

## Discovering structure: Person and accessibility

---

### Abstract

We probe grammatical person differences comparing 3sg with 1sg in actual language use, using Spanish variable subject pronoun expression. We reconfigure the familiar constraint of accessibility to distinguish between clause linking (prosodic and syntactic connectedness) in coreferential contexts and distance from the previous mention (intervening clauses) in non-coreferential contexts. This refinement reveals that accessibility impacts 1sg earlier than 3sg, for which the pronoun rate rises more slowly with increasing distance. At the same time, a greater proportion of 3sg than 1sg pronominal and unexpressed subjects occurs in coreferential contexts. 3sg pronominal and unexpressed subjects thus tend to cluster more closely. By these differences in the workings of accessibility and in contextual distribution, 3sg, unlike speech act participant 1sg, is a *transient person* in discourse.

### 1 Grammatical person in cross-linguistic perspective

Linguistic structure is not a given but “must be discovered through analysis”, as Ricardo Otheguy has urged (2002, p. 400). Grammatical categories espoused by linguists do not always coincide with actual usage, and though the category of person features in accounts of subject pronoun expression, its effects are not well understood. Admittedly, variable use of subject pronouns in Spanish has been so widely analyzed that we might wonder if there is anything more to be discovered about it. After all, “multiple studies ... across communities, across settings, and across the lifespan reveal the very consistent nature of structured variation” in subject pronoun expression (Carvalho, Orozco & Shin 2015, p. xxii).

Grammatical person is consistently found as a probabilistic constraint on subject expression, often the strongest, as in Ricardo Otheguy’s own analyses of Spanish in New York City (Otheguy & Zentella 2012; Otheguy, Zentella & Livert 2007). Person effects are reported for a range of other languages as well, for example, Arabic (Owens, Dodsworth & Kohn 2013, p. 268; Parkinson 1987, p. 356), Auslan (McKee, Schembri, McKee & Johnston 2011, p. 388), Bislama (Meyerhoff 2009, p. 311), Cantonese and Russian (Nagy, Aghdasi, Denis & Motut 2011, p. 141-142), and Turkish (Koban 2011, p. 362). Here we try to explain this effect.

A key distinction often made is between first and second person in contrast with (animate) third person, said to hold cross-linguistically. For example, in languages with split ergative marking based on person, the most common configuration is one in which first and second persons are treated differently from other NPs, including (animate) third person (e.g., Delancey 1981, p. 628). In languages with switch reference marking, this is sometimes applied only on third person verbs (e.g., Haiman

& Munro 1983, p. xi). And in agreement patterns, zero marking is more likely on the third person than on first and second (e.g., Bybee 1985, p. 53).

The status of third person follows from the “independent role that the distinction between speech act participants and third person referents plays in grammar and discourse” (Dahl 2008, p. 143). Indeed, much cited is Benveniste’s characterization of the third person as a “non-person”, “never being reflective of the instance of discourse” (Benveniste 1971, p. 221-222).

For an exploration of subject person in spontaneous speech, the most pertinent difference is that between first and third person singular (1sg and 3sg) human subjects, as these make up the bulk of the data. Together, they account for approximately two thirds of most Spanish datasets (e.g., Claes 2011, p. 199; Erker & Guy 2012, p. 540; Lastra & Butragueño 2015, p. 43; Otheguy et al. 2007, p. 790). In English conversational data also, 1sg and 3sg are the most frequent human subjects (Scheibman 2001, p. 68-80).

The most widely reported difference between these two persons is that 1sg subjects favor pronominal expression more so than 3sg (e.g., Bayley & Pease-Alvarez 1997, p. 363; Cameron 1992, p. 168; Enríquez 1984, p. 350; Morales 1986, p. 93-96; Orozco 2015, p. 27; Posio 2015, p. 72; Ranson 1991, p. 139). The same pattern has been observed not only in the typologically similar language Brazilian Portuguese (Silveira 2011, p. 48), but also in Mandarin Chinese (Jia & Bayley 2002, p. 110) and even in English, despite the rarity of unexpressed subjects in this “non-null subject” language (Torres Cacoullos & Travis To appear). Are subject pronoun rate differences manifestations of true differences between grammatical persons in the functions of subject forms?

Here we seek to better understand the 3sg vs. 1sg person difference by converting the question into one with a quantitative answer. Using prosodically transcribed spontaneous speech data (Section 2), we focus our attention on the role of accessibility in subject pronoun expression. Once we break down accessibility to consider both linking across clauses in coreferential contexts and distance from previous mention in non-coreferential contexts (Section 3), differences between 3sg and 1sg emerge. One lies in the implementation of the effect (Section 4). Another is found in the distribution of 3sg vs. 1sg subjects in relation to the degree of accessibility of their referents, as for 3sg, the distribution of pronominal and unexpressed subjects is impacted by the availability of lexical forms (Section 5). Contextual distribution of the data thus contributes to the overall rate of variants. In light of both contextual effects and contextual distributions related to accessibility, we revisit the variable context for subject expression to establish the alternative grammatical ways of “saying the same thing” (cf., Labov 1972, p. 72) (Section 6).

## 2 Spontaneous speech and prosodically transcribed data

To explore grammatical person differences in language use, we examine 1sg and 3sg subjects in spontaneous speech data. The generalizability of the analysis is bolstered by the examination of two corpora, representing different dialects and genres, and providing close to 8,000 tokens for observation. The New Mexico Spanish-English Bilingual corpus (NMSEB) was collected in the years 2010-2011, recording 40 *Nuevomexicano* speakers from northern New Mexico, USA (Torres Cacoullos & Travis 2018, , Chapters 2 & 3). NMSEB consists of sociolinguistic interviews conducted with extended family members and acquaintances, totaling

approximately 30 hours, or 300,000 words of speech. The Corpus of Conversational Colombian Spanish (CCCS) consists of 30 conversations between close friends and intimate family, such as spouses, recorded 1997-2004 (cf., Travis 2005, p. 9-25). It comprises a total of approximately 100,000 words, or nine hours of speech, from 37 speakers.

Initially extracted were all tokens of finite verbs with 1sg and 3sg (human, specific) subjects that are unexpressed or realized pronominally (*yo* ‘I’, *él* ‘he’, *ella* ‘she’). Outside the envelope of variation as defined are postverbal subject pronouns (Section 6 below). Contexts with negligible variation are also set aside (Otheguy et al. 2007, p. 776; Torres Cacoullos & Travis 2018, p. 119-121). This includes non-human and nonspecific human referents, because they are rarely realized as personal pronouns *él*, *ella* (in these data, never), and *wh*-interrogatives, in which subjects are either unexpressed or in postverbal position. Instances that cannot be reliably analyzed, such as truncated verbs, are also excluded.

This leaves 5,571 instances of variable pronoun expression in NMSEB, and 2,802 in CCCS.<sup>1</sup> A first finding is that, consistent with other studies, the subject pronoun rate is higher for 1sg than for 3sg in the two datasets, at 29% vs. 18% in NMSEB (N = 3,296 and 2,275) and 50% vs. 32% in CCCS (N = 1,389 and 1,413). Notice that, for both persons, rates are higher in CCCS than in NMSEB. It is through the analyses that we are able to test the role of genre in this blanket difference, and assess whether an overall rate difference is tantamount to a linguistically significant difference in the structure of variation patterns (cf., Travis 2007).

We chose these corpora because they are prosodically transcribed following the same precise protocols. Prosodically-based transcription sharpens the analysis of linguistic structure, as syntax tends to align with prosody. A manifestation of the syntax-prosody connection is that words in the same prosodic unit tend to have a tighter syntactic relationship than material across prosodic units (Croft 1995, p. 849-864). Relevant to the study of variable subject expression is that subject pronouns tend to occur in the same prosodic unit as the verb with which they occur. Prosody thus provides an objective criterion for determining the syntactic status (as bonafide subjects or not) of pronouns that are detached from a verb or that appear to be dislocated (Torres Cacoullos & Travis 2018, p. 55).

In the transcription, the speech stream is segmented into prosodic units, here the Intonation Unit (IU)—“a stretch of speech uttered under a single, coherent intonation contour” (Du Bois, Schuetze-Coburn, Cumming & Paolino 1993, p. 47). Each IU is transcribed on a distinct line, and is followed by punctuation representing “transitional continuity”, that is, the prosodic contour with which it ends. In example (1) the commas marking the first two IUs indicate “continuing” intonation, while the period of the third IU indicates “final” intonation (a fall to low pitch) (Du Bois et al. 1993, p. 53). The three clauses presented in a sequence of IUs in (1) form a “prosodic sentence” (Chafe 1994, p. 139-140) and illustrate the role of prosody in the linking of clauses.

