

Catherine E. Travis and Rena Torres Cacoullos

# What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation

**Abstract:** In languages with variable subject expression, or “pro-drop” languages, when do speakers use subject pronouns? We address this question by investigating the linguistic conditioning of Spanish first-person singular pronoun *yo* in conversational data, testing hypotheses about speakers’ choice of an expressed subject as factors in multivariate analysis. Our results indicate that, despite a widely held understanding of a contrastive role for subject pronouns, *yo* expression is primarily driven by cognitive, mechanical and constructional factors. In cognitive terms, we find that *yo* is favored in the presence of human subjects intervening between coreferential 1sg subjects (a refined measure of the well-described phenomenon of “switch-reference”). A mechanical effect is observed in two distinct manifestations of priming: the increased rate of *yo* when the previous coreferential first singular subject was realized as *yo* and when the subject of the immediately preceding clause was realized pronominally. And evidence for a particular *yo* + COGNITIVE VERB construction is provided by a speaker-turn effect, the favoring of *yo* in a turn-initial Intonation Unit, that is observed with cognitive (but not other) verbs, which form a category centered around high frequency *yo creo* ‘I think’.

**Keywords:** subject expression, variation, contrast, accessibility, priming, constructions, prefabs

---

**Catherine E. Travis:** E-mail: Catherine.Travis@anu.edu.au

**Rena Torres Cacoullos:** E-mail: rena@psu.edu

---

Both authors contributed equally to the work, though the order of authors is non-alphabetical. Thanks to Jenny Dumont and Leah Houle for assistance with the coding.

# 1 Introduction

In languages with variable subject expression, or “pro-drop” languages, when do speakers use subject pronouns? This topic has been widely studied and various cognitive, semantic, pragmatic and interactional factors have been called on to explain referent realization. However, such factors are often left without operational definitions and are applied without attention to differences between persons, syntactic roles and genres. In this paper, we tackle this issue through examining the use of the Spanish first-person singular subject pronoun (*yo*) in conversational data, testing hypotheses about speakers’ choice of an expressed subject as factors in multivariate analyses.

The paper is organized as follows. In the next section we present the corpus and variable context. In Section 3 we apply three operationalizations of contrast independently elaborated in previous literature, namely “double contrast” (Myhill and Xing 1996), juxtaposition of references to the speaker and interlocutor (Travis and Torres Cacoullos 2012), and Potential Referential Interference (Sun and Givón 1985), and find that none of these measures offers an account of *yo* expression in these data. Abandoning the attempt to individually classify tokens as contrastive, in Section 4 we use multivariate analysis and find that speakers are more likely to express *yo* in the presence of intervening human subjects between coreferential 1sg subjects, a cognitive effect; and when the previous coreferential first singular subject was realized as *yo* and when the subject of the immediately preceding clause was realized pronominally, two distinct manifestations of priming, a mechanical effect. In Section 5, independent multivariate analyses reveal that the linguistic conditioning of *yo* is distinct for cognitive verbs (e.g., *saber* ‘know’, *pensar* ‘think’), in particular in being sensitive to turn position, based on which we propose a particular *yo* + COGNITIVE VERB construction, which forms a category centered around high frequency *yo creo* ‘I think’. The actual usage patterns of *yo* thus do not support the widely held perception of a general contrastive function for expressed pronouns, and indicate instead a combination of mechanical, cognitive and construction effects as the major driving forces in this variation.

## 2 Corpus and variable context

The data are drawn from the Corpus of Conversational Colombian Spanish, a corpus of spontaneous conversations between spouses, families, and friends recorded in the city of Cali, Colombia, over a two-month period in 1997 (cf. Travis

2005a). For this study, we draw on four and a half hours of recordings (approximately 40,500 words). This comprises fourteen conversations (of between two and four participants), and 22 speakers (14 women and 8 men). All speakers are middle-class native Spanish-speaking Colombians, ranging between the ages of 20 and 60.

The data have been transcribed in accordance with the approach developed at the University of California, Santa Barbara (cf. Du Bois et al. 1993, see Appendix for transcription conventions). Fundamental to this approach is the notion of the Intonation Unit (IU), that is, “a stretch of speech uttered under a single coherent intonation contour” (Du Bois et al. 1993: 47), each of which is represented on a distinct line in the transcription (there are a total of 16,000 IUs in the material used for this study). This transcription method includes the annotation of features such as pauses, laughter, overlap, inhalation and so on, which may play a role in the interaction between the interlocutors.

Grammatical person has been found to be one of the strongest constraints on variable subject expression (cf. Otheguy, Zentella and Livert 2007) indicating that the different persons may exhibit different patterning and can be profitably analyzed independently (cf. also Silveira 2007; 2008, for Brazilian Portuguese). We choose to analyze first-person singular here for two reasons, (1) because it has only two realizations (pronominal *yo* or unexpressed), as compared with third person, which allows for lexical NPs, meaning that different issues arise in relation to priming effects; and (2) because it is first-person singular expression which is often thought to have a pragmatic or interactional function.

All tokens of first singular verbs were extracted, except for truncated tokens where the speaker cut off before completing the verb. We excluded from the analyses reported below all post-verbal tokens of *yo*, which represent just 5% of all expressed subjects (25/522), as post-verbal subject pronouns are unlikely to be subject to the same constraints as preverbal ones (e.g., López Meirama 2006: 38–46; Ocampo 1995: 443; Silva-Corvalán 2001: 165).<sup>1</sup>

Previous studies (e.g., Silva-Corvalán 2003) have excluded cases of morphologically marked contrast as non-variable contexts, such as with the emphatic *mismo*, lit. ‘same’ (e.g. *yo mismo* ‘I myself’), or *sí*, lit. ‘yes’ (e.g. *yo sí* ‘I yes’). The former did not occur in our data; (*yo*) *sí* did occur (N=11), and showed variable expression and thus was included in this study. Cases of unmarked contrast have similarly been excluded in past studies as contexts in which expressed subjects are presumed to be obligatory (e.g., Cameron 1992:

---

1 *DECIR + yo* may represent a set of formulas shaping patterns of postverbal *yo* (cf. Benevento and Dietrich 2011; Posio 2011: 794–795; Torres Cacoullos and Travis 2011: 249).

85; Silva-Corvalán 1982; 2003: 850; Toribio 2000), but it has also been noted that both expressed and unexpressed subjects occur in such environments (Amaral and Schwenter 2005; Otheguy et al. 2007: 775–776). It is such cases of unmarked contrast that we will attempt to identify in the following section. Following these protocols, a total of 1,020 tokens of first-person singular subjects was retained for quantitative analysis, for which the rate of expression was 49%.

### 3 Operationalizing and testing contrast

There is a widespread understanding that a function of expressed subjects in Spanish is to mark “contrast”, “emphasis” or “focus” (e.g., Aijón Oliva and Serrano 2010: 9; Butt and Benjamin 2004: 130; Gili y Gaya 1964: 23–34; Luján 1999: 1277; Posio 2011: 778). Silva-Corvalán, for example, writes that “la función contextual del pronombre sujeto es la de indicar al interlocutor que preste atención al referente del sujeto. No sorprende entonces que el pronombre represente entidades contrastivos, focales o topicales” (‘the contextual function of the subject pronoun is to indicate to the interlocutor to direct their attention to the subject referent. It is not surprising, then, that the pronoun represents contrastive, focal or topical entities’) (2003: 852). Indeed, this view is not restricted to Spanish, but is common cross-linguistically. According to Chafe (1976: 37), “in some languages, where the role of given nouns is captured primarily through agreement in the verb, independent pronouns appear to be used mainly to express a focus of contrast” (cf. also Givón 1983: 18; Levinson 1987: 384).

Despite this widespread view, rarely has the notion of contrast been operationally defined and tested, a problem that is recognized in the literature (e.g., Mayol 2010: 2497; Myhill and Xing 1996: 308; Silva-Corvalán 2003: 851). This lack of an operational definition renders analysts’ classification of instances of contrast nonreplicable, as illustrated in the following example. Here we have a case which could perhaps be considered “implicit” contrast (cf. Mayol 2010: 2501; Myhill and Xing 1996: 320), where María uses *yo no creo* ‘I don’t think’ to introduce what appears to be a difference of opinion with Angela. María’s response to Ángela’s comment that the referent should be studying at the Javeriana University suggests that María has understood Ángela to imply that she (the referent) is a good student, an assumption which María does not share. But classifying such tokens as contrastive, without independent evidence present in the discourse (except the *yo* itself), runs the risk of circular reasoning.

(1)

Ángela: *ese milagro que no está en la .. Javeriana.*

María: .. *Ah.*

... **Yo no creo** que ella es como muy buena estudiante no?<sup>2</sup>

(Calima1 670)

Ángela: ‘that miracle that she’s not at Javeriana (University).’

María: ‘.. Oh.

... **I don’t think** that she is like a very good student right?’

Our first step in the analysis is to attempt to identify contrastive vs. non-contrastive tokens in a replicable manner, and to assess whether the former do indeed favor expressed *yo*.

Contrast may be viewed as more of a semantic, interactional or cognitive notion.<sup>3</sup> Three studies that have sought to operationalize these different notions of contrast are Myhill and Xing (1996) based on the notion of “double contrast”, or pairs of clauses that differ in two ways, Travis and Torres Cacoullós (2012) based on the juxtaposition of speaker and interlocutor self-reference, and Sun and Givón (1985) based on “Potential Referential Interference”, or the mention of semantically compatible referents in the immediately preceding discourse. In order to determine to what degree the notion of contrast defined in these three ways is a determinant of speaker choice, we test how well these models account for *yo* expression in the corpus under study, as we describe below.

### 3.1 Double contrast: A semantic approach to contrast

The notion of “double contrast” was originally proposed, though not operationalized, by Chafe (1976: 35), and a very similar notion has been independently applied to work on Spanish subject expression (e.g., Cameron 1992: 88–94; Mayol 2010: 2499; Silva-Corvalán 2001: 174; Silva-Corvalán 2003: 853). A quantitative implementation of “double contrast” is found in Myhill and Xing (1996: 314), who set out to test the role it plays in object position in Chinese narratives. They operationalize “double contrast” in the following way: “in contrastive pairs there

<sup>2</sup> Note use of the indicative in the subordinate clause following *yo no creo*, where the subjunctive would be prescribed (cf. Thompson 2002, on the status of expressions such as ‘I don’t think’ as speaker stance frames, rather than “main clauses”).

<sup>3</sup> We apply the labels “semantic”, “interactional”, and “cognitive” merely to distinguish different approaches to contrast as based, respectively, more on the content of what is said in fairly narrow context, the positioning of interlocutors, and information flow in discourse.

are two or more elements which are different in two clauses (either verbs with opposite meanings or nonverbal elements in a set relationship)”.

We apply this operationalization here, classifying as manifestations of “double contrast” pairs of clauses that differ in two ways; in order for *yo* to be contrastive, one difference must be in the subjects, and the second in the predicates, which must be related, but in some sense converse, e.g. negated (*believe / not believe*), doing vs. not doing (*take along / leave behind*), opposite direction (*take/give*), and so on. Furthermore, the two clauses must be no more than three clauses apart (a slightly stricter measure than that of Myhill and Xing (1996: 319) of six clauses, but broader than Silva-Corvalán (2003: 853) and Cameron (1992: 86) who look just at consecutive clauses).

