

The role of pragmatics in shaping linguistic structures

Catherine E. Travis and Rena Torres Cacoullos

1 Introduction

Discourse pragmatics is the study of linguistic phenomena that are affected by broader textual considerations, such as information flow. As such, it is most compatible with a usage-based perspective, according to which grammatical structures are shaped by their frequency and contexts of use, in contrast to a formalist view of syntax as autonomous from discourse function and cognitive processes. In this chapter, we demonstrate how pragmatic notions can be replicably tested by devising operationalizations that convert pragmatic constructs into countable entities in discourse.

Spontaneously produced, sustained discourse, and not intuition- or cherry-picked example sentences, provides the most reliable data source. This is because pragmatic constructs invoke speaker-hearer cognitive operations that transcend propositional meaning at the level of the individual clause or sentence (cf. Givón, 2001, p. 437). And, as Hopper and Thompson note, “Only actual texts can . . . inform us about the real contexts for speech forms in a way that can give us clues to the motivations for these forms” (1993, p. 372). Above all, spontaneous speech corpora approximate the mode of everyday speech (the vernacular), thus providing a baseline against which to evaluate findings from elicited data, since variation is more patterned in unmonitored speech than in formal styles (Labov, 1984, p. 29).

We begin by reviewing the “given vs. new” distinction, to highlight that its scope is restricted by the discourse referentiality of noun phrases (Section 2). We then test two pragmatic notions that have received vast attention in the literature: ‘referent accessibility’ (Section 3.1) and ‘topicality’ (Section 3.2), as applied to Spanish noun phrase (NP) realization (lexical vs. pronominal vs. unexpressed subjects) and word order (pre- vs. postverbal subjects). We will see that the inherent variability in these subsystems provides a tool for assessing how pragmatic aspects

interplay with grammatical structure, even in a language contact situation (Section 3.3).

2 Review of existing research

2.1 Dimensions of information flow in natural speech: the role of discourse referentiality

Information flow concerns the “cognitive and social aspects of the way people ‘package’ ideational content” (Thompson, 1997, p. 65). The dimension of information flow that is most widely applied is the status of the referent in terms of its accessibility (Ariel, 2001, p. 33; Givón, 1983a, p. 17) or activation (Chafe, 1994, p. 75). While this notion is often interpreted in cognitive terms, as noted by Vázquez Rozas (2006), “el estatus cognitivo de las entidades en la mente de los hablantes (su grado de activación o accesibilidad) no es un dato directamente observable” (“the cognitive status of entities in the minds of speakers (their level of activation or accessibility) is not directly observable (we cannot get inside the minds of the interlocutors to determine whether a concept is activated or not)”) (2006, p. 394, our translation). A measure that has been developed to assess accessibility is distance from previous mention in the discourse; i.e., the more recently a referent has been mentioned, the more likely it is to be active in the hearer’s mind, or “given,” and the longer it hasn’t been mentioned the more likely it is to be “new” (cf. Givón, 1983a).

But there are other dimensions orthogonal to accessibility that must be taken into account in order to understand the role of information flow in shaping grammar; namely, identifiability, specificity, and, crucially, discourse referentiality.

Accessibility is often confused with identifiability, and though the two overlap, they do so only partially. While given NPs are necessarily identifiable, new NPs can be identifiable through association with another identifiable element, shared background, or presence in the extralinguistic context (cf. Prince, 1981). In Example (1), *la mayor de mis hijas* ‘my oldest daughter’ in line 1 is new, not having been mentioned previously in the current conversation, but it is identifiable through association with the

speaker, via the use of possessive pronoun (here, *mis* ‘my-pl’). We can think of new non-identifiable NPs as creating a new entity, new identifiable NPs as placing an existing entity into the discourse model, and given NPs as representing entities already in the discourse model (cf. Prince, 1981, pp. 235–236). In a computer-based metaphor, new non-identifiable NPs create a new cognitive file, new identifiable NPs open an existing file, and given NPs correspond to a currently open file (cf. Du Bois, 1980, p. 220ff) as in (1):

(1) Accessibility and identifiability (new, identifiable NP)

- 1 Trinidad: **la mayor de mis hijas**,
 2 uh te- tenía **amigas** en colegio de --
 3 en New Mexico State,
 4 ... que hablaban --
 5 que eran de México,
 6 ... de modo que **ella** aprendió !muy bien,
 7 a hablar.
 8 **Ø** lo hablaba muy bonito,
 9 como **los de México**.
 1 Trinidad: ‘**my oldest daughter**,
 2 ... uh had **friends** in college --
 3 in New Mexico State,
 4 ... who spoke --
 5 who were from Mexico,
 6 ... so she learned very well,
 7 to speak.
 8 **(she)** spoke it beautifully.
 9 like **people from Mexico**.’