---

<sup>1</sup> The CCCS 1sg dataset is from a portion of the corpus (see Travis & Torres Cacoullos 2012); 3sg was extracted from all transcripts. While this skews the relative frequency of the 1sg and 3sg subjects, it does not affect the linguistic conditioning of the variability within each person (the contribution of linguistic context to the selection of a pronominal over an unexpressed subject).

- (1)
- |    |  |                                  |
|----|--|----------------------------------|
| a. | Ivette: <i>Ella venía pa' mi casa,</i>       | ‘She would come to my house,     |
| b. | <i>parqueaba su carro y luego,</i>           | she would park her car and then, |
| c. | <i>nos íbamos las dos con el ~Rubencito.</i> | we would both go with Ruben.’    |
- (NMSEB 06 El Túnico, 52:42-52:47)<sup>2</sup>

### 3 Reconsidering accessibility: Clause linking and distance

A widely reported constraint on subject expression is accessibility. The cross-linguistic generalization is that less “coding material” (or less phonetic bulk) corresponds to more accessible referents and more coding material (here, pronouns) to less accessible referents (Givón 1983, p. 18) (see also Ariel 1988; Chafe 1994, p. Ch. 6; Levinson 1987). This cross-linguistic tendency has been widely demonstrated for Spanish since Silva-Corvalán’s (1982, p. 104) seminal work on the effect of subject continuity: subject pronouns are disfavored in coreferential contexts, that is, where the target subject has the same referent as the subject of the immediately preceding clause.

Here we probe accessibility, to consider the impact of structural connectedness and distance. In order to pinpoint the distinct behavior of 1sg and 3sg, we apply a finer measure of accessibility—distinguishing clause linking for coreferential contexts (same reference) and distance from previous mention for non-coreferential contexts (switch reference). On the one hand, the effect of coreferentiality of contiguous subjects is modulated by structural connectedness—both prosodic and syntactic—between clauses, or *linking*. On the other, the effect of non-coreferentiality with respect to the immediately preceding subject is modulated by how far back the nearest coreferential subject is, or *distance*. We begin with distance.

#### 3.1 Distance from previous mention: Refining non-coreferential contexts

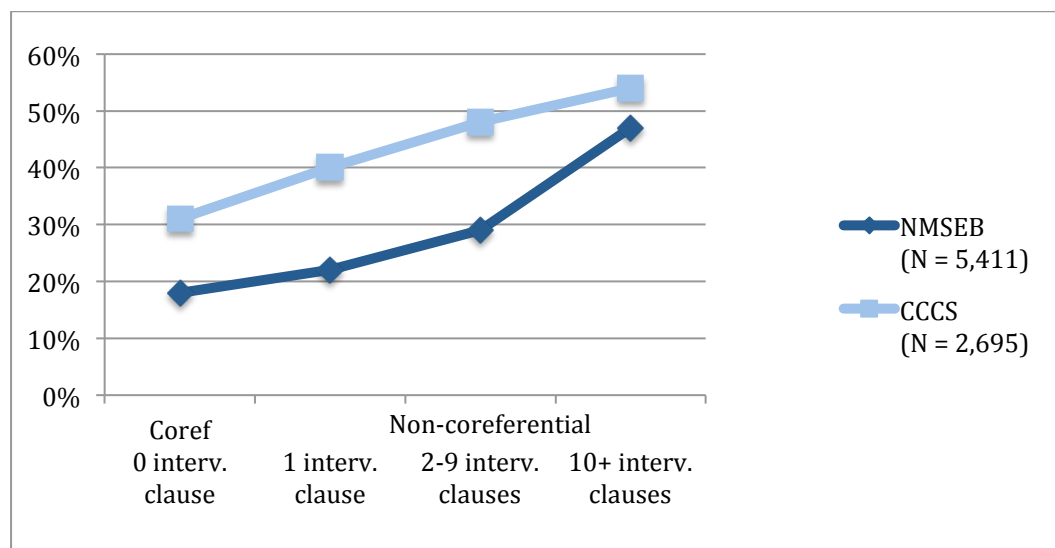
The gradient measure of distance as applied here considers the number of clauses intervening between the target subject and the *previous mention* of the same referent in subject position. This distance measure is inspired by the general measurement of topic continuity known as Referential Distance, which assesses the lapse from the previous mention of the referent (Givón 1983, p. 13; Myhill 2005, p. 473). For example, in (2), there are three intervening clauses (marked with dotted underlining) between the target in (g) and the previous mention in (a).

<sup>2</sup> Examples are reproduced verbatim from the transcripts, with the addition of underlining of pertinent material and insertion of Ø to indicate unexpressed subjects, marked in the English translation with parentheses around the pronoun. The information in parentheses following examples indicates the corpus, recording number and name, and identifiers of the lines reproduced (beginning-ending time stamps for NMSEB and line numbers for CCCS). See Appendix for transcription conventions.

- (2)
- |    |         |   |   |
|----|---------|---|---|
| a. | Betty:  | ... <i>Ø le fajé una nalgada.</i>                 | '... (I) <u>gave her</u> a slap on the bottom.'   |
| b. | Carrie: | .. @@@  | .. @@@  |
| c. | Betty:  | .. <i>she was starting to walk,</i>               | '.. <i>she was starting to walk,</i>              |
| d. |         | ... <i>and she wanted to go to the fireplace,</i> | ... <i>and she wanted to go to the fireplace,</i> |
| e. |         | <i>sit on the mantel,</i>                         | <i>sit on the mantel,</i>                         |
| f. |         | ... (1.0) <i>y no la dejaba la ~Bobbie.</i>       | ... (1.0) and Bobbie <u>wouldn't let her.</u>     |
| g. |         | <i>yo no sé qué estarían haciendo allá.</i>       | I <u>don't know</u> what they were doing there.'  |
- (NMSEB 13 La Acequia, 18:16-18:28)

As seen in Figure 1, in both datasets, the subject pronoun rate is lowest in coreferential contexts (with zero intervening clauses, i.e., where the target subject is coreferential with the subject of the immediately preceding clause), and highest at ten or more clauses. In between there are step-by-step increases at one intervening clause, and then at two–nine intervening clauses.<sup>3</sup> Note that while the pronoun rate in CCCS is consistently above that of NMSEB, the key point here is that the effect is the same in both datasets. The conclusion is that distance from the previous mention is pertinent for Spanish subject expression, beyond binary switch vs. same reference with respect to the subject of the immediately preceding clause.

Figure 1 Rate of pronoun expression according to distance from previous mention



NMSEB: 18%, 507/2,846; 22%, 168/766; 29%, 331/1,137; 47%, 311/662  
 CCCS: 31%, 372/1,187; 40%, 124/311; 48%, 262/549; 54%, 347/648<sup>4</sup>

<sup>3</sup> The difference between zero and one clause (coreferential vs. local switch reference contexts) turns out to be smaller in the absence than in the presence of intervening human subjects. How the accessibility measure of human switch reference (Travis & Torres Cacoullos 2012, p. 727-729) intersects with that of linking (see Section 3.2) remains to be investigated.

<sup>4</sup> The overall rise from 0 to 10+ intervening clauses is significant for both NMSEB ( $F(3, 5,407) = 93.37, p < .001$ ) and CCCS ( $F(3, 2,691) = 33.88, p < .001$ ). (Statistical results presented are based on Tukey post-hoc tests conducted on

### 3.2 Clause linking, prosodic and syntactic: Refining coreferential contexts

Structural linking applies to coreferential-subject clauses, and brings in the consideration of syntactic and/or prosodic linking. *Syntactic linking* is realized by means of a conjunction, whether a coordinating conjunction, usually *y* ‘and’, or a subordinating conjunction such as *que* ‘that (complementizer)’ or *porque* ‘because’. Syntactic linking is illustrated with a coordinate clause (operationalized here as a clause that is conjoined with a coreferential-subject clause via a coordinating conjunction) in (3) and with a subordinate clause in (4).