The following two examples illustrate our application of this operationalization. In (2), the subjects *yo* ‘I’ / *la otra* ‘the other’ (referring to the speaker’s wife) contrast, as do the predicates *despierto* ‘awake’ / *dormida* ‘asleep’. In (3), David’s assertion that “he” doesn’t litter anywhere is contrasted with Milena’s assertion that “one” doesn’t dare litter on the farm of the very environmentally aware friend they are discussing.

(2)

David: ... *Porque cuando yo estaba= --*

.. **despierto,**

*la otra estaba dormida.*

(Campaign 533)

‘... Because when **I was --**

**awake,**

the other was asleep.’

(3)

Milena: *A la finca de él,*

*uno no se atreve a botar un papel,*

*no es cierto?*

*Porque es como tan XX --*

.. *organizadita,*

*como,*

*todo tan --*

Rocío: *Hm.*

Milena: .. *Tan lindo ahí.*

*Y --*

David: .. ***Yo no boto en ninguna pa-*** --

(Contamination 1045)

Milena: ‘On his farm,

one daren’t even throw away one paper,

right?’

Because it's all so XX --  
 organized,  
 like,  
 everything so --'

Rocío: 'Hm.'

Milena: '. . So pretty there.  
 And --'

Davíd: '. . **I don't litter** anyw- --'

Notwithstanding the occurrence of some such apparently clear-cut examples, there were many issues in attempting to apply this operationalization to first-person singular subjects in the conversational data we are working with. Particularly problematic is the fact that “double contrast” includes tokens that do not intuitively seem to be contrastive. Example (4) meets the criteria in the operationalization in that we have distinct subjects (1pl vs. 1sg) and distinct direct objects (*areperías* vs. *sandwich cubano*). However, within the broader context of this conversation, which is about good places to go and eat in Cali, it seems that rather than contrasting where the speaker has eaten with where he and his wife have eaten, he is listing the different places he has been.

(4)

Fabio:  $\emptyset$  *Hemos comido de todo.*

. .  $\emptyset$  *hemos repetido areperías,*  
*sandwich cubano,*

**$\emptyset$  repetí varias veces.**

(Comida 540)

'(We) have eaten everything.

. . (We) have been back to *areperías,*

Cuban sandwich,

**(I) went back** to several times.'

Since the 1sg and 1pl subjects in (4) are partially coreferential, it could be argued that they should not be considered distinct. In (5), we have another case of 1sg and 1pl subjects, in this case with no overlap in referentiality, yet here there is another issue in applying this operationalization, which is in determining what constitutes a “converse predicate”: should *estamos aquí* ‘we are here’ and *voy pa' allá* ‘I will go there’ be interpreted as converse?

(5)

Davíd: <VOX No,

*es que  $\emptyset$  estamos aquí,*

*que VOX> --*  
*.. <VOX Ah,*  
*chévere,*  
*Ø ya voy pa' allá VOX>.* (Campaign 692)  
 '<VOX No,  
 it's that (we) are here VOX>,  
 .. <VOX Oh,  
 great.  
 (I)'ll go there now VOX>.'

Despite these complexities, we applied this operationalization to 162 tokens, and of these, found only 6 (or 4% of the data, 6/162, and 6%, 5/81, of all expressed 1sg subjects) that qualified as contrastive. Five of these six tokens occurred with an expressed *yo*, a rate which is markedly higher than the overall rate, suggesting that there may be a correlation between “double contrast” and subject expression. However, tokens meeting this operationalization are so few that it offers little in the way of a general account of subject expression in Spanish.

### 3.2 Speaker-interlocutor self-reference: An interactional approach to contrast

As another exploration of the potential manifestation of contrast in conversation, we now consider a more interactional approach, taking into account the speaker's role in relation to their interlocutor. An interactional contrastive function has been ascribed to expressed *yo*, of highlighting, or focusing, the role of the speaker (cf. Silva-Corvalán 2001: 166). This notion may be operationalized by considering the juxtaposition of speaker and interlocutor self-reference, as measured by the grammatical person and coreferentiality of the subject of the immediately preceding clause (cf. Travis and Torres Cacoullos 2012). If the rate of first-person singular subject expression is found to be higher when the preceding subject is a non-coreferential first- or second-person singular, this would support an interactional function. The more frequent possibilities for the preceding subject when it has a human referent are a 1sg, coreferential or non-coreferential (line 3, (6)) below; a 2sg, coreferential or non-coreferential (line 1, (6)); a 1pl, coreferential or non-coreferential; and a non-coreferential third person.<sup>4</sup>

<sup>4</sup> Though a coreferential 3<sup>rd</sup> person subject is possible in conversations of more than two people (where one interlocutor talks about another to a third party), there were only three such cases of the 139 tokens of 3<sup>rd</sup> person human preceding subjects.



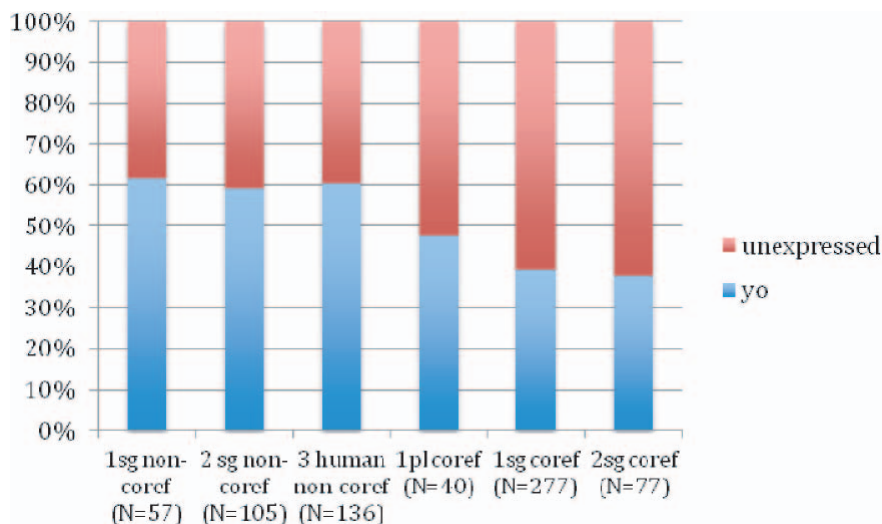


Fig. 1: Rate of *yo* expression by grammatical person and coreferentiality of preceding clause human subject (N = 692)

(6)

1. Ángela: *Ø te acordás que yo quería comprar un [congelador]?*
  2. Santi: *[Pues],*
  3. *yo pienso que el congelador de pronto no,*
  4. *mami --*
  5. Ángela: *yo también pienso que no.* (Almuerzo 245)
- Ángela: ‘(you) remember that **I wanted** to buy a [freezer]?’  
 Santi: ‘[Well],  
**I think** that the freezer maybe not,  
 honey --’  
 Ángela: ‘**I also think** not.’

Figure 1 gives the distribution of *yo* expression according to the subject of the immediately preceding clause, for human subjects.

If *yo* is contrastive in an interactional sense, we would expect a higher rate of expression when self-references to the speaker and interlocutor are juxtaposed (that is, with a preceding 1sg or 2sg) than when the preceding subject is in third person. What we find is that in coreferential contexts – whether the subject of the immediately preceding clause is a coreferential 1sg or 2sg – *yo* is expressed approximately 38% of the time, whereas in non-coreferential contexts – whether the

subject of the immediately preceding clause is a non-coreferential 1sg, 2sg or even 3<sup>rd</sup> person – *yo* is expressed nearly twice as often, approximately 60% of the time. Thus, rather than an interactional manifestation of contrast, consistent with a function of “emphasizing” or “focusing” the role of the speaker vis-à-vis the interlocutor, these results indicate a local coreferentiality effect.

### 3.3 Potential Referential Interference (PRI): A cognitive approach to contrast

An alternative approach to the notions of double contrast and speaker-interlocutor juxtaposition is a more cognitive view of contrast, based on how many referents the listener needs to retain in their consciousness (cf. Chafe 1994). We label considerations of accessibility and continuity of reference in discourse “cognitive” (cf. Ariel 1994), understanding cognitive factors as ones having to do with “information on states of affairs – on *what* is being said rather than on the manner or style of expression” (Labov 2010: 2).<sup>5</sup> A measure of accessibility proposed by Givón (1983) is “Potential Referential Interference” (PRI), defined by Sun and Givón as “the number of other referents in the directly preceding discourse environment – most commonly 3 clauses – that are semantically compatible with the predicate of the referent under consideration” (1985: 331) (cf. also Clancy 1980: 143). Sun and Givón note that PRI “is a good diagnostic for contrastive/emphatic constructions, since what is contrasted is normally a referent in the directly preceding discourse environment” (1985: 341).

In order to test whether PRI accounts for *yo* expression, we coded 446 tokens extracted from the data for the number of human referents that are semantically and pragmatically compatible with the subject of the target clause in the three clauses immediately preceding the target mention.<sup>6</sup> Singular and plural referents in any syntactic role were included, following Sun and Givón (1985), as were referents mentioned by any of the conversational participants, essential in this interactive data (an issue not addressed by Sun and Givón in their study based on narrative data). A ‘you’ coreferential with ‘I’ was not counted as “potentially interfering”, but a non-coreferential ‘you’ or ‘I’ was, as was ‘we’, regardless of whether it included the target 1sg subject since, even in such a case, it does not have identical reference to the target subject. Not considered to be potentially interfering referents were non-referential NPs (e.g., 3 pl impersonal subjects, *En*

<sup>5</sup> Such considerations have also been viewed as pragmatic (e.g., Levinson 1987).

<sup>6</sup> Out of a total of 562 tokens extracted, 116 could not be coded for PRI, either because the content was not clear, or because the target token or a preceding token occurred in quoted speech.

*esa novela que **pasaban, mostraban** mucho eso.* ‘In that soapie that (they) played, (they) showed that a lot.’), impersonal pronouns (e.g., **tú nunca vas a enriquecer tu vocabulario** ‘you will never enrich your vocabulary’, re language classes that involve only repetition), fixed expressions (e.g., *será que* ‘it might be that’, *es que* ‘the thing is that’, *oyó* ‘you hear’), indefinite pronouns (e.g., **quién sabe** ‘who knows’); subjects of imperatives (many of which are discourse markers, e.g. *ve, mira* ‘look’, *venga* ‘come on’); and vocatives. Finally, clauses produced in overlap with the target clause were not counted as part of the preceding three clauses.<sup>7</sup>

The following example illustrates the application of this protocol, where the target tokens appear in lines 6 and 7, and are marked in bold. In line 6, the preceding three clauses (in lines 2, 4 and 5) present two “potentially interfering” referents (both in line 2): *nos* ‘us’, as indirect object, which refers to both Santi and Ángela; and the unexpressed subject of *dar* ‘give’, which refers to Ángela’s mother. *Quién* ‘who’, as an indefinite pronoun, is not considered potentially interfering. The PRI score, then, is two. For the token in line 7, we disregard Ángela’s *yo creo que sí* ‘I think so’ because it was produced in overlap and thus the three preceding clauses are identical to those of *yo* in line 6. In this case, there is one more potentially interfering referent, which is Ángela’s *yo* in line 4 (*yo no sé* ‘I don’t know’), and thus we have a PRI score of three.