[21 Demerits, 03:20–03:30]

Accessibility cannot be measured without first taking into account the information flow dimension of discourse referentiality, which concerns how a noun is deployed in the broader discourse. According to this understanding, discourse referential, or “manipulable,” NPs (Hopper & Thompson, 1984, p. 711) are “used to speak about an object as an object, with continuous identity over time” (Du Bois, 1980, p. 208). A practical way of thinking about discourse referentiality is as referent “tracking,” whereby a discourse referential NP is used “for the purpose of either introducing [a referent] for further tracking or continuing tracking an earlier mention” (Thompson, 1997, p. 69). The three subjects in

lines 1, 6, and 8 in Example (1) are all discourse referential, in that they represent a referent that is being tracked: *la mayor de mis hijas* ‘my oldest daughter’ is what the speaker is talking about in this stretch of discourse. Similarly, *amigas* ‘friends’ in line 2 is also tracked (albeit only briefly). On the other hand, *los de México* ‘people from Mexico,’ which occurs in the oblique phrase with the preposition *como* ‘like’ in line 9, does not introduce a referent to talk about it, but serves to describe the manner of the verb (the way she speaks); this is a non-discourse referential use.

Non-discourse referential NPs abound in everyday speech. These include NPs used to orient an event, as in line 9 in (1), and in (2). Also non-referential are NPs used to classify another referent, as in predicate nominals, illustrated in (3). A third common non-discourse-referential use is in verb-noun units in which the noun is absorbed into the predicate, illustrated in (4) (cf. Hopper & Thompson, 1984, p. 711ff; Thompson, 1997, p. 69). In Spanish, this includes verb-object combinations (as in *tener hambre* ‘have hunger/be hungry’ or *tener problemas* ‘have problems’), as well as in dative experiencer constructions (as in *dar miedo* ‘give someone fear/be afraid’; cf. Bentivoglio & Sedano, 2007, p. 203).

- (2) Non-referential orienting use
 eso haces **todos los días**,
 ‘you do that every day,’
 (Dumont, 2016, p. 58)
- (3) Non-referential classifying use
 pero ella es **profesora**,
 ‘but she is a teacher,’
 (Dumont, 2016, p. 57)
- (4) Non-referential predicate-forming use in verb-noun unit with light verb
ellos siempre tenían hambre.
 ‘they were always hungry.’
 [23 El Pacific, 12:40–12:42]
yo no tengo problema en que se quede ahí.
 ‘I don’t have a problem with it staying there.’
 [CCCS, 02 Restaurante, 328]1
y a él le dio mucho miedo,
 ‘and he was very scared,’
 [18 Las Minas, 1:12:08–1:12:09]

As these examples illustrate, non-discourse-referential NPs often combine with light verbs, such as *tener* ‘have’ or *dar* ‘give,’ to form units, or chunks (cf. Bybee, 2010, p. 34). Verb-noun units need not be entirely fixed, however, allowing for noun modifiers (for example, *dar miedo/dar mucho miedo* ‘give someone fear/give someone a lot of fear’), or variant verbs (for example, *tener/pasar/traer hambre*, literally ‘have/go through/carry hunger’). Even so, the noun remains non-referential; it contributes to the meaning of the predicate, naming an activity or state, rather than representing a referent for the purposes of saying something about it (Thompson & Hopper, 2001, p. 33). That is, *hambre* ‘hunger’ in *pasar hambre* or *traer hambre* is no more a true object than it is with the light verb in the more frequent *tener hambre*, and no more referential than the predicate adjective in *estar hambriento* ‘be hungry.’ Note also the common use of bare nouns in this context, consistent with the lack of discourse referentiality, as observed by Torres Cacoullós & Aaron (2003, pp. 306–307) (cf. Alarcos Llorach, 1972, p. 176; Alonso, 1951, pp. 161–162; Lapesa, 1975, p. 129).

Verb-noun units are not restricted to abstract nouns, nor to light verbs. An example is the expression *montar bicicleta* ‘ride a bicycle,’ where *bicicleta* does not refer to an object but, adjoined to the verb, represents a situation (akin to an incorporated noun, as it would be in the constructed ‘bicycle ride’). Similarly, with *montarse en canoa* ‘get in a canoe’ in (5), ‘canoe,’ here a prepositional object, is used non-referentially (‘to canoe’).

- (5) Non-referential predicate-forming use as prepositional object, with non-light verb (re: escaping from a tidal wave in Chocó, Colombia)
 Ángela: Y ahí mismo pues la gente empezó a levantar a todo el mundo a montarse en **canoas** y,
 y a huir hacia las fincas.
 pero pues no les dio tiempo.
 ‘And right then well the people started to wake everyone up
 to get in the **canoas** and,
 and escape to the farms.
 but well they didn’t have time.’

[CCCS, 30 Terremoto, 53–55]

Recognizing non-discourse referential uses is crucial for analyses of accessibility, because accessibility applies only to discourse-referential NPs. In *todos los días* ‘every day,’ *ella es profesora* ‘she is a teacher,’ or *montar bicicleta* ‘to ride a bicycle,’ for example, we cannot ask whether *días*, *profesora*, or *bicicleta* have been activated in the hearer’s mind or whether they are being newly activated; they simply are not candidates for activation at all because the referent is not being talked about as an entity. Non-discourse referential NPs must be set aside in coding for accessibility, since a wrong interpretation of them as “new” (due to their tendency to occur without a previous mention; cf. Dumont, 2016, p. 89) would inflate the proportion of “new” referents and mask true accessibility effects.