- (3)
- |    |         |  |   |
|----|---------|--|---|
| a. | Ángela: | ... <i>Yo a los dos quiero,</i>          | ‘... I <u>love</u> both of you,         |
| b. |         | <i>y=,</i>                               | and,                                    |
| c. |         | ... <i>y yo tengo que ser imparcial.</i> | ... and I <u>have to be</u> impartial.’ |
- (CCCS 02 Restaurante, 356-358)

- (4)
- |    |           |                              |   |
|----|-----------|------------------------------|---|
| a. | Patricia: | <i>Ø no lo voy a botar,</i>  | ‘(I) <u>am not going to throw</u> it out, |
| b. |           | <i>porque Ø lo necesito.</i> | because (I) <u>need it.</u> ’             |
- (CCCS 09 Diccionario, 259-260)

*Prosodic linking* of grammatical elements in speech may be assessed in a replicable manner through prosodically-based transcription (see Section 2, above). Prosodically linked clauses are those produced by the same speaker that occur either in adjacent IUs connected by continuing intonation (indicated by a comma), as in (5), or in the same Intonation Unit (IU), as in (6) (where they are also syntactically linked).

- (5)
- |    |           |  |                                      |
|----|-----------|--|--------------------------------------|
| a. | Santiago: | <i>Ø <u>Me fui</u> pa’ mi casa,</i>          | ‘(I) <u>went</u> home,               |
| b. |           | .. <i>Ø <u>dormí</u> perfectamente bien.</i> | .. (I) <u>slept</u> perfectly well.’ |
- (CCCS 15 Clínica, 389-390)

- (6)
- |    |        |   |  |
|----|--------|---|--|
| a. | Nancy: | <i>Vanessa se va a ir y Ø me va a <u>dejar so=la.</u></i> | ‘Vanessa <u>is going to leave</u> and (she) <u>is going to leave me all alone.</u> ’ |
|----|--------|---|--|
- (CCCS 23 Blusas, 294)

All other cases are considered not prosodically linked, be that due to non-continuing intonation (final, appeal or truncation), as in (7); non-adjacency because of the presence of one or more intervening prosodic units, as in (8); or because the immediately preceding coreferential-subject clause is produced by another speaker, as in (9).

---

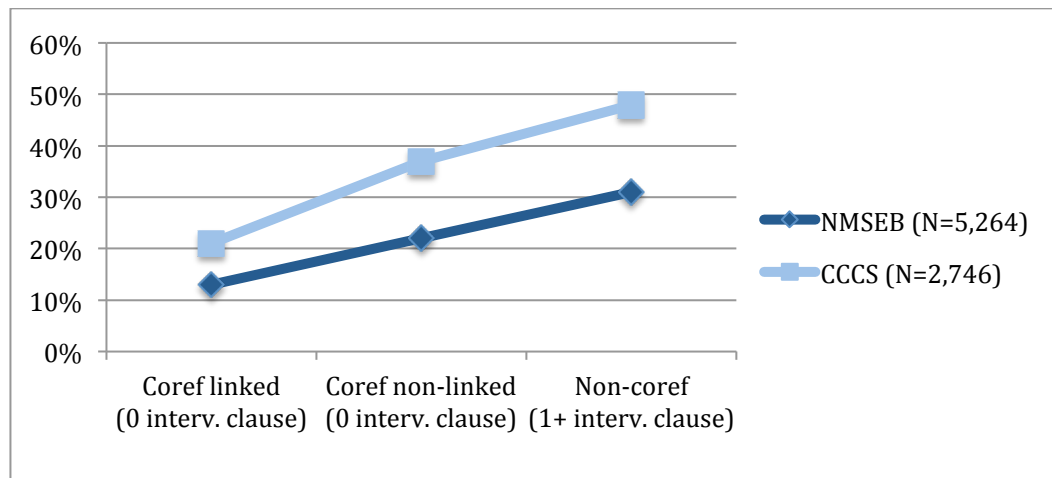
separate one-way Anovas fit to each dataset.) Due to analysis-specific exclusions, token numbers in the cross-tabulations vary; excluded here are cases where it was not possible to determine the exact number of intervening clauses, for example, due to unclear speech.

- (7)
- |    |           |   |   |
|----|-----------|---|---|
| a. | Santiago: | <i>Ø <u>Me cortó</u> en todos esos tres<br/>punticos.</i> | ‘(He) <u>cut me</u> on these three spots.                                   |
| b. |           | <i>.. <u>Y este dedo Ø me lo apretó,</u></i>              | <i>.. And (he) <u>squeezed this finger.</u></i><br>(CCCS 16 Chuzo, 201-202) |
- (8)
- |    |        |   |  |
|----|--------|---|--|
| a. | Julia: | <i>... <u>Entonces,</u></i>                                     | ‘... So,   |
| b. |        | <i>.. <u>Que Ø quería ir a trabajar al<br/>restaurante,</u></i> | <i>.. (he) <u>wanted to go</u> and work at the<br/>restaurant,</i>                               |
| c. |        | <i>entonces,</i>  | so,  |
| d. |        | <i>supuestamente,</i>   | supposedly,  |
| e. |        | <i>en un mes,</i>   | in one month,  |
| f. |        | <i>Ø <u>iba a ganar un millón y pico,</u></i>                   | <i>(he) <u>was going to earn</u> over a<br/>million ((pesos)),</i><br>(CCCS 03 Familia, 688-693) |
- (9)
- |    |          |                                      |  |
|----|----------|--------------------------------------|--|
| a. | Fabiola: | <i>Ø <u>tenía mucho</u> cabello,</i> | ‘(He) <u>had</u> a lot of hair,                                      |
| b. |          | <i>o no?</i>                         | or not?’   |
| c. | Molly:   | <i>...(1.0) Ø <u>tenía poco,</u></i> | ‘...(1.0) (he) <u>had</u> little,                                    |
| d. |          | <i>.. no no muy muchote no.</i>      | <i>.. no not a whole lot no.’</i><br>(NMSEB 09 La Salvia, 2:00-2:04) |

In the proposed refinement, subject continuity comprises a combination of semantic and structural features, with the semantic link of coreferentiality being broken into two categories based on the presence or absence of structural linking. The hypothesis is that, within coreferential contexts, unexpressed subjects are favored more in clauses that are structurally linked—syntactically and/or prosodically—than in those that have no such linking.

Figure 2 shows that structural linking does make a difference: a pronoun is least likely if the subject is both coreferential with the preceding subject and is prosodically and/or syntactically linked to it; it is most likely in non-coreferential contexts. Merely coreferential contexts, in the absence of structural linking, are intermediate. Again, though the rate is higher overall in CCCS, this holds for both datasets.

Figure 2 Rate of pronoun expression according to structural linking between clauses—prosodic and syntactic



NMSEB: 13%, 178/1,384; 22%, 270/1,238; 31%, 829/2,642  
 CCCS: 21%, 98/458; 37%, 261/700; 48%, 767/1,588<sup>5</sup>

It is important to recognize that the linking effect is not a surrogate for clause type effects (whereby coordinate clauses and, in some studies, subordinate clauses, disfavor pronominal subjects).<sup>6</sup> First of all, there can be syntactic linking (via a conjunction) without prosodic linking, as in (7). Furthermore, prosodic linking also applies to non-coordinate main clauses (as in (5) above). Indeed, one third of main clause structurally linked tokens are linked via prosody alone (312/946 in NMSEB and 102/291 in CCCS), and the rate of pronoun expression is equally as low as that in clauses with a coordinating or subordinating conjunction (at 15% in NMSEB and 21% in CCCS).

#### 4 The workings of accessibility for 3sg vs. 1sg

It makes sense that there would be a true difference between first and third person in relation to their accessibility. While a first person referent is thought to be always accessible as a discourse participant the same is not so of third person referents which are typically made accessible through a mention in the discourse (Chafe 1994, p. 78-79; Dahl 2000, p. 64-66; Prince 1981). But in fact, both first and third person are conditioned by accessibility.