(7)

1. Santi: *Cuánto nos irá a dar tu mamá,*
  2. *será que Ø nos da --*
  3. *.. [quinientos o qué].*
  4. Ángela: [*<WH Yo no sé WH>*].
  5. *... Quién sabe.*
  6. *.. [**Yo creo que sí**].* PRI = 2
  7. Santi: [***Yo creo***] -- PRI = 3 (Almuerzo 307)
- Santi: ‘How much might your mum give us,  
might (she) give us --  
.. [five hundred (pesos)].’
- Ángela: [*<WH ‘I don’t know WH>*].  
... Who knows.  
.. [**I think so**].’
- Santi: ‘[**I think**] --’

<sup>7</sup> Clauses produced in overlap with each other, but not with the target clause, were counted as distinct clauses.

Example (8) illustrates zero PRI: the only human referents appearing in the three clauses preceding the target *yo* in line 6 are a coreferential second person subject, produced by the interlocutor (line 5) and a coreferential first-person subject (line 4). The non-human subject in line 1 is not semantically compatible.

(8)

1. Ángela: *Está más bueno ese libro.*
  2. . . . *El caballo de Troya.*
  3. @@
  4. Santi: ***Yo ayer no leí.***
  5. Ángela: *Ø No has [leído]?*
  6. Santi: ***[Yo caí] -- (muerto)*** (Almuerzo 419)
- Ángela: 'It's so good that book.  
 . . The Trojan Horse.  
 @@'  
 Santi: '**I didn't read** yesterday.'  
 Ángela: '(You) haven't [read]?'  
 Santi: '**[I fell] --**' (fast asleep)

The degrees of PRI that occur in the data range from zero to four. Tokens with zero-PRI make up close to one third of the data, tokens with just one intervening referent make up close to one half, and those with a PRI score of two constitute another fifth. If expressed *yo* were to meet this operationalization of “contrastive”, we would expect to see a higher rate of *yo* expression with a higher PRI score. This is not the case, however, as seen in Table 1. There is in fact no predictable change in the rate of *yo* as the number of intervening referents increases (given on the left half of the table). Nor is there a correlation between the rate of *yo* and the binary distinction between absence and presence of any interfering referent (given on the right half of the table), which is the crucial distinction, according to Sun and Givón (1985: 341). When we compare tokens with no interven-

| PRI | % expressed <i>yo</i> (N):<br>different degrees of PRI | % expressed <i>yo</i> (N):<br>presence vs. absence of PRI | % data |
|-----|--|---|--------|
| 0   | 51% (70/138)   | 51% (70/138)  | 31%    |
| 1   | 43% (86/200)   |   | 45%    |
| 2   | 60% (58/96)  | 48% (148/308)   | 22%    |
| 3–4 | 33% (4/12)   |   | 3%     |

**Table 1:** Expressed *yo* according to PRI (Potential Referential Interference) in the 3 preceding clauses (N = 446; rate expressed *yo* = 49%)

ing referent (PRI = 0) with those where there is one or more, we still fail to obtain a significant difference ( $p = 0.61$ ) and the direction is contrary to prediction, with a marginally lower rate of expressed *yo* with a positive PRI score (51% vs. 48%).

How might we account for the difference between these results and those of Sun and Givón (1985), who did find an effect of PRI in their study of preverbal objects in Biblical Hebrew narratives? It is striking that the incidence of PRI was much lower for objects in narrative, at approximately 15% (113/943 for the written and 48/320 for the spoken data) (1985: 343), than for our 1sg subjects, of which a little over two thirds had potentially interfering referents in the preceding clauses (308/446).<sup>8</sup> One issue may be the genre, as it may be that monologic narratives, with greater topic continuity, have lower rates of PRI than interactive conversation, where topics change rapidly and frequently (cf., Travis 2007). It may also be related to the notion of semantic compatibility and person: while first person may be compatible with all referential, human subjects, for third person, a distinction must be made between human and non-human referents, which may limit the pool of potentially interfering referents. But perhaps most important is the issue of subjects vs. objects, as objects are more constrained than subjects by the verb with which they occur (and are considered “internal” to the Verb Phrase in formalist models). Examples such as “knocking (on) the door”, “learning stuff”, “cut the boneless part” from Sun and Givón (1985) are all cases where a relatively limited set of referents could fill the object slot compared with what could fill the subject slot. What this demonstrates is that care needs to be taken in comparing across genres, persons and syntactic roles, and that a PRI measure may need to be adapted in order to provide an appropriate measure for the variable under study. Below (Section 4.1) we will propose one adaptation, focusing on just human subjects occurring between coreferential 1sg subjects, which we do find to play a significant role in *yo* realization.

### 3.4 The failure of contrast accounts

To summarize this section: of the three operationalizations of contrast found to have an effect on other linguistic variables, double contrast (Myhill and Xing 1996) applied to a miniscule portion of the data, while speaker-interlocutor self-reference (Travis and Torres Cacoullos 2012) was applicable but ineffectual in accounting for *yo* expression rates, as was PRI (Sun and Givón 1985). In short, the

---

<sup>8</sup> A higher rate of PRI for 1sg subjects in conversation appears not to be unique to these data: applying PRI to stressed vs. unstressed / in English, Travis and Torres Cacoullos (2012) found that 70%, 330/470, of the analyzed tokens of / had a PRI score of one or more.

attempt to assess the role of contrast in subject expression based on ascribing a global contrastive vs. non-contrastive function to each token of expressed and unexpressed *yo* failed.

This failure is consistent with other studies that have attempted to quantify notions of “emphasis” or “contrast”. Paredes Silva (1993: 41–43), whose criteria for contrastiveness included verbs with opposite meanings, negative versus affirmative, and different complements for the same verb, found that such “emphasis” favored pronoun expression, but that just 12% (204/1,650) of 1sg subjects in Brazilian Portuguese letters met the criteria. Morphologically marked contrast appears to be similarly infrequent in spontaneous discourse. Bentivoglio (1987: 46–48) counted instances of 1sg and 1pl marked with *mesmo*, connectives such as *pero* ‘but’, *mas* ‘yet’, *sin embargo* ‘however’, or adverbs such as *justamente* ‘precisely’, in a corpus of sociolinguistic interviews. Though the subject pronoun rate for these contrastive tokens was 56% (20/36), compared with 39% (337/856) in the absence of such markers, contrastive tokens thus classified accounted for only 4% of her data (36/892).<sup>9</sup>

If expressed pronouns are not contrastive in general, the question remains of just what they do in discourse. In order to answer this question, we adopt the variationist method, relying on evidence in the linguistic context of each token to operationalize hypotheses about what motivates variant choice (cf. Labov 1966; Sankoff 1988b).

## 4 The structure of variable 1sg pronoun expression

The variationist method seeks to account for speakers’ choices among different forms serving a similar discourse function (Labov 1969; Sankoff 1988a), here, expressed and unexpressed first-person singular subjects, which both serve the general function of indexing the speaker. Linguistic variability is structured, conditioned by elements of the linguistic and extralinguistic context, which contribute to speaker choices among variant forms. The more particular function(s) of variant forms (such as cognitive, semantic or interactional functions), or the lack thereof, are quantitatively observable in the **linguistic conditioning** of variant selection, that is, probabilistic statements about the relative frequency of occurrence of variant forms and elements of the linguistic context in which they

<sup>9</sup> In these data we find no indication of a higher rate of *yo* with the potentially “contrastive” conjunction *pero* ‘but’: *yo* is expressed at a similar rate in clauses conjoined with *pero* (52%, 15/29), with *y* ‘and’ (46%, 28/61) and in main clauses without a conjunction (49%, 400/822).

appear (Poplack and Malvar 2007: 137–143; Poplack and Tagliamonte 2001: 88–94; Silva-Corvalán 2001: 133–138). Here we operationalize hypotheses about the function(s) of expressed *yo* based on contextual elements, which define factor groups. Building on the findings presented thus far, as well as findings from previous studies of subject expression, we consider seven factor groups (predictor variables) for multivariate analysis: (1) intervening human subjects, (2) verb class and (3) tense, (4) realization of previous coreferential subject and (5) of immediately preceding subject, (6) polarity, and (7) turn position.

Table 2 presents the results from a multivariate (Variable-rule) analysis using Goldvarb Lion (Sankoff, Tagliamonte and Smith 2012).<sup>10</sup> In this and following tables, the “input” (here .48) indicates the overall likelihood that expressed *yo* will occur (the overall rate of expression is 49%). The factor groups that jointly account for the largest amount of variation in a significant way are listed in bold in the table: here, semantic class of verb, intervening human subjects, realization of previous coreferential 1sg subject and of the subject of the immediately preceding clause, and verb tense. Shown in the first column are the probabilities, or factor weights: factors (contexts) with a higher probability (closer to 1) can be said to *favor*, and those with a lower probability (closer to 0) to *disfavor*, *yo* expression, or conversely, to favor unexpressed. The table also shows in subsequent columns, for each factor, the rate of expressed *yo*, the number of tokens, and the percentage of the data the factor makes up.

The results for the factor group appearing first on the table replicate the strong effect of semantic class of verb, reported in numerous previous studies across different varieties (e.g., Bentivoglio 1987: 60; Enríquez 1984: 240; Silva-Corvalán 1994: 162; Torres Cacoullós and Travis 2011: 252; Travis 2007: 116–117). Cognitive verbs favor *yo* expression and show the highest factor weight at .66. This raises the question of whether there is a class of cognitive verbs that behaves differently from others. We return to this question in Section 5, where we discuss in detail the behavior of the cognitive verbs as a class in comparison with other (non-cognitive) verbs.

Other environments in which *yo* is favored are when there are intervening human subjects between coreferential mentions; when the previous coreferential subject was realized pronominally; when the subject of the immediately preceding clause was realized pronominally; and with backgrounded tense-aspect-moods. We do not observe a significant effect for polarity, nor for turn position. Below we discuss each of these results individually.