Discourse referentiality is not to be confused with specificity. Non-specific NPs refer to any member of a class, as in *amigas* ‘friends’ and *los de México* ‘people from Mexico’ in lines 2 and 9 in Example (1); specific NPs refer to an entity that is not interchangeable, as in *la mayor de mis hijas* ‘my oldest daughter’ in line 1 (Ashby & Bentivoglio, 1993, pp. 60–70; Torres Cacoullos & Aaron, 2003, p. 307). Discourse referential NPs tend to be specific (as is the case for *la mayor de mis hijas*) while non-discourse referential NPs tend to be non-specific (as in *los de México*). The two dimensions are not, however, conflatable. For example, *amigas* is not specific but it is discourse-referential—it refers to any of the friends from Mexico, and is used to talk about them. In the following example, *trabajo* ‘work’ is specific but non-discourse referential (it refers to a specific workplace and serves an orienting function in the oblique phrase *en el trabajo* ‘at work’).

- (6) Specific but non-referential use
 Y le darían permiso en el **trabajo**?
 ‘And will he be given permission at work?’

(Dumont, 2016, p. 54)

An important conclusion to draw here is that accessibility intersects with other dimensions of information flow. In particular, its scope is limited by discourse referentiality, such that the given/new distinction applies to only a portion of NPs in a corpus; specifically, to those that are discourse referential. Let us now

consider the impact of these dimensions of information flow on grammar.

2.2 Role of information flow in shaping the grammar of NPs

The impact of accessibility on NP realization has been widely observed cross-linguistically: greater accessibility is associated with more minimal formal expression, and vice versa (Ariel, 2001, p. 33; Givón, 1983a, p. 18; Levinson, 1987, p. 384). Example (1) serves to illustrate where a subject referent is introduced into the discourse for the first time as a lexical NP (*la mayor de mis hijas* ‘my oldest daughter’), followed by a pronominal mention in line 6 (*ella* ‘she’), and then an unexpressed (null) subject in line 8. Similarly in Example (7), the referent is introduced in line 1 with a lexical NP, and subsequently referred to in line 2 with an unexpressed subject.

- (7) New lexical subject and given unexpressed (\emptyset) subject
- | | | |
|---|------------|--|
| 1 | Bartolomé: | porque m- mi papá era ranchero y, |
| 2 | | ... \emptyset vendía verduras, |
| 1 | Bartolomé: | ‘because m- my dad was a rancher and, |
| 2 | | ... (he) sold vegetables,’ |
- [02 La Marina, 38:30–38:44]

Thus, pronouns and unexpressed arguments tend to be given. New referents are mostly introduced with lexical NPs (e.g., Travis & Torres Cacoulios, 2018, p. 80). The reverse, however, is not the case, in that most lexical NPs do not represent new referents (e.g., Bentivoglio, 1993, pp. 218–219; Dumont, 2016, pp. 84, 88). Furthermore, most “new” referents are identifiable, be that through anchoring to a given referent such as the speaker, as with *la mayor de mis hijas* ‘my oldest daughter’ in (1) and *mi papá* ‘my dad’ in (7), or through a frame, as is the case for *la profesora* ‘the teacher’ in (8), due to the mention of *clases* ‘classes’ in the previous clause.

- (8) New, identifiable lexical subject
- También habían unas **clases** de música interesantes,
Vale, la **profesora** daba canciones así del maestro Sojo
‘There also were some interesting music lessons,

the teacher used to present songs by Maestro Sojo’
(Bentivoglio, 1993, p. 216)

The asymmetrical association between lexical forms and new referents becomes apparent from a set of studies reporting that one-third of lexical NPs (across syntactic roles) represent new information, and even fewer are “brand new information”, that is non-identifiable, as per Prince, 1981) (e.g., Ashby & Bentivoglio, 1993, p. 71; Bentivoglio, 1993, p. 221).

Also impacting NP realization is discourse referentiality. Personal pronouns overwhelmingly have discourse referential uses, but the same is not true of lexical NPs (Dumont, 2016, p. 71). Discourse referentiality bears on syntactic role as well. While discourse referential NPs predominate in subject position, and are found, though to a lesser degree, in object position, they are less common in obliques and predicate nominals, as seen in Examples (2) and (3) (Dumont, 2016, p. 76). The intersection of discourse referentiality with syntactic role is linked, in turn, with the impact on NP realization. Lexical NPs occur more in obliques and predicate nominals, while pronouns occur more in the core roles of subject and object, and unexpressed mentions are largely restricted to subjects in Spanish (Thompson, 1997, p. 83).²

These associations have a methodological consequence. Because subjects are more likely than objects to have given human referents (e.g., Prince, 1981, p. 243; Vázquez Rozas, 2006, p. 405), the syntactic and semantic combination of human-specific subjects serves as a proxy for discourse referentiality (cf. Dumont, 2016, p. 81). That is, while other subject types and other syntactic roles will often host non-referential NPs, human-specific subjects will be overwhelmingly discourse referential. In the next section, we therefore rely on human-specific subjects to illustrate how pragmatic notions related to information flow may be operationalized in natural speech data.

3 Methodological considerations

How may accessibility and topicality be operationalized for quantitative tests? For a quantitative exploration of how discourse-pragmatic aspects shape linguistic structures, a heuristic device is

the linguistic variable: forms with a shared grammatical function that exist as competing variants between which speakers have a choice (cf. Labov, 1972, p. 72). The working hypothesis of the variationist method is that “within a given locus of variability, or variable context, . . . [the] competing variants will occur at greater or lesser rates depending on the features that constitute the context” (Poplack, 2001, p. 405, italics in original). The impact of pragmatic aspects on the competing variants may be discerned in their rates of occurrence in contexts that are hypothesized to be associated with the pragmatic aspects being tested. Tokens of the variant forms are coded for co-occurring elements to determine if one variant is more likely than an alternative to co-occur with particular elements as hypothesized.