Figure 3 shows, in both datasets, the rate of expression of 1sg and 3sg subjects by linking and distance. The rate of expression is confirmed to be higher for 1sg than for 3sg at every degree of distance. Nevertheless, the similarities are striking. On the

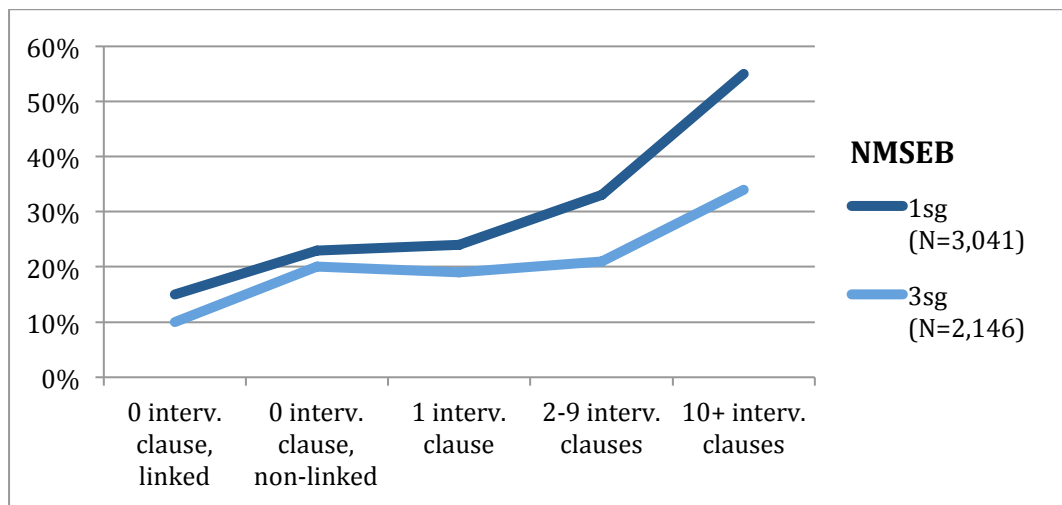
<sup>5</sup> The rise from coreferential-linked to non-coreferential is significant for both datasets (NMSEB  $F(2, 5,261) = 90.33, p < .001$ ; CCCS  $F(2, 2,743) = 58.17, p < .001$ ).

<sup>6</sup> Coordinate clauses have a lower pronoun rate than non-coordinate main clauses; reports on subordinate clauses are inconsistent (see, e.g., Enríquez 1984, p. 257; Orozco 2015, p. 22; Otheguy & Zentella 2012, p. 164; Shin 2014, p. 211; Travis 2007, p. 115).

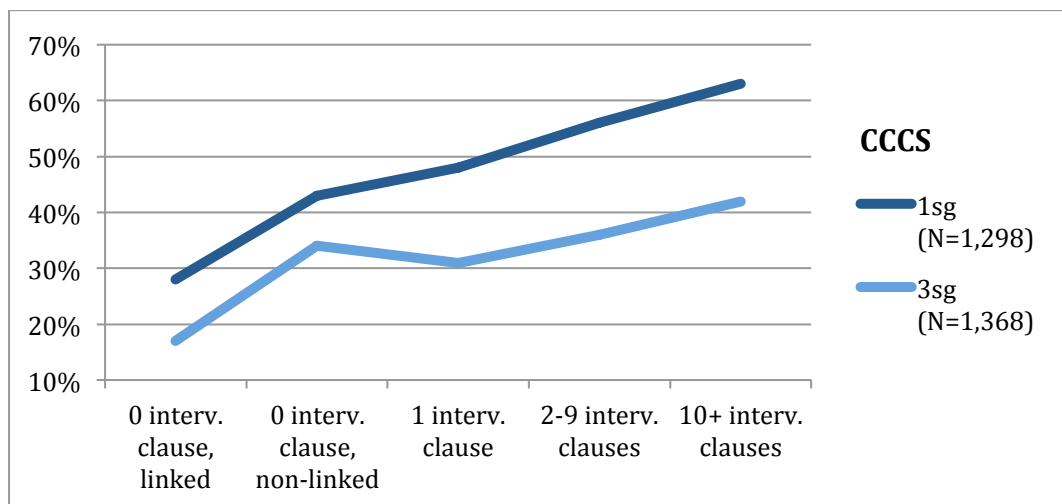


one hand, the same linking effect described in the previous section applies to both persons. This is seen in two ways. First, the subject pronoun rate increases from linked to non-linked coreferential contexts by a factor of approximately 1.5 to 2. Second, there is little difference between non-linked coreferential contexts and non-coreferential contexts with a distance of one intervening clause (increasing at most by a factor of 1.1). Thus, structural linking is relevant in the application of accessibility for both persons. On the other hand, also apparent for both persons is a distance effect, with the rate of subject expression greater when there are ten or more intervening clauses than when there is just one intervening clause from the previous mention. In sum, for both persons, effects of structural linking and of distance are clearly apparent.

Figure 3 Rate of pronoun expression according to structural linking and distance, for 1sg vs. 3sg



1sg: 15%, 112/746; 23%, 158/678; 24%, 109/458; 33%, 247/739; 55%, 229/420  
 3sg: 10%, 66/638; 20%, 112/560; 19%, 59/308; 21%, 84/398; 34%, 82/242



1sg: 28%, 50/179; 43%, 116/268; 48%, 76/157; 56%, 180/323; 63%, 232/371  
 3sg: 17%, 48/279; 34%, 145/432; 31%, 48/154; 36%, 82/226; 42%, 115/277

For the first person, the application of the accessibility effect (that is, the use of pronouns for less accessible referents) has been interpreted in the sense of speakers “bringing the ideas of themselves back into the active consciousness of the listeners”

(Chafe 1994, p. 87). For the third person, it is often assumed that, as third person referents are outside “the instance of discourse” (as Benveniste (1971) put it) and have to be introduced, they are “by default of rather low accessibility” (Ariel 2000, p. 218). Thus, For Spanish, it has been predicted that “contextual salience or discourse-induced accessibility should play a smaller role [for first person] than in the case of anaphoric third person” (Posio 2013, p. 257).

Within the variable context for subject pronoun expression, though both persons are sensitive to accessibility, they do indeed differ—but not quite as predicted. Once we refine accessibility by the discerning measures of clause linking and distance, we are able to ascertain that the difference lies in the way the effect applies. The effect begins to impact pronoun rate earlier for 1sg, actually, than for 3sg. This is depicted in Figure 3, where, for 1sg, we observe a rise from one through to two-nine and then again at 10 or more intervening clauses for both NMSEB (top panel) and CCCS (bottom panel). For 3sg, on the other hand, in both data sets, a rise begins later and is smaller than for 1sg.<sup>7</sup>

In sum, the differential implementation of accessibility effects with the two grammatical persons is similar in the two datasets. It appears, then, that unexpressed and pronominal 3sg subject referents are treated as remaining locally accessible—the pronoun rate stays more or less flat—for a longer chunk of discourse than 1sg are. For an explanation, we look to the contextual distributions of the grammatical persons, which will bring us to the variable context for subject expression.

## 5 Contextual distribution

### 5.1 *Lexical vs. pronominal and unexpressed 3sg subjects*

We have seen that the rate of subject pronouns is higher for 1sg than 3sg when comparing the personal pronoun vs. unexpressed variants (overall, 29% vs. 18% in NMSEB). There is, however, a third option for 3sg subjects, namely a lexical, or full noun phrase subject, as in line (a) in (10).

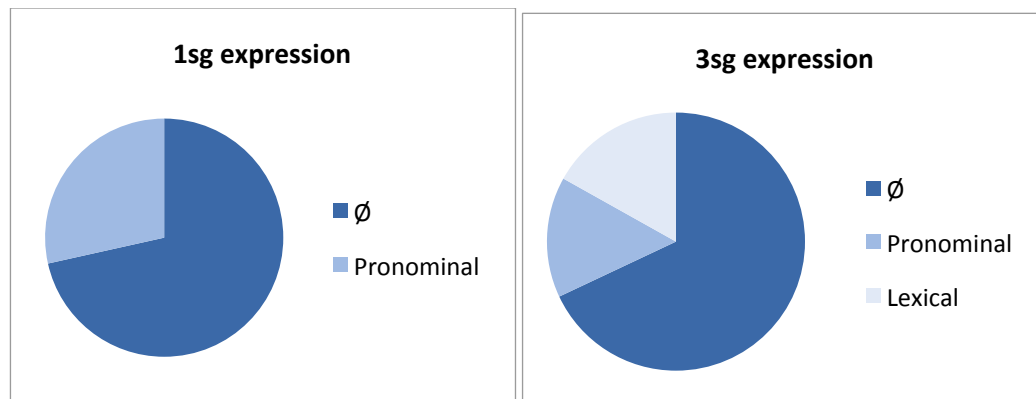
---

<sup>7</sup> For 3sg, there is no significant difference in the rate of pronominal expression from coreferential non-linked contexts to 2-9 intervening clauses in either dataset (NMSEB  $F(2, 1,263) = 0.21, p = .81$ ; CCCS  $F(2,809) = 0.35, p = .71$ ), while for 1sg the difference is significant (NMSEB  $F(2, 1,872) = 11.20, p < .001$ ; CCCS  $F(2, 745) = 4.65, p < .001$ ).