---

<sup>10</sup> Variable-rule analysis uses logistic regression to perform binomial multivariate analysis for a choice of the “1” variant (here, expressed 1sg pronoun *yo*) vs. the “0” variant (unexpressed) (Sankoff 1988b).

| N = 1,020; Input: .48 (Overall rate: 49%)  |       |             |     |        |
|--|-------|-------------|-----|--------|
|  | Prob  | % <i>yo</i> | N   | % data |
| <b>Semantic class of verb</b>  |       |             |     |        |
| Cognitive  | .66   | 67%         | 201 | 20%    |
| Other (motion, copula, speech, perception, other)  | .46   | 44%         | 819 | 80%    |
| <b>Intervening human subjects</b>  |       |             |     |        |
| Present  | .56   | 55%         | 519 | 59%    |
| Absent   | .42   | 40%         | 365 | 41%    |
| <b>Realization of previous coref 1sg subject*</b>  |       |             |     |        |
| Pronoun ( <i>yo</i> )  | .58   | 56%         | 372 | 53%    |
| Unexpressed  | .41   | 34%         | 330 | 47%    |
| <b>Realization of subject of immed. preceding clause</b>   |       |             |     |        |
| Personal pronoun   | .57   | 56%         | 238 | 24%    |
| Other (non-human, lexical, impersonal, etc.)   | .50   | 53%         | 409 | 39%    |
| Unexpressed  | .46   | 40%         | 369 | 37%    |
| <b>Tense-aspect-mood</b>   |       |             |     |        |
| Imperfect (backgrounded)   | .62   | 61%         | 116 | 13%    |
| Present  | .48   | 50%         | 573 | 66%    |
| Preterit (foregrounded)  | .48   | 46%         | 186 | 21%    |
| <b>Polarity</b>  |       |             |     |        |
| Negative   | [.56] | 59%         | 216 | 21%    |
| Positive   | [.49] | 46%         | 795 | 79%    |
| <b>Turn position</b>   |       |             |     |        |
| Turn-initial intonation unit   | [.49] | 50%         | 332 | 33%    |
| Non-turn-initial intonation unit   | [.51] | 48%         | 686 | 67%    |
| Non-significant factor groups (probabilities given in square brackets): Polarity, Turn position.   |       |             |     |        |
| * Realization of previous coreferential 1sg subject applies to coreferential subjects that occur at a distance of two or fewer intervening human subjects. |       |             |     |        |

**Table 2:** Variable-rule analyses of the contribution of factors selected as significant to the choice of expressed *yo* in conversational Colombian Spanish

## 4.1 Intervening human subjects

One of the most robust findings regarding Spanish subject expression relates to switch reference, with expressed subjects being consistently found to be favored in contexts where there is a change in subject from the preceding clause, and disfavored where there is no such change (cf. Bentivoglio 1987: 37; Cameron 1994: 32; Otheguy et al. 2007: 789; Silva-Corvalán 1982: 104; Torres Cacoullós and Travis



2011: 251; Travis 2007: 120, *inter alia*). This result is consistent with the widely discussed notion of accessibility (Ariel 1990: 73; Chafe 1994: 75; Givón 1983: 18), whereby cross-linguistically we find less coding material (in this case, unexpressed subjects) in contexts of greater accessibility (such as when the referent has been recently mentioned), and more coding material (in this case, pronouns) in contexts of lesser accessibility. It is this concept that Sun and Givón (1985) built on in developing the PRI measure discussed above (Section 3.3), which, however, we found did not have an effect on *yo* expression.

Here, we reconfigure Potential Referential Interference (PRI), and propose that for 1sg subjects in conversation, a more appropriate measure of accessibility than the presence of semantically compatible referents in any syntactic role is the number of semantically compatible **intervening subjects** between coreferential mentions. Given that subjects tend to be more topical in that they tend to have greater continuity of reference (Givón 1983: 22), they may “interfere” more with other subjects than objects do. And secondly, rather than an arbitrarily chosen measure of three clauses, potentially interfering referents occurring between coreferential mentions may have a greater effect on accessibility.

The following example illustrates the difference between switch reference, PRI and our proposed measure of Intervening Human Subjects. The target token in line 6 occurs in a switch reference context, and in terms of PRI, in the three clauses preceding the target token (in lines 1, 3 and 5), there is one potentially interfering referent, the 2sg indirect object in line 3, *te*. However, if we are to consider only the subjects between the target clause in line 5 and the previous coreferential subject in line 1, there are no potentially interfering subjects. Thus, although this token occurs in a non-coreferential context and with a positive PRI score, it has an Intervening Human Subjects score of 0.

(9)

1. Ángela: *Ø Voy a grabar.*
  2. Sara: ... [*Sí*]?
  3. Ángela: [*Si no*] *te importa.*
  4. Sara: [*2@@@2*]
  5. Ángela: [*2Pero no es por nada2*],
  6.           *ahorita Ø te explico,* (Insurance 6)
- Ángela: ‘**(I) am going to record.**’  
 Sara: ‘... [Yes]?’  
 Ángela: ‘[If you don’t mind].’  
 Sara: ‘[*2@@@2*]’  
 Ángela: ‘[*2It’s nothing2*],  
**(I)’ll explain to you** in a minute.’

| Intervening human subjects | % expressed <i>yo</i> (N)<br>Coreferential (0 interv. clauses) | % expressed <i>yo</i> (N)<br>Switch reference (1+ interv. clauses) |
|----------------------------|--|--|
| 0 (absent)                 | 39% (113/289)  | 41% (31/76)  |
| 1+ (present)               | No cases   | 54% (283/519)  |

**Table 3:** Expressed *yo* according to switch reference and the presence of human subjects intervening between coreferential 1sg mentions (N = 884)

To test this new measure of accessibility, we coded for the number of intervening human subjects between coreferential mentions.<sup>11</sup> Table 2 shows that expressed *yo* is favored (.56) in contexts where there are other human subjects intervening between coreferential mentions, as in line 5 in (6), and disfavored – or, conversely, unexpressed is favored (.42) – when there are not, as in (9).

While the presence of intervening human subjects and switch reference largely overlap – nearly 80% (289/365) of the tokens having no intervening human subjects are coreferential – Intervening Human Subjects provides a better account of variable *yo* expression than does coreferentiality. Table 3 shows that when there are no intervening human subjects between coreferential mentions, the difference between coreferential and switch reference contexts vanishes (with rates of expression of 39% and 41% respectively). On the other hand, when we consider switch reference contexts, we do find a significantly higher rate of *yo* expression if there is a human subject in the intervening clause(s) than if there isn't (54% vs. 41% respectively,  $p = .04$ ).<sup>12</sup>

The following example illustrates this phenomenon. Only *te digo* '(I) will tell you' (line 3) occurs in what is traditionally considered a coreferential context, while *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 none of these tokens have any human subjects intervening between coreferential mentions, as the two "switch reference" tokens are preceded by a clause with a non-human subject. (And all three clauses appear without a subject pronoun.)

**11** We have excluded tokens for which this is not identifiable, including those with unclear material between coreferential mentions, those occurring in quoted speech and those that had over ten clauses intervening between coreferential mentions, as well as first mentions of a referent in the transcript.

**12** While with the measure of Intervening Human Subjects between coreferential mentions the pertinent distinction is presence versus absence, with that of intervening clauses (any non-coreferential subject) there may be a graded effect (in accordance with the notion of accessibility). We observe a *yo* rate of 48% (65/135) with one intervening clause, of 52% (150/287) with between two and nine, and of 56% (142/254) at distances of ten or more.

(10)

1. Santi: *Yo me averiguo,*
  2. *y,*
  3. *Y Ø te digo si hay algún apartamento,*
  4. *Entonces,*
  5. *Ø lo miro,*
  6. *.. si tiene alcoba del servicio,*
  7. *.. entonces te --*
  8. *.. Ø Te aviso.* (Pizza 1369)
- ‘I’ll find out,  
and,  
And (I)’ll tell you if there’s an apartment,  
So,  
(I)’ll look at it,  
.. if it has a service room,  
.. so --  
.. (I)’ll let you know.’

In summary, in exploring issues related to accessibility, both in Spanish and cross-linguistically, a large amount of attention has been dedicated to patterns of switch reference. Other factors that have been described as playing a role in “accessibility” include the distance between coreferential mentions, or the related measure of “discourse connectedness” (Balasch 2008; Bayley and Pease-Alvarez 1997: 361; Paredes Silva 1993; Travis 2007: 120); “referential importance”, or topicality (Chafe 1994: 88–91; Clancy 1980: 178); and episode boundaries (Clancy 1980: 171; Comajoan 2006; Fox 1987) (see Ariel 1990: 22–30; Blackwell 2003, Ch 3, for an overview of these and other such factors). Our study has shown that one factor that plays a role in the degree of accessibility of a subject referent is the presence of another human subject between coreferential mentions, and indeed it is this, rather than switch reference in and of itself, that has an effect on *yo* expression. We hope that future studies of subject expression more broadly test this measure of accessibility.

## 4.2 Priming: Realization of previous coreferential subject and of subject of immediately preceding clause

With these two factor groups we test for a structural priming or perseveration effect, whereby the use of a certain structure in one utterance functions as a prime

on a subsequent utterance, such that that same structure is repeated (cf. Bock 1986; Labov 1994: 547–568; Poplack 1980; Scherre and Naro 1991; Weiner and Labov 1983, *inter alia*).

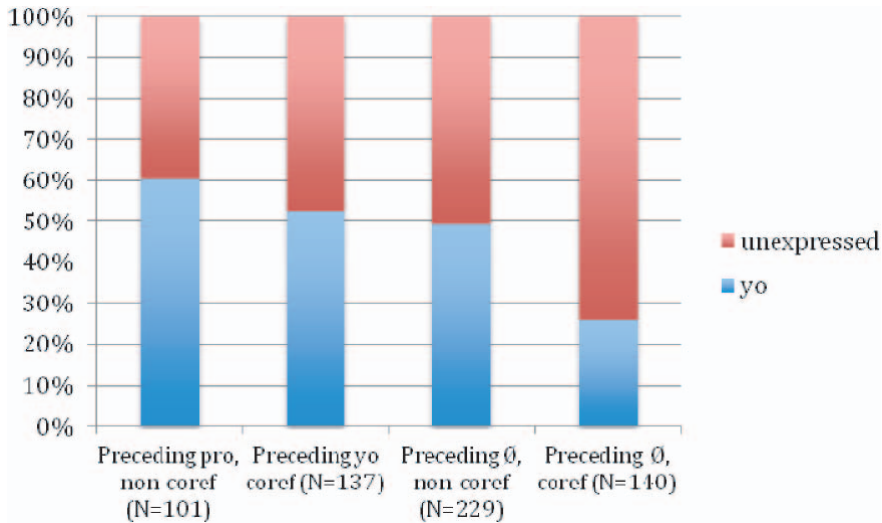
A priming effect for the form of the previous coreferential subject has been observed for subject expression in several studies of Spanish (cf. Cameron and Flores-Ferrán 2003: 50–54; Torres Cacoullos and Travis 2011: 251–252; Travis 2007: 120–121), whereby a preceding coreferential pronominal mention favors a subsequent pronominal mention, and a preceding unexpressed mention favors a subsequent unexpressed mention. The factor group “Realization of previous coreferential 1sg subject” measures this effect for coreferential subjects that occur at a distance of two or fewer intervening human subjects.<sup>13</sup> The factor weights for the two factors, Pronoun (*yo*) and Unexpressed, in Table 2 demonstrate precisely what is expected: speakers have a strong tendency to repeat the form of the previous coreferential subject.

One question that arises here is whether this is a structural or a lexical effect, where one *yo* favors another *yo* – a *yo-yo* effect, as it was dubbed by Travis (2005b). While structural priming does not rely on lexical repetition, psycholinguistic studies have found that it can be enhanced by lexical repetition (e.g., Pickering and Branigan 1998). Note, though, that we observe not only that previous *yo* favors *yo* (.58), as for the token in line 6 in example (7), but the converse as well, that is, previous coreferential unexpressed favors unexpressed, with the lowest factor weight (.41), as illustrated in lines 5 and 8 in (10), an effect that cannot be considered lexical (as there is no lexical content).