For these analyses, we draw on the New Mexico Spanish-English Bilingual (NMSEB) corpus. NMSEB is a community-based corpus of sociolinguistic interviews with Hispanic New Mexicans who regularly use Spanish and English in their daily lives (Torres Cacoullos & Travis, 2018, Chapters 2 & 3). The transcription is prosodically based, such that each line represents an Intonation Unit (IU), a segment of speech produced with “a single, coherent intonation contour” (Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993, p. 47) (cf. Chafe, 1994, pp. 58–60).

Here we look at the linguistic variables of Spanish subject realization (lexical, pronominal, and unexpressed subjects) and position (pre- and postverbal subjects), focusing on first-person and third-person singular (1sg and 3sg). The dataset for analysis consists of 4,218 unexpressed and 1,598 pronominal subjects, and, for 3sg, 807 lexical NPs.

3.1 Accessibility and subject forms: distance

We first consider pronominal, unexpressed, and lexical subjects in the context of greater vs. lesser distance from previous mention, as an operationalization of lesser vs. greater accessibility.

3.1.1 Accessibility and subject realization

For Spanish subject realization, accessibility is often examined in terms of “switch reference”; i.e., whether the subject referent of

the target clause differs from that of the previous clause, as in lines 3, 6, and 7 in (9), or whether it is the same, as for the target subject in line 4, showing continuity of reference with the subject of line 3. Subject pronoun expression in Spanish has long been known to be sensitive to switch reference (since Silva-Corvalán, 1982, p. 104; see also Cameron, 1994). Still, accessibility, and the sensitivity of subject expression, extends beyond the immediately preceding clause (Travis & Torres Cacoullos, 2018, p. 72).

A heuristic for accessibility in a stretch of discourse is the anaphoric measure of distance from previous discourse mention (initially explored across languages as Referential Distance in Givón (1983b), including for Spanish (Bentivoglio, 1983); cf. Givón (2001, p. 463)). Distance, as applied here, is the number of intervening clauses between the target subject and the previous mention, which we limit to mention of that same referent in subject role.³ For example, in (9), the target subject in line 7 *yo* ‘I’ has a distance value of 3, because there are three clauses intervening between it and the previous subject mention in line 1; the target subject *she* in line 4, in turn, has a distance of 0, as there is a subject mention in the immediately preceding clause (as a bilingual corpus, clauses are Spanish or English; see Section 3.3; speech originally produced in English is given in italics in the translation that follows).

(9) Distance (intervening clauses between subject mentions)

- 1 Betty: ... Ø le fajé una nalgada.
 2 Carrie: .. @@@
 3 Betty: .. she was starting to walk,
 4 ... and she wanted to go to the fireplace,
 5 sit on the mantel,
 6 ... (1.0) y no la dejaba la ~Bobbie.
 7 **yo** no sé qué estarían haciendo allá.
 1 Betty: ‘... (I) gave her a slap on the bottom.
 2 Carrie: .. @@@
 3 Betty: .. she was starting to walk,
 4 ... and she wanted to go to the fireplace,
 5 sit on the mantel,
 6 ... (1.0) and Bobbie wouldn’t let her.
 7 I don’t know what they were doing there.’

[13 La Acequia, 18:16–18:28]

As a manifestation of accessibility, we would predict a correlation between distance and subject realization. To determine this, for each clause in the dataset, we counted the distance in clauses between the target and previous coreferential subject, from 0 (for a previous mention in the immediately preceding clause) up to 10 (counting as 10 all instances at a distance of 10 or more intervening clauses, including newly introduced referents).⁴ Figure 7.1 shows the mean distance for 1sg and 3sg subjects according to realization as unexpressed or as preverbal pronominal or lexical NP (we consider postverbal subjects in Section 3.2). Consistent with the predicted pattern, unexpressed subjects have the lowest distance (with the previous coreferential mention occurring on average approximately two clauses back), followed by pronominal subjects. This pattern holds for both 1sg and 3sg subjects, despite the view of 1sg as always given (e.g., Chafe, 1994, p. 79; Givón, 2001, p. 460); we return to grammatical person in the following section. Lexical subjects, which only apply to 3sg, have a notably higher distance than both unexpressed and pronominal subjects (with the previous mention on average nearly seven clauses back, over twice the distance of 3sg pronouns, and over three times that of 3sg unexpressed subjects).

Breaking down these figures, we find that the majority of the instances of 1sg and 3sg unexpressed and pronominal subjects occur in a coreferential context or with just one clause intervening, whereas the majority of the lexical subjects occur with 10 or more clauses intervening. We can confirm, then, that the preferred form for new 3sg referents are lexical NPs (Torres Cacoullos & Travis, 2018, p. 197).