- (10)
- |    |            |  |   |
|----|------------|--|---|
| a. | Bartolomé: | <i>porque m- mi papá era rancharo y,</i>           | ‘becuase m- <u>my dad</u> was a rancher and,’         |
| b. |            | <i>... Ø vendía verduras,</i>                      | ... (he) sold vegetables,                             |
| c. |            | <i>y=,</i>   | and,’   |
| d. | Gabriel:   | <i>sí=.</i>  | ‘yes.’  |
| e. | Bartolomé: | <i>y= ... y veníamos a la plaza aquí a vender,</i> | ‘and ... and we would come to the plaza here to sell, |
| f. |            | <i>...(1.6) los blanquillos .. los huevos.</i>     | ...(1.6) the little white ones .. the eggs.           |
| g. |            | <i>... a diez centavos la docena.</i>              | ... for ten cents a dozen.’                           |
| h. | Gabriel:   | <i>.. uh huh &lt;@ diez centavos @&gt;?</i>        | ‘.. uh huh <@ ten cents @>?’                          |
| i. |            | <i>no me digas.</i>                                | you don’t say.’                                       |
- (NMSEB 02 La Marina, 38:30-38:44)

Figure 4 gives the distribution of subject forms with this three-way breakdown for 3sg, corresponding to the two-way breakdown for 1sg (here, as above, for human, specific preverbal subjects).<sup>8</sup> Pronominal tokens still make up proportionally twice as much of the 1sg data (29%) than of the expanded 3sg data (15%) once we add in lexical subjects (17%). However, the proportion of unexpressed subjects is now virtually the same for the two persons: 72% for 1sg, compared with 68% for 3sg. Thus, the widely reported higher rate of unexpressed 3sg subjects is at least in part due to the fact that there exists a third option that, at first blush, can be taken to compete with pronominal and unexpressed subjects (cf., Gudmestad, House & Geeslin 2013; Posio 2015, p. 72).

Figure 4 Distribution of 1sg and 3sg subject forms (NMSEB)



1sg: 71% unexpressed, 29% pronoun (N = 3,296)

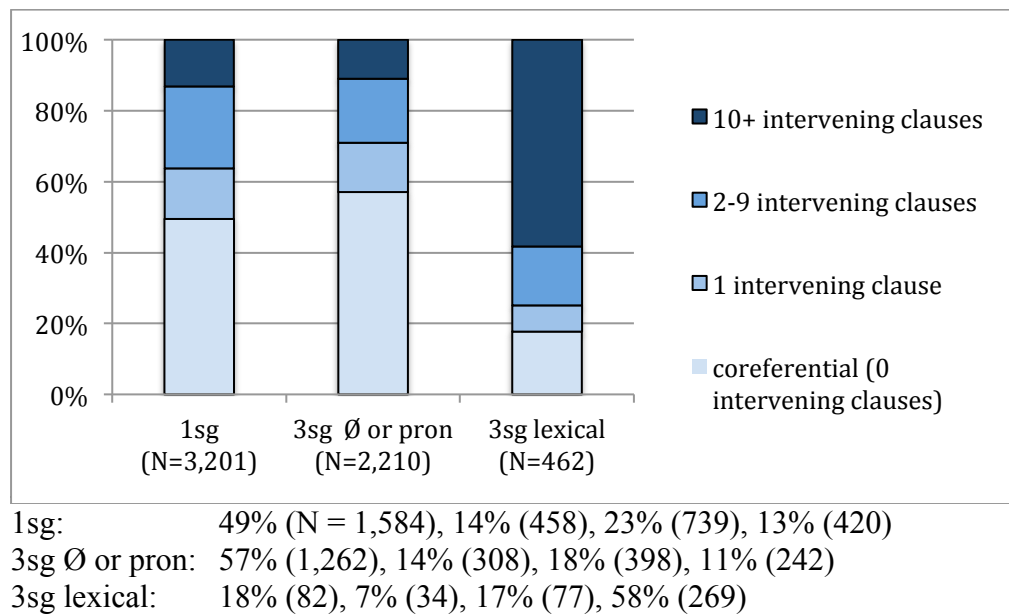
3sg: 68% unexpressed, 15% pronoun, 17% lexical (N = 2,737)

What impact might the use of lexical subjects have on the workings of accessibility in the choice between 3sg pronominal and unexpressed subjects? Observe the distribution of subject forms according to distance from the previous mention, in Figure 5. The first two bars give the breakdown for 1sg and 3sg subjects, considering

<sup>8</sup> Comparison with CCCS is not available, because lexical subjects were not extracted.

pronominal and unexpressed forms; the third bar gives the breakdown for 3sg lexical subjects. Comparing the first two bars, we see that pronominal and unexpressed 3sg subjects tend to occur closer together than 1sg subjects, appearing proportionally more in coreferential contexts and less with a previous mention at two or more intervening clauses (at one intervening clause, the proportion is the same for both persons).<sup>9</sup> 3sg lexical subjects, in turn, differ from both 3sg pronominal and unexpressed and 1sg, as they occur over half the time at a distance of ten or more intervening clauses from their previous mention.<sup>10</sup>

Figure 5 Distribution of 1sg and 3sg subject forms according to distance from previous mention ( $\emptyset$  and pronominal vs. lexical) (NMSEB)



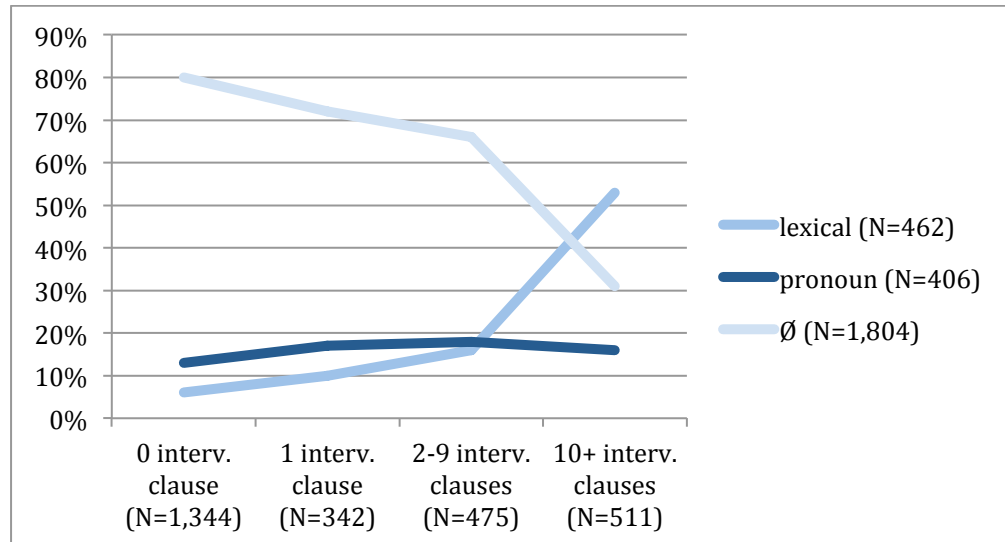
Full NPs are the favored 3sg form to introduce a new referent, or to return to a prior referent that hasn't been mentioned for some time (cf., Dumont 2016, p. 84). This we confirm in Figure 6, which gives the proportion of  $\emptyset$ , pronominal, and lexical subjects at different degrees of distance, for 3sg. At a distance of 0 intervening clauses, unexpressed subjects predominate (80% of subjects in coreferential contexts are unexpressed). At the other end, at a distance of 10+, that is, for referents without a previous mention in subject position within the preceding ten clauses, a lexical form is

<sup>9</sup> Similarly in CCCS, the distribution of  $\emptyset$  and pronominal subjects according to distance from previous mention indicates that 3sg subjects occur closer together than 1sg subjects (see token counts in Figure 3).

<sup>10</sup> In comparing 3sg lexical forms vs. 3sg pronouns and  $\emptyset$  (third and second bars), a generalized linear regression with a logit link function demonstrates that lexical forms are significantly favored more at greater degrees of distance between mentions (0 vs. 1:  $z = -2.48, p < .05$ ; 0 vs. 2-9:  $z = -6.47, p < .001$ ; 0 vs. 10:  $z = -19.67, p < .001$ ). In comparing 1sg pronouns and  $\emptyset$  vs. 3sg pronouns and  $\emptyset$  (first and second bar), 1sg pronouns and  $\emptyset$  are favored more the greater the distance (0 vs. 1:  $z = -2.05, p = .05$ ; 0 vs. 2-9:  $z = -5.38, p < .001$ ; 0 vs. 10:  $z = -3.64, p < .001$ ).

much more likely (53%) than either a pronominal (16%) or unexpressed (31%) subject.<sup>11</sup>

Figure 6 Proportions of 3sg subject forms according to distance (∅ vs. pronominal vs. lexical) (NMSEB)



0: 80% ∅, 13% pronoun, 6% lexical  
 1: 72% ∅, 17% pronoun, 10% lexical  
 2-9: 66% ∅, 18% pronoun, 16% lexical  
 10+: 31% ∅, 16% pronoun, 53% lexical

## 5.2 The clustering of 3sg pronominal and unexpressed subjects

Precisely because speakers tend to use a lexical form for less accessible 3sg subjects, the occurrence of unexpressed and pronominal targets is depressed at greater degrees of distance from the previous mention. Instead, as we saw in Figure 5, 3sg *personal pronoun and unexpressed* subjects tend to *cluster* together (occur with no intervening clauses from their previous mention) more than 1sg subjects do.