An apparently purely structural priming effect in subject expression has been observed in Cameron (1994), who considers the effect of the form of the subject of the immediately preceding clause, and finds that in both switch reference and coreferential contexts, “pronouns lead to pronouns, and null subjects lead to null subjects” (1994: 40). The factor group “realization of subject of immediately preceding clause” replicates this effect, with a favoring of *yo* in the context of preceding personal pronouns (.57), and of unexpressed subjects in the context of preceding unexpressed subjects (.46). (See examples of instances of

---

**13** When there are more than two intervening human subjects between coreferential mentions, the priming effect greatly dissipates: comparing the rate of *yo* in tokens with a previous realization as *yo* vs. ones with a previous unexpressed mention, we observe, at 0 intervening human subjects, 52% (100/192) vs. 25% (44/173); at one, 62% (81/131) vs. 46% (49/106); at two, 57% (28/49) vs. 37% (19/51); and at three or more 60% (52/86) vs. 56% (53/95). This corresponds with the patterning observed in Travis (2007: 121), who found that coreferential priming only had a significant effect on subject realization when the coreferential subject occurred in the immediately preceding clause, or with one intervening clause (not distinguishing between human and non-human intervening subjects).



**Fig. 2:** Rate of *yo* expression by realization of subject of the immediately preceding clause, in non-coreferential and coreferential contexts (0 clauses intervening between coreferential mentions)

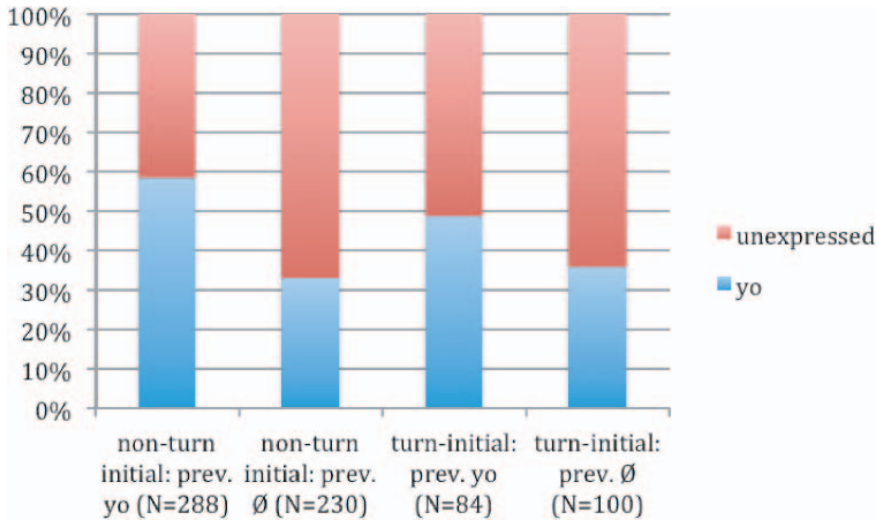
the variable with preceding subjects of different grammatical persons in (4)–(7) above.)

Note that the factor of “personal pronoun” in the group “realization of the subject of the immediately preceding clause” includes both tokens when that subject was *yo*, as well as when it was another pronoun (with a non-coreferential referent), and likewise for the unexpressed subjects. Thus there is some overlap between these two priming measures.

Figure 2 breaks this down, by presenting the rate of expression according to the realization and coreferentiality of the preceding subject.<sup>14</sup>

These results demonstrate the interplay between priming and coreferentiality. In coreferential contexts (columns 2 and 4), the form of the preceding subject has a significant effect on subject realization (53% *yo* when the immediately preceding subject was pronominal vs. 26% *yo* when it was unexpressed;  $p < 0.0001$ ). In non-coreferential contexts (columns 1 and 3), the same pattern emerges, but it is not quite significant (60% *yo* when the immediately preceding subject was pronominal vs. 49% *yo* when it was unexpressed;  $p = 0.073$ ) (cf. Cameron and

<sup>14</sup> In order to facilitate comparison with Cameron (1994), in Figure 2 we apply the measure of switch reference, rather than that of Intervening Human Subjects.



**Fig. 3:** Rate of *yo* expression by realization of previous coreferential 1sg subject (at a distance of 2 or fewer clauses), in non-turn-initial and turn-initial position

Flores-Ferrán 2003: 49). We can say that priming operates in coreferential contexts but is neutralized under switch reference, while switch reference operates when the preceding subject is unexpressed but is neutralized by priming, whether the preceding subject is *yo* or another personal pronoun.

Viewing all four columns, the significant difference is between the fourth column, that of a preceding unexpressed coreferential subject, which shows the lowest rate of expression (26%), and all three other columns. This can be accounted for by the fact that in certain environments, priming and coreferentiality work synergistically, and in others, antagonistically (cf. Travis 2007: 124). We get the lowest rate of *yo* when these two work together, i.e., in coreferential contexts with an immediately preceding unexpressed subject (two factors which independently disfavor *yo*).

A further issue to consider for priming in the conversational discourse under study here is the way it interacts with position in the turn: both priming effects operate turn-internally, but are greatly degraded across turns. For the coreferential priming effect – realization of the previous coreferential 1sg subject – shown in Figure 3, *yo* favors *yo*, but only within a speaker turn, that is when the 1sg token occurs in a non-turn-initial intonation unit (58%, 168/288, when previous realization is expressed vs. 33%, 76/230, when it is unexpressed,  $p < 0.0001$ , columns 1 and 2 in Figure 3). When the target 1sg token occurs in a turn-initial IU, previous realization makes a smaller, if any, difference (49%, 41/84 vs. 36%,

36/100,  $p = 0.099$ ).<sup>15</sup> We find a similar weakening across speaker turns for the more structural priming effect (realization of the subject of the immediately preceding clause, regardless of coreferentiality): a preceding pronoun favors *yo* within a speaker turn (59%, 103/175, with a preceding pronoun vs. 37%, 92/247, when the preceding subject is unexpressed,  $p < 0.0001$ ) but not when the 1sg token occurs in a turn-initial IU (48%, 30/63 vs. 47%, 57/122,  $p = 1.00$ ).

What do these priming effects tell us about the role of *yo* in discourse? Priming, or perseveration, has been characterized as “mechanical” rather than “functional” (Labov 1994: 547–568), being defined by Bock and Griffin (2000: 177) as “the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of an utterance”. The lack of a pragmatic motivation is supported by the fact that priming from previous experience is also found outside of language, such as in human motor control (van der Wel et al. 2007). Thus, these priming effects are unrelated to the expression of notions such as emphasis or contrast.

### 4.3 Tense-aspect-mood

The remaining significant factor group is tense-aspect-mood. It is often assumed that unexpressed subjects are allowed in Spanish because verbs carry person and number marking, and therefore in many contexts an explicit subject is redundant (cf. discussion in Toribio 1996: 409–411). In the Imperfect, Conditional and Subjunctive, however, first-person and third-person singular take the same form, and thus are potentially ambiguous, an ambiguity that expressed subjects would resolve (Hochberg 1986). A number of studies have found a correlation between ambiguous verb forms and expressed subjects (Cameron 1994; Hochberg 1986), but others have found no such correlation (Bentivoglio 1987; Enríquez 1984; Ranson 1991).

Silva-Corvalán (1997, 2001) has proposed that it is not the ambiguity but the discourse function of the different TAMs that motivates their use with expressed or unexpressed subjects. Pronominal subjects are more likely with the Imperfect, Conditional and Subjunctive because of the backgrounded nature of the verbal

---

**15** If the weakening or lack of a coreferential priming effect in turn-initial position were due to the presence of intervening human subjects or to coreferentiality (turn-initial IU tokens mostly also occur in intervening human subjects or switch reference contexts), we should have observed a higher, not lower, *yo* rate in turn-initial (49%) than in non-turn-initial IU position (59%) when the previous realization is *yo*. A larger data set is required for further examination of interactions between turn position, switch reference, and coreferential priming.

situation expressed with these TAMs; they are less likely with the Preterit because perfectives foreground events; and the Present is expected to show little effect (2001: 161–163) (cf. also Bayley and Pease-Alvarez 1997). In these data, the Imperfect, Conditional and Subjunctive favor *yo* expression (.62), as predicted; the Present tense, also as predicted, neither favors nor disfavors (.48); but the Preterit does not act as predicted, in showing little effect (.48). This may be because of the conversational nature of these data, as opposed to the narrative data considered in previous studies. As Travis (2007: 119) suggests, TAM effects are likely sensitive to genre.

While these results indicate that backgrounding TAMs do have an effect on *yo* expression, it is worth noting that they only account for 13% of the tokens studied here (as opposed to the Present, which makes up two-thirds [66%] of the tokens, as we would expect for 1sg in conversational data [cf. Scheibman 2002: 62, on American English conversations]). Thus, as well as being sensitive to genre this widely discussed TAM effect may also depend on grammatical person, applying more to third person subjects.

In summary, the evidence of distribution patterns refutes the notion of subject pronoun *yo* as generally playing a contrastive role, and indicates that the realization chosen is affected by (1) the presence of intervening human subjects, a cognitive factor and (2) speakers' tendency to repeat forms that they have recently produced, a mechanical effect. We return now to the effect of semantic class of verb.

## 5 Evidence for a cognitive verb class and particular constructions

As we saw in the multivariate analysis in Table 2, cognitive verbs favor *yo* expression. Other groupings of verbs (such as copulas, *ser*, *estar*, and verbs of speech, e.g., *decir* 'say, tell', *llamar* 'call', *preguntar* 'ask',) favor *yo*, though not as highly as do cognitive verbs (Bentivoglio 1987: 50–54; Travis 2007: 116–117). Note that the rate of expression for the cognitive verbs is 67%, while for other verbs it is just 44% (Table 2). It might be asked whether this is an epiphenomenon of the environments in which cognitive verbs occur, as Cameron (1995) found to be the case in part for the number effect, whereby singular subjects favor expression more than plurals. Cameron (1995: 17–21) showed that plural subjects overwhelmingly occur in contexts in which the referent set or members of the set have been mentioned within the previous five clauses; in other words, apparent number effects are largely due to switch reference. Cognitive verbs are in fact more likely than non-cognitive verbs to occur in the presence of intervening human subjects (71%,



|  | Cognitive verbs N = 201;<br>Input: .65 (Overall rate: 67%) |             |     |        | Non-Cognitive verbs N = 819;<br>Input: .44 (Overall rate: 44%) |             |     |        |
|--|--|-------------|-----|--------|--|-------------|-----|--------|
|  | Prob   | % <i>yo</i> | N   | % data | Prob   | % <i>yo</i> | N   | % data |
| <b>Turn position</b>                                     |  |             |     |        |  |             |     |        |
| Turn-initial IU  | .69  | 83%         | 70  | 35%    | [.45]  | 41%         | 262 | 32%    |
| Non-turn-initial IU                                      | .40  | 58%         | 131 | 65%    | [.53]  | 46%         | 555 | 68%    |
| <b>Realization of previous coref 1sg subject</b>         |  |             |     |        |  |             |     |        |
| Pronoun ( <i>yo</i> )                                    | .57  | 68%         | 81  | 63%    | .59  | 53%         | 291 | 51%    |
| Unexpressed  | .38  | 50%         | 48  | 37%    | .41  | 31%         | 282 | 49%    |
| <b>Intervening human subjects</b>                        |  |             |     |        |  |             |     |        |
| Present  | [.53]  | 73%         | 126 | 71%    | .56  | 49%         | 393 | 56%    |
| Absent   | [.42]  | 54%         | 52  | 29%    | .42  | 37%         | 313 | 44%    |
| <b>Realization of subject of immed. preceding clause</b> |  |             |     |        |  |             |     |        |
| Pronoun  | [.54]  | 69%         | 48  | 24%    | .57  | 53%         | 190 | 23%    |
| Other (non-hum.,<br>lex., impers., etc.)                 | [.45]  | 65%         | 95  | 47%    | .51  | 48%         | 314 | 39%    |
| Unexpressed  | [.56]  | 67%         | 58  | 29%    | .44  | 35%         | 311 | 38%    |
| <b>Tense-Aspect-Mood</b>                                 |  |             |     |        |  |             |     |        |
| Imperfect  | [.53]  | 72%         | 18  | 9%     | .62  | 59%         | 98  | 15%    |
| Present  | [.48]  | 65%         | 163 | 82%    | .48  | 44%         | 410 | 61%    |
| Preterit   | [.64]  | 82%         | 17  | 9%     | .47  | 43%         | 169 | 25%    |
| <b>Polarity</b>  |  |             |     |        |  |             |     |        |
| Negative   | [.45]  | 63%         | 97  | 49%    | .63  | 56%         | 119 | 15%    |
| Positive   | [.55]  | 71%         | 101 | 51%    | .48  | 42%         | 694 | 85%    |

**Table 4:** Two independent Variable-rule analyses of the contribution of factors selected as significant to the choice of expressed *yo* in conversational Colombian Spanish, for cognitive and non-cognitive verbs

126/178, vs. 56%, 393/706,  $p = 0.0002$ , Table 4). However, the favoring of *yo* by cognitive verbs remains even in the absence of intervening human subjects (54%, 28/52, for cognitive verbs vs. 37%, 116/313 for non-cognitive verbs,  $p = 0.03$ ).<sup>16</sup> The verb class effect therefore is genuine, and warrants further analysis. Do cognitive verbs form a category for 1sg subject expression, behaving differently from other verbs?

