than with the positioning of interlocutors (cf. Thompson 1998: 325). Evidence is that instances of negated $I$ are not found disproportionately in a turn-initial Intonation Unit, nor is turn-initial $I$ more likely to be negated, and the polarity effect is weaker in turn-initial position.

By these operationalizations of contrastive functions, Spanish $yo$ differs. The analysis of variable $yo$ expression does not provide evidence for $yo$ generally serving in contrastive functions, as neither Polarity nor Turn position achieve significance. Instead, the effect of Turn position is operative for cognitive verbs in Spanish (§ 3.1).

5.3 Lack of stress on $I$ in ‘cognitive’-verb prefabs

An evident conflict site with expressed $yo$ is the absence of a cognitive-verb class uniformly favoring stressed $I$. Four ‘cognitive’ verbs—$mean$, $think$, $know$, $guess$—constituting close to one-third of all 1sg-verb tokens (593/1,861), exhibit disparate rates of stressed $I$. The rate of stressed $I$ is 23% with $know$ ($N = 153$), but nil with $I$ $mean$ ($N=210$) and $I$ $guess$ ($N=41$) (which were excluded from the analysis in Table 1). $I$ $+$ $think$ ($N = 189$), the putative translation counterpart of $yo$ $creo$, displays a stressed $I$ rate identical to the average of 13%.

Instead of a uniform cognitive verb class, there are lexically particular constructions, with tendencies for lower or higher rates of stressed $I$. $I$ is generally unstressed in formulaic expressions constituting ‘chunks’ (Bybee 2010: 34). Counted as discourse-marker (non-clausal) uses of $I$ $know$, $I$ $think$, $I$ $remember$ and other collocations were those tokens that were produced prosodically independently from other clausal material, as in line 2 in (8), or as parentheticals, as in (9). Also generally unstressed is $I$ in quotatives $be$ $like$, $say$, $think$ introducing quoted speech or internal thought. These prefabs show a stressed $I$ rate of 4% ($N=110$).

(8) Discourse-marker prefabs
1. Pamela: $Well,$
2. $I$ $don't$ $know$,
3. $I$ $guess$ $it$ $must$,

(SBCSAE, 05 A book about death, 239)

(9) Discourse-marker prefabs
Corinna: $...$ $he's$ $gonna$ $divorce$ $her$ $I$ $think$.

(SBCSAE, 45 The classic hooker: 519)

On the other hand, $I$ $don't$ $know$ (CLAUSE) presents a higher than average stressed $I$ rate (27%, $N = 84$) even compared with other cases of negation and even when it is formulaic (as strictly defined above, for example, as in (8)). A second particular construction displaying a higher rate of stressed $I$ beyond the predictions of the general constraints identified in the multivariate analysis, is conditional $I$ $would$ (NEG) VERB.

In summary, with stressed $I$ we find the by now familiar coreferential subject
priming effect. Subject continuity works in terms of distance rather than local coreferentiality. A further difference from expressed yo is the role of contrastive functions in a semantic sense, as measured by a polarity effect, and interactionally, as per a speaker-turn effect. Finally, contributing to the structure of variation are lexically particular constructions. In Otheguy’s (2004: 177) words, we must “discard the excessively reductionist equivalence between Spanish oters and English stressed”.

6 Conclusion

Subject pronoun variants in Spanish and English are conditioned by parallel accessibility and priming effects, candidate cross-linguistic constraints on subject realization. However, we have found that for subject expression, both Spanish yo and English I, the subject continuity effect is one of local coreferentiality (whether in terms of Intervening Human Subjects or Switch Reference), while for English subject pronoun stress it is one of distance from the previous mention as subject. As concerns priming, we find an effect for coreferential subject priming, that is, the realization of the previous coreferential subject. Such a shared effect may not be surprising, if priming is ubiquitous in language variation, but is nevertheless important to include, because subject continuity effects are moderated by priming.

Having conducted these comparisons, we are now in a position to answer the question of whether Spanish subject expression in Spanish-English bilingual speech is an appropriate linguistic variable to ascertain convergence. Yes, since the linguistic conditioning of English I realization presents clear differences from that of yo expression. With respect to expressed I, the difference lies in the restriction of unexpressed subject pronouns to and-coordinating constructions and absolute prosodic-initial position. Grammatical differences with stressed I, in turn, are instantiated in the contributions of polarity and turn-position, which apply generally to English 1sg verbs but define a particular ‘cognitive’ verb construction in Spanish. We have shown, too, that particular constructions—from prefabs such as yo creo and I think, to more schematic units, such as (yo) + COGNITIVE VERB1SG and [I VERB1SG and Ø VERB-OF-SPEECH1SG]—may be profitably identified as loci of grammatical difference.

The analysis of intra-linguistic variability in this chapter has allowed us to distinguish superficial from structural inter-linguistic similarity, and calls into question assumptions of blanket equivalence between English and Spanish subject realization. In sum, while cognitive (subject continuity) and mechanical (priming) effects go in the same direction, prosodic and contrastive-function constraints provide “conflict sites” (Poplack and Meechan 1998b), as do language-specific constructions. Beyond ‘higher’ or ‘lower’ overall rates, the conflict sites in the linguistic conditioning of variant choice identified via systematic quantitative analysis will enable us to gauge grammatical (dis)similarity.
Appendix: Transcription Conventions (Du Bois et al. 1993)

. final intonation contour = lengthening
, continuing intonation contour [ ] speech overlap
? appeal intonation contour (H) in-breath
... medium pause (> 0.7 secs) (Hx) out-breath
.. short pause (about 0.5 secs) (TSK) click
-- truncated intonation contour (SNIF) speaker sniffs
- truncated word

References