The greater clustering of 3sg (unexpressed and pronominal) than 1sg subjects can be explained by the deictic properties of these subject persons: as external parties, 3sg subject referents are introduced, continue in the discourse briefly as topics of discussion, and are then abandoned (as in (10)), while 1sg subject referents can be referred to at any time in the conversation without introduction. Dahl similarly finds greater “clustering” for third person pronouns than for first and second person pronouns in Swedish conversation and remarks that “once you have started talking about a third person referent, the chance that you will continue doing so also in the following clause is much higher than in the case of egophoric [1sg and 2sg] referents, other things being equal” (2000, p. 65). That is, “the typical behavior of third person referents is to stay on the scene once they have been introduced while egophoric

<sup>11</sup> The lexical subject proportion in NMSEB is one-tenth (11%, 36/322) at a distance of 2-4 intervening clauses between the target subject and previous mention but reaches a proportion of one fourth (27%, 41/153) at a distance of 5-9 intervening clauses. How far back to look for a previous mention in subject expression studies remains an empirical question.

referents pop in and out all the time” (2000, p. 66). As Dahl (2000, p. 66) points out, 3sg clustering may be part of the explanation for zero-marked 3sg agreement (alongside frequency-based explanations, see Bickel, Witzlack-Makarevich, Zakharko & Iemmolo 2015, p. 47).

We have verified here that the person difference is found in the greater proportion of 3sg tokens occurring in coreferential contexts (see Figure 5). In other words, the “clustering” of 3sg is indeed the higher proportion of occurrences in adjacent clauses. Consistent with this trend is that a larger proportion of the 3sg than 1sg data occur in structurally linked contexts (see Figure 3). The clustering of 3sg is thus quite tight, and short lived in discourse. In this sense, we can characterize 3sg as a *transient* person.

The contextual distributions contribute to the higher rate of pronouns with 1sg subjects, since 3sg (unexpressed and pronominal) subjects occur proportionally more often in coreferential contexts which, by the effect of subject continuity, disfavor pronominal subjects. Contextual distribution goes a long way to explaining pronoun rate differences by grammatical number as well. One consistent finding for Spanish is that plural subjects have a lower rate of pronominal vs. unexpressed subjects (e.g., Bentivoglio 1987, p. 36; Otheguy et al. 2007, p. 791). This result is a byproduct of the greater propensity of plural than singular subjects to occur in the context of a previous partially coreferential mention (Cameron 1995, p. 21). Contextual distribution is typically controlled in lab-based studies and often neglected in corpus studies focused on significance of predictor effects. But the lesson here is that contextual distributions can illuminate quantitative usage patterns.

### 5.3 Contextual distribution and genre

The same clustering of 3sg in the NMSEB sociolinguistic interviews is evident in the CCCS conversations among family and friends. Where there is a difference between the corpora is in aggregate data distribution, with proportionally fewer CCCS subjects overall in coreferential contexts (36% and 52% for 1sg and 3sg respectively, vs. 49% and 57% for NMSEB) and more at greater degrees of distance of 10+ clauses (28% and 20% for 1sg and 3sg respectively, about double the proportions in NMSEB, at 13% and 11%) (see Figure 5 and Footnote 9).

The difference is attributable to genre: more dialogic interaction involves dense switching back and forth between interlocutors.<sup>12</sup> There is *less subject continuity*—seen in the lower proportion of clauses in which the subject is coreferential with the immediately preceding clause—and *more shifting of topics*, seen in the greater proportion of tokens in non-coreferential contexts at precisely longer distances from previous mention (43%, 648/1,508, of all non-coreferential subjects in CCCS occur at a distance of 10 or more intervening clauses, compared with just 26% (662/2,565) in NMSEB,  $p < .001$ ). Note that the genre difference lies in data distribution rather than

---

<sup>12</sup> In Figure 3, there is an increase in the subject pronoun rate between non-linked coreferential contexts (0 intervening clauses) and 1 intervening clause for 1sg but not for 3sg in CCCS. This is partly attributable to immediately preceding coreferential-subject clauses produced by the interlocutor (by definition, not prosodically linked), which account for 19% and 30% of 1sg and 3sg coreferential tokens, respectively, in the more dialogic CCCS (compared with 11% and 13% only, in NMSEB) (on interlocutor-produced previous mentions, see Torres Cacoullos & Travis 2018, p. 87).

in conditioning of variant choice, or distinct grammars (Travis 2007; Travis & Lindstrom 2016). More generally, contextual distributions of the data may give rise to aggregate rate differences that may in turn mask similarities in the conditioning of variant choice.

In sum, grammatical person differences in rates of pronominal vs. unexpressed subjects are in part attributable to the availability of 3sg full NPs. Evident are different distributions according to contexts impinging on subject expression, with 1sg subjects tending to occur more than 3sg subjects (setting aside full NPs) in environments that are propitious to subject pronouns, namely in non-coreferential contexts.<sup>13</sup> In contrast, 3sg subjects (again setting aside full NPs) tend to cluster together (Figure 5). Beyond contextual distributions, accessibility via previous mention has minimal impact at low degrees of distance for 3sg, such that the rate of subject pronouns remains comparatively flat (here, up to 10 clauses from the previous mention) (Figure 3, Section 4).

## 6 Inside and outside the variable context

Contextual distributions of unexpressed, pronominal, and lexical forms with respect to distance give a first indication that lexical subjects are the odd man out: lexical subjects are a very minor player when the previous mention occurs at a distance of under ten clauses, but at ten clauses, they are a robust contender, in fact, the preferred variant (Figure 5 and Figure 6). We can say, then, that *pronominal and unexpressed subjects are grammatical means of referring to an accessible subject*. This is the definition of the variable context for subject expression. As full NPs do something else, they are outside this envelope of variation.

True, subject pronouns are taken to be semantically distinct from unexpressed subjects, being used to mark “contrast”, “emphasis” or “focus” (e.g., Chafe 1994, p. 37; Payne 1997, p. 43). Such functions have been widely claimed not only for Spanish (e.g., Davidson 1996; Posio 2013; Serrano 2014) and other Romance languages (Mayol 2010), but for a range of languages in which subject expression is variable, including Finnish (Helasvuo 2014, p. 454), Japanese (Lee & Yonezawa 2008, p. 741-739), and Javanese (Ewing 2014, p. 55-56).

In treating pronominal and unexpressed subjects as competing variants of the same linguistic variable, we do not deny that meaning differences may come into play in some contexts. Studies that have operationally tested for contrast in everyday speech indeed report a favoring of pronouns in particular constructions with contrastive elements (such as *sólo* ‘only’) but also that contrastive contexts arise so rarely that they account for a minor portion of the data (e.g., Bentivoglio 1987, p. 46-48; Paredes Silva 1993, p. 41-43; Travis & Torres Cacoullos 2012, p. 714-723). Thus, rather than assume that potential differences in meaning drive the choice between forms in every given instance, we work on the basis of the hypothesis that differences in grammatical function can be neutralized in discourse (Sankoff 1988, p. 153). Neutralization-in-

---

<sup>13</sup> The second contextual distribution difference between the persons is that 1sg, but not 3sg, subjects are associated with cognition verbs (Bentivoglio 1987, p. 51; Torres Cacoullos & Travis 2018, p. 102). In Spanish, separate regression analyses by person confirm that semantic class has a significant impact for 1sg—cognition verbs favoring pronominal subjects (Travis & Torres Cacoullos 2012, p. 734-742), but is “not relevant” for 3sg (Shin 2014, p. 325).

discourse is a requirement for (most) language change, since for one form to take over from another there must be a period of overlap in use. It is also the hypothesis for stable variation as in the case of Spanish subject expression; the antithesis, that grammatical alternatives necessarily always reflect communicative differences, must also remain a hypothesis (Sankoff 1988, p. 154).