<sup>16</sup> Cognitive verbs are also more likely than other verbs to occur in switch reference contexts (71%, 94/133, vs. 43%, 251/579,  $p = 0.0032$ ); the rate of *yo* for cognitive vs. non-cognitive verbs in coreferential contexts is 49%, 19/39, vs. 38%, 94/251, a difference which does not achieve significance,  $p = 0.2169$ , but is in the right direction.

Table 4 gives the results for the same set of data seen in Table 2, but here in the form of two independent analyses, one of the cognitive verbs only and the second of the remaining verbs. In comparing the results across the two groups of verbs, we see that for both, *yo* is favored when the previous coreferential 1sg subject was realized pronominally. It is also favored in both analyses when there are intervening human subjects between coreferential mentions. Although Intervening Human Subjects does not achieve significance for the cognitive verbs, most revealing of linguistic conditioning is the direction of effect, which is the same in both analyses.<sup>17</sup>

Thus, for the coreferential priming and Intervening Human Subjects effects, the linguistic conditioning is parallel across both groups of verbs. However, we note that in the cognitive verbs, there is no effect for the realization of the subject of the immediately preceding clause, nor for tense-aspect-mood (not surprisingly, a larger proportion of the cognitive verbs is present tense; 82% (163/201) vs. 61% (410/819)). More importantly, we observe a striking reversal in direction of effect in relation to turn position and polarity. This divergence in the linguistic conditioning of *yo* is evidence for a separate class of cognitive verbs.

## 5.1 The turn-position effect for cognitive verbs

Some scholars have proposed that expressed Spanish first-person singular subjects serve to “signal a speaker’s intention to take the floor” (Davidson 1996: 561), and provide a means for the speaker to “hacerse presente en la escena” [‘make themselves present on the scene’] (Aijón Oliva and Serrano 2010: 16). Consonant with such proposals is a turn-position effect reported by Bentivoglio (1987: 38–40), whereby first-person subject pronouns were favored when the immediately preceding clause was not part of the same speaker turn.

Here, we considered where in a speaker’s turn the target verb with a 1sg subject occurred, using Intonation Units. Tokens that occurred in a turn-initial IU (regardless of where in the IU they occurred) were coded as turn initial. Thus, as well as instances of *yo* (or the corresponding verb, for unexpressed subjects) occurring as the first word in a turn (as in lines 1 and 5 in (6)), we included as “turn initial” instances where the token was not the first word in IU (e.g. in (2)

---

<sup>17</sup> Weights for non-significant factor groups are from the first “stepdown” run in GoldVarb, in which all factors are included in the regression. When comparing two data sets, direction of effect (the order of the factors within a group from highest to lowest factor weight) gives “the detailed structure of the relationship between variant and context”, whereas significance of a factor group may not be achieved in a smaller data set (Poplack and Tagliamonte 2001: 93–94).

*porque cuando yo estaba despierto*), and where it occurred in the second IU in the turn, but the first IU contained a discourse marker with continuing intonation, marked with a comma in this transcription method (as in line 3 in (6), *pues, yo pienso . . .*).

As shown in Table 4, we find that for the cognitive verbs, occurrence in a turn-initial intonation unit strongly favors *yo* (.69), while for the non-cognitive verbs this is not only non-significant, but also tends toward the opposite direction.

It is important to note that turn position is not coterminous with the presence of intervening human subjects: not surprisingly, most turn-initial tokens occur in the presence of intervening human subjects (only 17%, 55/332, occur in the absence of intervening human subjects). But for cognitive verbs, in the presence of intervening human subjects, the rate of *yo* is still higher in a turn-initial (87%, 46/53) than in a non-turn-initial IU (63%, 46/73,  $p = 0.0041$ ). The same holds for the traditional notion of switch reference. Tokens in turn initial position are overwhelmingly in switch reference contexts (only 8%, 27/332, occur in coreferential contexts). Nevertheless, for cognitive verbs, within switch reference contexts, the effect of turn position remains: the rate of *yo* is higher turn initially (83%, 54/65) than non-turn initially (63%, 56/89;  $p = 0.0069$ ).

The strong turn-position effect for cognitive but not other verbs is indicative of a *yo* + COGNITIVE VERB construction. Since this turn position effect is independent of intervening human subjects (and of switch reference), we interpret it as an interactional effect having to do with turn taking, rather than with referent accessibility. The strength of this local interactional effect and this particular construction may have contributed to scholars' global perception of what has been called a contrastive, emphatic or focal function of subject pronouns in the voluminous literature on the topic.

## 5.2 Negative polarity in non-cognitive verbs

For the non-cognitive verbs, negative polarity favors *yo* (.63), while for the cognitive verbs this is non-significant, and tends in the opposite direction. Should the higher rate in negative clauses be interpreted as an independent measure of a contrastive function for *yo* with non-cognitive verbs? One might be led to believe so, based on examples of negated clauses contrasting with a previous affirmative, as in (3) above, where the speaker's assertion that "I don't litter anywhere" contrasts (in Myhill and Xing's (1996) sense) with the interlocutor's assertion that "one" doesn't dare litter on the farm. But it does not appear that most of the time negation marks a contrast to a corresponding affirmative proposition, neither the speaker's nor interlocutor's. The polarity effect is maintained in turn-initial

position, an environment that lends itself to setting up contrast in an interactional sense of countering a proposition put forward by the interlocutor, but also in non-initial position, which would not appear to be associated with contrast vis-à-vis the interlocutor.

Beyond these data, despite a widely held understanding of negation as contrastive (e.g., Flores-Ferrán 2010: 70; Sun and Givón 1985: 346), we have no independent evidence that this is the case, at least in an interactional sense. In her study of negation in conversational English, Tottie (1991: 21, 35) found that 81% (N = 427) of negative clauses were denials of something which is “not explicitly present in the conversation” (Thompson 1998: 325), rather than of something that was previously uttered or is presupposed, as had been claimed in the early literature on negation (cf. Givón 1975: 104; Horn 1985: 143).<sup>18</sup> Thompson stresses that, unlike interrogatives, which serve the interactional function of questioning,<sup>19</sup> negative clauses participate neither in adjacency pairs nor in any other interactional pattern, but “serve to deny some state or event which is usually neither explicit nor ‘implicit’ in the context” (1998: 325–326). Thus, the favoring effect of negation on *yo* expression, found with non-cognitive verbs in the present conversational data, cannot be taken in and of itself to indicate an interactional contrastive function. Another explanation will therefore have to be sought in future work.

### 5.3 Constructions and prefabs

In sum, the patterning of the cognitive verbs in relation to turn position is evidence for a particular construction with respect to 1sg subject expression. Constructions are form-function pairings, where function includes stored information about linguistic and extralinguistic contexts of use (for an overview, see Croft and Cruse 2004: 225–290; Goldberg 2006: 3–17). These units of grammar range on a continuum from fixed expressions to productive morphosyntactic structures that are at least partially schematic. Here, within the highly schematic (SUBJECT PRONOUN) + VERB construction (where the parentheses are intended to capture variable expression), there exists a more specific (*yo*) + COGNITIVE VERB construc-

<sup>18</sup> Givón (1979: 104), for example, stated that “the speaker uttering a negative sentence assumes that the hearer knows that the corresponding affirmative was likely or has been previously mentioned”, and Horn (1985: 143) claimed that negative assertions often “presuppose a context in which the affirmative proposition has been asserted or at least entertained”.

<sup>19</sup> According to Thompson (1998: 321, based on Freed 1994), 80% (N = 816) of the time interrogative clauses seek information.

tion, which not only shows a greater tendency to the expression of *yo*, but a variability which does not pattern in precisely the same way as it does in the more general construction.

It should be noted that just two verbs make up nearly three-quarters (71%, 142/201) of the cognitive verb tokens, namely *creer* ‘think/believe’ (N = 55) and *saber* ‘know’ (N = 87). The next most frequent, *pensar* ‘think’ (N = 27), has half the token frequency of *creer*, followed by *acordarse* ‘remember’ (N = 14), *imaginarse* ‘imagine’ and *entender* ‘understand’ (N = 6 each). Despite the large proportion of tokens that *creer* and *saber* represent, examination of the non-frequent lexical types separately (i.e. the remaining 29% of the tokens), reveals the same tendencies. This uniformly distinct linguistic conditioning between cognitive and non-cognitive verbs thus indicates that there is a genuine category of cognitive verbs, which is formed by the items that appear in the verb slot in the (*yo*) + COGNITIVE VERB construction (cf. Bybee 2010: 80–81).

What is the relationship of the more frequent verbs to the category? In a usage-based view of grammar, frequent expressions serve as the central members of categories (Bybee 2010, ch. 4). In other words, we may think of *creer* and *saber* as central members of a cognitive verb class, in which other verbs are included by analogical local comparisons with these frequent verbs, taking into account semantic similarity.

In fact, it is particular forms within each of these two verb types that are frequent, namely *yo creo* ‘I think’ and *no sé* ‘(I) don’t know’. These may be viewed as prefabs, or prefabricated units (Bolinger 1976: 1) that is, single units rather than analyzable combinations of pronoun and verb or negation and verb. Evidence for prefab status comes from the high token and relative frequency of these expressions and from their distinct variation patterns.

With respect to token and relative frequency, we use the oral portion of the 100-million word Corpus del Español (Davies 2002-) as an ancillary source to the present Corpus of Conversational Colombian Spanish, though we recognize that, since prefabs “represent the conventional way of expressing an idea” (Bybee 2010: 81), they may be specific to a speech community, and thus may not be evident in the wide range of genres and dialects represented in large corpora.