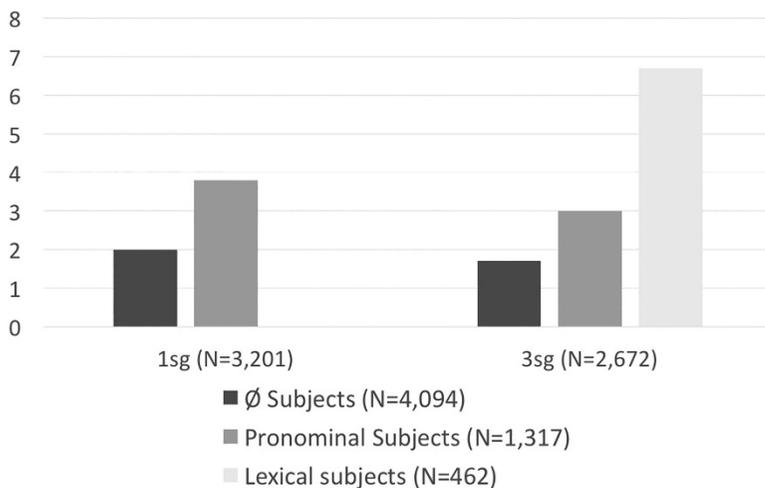


Figure 1 Distance (intervening clauses between subject mentions) of preverbal subjects by form.

3.1.2 Accessibility and grammatical person

Much has been made of the difference between first- and third-person subjects. For example, the distinction between speech act participants (first- and second-person) and all other NPs (including 3sg animate NPs) often has greater impact on grammatical structures than does that between animate and inanimate NPs (cf. Dahl, 2008, p. 143).⁵ From a discourse-pragmatic perspective, as a discourse participant, a first-person referent is thought to be always accessible; third person referents, on the other hand, are typically made accessible through a mention in the discourse (e.g., Chafe, 1994, pp. 78–79; Dahl, 2000, pp. 64–66; Prince, 1981, p. 236). For this reason, it is often understood that the first person is used deictically (to point out a referent in the discourse setting), whereas third person is used anaphorically (to refer back to a previous mention) (e.g., Grundy, 2000, p. 78). We might then expect accessibility to be irrelevant for first-person subject expression as would be manifested, for example, in a lack of sensitivity to distance. However, as seen in Figure 7.1, 1sg subjects do show a tendency for pronouns to be favored over unexpressed subjects at greater degrees of distance (Travis & Torres Cacoullos, 2018,

p. 76). Thus, there are degrees of accessibility even for discourse participants.

In fact, while the same general tendency for distance holds for 1sg and 3sg pronominal and unexpressed subjects, 3sg exhibits greater clustering than 1sg, as seen in the shorter average distance (3 for 3sg pronouns, and 3.8 for 1sg). Such a clustering pattern is consistent with the observation that speech act participants “pop in and out all the time,” while third-person subjects, once introduced into the discourse, “stay on the scene” (Dahl, 2000, p. 66), though they tend to do so only for a short time, as is the case for *mi papá* ‘my dad’ in (10) (Travis & Torres Cacoullos, 2018, p. 79).

(10) Short-term persistence of 3sg subjects

Bartolomé: porque m- **mi papá** era ranchero y,
... Ø vendía verduras,

y=,

Gabriel: sí=.

Bartolomé: y= ... y veníamos a la plaza aquí a vender,
... (1.6) los blanquillos ... los huevos.
... a diez centavos la docena.

Gabriel: .. uh huh <@ diez centavos @>?
no me digas.

Bartolomé: ‘because m- my dad was a rancher and,
... (he) sold vegetables,
and,

Gabriel: yes.

Bartolomé: and... and we would come to the plaza here to sell,
... (1.6) the little white ones .. the eggs.
... for ten cents a dozen.

Gabriel: .. uh huh <@ ten cents @>?
you don’t say.’

[02 La Marina, 38:30–38:44]

To summarize, by operationalizing accessibility via distance from previous subject mention, we confirm that lexical NPs are the preferred form to introduce a new referent into the discourse, while unexpressed and pronominal subjects tend to be used to refer to more accessible subjects (with lesser distance between mentions). Accessibility also affects grammatical person. Despite first person being present in the discourse context, it, too, is susceptible to distance. Considering unexpressed and pronominal subjects, for 1sg these tend to occur with greater

distance between mentions than is the case for 3sg, reflecting greater clustering of the latter.

3.2 Topicality and word order

Another pragmatic notion that has been understood to play a role in grammatical structure is topicality, which we may intuitively understand as related "aboutness" or "the frame within which the sentence holds" (Chafe 1976:50). The definitions for topicality, however, are nebulous and discrepant. We illustrate here two discourse measurements for two psychological aspects of topicality: the accessibility and the importance of a referent (Givón, 1988). Accessibility, as discussed earlier, relates to the prior mention, and applies to topicality in the sense that a topical referent is likely to have already been mentioned in the preceding discourse. Importance, on the other hand, relates to subsequent mentions, and applies to topicality in the sense that a topical referent is likely to be mentioned again (Givón, 2001, p. 198). The linguistic variable we examine here is variable word order. Subjects in Spanish occur most often preverbally, and as such, verb-subject (VS) word order is generally thought to be pragmatically motivated, with topicality implicated as a motivating factor, though accounts vary as to whether it is post- or preverbal subjects that are topical.

3.2.1 *Topicality and word order: distance*

Let us first consider topicality as accessibility, using distance (Section 3.1) as an anaphoric measure. Topics (often contrasted with focus or comment) are typically considered to be given (Ocampo, 2003, p. 199); according to this understanding, NPs with lower average distance can be considered more topical. How does this apply to word order?