The neutralization-in-discourse hypothesis permits us to circumscribe a variable context, or locus of variability, which in turn permits us to state differences between the variants. If linguistic forms tend to co-occur with contextual elements harmonious with their meaning, forms with different meanings should be preferred in different (sub)contexts (Aaron & Torres Cacoullos 2005, p. 615; Poplack 2001, p. 405). Differences in how the variants are deployed within the variable context are manifested in the direction of effect of conditioning factors—contextual elements probabilistically constraining the choice—favoring one variant, disfavoring the other.

Forms whose use is affected by different conditioning factors do not belong to the same variable context. Consider postverbal subject pronouns, illustrated in (11). These are properly outside the variable context for subject pronoun expression (e.g., Cameron 1992, p. 116). The reason is that subject pronoun position is sensitive to different contextual features from those affecting expression. Postverbal placement of the subject pronoun is strongly favored in the presence of preverbal elements (adverbs, objects) in the clause and in the prosodic unit (Benevento & Dietrich 2015, p. 415; Silva-Corvalán 1982, p. 113), features that are not, as far as we know, relevant to subject expression.

(11)

Mariana:  *puras mujercitas tengo yo.*

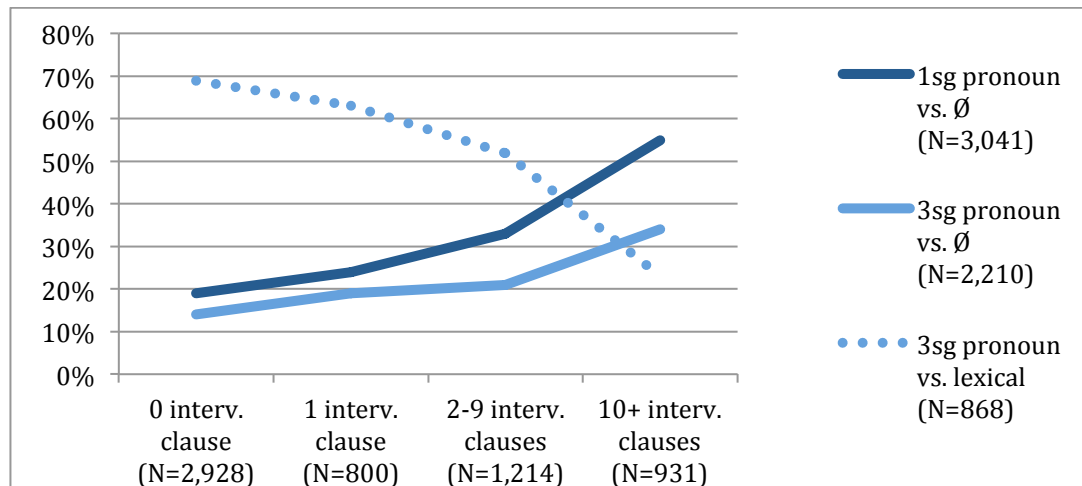
‘I<sub>POST</sub> have all girls.’  
(NMSEB 19 School Bus, 40:05-40:07)

Confirmation that lexical subjects are not part of the same envelope of variation as unexpressed and pronominal subjects is that their inclusion provokes a contrary accessibility effect. Returning to Figure 6 above, which depicts proportions of 3sg subject forms by distance, notice that the proportion of unexpressed 3sg subjects gradually diminishes from zero to nine intervening clauses from the previous mention, and then radically drops once the distance reaches ten clauses. The proportion of pronouns, however, hardly changes according to distance, presenting a flat line at all degrees of distance (hovering around 16%); the difference is that at 0-9 intervening clauses unexpressed subjects predominate while at 10+ lexical forms do. The steady proportion of pronouns regardless of distance from the previous mention—when lexical subjects are counted—is consistent with the report that when 3sg lexical, pronominal, and unexpressed subject forms are all considered in a single analysis, switch reference is pertinent for the choice of full NPs, but not personal pronouns, over unexpressed subjects (Gudmestad et al. 2013, p. 287).

Figure 7 now depicts subject pronoun rates as a choice over unexpressed subjects for 1sg and for 3sg (in the solid lines, seen previously in Figure 3), and over a lexical form for 3sg pronouns (in the dotted line). In opposition to unexpressed subjects, the 3sg pronoun rate rises with distance, as it does for 1sg; in opposition to a lexical form, it declines. What Figure 7 clearly confirms, then, is that the relevant comparison across the different grammatical persons is pronoun vs. Ø: within this variable context, 1sg and 3sg respond similarly to accessibility. At the same time, by establishing refined measures of subject continuity, a difference has been discovered in the workings of accessibility, namely the earlier and sharper pronoun rise with increasing distance for 1sg as compared with 3sg (Section 4 above).



Figure 7 Rate of subject pronouns according to distance (NMSEB):  
 1sg pronoun vs.  $\emptyset$ ,  
 3sg pronoun vs.  $\emptyset$ ,  
 3sg pronoun vs. lexical



Ricardo Otheguy has called for “handling data without relying on antecedently given formal or relational structure” (2002, p. 401). What are unexpressed subjects the non-expression of? Here we have scrutinized 3sg subjects inside and outside the variable context for subject expression as it is usually and often unreflectingly circumscribed. Eschewing analytical reliance on a property of linguistic structure taken a priori as a given, we relied on actual usage to interpret the workings of referent accessibility in discourse.

The conclusion is that lexical subjects affect patterns within the variable context for subject expression but are appropriately left outside it. We began by treating the presence of a subject pronoun and the absence of any subject noun phrase, commonly termed an unexpressed or null ( $\emptyset$ ) subject, as “alternative ways of ‘saying the same thing’” (cf., Labov 1972, p. 72). The analyses lead us to say that the “same thing” here is grammatical means of mentioning an accessible subject. The variants are appropriately labeled in terms of *expression* or presence vs. absence of the *pronoun*, because the absence of any subject noun phrase is truly in variation with pronominal, not lexical, subjects.

This exercise confirms that linguistic meaning “is [not] the beginning of analysis; ... it is the end result” (Otheguy 2002, p. 388). In the analysis of variation, “the definition of a linguistic variable is the first and also the last step” (Labov 2004, p. 7).

## 7 Conclusion

Let us summarize the findings on contextual effects and distributions. Accessibility is operationalized in the data of discourse as clause linking—prosodic and syntactic—and distance from the previous mention, measured here by the number of intervening clauses from the previous coreferential subject. The use of lexical subjects at greater degrees of distance contributes to distinct contextual

distributions of unexpressed and pronominal subjects for 3<sup>rd</sup> person. As compared with 1sg, 3sg unexpressed and pronominal subjects are more likely to occur in a coreferential context. They are also less impacted by accessibility at shorter distances from the previous mention. The clustering of 3sg unexpressed and pronominal subjects means that their referents tend to be talked about in adjacent clauses.

On this basis, we offer an amendment of the characterization of the third person as a “non-person” (Benveniste 1971). Third person is a transient person, in Spanish, in the form of clustered pronominal and, especially, unexpressed subjects. The amendment generates a hypothesis, testable in the data of language use.

## Acknowledgements

We gratefully acknowledge support from the National Science Foundation (<http://nmcodeswitching.la.psu.edu>), the ARC Centre of Excellence for the Dynamics of Language, and the project teams at the Pennsylvania State University and Australian National University. We thank James Grama and Grant Berry for help with the statistical analyses, the anonymous reviewers, and especially the editors of the volume, Naomi Shin and Danny Erker, for their comments on the paper.

## References

- Aaron, J. E., & Torres Cacoulios, R. (2005). Quantitative measures of subjectification: A variationist study of Spanish *salir(se)*. *Cognitive Linguistics*, 16(4), 607-633.
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24(1), 65-87.
- Ariel, M. (2000). The development of person agreement markers: From pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (pp. 197-260). Stanford: CSLI.
- Bayley, R., & Pease-Alvarez, L. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, 9(3), 349-371.
- Benevento, N., & Dietrich, A. (2015). I think, therefore *digo yo*: Variable position of the 1sg subject pronoun in New Mexican Spanish-English code-switching. *International Journal of Bilingualism*, 19(4), 407-422.
- Bentivoglio, P. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Benveniste, É. (1971). *Problems in General Linguistics* (trans. by M. E. Meek). Coral Gables, Fa: University of Miami Press.
- Bickel, B., Witzlack-Makarevich, A., Zakharko, T., & Iemmolo, G. (2015). Exploring diachronic universals of agreement: Alignment patterns and zero marking across person categories. In J. r. Fleischer; E. Rieken; & P. Widmer (Eds.), *Agreement from a diachronic perspective* (pp. 29-51). Berlin: De Gruyter.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins.
- Cameron, R. (1992). Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Cameron, R. (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics*, 6-7, 1-27.
- Carvalho, A. M., Orozco, R., & Shin, N. L. (2015). Introduction. In A. M. Carvalho; R. Orozco; & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. xiii-xxvi). Georgetown: Georgetown University Press.
- Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Claes, J. (2011). ¿Constituyen las Antillas y el Caribe continental una sola zona dialectal? Datos de la variable expresión del sujeto pronominal en San Juan de Puerto Rico y Barranquilla, Colombia. *Spanish in Context*, 8(2), 191-212.
- Croft, W. (1995). Intonation units and grammatical structure. *Linguistics*, 33, 839-882.
- Dahl, Ö. (2000). Egophoricity in discourse and syntax. *Functions of Language*, 7(1), 37-77.
- Dahl, Ö. (2008). Animacy and egophoricity: Grammar, ontology and phylogeny. *Lingua*, 118, 141-150.
- Davidson, B. (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.
- Delancey, S. (1981). An interpretation of split ergativity and related patterns. *Language*, 57(3), 626-657.
- Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (Eds.), *Talking data: Transcription and coding in discourse* (pp. 45-89). Hillsdale: Lawrence Erlbaum Associates.