In the present data, *creo* (N = 51) and *sé* (N = 79) are the most frequent 1SG SUBJECT + VERB forms to occur, and in the oral portion of the Corpus del Español, considering just the present indicative, *sé* and *creo* are again the most frequent 1sg verb forms (*creo* N = 9,215, *sé* N = 5,885).<sup>20</sup> Furthermore, the combination

<sup>20</sup> The next most frequent are *he* ‘I have [auxiliary]’ (N = 5,807), *tengo* ‘I have [possessive]’ (N = 4,033) and *digo* ‘I say’ (N = 3,032).

*yo + creo* is also highly frequent: in our data, *yo creo* constitutes 47% (24/51) of the tokens of *creo*, and in the oral portion of the Corpus del Español, *yo creo* constitutes 45% (4,165/10,986) of all occurrences of *creo* and 14% (4,165/30,846) of all occurrences of the pronoun *yo*.<sup>21</sup> Similarly, *no sé* constitutes 87% (69/79) of the tokens of *sé* in the present data and 75% (4,409/5,885) of the tokens of *sé* in the oral portion of the Corpus del Español, where it also comprises 5% (4,409/91,670) of the tokens of *no*.

Both 1sg Present forms constitute a large proportion of their respective lexical types. In the data under study here, *creo* comprises 93% (51/55) of all first singular tokens of the lexical type *creer*, and in the Corpus del Español (Davies 2002-), *creo* makes up 84% (9,215/10,986) of all tokens – all persons and tenses – of *creer*. Present indicative *sé* ‘I know’, too, makes up a substantial proportion of its lexical type, 91% (79/87) all 1sg *saber* tokens in the present data, and 47% (5,885/12,390) of all *saber* tokens in the oral portion of the Corpus del Español. These frequency measures provide evidence for viewing *yo creo* and *no sé* as prefabs.

Additional evidence for prefab status comes from distribution patterns. Comparing negated *no sé* to affirmative *sé* in an expanded data set from the Corpus of Conversational Colombian Spanish, we find differences in the presence and types of objects it occurs with. The majority of the affirmative tokens (60%, 13/22), *sé*, occur with a complement clause, while only one quarter (44/168) of the negated tokens do so. On the other hand, over one half (89/168) of the tokens of *no sé* occur with no direct object (as in line 4 in (7) above), while there is only one token of *sé* that occurs in such an environment. Where does *yo* variation sit in this prefab? While the rate of expression (56%, 94/168) is somewhat higher than the overall rate of 49%, an independent multivariate analysis of *no sé* replicates the effects we found for the cognitive verbs overall (Table 4).<sup>22</sup>

However, *yo creo*, for which the rate of expression in this expanded corpus is 62% (75/121), shows some divergence in the linguistic conditioning of variable *yo* expression. Independent multivariate analysis of Present-tense *creo* shows that

**21** In Corpus del Español, *yo creo* is the most frequent *yo + V<sub>Present Indicative</sub>* combination, appearing over five times as often as the next most frequent combinations (*yo he* ‘I have [auxiliary]’ and *yo tengo* ‘I have [possessive]’, both of which occur just over 700 times). In the data for this study, *yo (no) sé* is the most frequent (N = 58), followed by *yo creo* (27), *yo tengo* (26) and *yo digo* ‘I say’ (20). For *no sé*, *sé* is by far the most frequent 1sg Present Indicative form to occur following *no* in Corpus del Español (occurring approximately seven times more than the next most frequent forms, *no tengo* ‘I don’t have’ [N = 645] and *no creo* ‘I don’t think’ [N = 580]).

**22** Results of the analysis of *no sé* are as follows: N = 168, 56% *yo*, input .53: Previous realization, *yo* .63 – unexpressed .30; Turn position, initial .62 – non-initial .41. Non-significant factor groups: Realization of preceding subject, pronominal 60% – unexpressed 50% – other 60%; Intervening human subjects, present 62% – absent 43%.

*yo* is favored in turn-initial position and when the previous realization was *yo*, as in the general cognitive verb analysis (Table 4). The lack of structural priming from the realization of the subject of the immediately preceding clause is also replicated, which is evidence for the reduced analyzability of *yo creo* and potentially for other instances of the *yo* + COGNITIVE VERB construction. However, Intervening Human Subjects was not significant, and the opposite tendency to that in the general cognitive verb analysis was observed, namely a higher *yo* rate in the absence of an intervening human subject (and likewise, a higher *yo* rate in coreferential [same reference] contexts).<sup>23</sup>

The lack of a coreferentiality effect for *yo creo* makes a good case for its status as a particular construction that is a single unit rather than an analyzable combination of subject pronoun and verb. In Bybee's (2010: 48) terms, *yo creo* is largely autonomous from other instances of a more schematic (SUBJECT PRONOUN) + VERB construction. Contributing to the status of *yo creo* as an autonomous unit is (1) the token frequency of the string; (2) the high proportion it comprises of all occurrence both of the lexical type *creer* and of the pronoun *yo*; and (3) its immunity to the effects of priming from the preceding subject and of intervening human subjects (or switch reference).

At the same time, in other ways, the linguistic conditioning of *yo creo* is largely parallel to that of *yo* expression with other first singular verbs. The multivariate analysis shows that the strong coreferential priming effect observed in the analysis of all verbs is retained, as is the turn-position effect operative with cognitive verbs. That is, *yo creo* is not completely autonomous, but retains associations with the more general (*yo*) + COGNITIVE VERB construction, as well as the even more schematic (*yo*) + VERB construction.

These findings corroborate what previous research has shown, namely that prefabs, or particular constructions, need not be completely autonomous from other instances of the more general construction. Evidence for the interaction of the particular and the general has been adduced in studies of diachronic grammaticalization (Bybee and Torres Cacoullós 2009), language acquisition (e.g., Dąbrowska and Lieven 2005), and, as is the case here, synchronic variation in speech production (e.g., Torres Cacoullós and Walker 2009).

In summary, cognitive verbs not only present higher rates of *yo* expression, as is well known from previous studies, but also distinct linguistic conditioning

---

**23** Results of the analysis of *creo* are as follows: N = 121, 62% *yo*, input .56: Polarity, affirmative .57 – negative .16; Previous realization, *yo* .65 – unexpressed .37; Turn position, initial .64 – non-initial .42. Non-significant factor groups: Realization of preceding subject, pronominal 73% – unexpressed 68% – other 54%; Intervening human subjects, present 52% – absent 60% [Switch reference: switch 51% – coreferential 67%].

of the variation, something which had not been previously observed and was revealed here by independent analyses. The turn-position effect found with cognitive, but not other, verbs is interpretable as an interactional factor. The class of cognitive verbs is centered around frequent expressions, and in particular, *yo creo* stands out as a *yo* prefab, by virtue of its high token and relative frequency and the absence of a coreferentiality effect.

## 6 Conclusion

We began by considering three different operationalizations of contrast, and found no evidence for a general contrastive function for *yo*. Where, then, does the attribution of a general contrastive function to subject pronouns come from?

One contributing factor may be “double contrast” contexts, an environment where we do find a higher rate of *yo* expression (five out of the six tokens that met the operationalization did occur with expressed *yo*). In the absence of quantitative study, such contexts may be particularly salient, even though in actual usage they are rare (in the data under study here accounting for only 4% [6/162] of the first-person singular subjects analyzed). The same may be so for constructions with adversative conjunctions such as *pero* ‘but’ or with emphatic *mismo* ‘very same, self’, which have been found to occur with relatively higher rates of expressed subject pronouns in other studies, but also to be infrequent (e.g., Bentivoglio 1987: 46–48; Paredes Silva 1993: 41–43).

A general contrastive function may also have been inferred from the well-recognized role of accessibility in subject realization. For example, Sun and Givón’s (1985) appeal to Potential Referential Interference as a measure of “contrast” implies a link between the two, and it may well be thought that contrast is associated with switch reference (cf. Cameron and Schwenter Forthcoming). Our results, however, show that switch reference is distinct from contrast. Only a small subset of switch reference tokens would in fact be contrastive, once tokens are counted based on operational definitions such as double contrast or adversative conjunctions (Sections 3.1 and 3.4). Furthermore, switch reference uniformly raises *yo* rates, not distinguishing between third person and speaker-interlocutor self reference, which was one measure of an interactional function considered here (Section 3.2). And switch reference itself – whatever associations it may or may not have with contrast – is neutralized by *yo*-to-*yo* priming (Section 4.2), which most earlier studies did not include as a constraint.

A further possibility is that a notion of contrast for *yo* has been assumed from the interactional role of high frequency *yo creo*. Our study finds an interactional



function – measured by the favoring effect of occurrence in a turn-initial Intonation Unit (independently of the presence of intervening human subjects) – but this pertains to *yo creo* and the particular *yo* + COGNITIVE VERB construction, rather than being evidence of an overarching abstract meaning-feature of contrast for a general (SUBJECT PRONOUN) + VERB construction.

In the light of a resounding lack of evidence for a contrastive role for expressed subjects, we have addressed the question of what subject pronouns do in discourse through multivariate analyses of the linguistic conditioning of expressed Spanish first singular pronoun *yo*, and have found that cognitive and mechanical constraints are paramount. The multivariate analysis demonstrated that speakers' choice of expressed *yo* is statistically more probable in the presence of another human subject intervening between coreferential 1sg mentions; with previous realization as *yo*; and following a personal subject pronoun – coreferential or not – in the immediately preceding clause.

Alongside these cognitive and mechanical effects distinct patterns of *yo* expression for cognitive verbs suggest constructional effects in variable subject expression. The multivariate analysis identified a verb class constraint, such that cognitive verbs are more favorable than other verbs to *yo*. Evidence for a construction specific to cognitive verbs, beyond the higher rate of *yo* observed, is the distinct linguistic conditioning of *yo* expression, such that for cognitive verbs, but not others, *yo* is favored when the target 1sg token occurs in the first Intonation Unit of a speaker's turn. The verb slot defines a category of cognitive verbs, which, in a usage-based approach to grammar, may be viewed as being constituted by local analogy with frequent instances, most prominently *yo creo*.

The study of referent realization can be advanced through the pursuit of accountable quantitative studies, in different language varieties; taking into consideration different genres, persons and syntactic roles; employing replicable operationalizations of notions to be tested; exploring further the workings of accessibility and the strength and interactions of priming effects; and identifying fixed constructions which may exhibit distinct behavior.

## References

- Aijón Oliva, Miguel Angel & María José Serrano. 2010. El hablante en su discurso: Expresión y omisión del sujeto de *creo*. *Oralia* 13. 7–38.
- Amaral, Patricia Matos & Scott A. Schwenter. 2005. Contrast and the (non-)occurrence of subject pronouns. In David Eddington (ed.), *Selected proceedings of the 7th Hispanic Linguistics Symposium*, 116–127. Somerville, MA: Cascadilla Press.
- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. New York, NY: Routledge, Chapman & Hall.