The literature yields two opposing scenarios. If postverbal subjects are associated with low topicality (Delbecque, 1988, p. 182), they should be favored with new referents; this is consistent with an effect for first mention reported in some studies (e.g., Bentivoglio & Weber, 1986, p. 25; Rivas, 2013, p. 105). We would thus expect postverbal subjects to display a higher distance

than preverbal ones. A contradictory view is that postverbal subjects are associated with “higher referential predictability” (Givón, 1988, p. 251), from which we might infer that they are high in topicality. Accordingly, we would expect postverbal subjects to display a lower distance.

Figure 7.2 compares distance for post- vs. preverbal pronominal and lexical subjects. First we observe that the impact of distance on subject realization noted earlier (Section 3.1.1) holds independently of word order; whether pre- or postverbal, lexical subjects have a higher value than both 1sg and 3sg subject pronouns. For subject position, however, there is no difference in average distance for pronouns, nor for lexical subjects (4.1 for postverbal and 3.8 for preverbal 1sg subjects, 3.3 and 3 for 3sg pronouns, and 7 and 6.7 for lexical subjects). Similarly, a minimal distance difference has been reported for definite NP human subjects, which, like the subjects studied here, can be considered to be discourse referential (average distance is 6.28 for postverbal and 7.98 for preverbal subjects, though the sample is modest, n = 132) (Bentivoglio, 1983, p. 299, Table 10).

As an interim conclusion, topicality operationalized as accessibility is not an overarching motivation for the choice between pre- vs. postverbal human-specific subjects. Let us, then, consider an alternative operationalization of topicality; that based on importance.

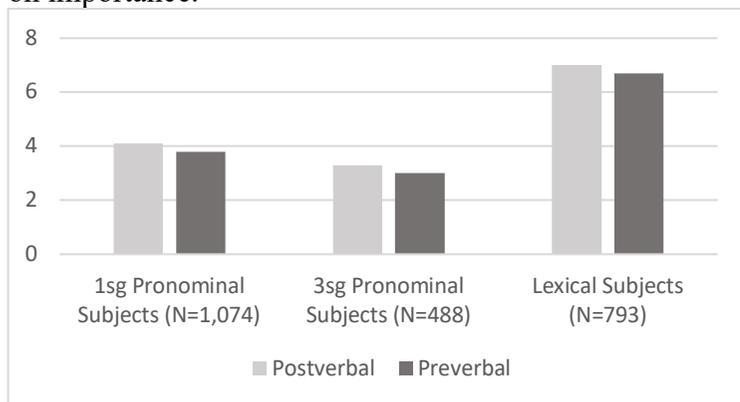


Figure 2 Distance (intervening clauses between subject mentions) of subjects by position and form.

3.2.2 Topicality and word order: persistence

A text-based correlate of topicality as the importance of a referent in the discourse is the measure of persistence. This cataphoric measure assesses the subsequent mentions of a referent once it has been introduced (Givón, 1988, p. 248). If a subject is placed postverbally when its referent is important or, put differently, when it is the “point of the utterance” (Bolinger, 1954, p. 49), we would expect a higher average rate of postverbal subjects with persistent than non-persistent referents. But it has also been proposed that VS is associated with “communicative peripherality” (Naro & Votre, 1999, p. 75), which leads to the contrary prediction of a lower average rate of postverbal position with persistent subjects. Here we counted the number of mentions of the referent in subject role in the three clauses subsequent to the target subject and categorized as persistent those with at least one subject mention in the following three clauses. For example, in (11), *mi mamá* ‘my mother’ is persistent; it is mentioned twice in the following three clauses:

(11) Persistence

Javier: ... y= unos panes que se ha --
unos --

... pasteles que hacía **mi mamá**.
... (1.0) pu=ras calabazas había.

Gabriel: .. hm.

Javier: ... poquita canela \emptyset le echaba ahí no más,
y poquita azúcar,
porque \emptyset no tenía suficiente.

Javier: ‘and some breads that --
some --
... sweets that my mother_{post} made.
... (1.0) it was all pumpkin.

Gabriel: .. hm.

Javier: ... just a bit of cinnamon (she) would put,
and a bit of sugar,
because (she) didn’t have enough.’

[17 La Comadreja, 29:15–29:28]

Figure 7.3 depicts persistence of pronominal and lexical subjects in the height of the columns, and VS rate in the light shading within each column (also indicated by the percentage). We

first observe that both 1sg and 3sg pronominal subjects have a higher proportion of persistent subject referents than do lexical subjects (seen in the relative heights of the columns within each pair; for pronouns, the number of persistent doubles that of non-persistent referents, but for lexical subjects it is just slightly greater). This confirms, then, that pronouns are more topical than lexical subjects in terms of importance, consistent with the result for topicality measured as accessibility.