- Dumont, J. (2016). *Third person references: Forms and functions in two spoken genres of Spanish*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Enriquez, E. V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto Miguel de Cervantes.
- Erker, D., & Guy, G. R. (2012). The role of lexical frequency in syntactic variability: Variation subject personal pronoun expression in Spanish. *Language*, 88(3), 526-557.
- Ewing, M. C. (2014). Motivations for first and second person subject expression and ellipsis in Javanese conversation. *Journal of Pragmatics*, 63, 48-62.
- Givón, T. (1983). Topic continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-linguistic study* (pp. 1-41). Amsterdam: John Benjamins.
- Gudmestad, A., House, L., & Geeslin, K. (2013). What a Bayesian analysis can do for SLA: New tools for the sociolinguistic study of subject expression in L2 Spanish. *Language Learning*, 63(3), 371-399.
- Haiman, J., & Munro, P. (1983). Introduction. *Switch reference and universal grammar: Proceedings of a symposium on switch reference and universal grammar* (pp. ix-xv). Amsterdam: John Benjamins.
- Helasvuo, M.-L. (2014). Searching for motivations for grammatical patternings. *Pragmatics*, 24(3), 453-476.
- Jia, L., & Bayley, R. (2002). Null pronoun variation in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics*, 8(3), 103-116.
- Koban, D. (2011). Continuity of reference and subject personal pronoun variation in the Turkish spoken in Turkey and in New York City. *Australian Journal of Linguistics*, 31(3), 351-389.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Basil Blackwell.
- Labov, W. (2004). Quantitative reasoning in linguistics. In U. Ammon; N. Dittmar; K. J. Mattheier; & P. Trudgill (Eds.), *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society*, vol. 1 (pp. 6-22). Berlin: Mouton de Gruyter.
- Lastra, Y., & Butragueño, P. M. (2015). Subject pronoun expression in oral Mexican Spanish. In A. M. Carvalho; R. Orozco; & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 39-57). Georgetown: Georgetown University Press.
- Lee, D.-Y., & Yonezawa, Y. (2008). The role of the overt expression of first and second person subject in Japanese. *Journal of Pragmatics*, 40(4), 733-767.
- Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, 23(2), 379-434.
- Mayol, L. (2010). Contrastive pronouns in null subject Romance languages. *Lingua*, 120, 2497-2514.
- McKee, R., Schembri, A., McKee, D., & Johnston, T. (2011). Variable "subject" presence in Australian Sign Language and New Zealand Sign Language. *Language Variation and Change*, 23(3), 375-398.
- Meyerhoff, M. (2009). Replication, transfer, and calquing: Using variation as a tool in the study of language contact. *Language Variation and Change*, 21(3), 297-317.
- Morales, A. (1986). *Gramáticas en contacto: Análisis sintáctico sobre el español de Puerto Rico*. San Juan: Editorial Playor.

- Myhill, J. (2005). Quantitative methods of discourse analysis. In R. Köhler; G. Altmann; & R. Piotrowski (Eds.), *Quantitative linguistik: Ein internationales handbuch* (pp. 471-497). Berlin / New York: Walter de Gruyter.
- Nagy, N., Aghdasi, N., Denis, D., & Motut, A. (2011). Null subjects in heritage languages: Contact effects in a cross-linguistic context. *University of Pennsylvania Working Papers in Linguistics*, 17(2), 135-144.
- Orozco, R. (2015). Pronominal variation in Colombian Costeño Spanish. In A. M. Carvalho; R. Orozco; & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 17-37). Georgetown: Georgetown University Press.
- Otheguy, R. (2002). Saussurean anti-nomenclaturism in grammatical analysis: A comparative theoretical perspective. In W. Reid; R. Otheguy; & N. Stern (Eds.), *Signal, Meaning, and Message: Perspectives on sign-based linguistics* (pp. 373-403). Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language contact, dialect levelling, and structural continuity*. Oxford: Oxford University Press.
- Otheguy, R., Zentella, A. C., & Livert, D. (2007). Language and dialect contact in Spanish of New York: Toward the formation of a speech community. *Language*, 83(4), 770-802.
- Owens, J., Dodsworth, R., & Kohn, M. (2013). Subject expression and discourse embeddedness in Emirati Arabic. *Language Variation and Change*, 25(2), 255-285.
- Paredes Silva, V. L. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, 5(1), 35-49.
- Parkinson, D. B. (1987). Constraints on the presence/absence of 'optional' subject pronouns in Egyptian Arabic. *15th Annual Conference on New Ways of Analyzing Variation*, 348-360.
- Payne, T. E. (1997). *Describing morphosyntax: A guide to field linguists*. Cambridge: Cambridge University Press.
- Poplack, S. (2001). Variability, frequency, and productivity in the irrealis domain in French. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 405-428). Amsterdam / Philadelphia: John Benjamins.
- Posio, P. (2013). The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica*, 47(1), 253-291.
- Posio, P. (2015). Subject pronoun usage in formulaic sequences: Evidence from Peninsular Spanish. In A. M. Carvalho; R. Orozco; & N. L. Shin (Eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective* (pp. 59-78). Washington, D.C.: Georgetown University Press.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223-255). New York: Academic press.
- Ranson, D. L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change*, 3(2), 133-152.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. Newmeyer (Ed.), *Linguistics: The Cambridge survey (Vol. 4, Language: The socio-cultural context)* (pp. 140-161). Cambridge: Cambridge University Press.
- Scheibman, J. (2001). Local patterns of subjectivity in person and verb type in American English conversation. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 61-89). Amsterdam: John Benjamins.
- Serrano, M. J. (2014). El sujeto y la subjetividad: Variación del pronombre 'yo' en géneros conversacionales y de los medios de comunicación del español de Canarias

- (The subject and the subjectivity: Variation of ‘yo’ in conversational and mass-media genres of Canarian Spanish). *Revista Signos (Estudios de Lingüística)*, 47(85), 321-343.
- Shin, N. L. (2014). Grammatical complexification in Spanish in New York: 3sg pronoun expression and verbal ambiguity. *Language Variation and Change*, 26(3), 303-330.
- Silva-Corvalán, C. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elías Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93-120). New York: Cambridge University Press.
- Silveira, A. S. (2011). Subject expression in Brazilian Portuguese: Construction and frequency effects. Doctoral dissertation, University of New Mexico.
- Torres Cacoullos, R., & Travis, C. E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Torres Cacoullos, R., & Travis, C. E. (To appear). Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics*.
- Travis, C. E. (2005). *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin / New York: Mouton de Gruyter.
- Travis, C. E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19(2), 101-135.
- Travis, C. E., & Lindstrom, A. M. (2016). Different registers, different grammars? Subject expression in English conversation and narrative. *Language Variation and Change*, 28(1), 103-128.
- Travis, C. E., & Torres Cacoullos, R. (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics*, 23(4), 711-748.

## Appendix Transcription conventions

Transcription conventions used in this chapter (cf., Du Bois et al. 1993)

Carriage return	new Intonation Unit
	(where the IU does not fit on one line, the second line is indented)
.	final intonation contour
,	continuing intonation contour
?	appeal intonation contour
--	truncated intonation contour
..	short pause (0.2 secs)
...	medium pause (0.3-0.6 secs)
...( )	timed pause (0.7 secs or longer)
-	truncated word
=	lengthened syllable
[ ]	overlapped speech
@	one syllable of laughter
<@ @>	speech uttered while laughing
~	pseudonymized proper noun