- Ariel, Mira. 1994. Interpreting anaphoric expressions: A cognitive versus a pragmatic approach. *Journal of Linguistics* 30(1). 3–42.
- Balasz, Sonia. 2008. La conectividad discursiva en el discurso interactivo. In Joyce Bruhn de Garavito and Elena Valenzuela (eds.), *Selected Proceedings of the 2006 Hispanic Linguistics Symposium*, 300–311. Somerville, MA: Cascadilla Proceedings Project.
- Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9(3). 349–371.
- Benevento, Nicole & Amelia Dietrich. 2011. Variable position of 1sg subject pronoun in New Mexico Spanish-English Bilingual Corpus (NMSEB). The Pennsylvania State University. Ms.
- Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Blackwell, Sarah E. 2003. *Implicatures in discourse: The case of Spanish NP anaphora*. Philadelphia, PA: John Benjamins.
- Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–387.
- Bock, J. Kathryn & Zeni M. Griffin. 2000. The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General* 129(2). 177–192.
- Bolinger, Dwight. 1976. Meaning and memory. *Forum Linguisticum* 1(1). 1–14.
- Butt, John & Carmen Benjamin. 2004. *A new reference grammar of modern Spanish*, 4th ed. London: Edward Arnold.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan & Rena Torres Cacoullos. 2009. The role of prefabs in grammaticization: How the particular and the general interact in language change. In Roberta L. Corrigan, Edith Moravcsik, Hamid Ouali, and Kathleen Wheatley (eds.), *Formulaic language*, vol. 1: Distribution and historical change, 187–217. Amsterdam: John Benjamins.
- Cameron, Richard. 1992. *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Philadelphia, PA: University of Pennsylvania dissertation.
- 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. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6–7. 1–27.
- Cameron, Richard & Nydia Flores-Ferrán. 2003. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1(1). 41–65.
- Cameron, Richard & Scott Schwenter. Forthcoming. Pragmatics and variationist sociolinguistics. In Robert Bayley, Richard Cameron, and Ceil Lucas (eds.), *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chafe, Wallace. 1994. *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Clancy, Patricia M. 1980. Referential choice in English and Japanese discourse. In Wallace Chafe (ed.), *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*, 127–202. Norwood, NJ: Ablex.

- Comajoan, Llorenç. 2006. Continuity and episodic structure in Spanish subject reference. In J. Clancy Clements and Jiyoung Yoon (eds.), *Functional approaches to Spanish syntax: Lexical semantics, discourse and transitivity*, 53–79. New York: Palgrave MacMillan.
- Croft, William & D. Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Dąbrowska, Eva & Elena Lieven. 2005. Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16. 437–474.
- Davidson, Brad. 1996. 'Pragmatic weight' and Spanish subject pronouns: The pragmatic and discourse uses of *tú* and *yo* in spoken Madrid Spanish. *Journal of Pragmatics* 26(4). 543–565.
- Davies, Mark. 2002-. Corpus del Español: 100 million words, 1200s–1900s: Available online at <http://www.corpusdelespanol.org>.
- Du Bois, John W., Stephan Schuetze-Coburn, Susanna Cumming, & Danae Paolino. 1993. Outline of discourse transcription. In Jane Edwards and Martin Lampert (eds.), *Talking data: Transcription and coding in discourse*, 45–89. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto Miguel de Cervantes.
- Flores-Ferrán, Nydia. 2010. ¡Tú no me hables! Pronoun expression in conflict narratives. *International Journal of the Sociology of Language* 203. 61–82.
- Fox, Barbara. 1987. *Discourse structure and anaphora: Written and conversational English*. Cambridge: Cambridge University Press.
- Freed, Alice F. 1994. The form and function of questions in informal dyadic conversation. *Journal of Pragmatics* 21. 621–644.
- Gili y Gaya, Samuel. 1964. *Curso superior de sintaxis española*, 9th ed. Barcelona: Bibliograf.
- Givón, T. 1979. *On understanding grammar*. New York: Academic Press.
- Givón, T. 1983. Topic continuity in discourse: An introduction. In T. Givón (ed.), *Topic continuity in discourse: A quantitative cross-linguistic study*, 1–41. Amsterdam: John Benjamins.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Hochberg, Judith G. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3). 609–621.
- Horn, Laurence R. 1985. Metalinguistic negation and pragmatic ambiguity. *Language* 61(1). 121–174.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45(4). 715–762.
- Labov, William. 1994. *Principles of linguistic change: Internal factors*, vol. 1, 3 vols. Oxford: Basil Blackwell.
- Labov, William. 2010. *Principles of linguistic change: Cognitive and cultural factors*, vol. 3, 3 vols. Oxford: Blackwell.
- 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.

- López Meirama, Belén. 2006. Semantic and discourse-pragmatic factors in Spanish word order. In J. Clancy Clements and Jiyoung Yoon (eds.), *Functional approaches to Spanish syntax: Lexical semantics, discourse and transitivity*, 10–79. New York: Palgrave MacMillan.
- Luján, Marta. 1999. Expresión y omisión del pronombre personal. In Ignacio Bosque and Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, vol. 1, 1275–1315. Madrid: Espasa-Calpe.
- Mayol, Laia. 2010. Contrastive pronouns in null subject Romance languages. *Lingua* 120. 2497–2514.
- Myhill, John & Zhiqun Xing. 1996. Towards an operational definition of contrast. *Studies in Language* 20(2). 303–360.
- Ocampo, Francisco. 1995. The word order of two-constituent constructions in spoken Spanish. In Pamela Downing and Michael Noonan (eds.), *Word order in discourse*, 425–447. Amsterdam/Philadelphia: John Benjamins.
- Otheguy, Ricardo, Ana Cecilia Zentella, & David Livert. 2007. Language and dialect contact in Spanish of New York: Toward the formation of a speech community. *Language* 83(4). 770–802.
- Paredes Silva, Vera Lucia. 1993. Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5(1). 35–49.
- Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4). 633–651.
- Poplack, Shana. 1980. The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In William Labov (ed.), *Locating language in time and space*, 55–67. New York: Academic Press.
- Poplack, Shana & Elisabete Malvar. 2007. Elucidating the transition period in linguistic change: The expression of the future in Brazilian Portuguese. *Probus* 19. 121–169.
- Poplack, Shana & Sali Tagliamonte. 2001. *African American English in the diaspora*. Massachusetts: Blackwell Publishers.
- Posio, Pekka. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3). 777–798.
- Ranson, Diana L. 1991. Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3(2). 133–152.
- Sankoff, David. 1988a. Sociolinguistics and syntactic variation. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge survey (Vol. 4, Language: The socio-cultural context)*, 140–161. Cambridge: Cambridge University Press.
- Sankoff, David. 1988b. Variable rules. In Ulrich Ammon, Norbert Dittmar, and Klaus J. Mattheier (eds.), *Sociolinguistics: An international handbook of the science of language and society*, vol. 2, 984–997. Berlin/New York: Walter de Gruyter.
- Sankoff, David, Sali Tagliamonte, & Eric Smith. 2012. Goldvarb LION: A variable rule application for Macintosh. University of Toronto. URL <http://individual.utoronto.ca/tagliamonte/goldvarb.htm>.
- Scheibman, Joanne. 2002. *Point of view and grammar: Structural patterns of subjectivity in American English conversation*. Amsterdam/Philadelphia: John Benjamins.
- Scherre, Maria Marta Pereira & Anthony J. Naro. 1991. Marking in discourse: ‘Birds of a feather’. *Language Variation and Change* 3(1). 23–32.

- Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. In Jon Amastae and Lucía Elías Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 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. Variación sintáctica en el discurso oral: Problemas metodológicos. In Francisco Moreno Fernández (ed.), *Trabajos de sociolingüística hispánica*, 115–135. Alcalá de Henares, España: Universidad de Alcalá.
- Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.
- Silva-Corvalán, Carmen. 2003. Otra mirada a la expresión del sujeto como variable sintáctica. In Francisco Moreno Fernández, Francisco Gimeno Menéndez, José Antonio Samper, María Luz Gutiérrez Araua, María Vaquero, and César Hernández (eds.), *Lengua, Variación y contexto: Estudios dedicados a Humberto López Morales*, vol. 2, 849–860. Madrid: Arco Libros.
- Silveira, Agripino S. 2007. Patterns of subjectivity in spoken Brazilian Portuguese. In Clemliton Lopes Pinheiro and Kilpatric M. R. Campelo (eds.), *Ensaio provindos do corpus oral do Português de Fortaleza*, 219–264. Fortaleza, Brazil: Eduece.
- Silveira, Agripino S. 2008. Frequency effects, specialization of forms, and subject expression in Brazilian Portuguese. In Evan Ashworth and Daniel Sanford (eds.), *Proceedings of the VII High Desert Linguistics Society Conference*, 1–19. Albuquerque, NM: Department of Linguistics, University of New Mexico.
- Sun, Chao Fen & T. Givón. 1985. On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language* 61(2). 329–351.
- Thompson, Sandra A. 1998. A discourse explanation for the cross-linguistic differences in the grammar of interrogation and negation. In Anna Siewierska and Jae Jung Song (eds.), *Case, typology and grammar: In honor of Barry J. Blake*, 309–341. Amsterdam: John Benjamins.
- Thompson, Sandra A. 2002. 'Object Complements' and conversation: Towards a realistic account. *Studies in Language* 26(1). 125–163.
- Toribio, Almeida Jacqueline. 1996. Dialectal variation in the licensing of null referential and expletive subjects. In Claudia Parodi, Carlos Quicoli, Mario Saltarelli, and Maria Luisa Zubizarreta (eds.), *Aspects of Romance linguistics: Selected papers from the 24th Linguistic Symposium on Romance Languages, 1994*, 409–432. Washington, DC: Georgetown University Press.
- Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 110(5). 315–341.
- Torres Cacoullous, Rena & Catherine E. Travis. 2011. Using structural variability to evaluate convergence via code-switching. *International Journal of Bilingualism* 15(3). 241–267.
- Torres Cacoullous, Rena & James A. Walker. 2009. The present of the English future: Grammatical variation and collocations in discourse. *Language* 85(2). 321–354.
- Tottie, Gunnel. 1991. *Negation in English Speech and Writing*. New York: Academic Press.
- Travis, Catherine E. 2005a. *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin/New York: Mouton de Gruyter.
- Travis, Catherine E. 2005b. The yo-yo effect: Priming in subject expression in Colombian Spanish. In Randall Gess and Edward J. Rubin (eds.), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages, 2004*, 329–349. Amsterdam/Philadelphia: John Benjamins.

- 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. & Rena Torres Cacoullos. 2012. On the patterning of stressed *I*. Submitted.
- van der Wel, Robrecht P.R.D., Robin M. Fleckenstein, Steven A. Jax, & David A. Rosenbaum. 2007. Hand path priming in manual obstacle avoidance: Evidence for abstract spatiotemporal forms in human motor control. *Journal of Experimental Psychology: Human Perception and Performance* 33(5). 1117–1126.
- Weiner, E. Judith & William Labov. 1983. Constraints on the agentless passive. *Journal of Linguistics* 19(1). 29–58.

## Appendix

### Transcription Conventions (Du Bois et al. 1993)

|                  |   |
|------------------|---|
| Carriage return: | new Intonation Unit   |
| .                | final intonation contour                                    |
| ,                | continuing intonation contour                               |
| ?                | appeal intonation contour                                   |
| --               | truncated intonation contour                                |
| -                | truncated word  |
| ...              | medium pause (>0.7 secs)                                    |
| ..               | short pause (about 0.5 secs)                                |
| [ ]              | overlapped speech   |
| [2 2]            | overlapped speech, used to distinguish consecutive overlaps |
| @                | one syllable of laughter                                    |
| X                | unclear syllable  |
| <VOX VOX>        | marked speech quality                                       |
| <WH WH>          | whispered speech  |
| (( ))            | researcher comment  |