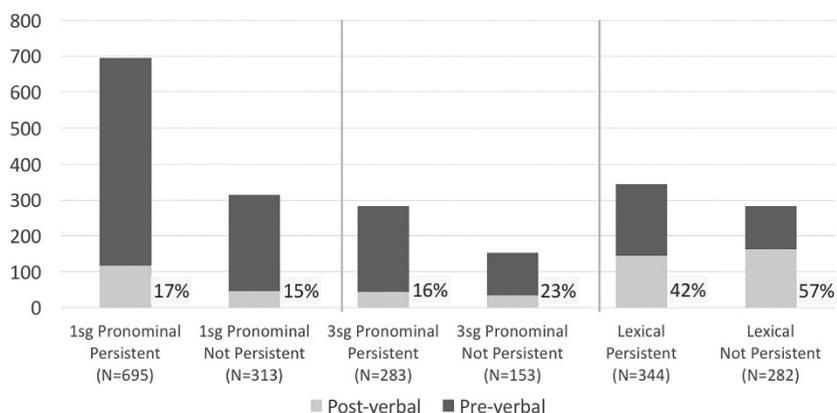


Figure 3 Distribution of tokens and postverbal subject rate according to topic persistence

As to subject position, VS rates for 1sg subjects are the same with persistent and non-persistent referents (17% and 15%, respectively), and for 3sg subjects, they appear to be somewhat lower with persistent referents (16% vs. 23% for pronouns, and 42% vs. 57% for lexical NPs). However the slight difference found here, again matching reports for other varieties of Spanish (Bentivoglio, 1983, p. 299, Table 10), is in keeping with the cross-linguistic absence of a “clear pattern relating [Persistence] and word order” (Myhill, 2005, p. 480).

3.2.3 Topicality and constructions

Another way to think about how topicality may affect word order is not as a general characterization of post- versus preverbal subjects, but rather as pertinent to specific constructions. Constructions are generally defined as pairings of form and meaning (e.g., Goldberg, 2013). In discourse, we may define

constructions as fixed parts and schematic slots co-occurring in particular discourse-pragmatic contexts. An example is the existential-presentative VS construction represented in (12), and illustrated with the indefinite NP subject *una bruja* ‘a witch’ (cf. Bentivoglio, 1983, pp. 279, 297–298; Ocampo, 1995, p. 427). The indefinite article is a fixed element, the verb slot is schematic but is circumscribed to certain classes, and the subject slot is more schematic, open to any human-specific referent.

(12) [*estar/vivir/motion verb + un(a) X_{human,specific}*]_{new and persistent X}

Rocío: ... pues ahí,
 .. dicen,
 que vivía **una bruja**.
 ... (2.0) y que,
 .. que esa bruja,
 se cambiaba de cuerpo.

Rocío: ‘... well there,
 .. they say,
 that there lived a witch_{post}.
 ... (2.0) and that,
 .. that that witch,
 changed bodies.’

[05 Las Tortillas 30:59–31:08]

The quantitative patterns of the co-occurring elements that constitute the construction show the relevance of accessibility and importance in the discourse. Lexical subjects with an indefinite article lack a previous mention within the preceding 10 clauses most of the time (96%, 27/29), compared with just over half the time for other lexical subjects (427/735). That is, indefinite NP subject referents are less accessible or predictable. At the same time, they are more important as measured by persistence, displaying a higher proportion of persistent instances (75%, compared with 54% for all others). These new referents that persist tend to co-occur with verbs of location (*estar* ‘be located’), existence (*ser* ‘be,’ *vivir* ‘live’) or intransitive motion (*llegar* ‘arrive’) (53% of indefinite NP subjects, compared with 26% for all others). And, they favor VS (at a rate of 53%, compared with 39% for all others, a tendency similar to that reported in Sedano & Bentivoglio, 2014, p. 152 for another variety).

Constructions show more consistent subject position patterns than do individual verbs. For example, locative *estar* ‘be,’ seen in (13), has a VS rate about triple that of predicative [*estar* + Adj], seen in (14) (43%, 55/128 vs. 14%, 5/35). As Myhill (2005, p. 480) suggests, existential-presentative constructions in SVO languages likely lie behind the idea that old information precedes new information, and may account for the reported favoring of postverbal subjects with new referents in some studies of Spanish.

- (13) *Estar* in locative construction
 Miguel: pues ahí **estaba** **Valeria**,
 ‘well there was Valeria-post,’
 [04 Piedras y Gallinas, 1:03:33–1:03:34]
- (14) *Estar* in predicate adjective construction
 Miguel: ... y **Valeria** también **estaba** más o menos flaca también.
 ‘.. and Valeria-pre was also more or less thin too.’
 [04 Piedras y Gallinas, 1:04:12–1:04:15]

In sum, we have seen how topicality may be operationalized through the measures of distance and persistence. Though it does not seem to play a global role in subject position, we can identify specific topicality constructions. In the existential-presentative construction, postverbal subjects have low accessibility but high persistence, and the construction is thus associated with topicality as importance in the discourse, for new referents.

3.3 Pragmatics in language contact

As has been indicated in this chapter, the findings for distance and persistence presented here, from a bilingual Spanish-English corpus, are not unique to these bilingual data, but are consistent with findings for other varieties of Spanish. The bilinguals in this community maintain the grammatical patterns of subject expression, as they do of other Spanish variables hypothesized to be sensitive to pragmatic constraints, such as differential object marking (Sankoff et al. 2015) and subjunctive use (LaCasse 2018). These results are particularly interesting, as pragmatics has been implicated in contact-induced grammatical change. A prominent

conjecture is that grammatical subsystems at “the syntax-discourse interface” are “vulnerable” to contact-induced change (Sorace, 2004, p. 143).

An example of proposed “loss of semantic-pragmatic constraints” is word order in the Spanish of minority-language communities in contact with English (Silva-Corvalán, 1994, p. 144). The hypothesis is that contact with English lowers the rate of postverbal subjects. A bilingual corpus supplies a compelling test of contact-induced pragmatic loss, because it features the most intimate language contact, in the spontaneous use of both languages by the same speakers. This is captured in the NMSEB corpus through ample code-switching, or alternating between English and Spanish multi-word sequences. The bilingual corpus allows us to apply a stringent test, comparing postverbal subject rates in the presence vs. absence of code-switching, as a measure of the online consequences of activation of English. If speakers’ English is affecting their Spanish, VS rates should tend to be lower when they are code-switching (Torres Cacoullos & Travis, 2018, Chapter 9). Proximate code-switching is operationalized as the use of English multi-word strings within the same or immediately preceding clause as the target Spanish subject, as in (15). The comparison is with absence of proximate code-switching is illustrated in (16).

(15) Subject position with proximate code-switching

Fabiola: and when they came again,
seguro ya no **vino la Nancy**.
‘and when they came again,
Nancy-POST didn’t come.’

[09 La Salvia, 43:40–43:43]

(16) Subject position in the absence of proximate code-switching

Fabiola: trujeron casi a todas,
pero esa **Nancy** se me hace que no **vino**.
‘they brought nearly all of them,
but that Nancy-PRE I think didn’t come.’

[09 La Salvia, 43:00–43:03]

Figure 7.4 shows that VS rates are no lower in the presence than in the absence of proximate code-switching in the NMSEB dataset. In addition, the disparity between subject forms in VS

rates seen earlier (Section 3.2) is maintained in the presence and absence of code-switching, with VS rates under 20% for pronominal subjects, and close to 50% for lexical subjects.

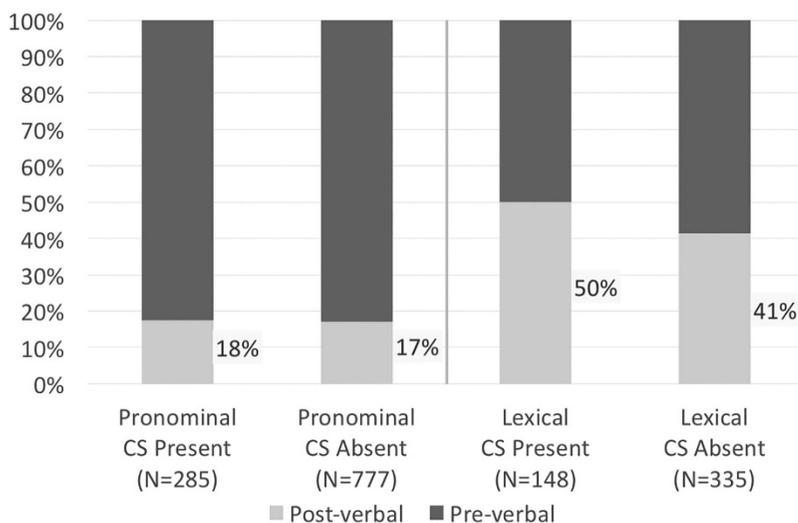


Figure 7.4 Postverbal subject rate in the presence vs. absence of code-switching (CS).

Furthermore, the hypothesis of contact-induced change is refuted by comparing factors conditioning variable subject placement in bilingual and monolingual varieties. Previous work has found no support for this hypothesis for 1sg subject placement in New Mexican Spanish, where contexts favoring the postverbal pronoun *yo* ‘I’ (according to verb type and the presence of preverbal structural elements) are similar to those reported for monolingual varieties (Benevento & Dietrich, 2015, p. 415). In sum, as to loss of pragmatic constraints being particularly vulnerable to change, thanks to operationalizations that permit quantitative comparisons, we have seen here that these bilinguals maintain existential-presentative constructions, and in this local topicality effect on word order, there is continuity, not change.

4 Future directions and conclusion

Operationalizing pragmatic notions will be a profitable avenue for quantitative explorations of the role of pragmatics in shaping

linguistic structures. By converting pragmatic constructs into countable entities in discourse and measuring their impact in a corpus of natural speech, we can test their interplay with grammatical structures. This can be done through the analysis of quantitative co-occurrence patterns produced by speakers' repeated choices between structural alternatives, within the set of contexts where a choice is available to speakers (the variable context). Only through such operationalization can pragmatic functions proposed in the literature confront distribution tendencies observed in everyday speech.

The impact of accessibility in discourse will be accurately assessed only if discourse referential deployment of NPs is distinguished from uses in which there is no referent tracking. Thus, orienting, classifying and predicate-forming uses of nouns must be set aside in analyses of accessibility. Moving forward in coding NP tokens, distributions according to information flow dimensions in conjunction with syntactic role and referent animacy point to human-specific subjects as a proxy for discourse referentiality. Accessibility can then be operationalized as distance from previous mention. Here this operationalization showed how accessibility affects variable subject realization, with some observable grammatical person differences associated with how first vs. third person is deployed in discourse.

The intuitive but slippery notion of topicality can be the object of replicable study by decomposing it into referent predictability (utilizing the corresponding anaphoric measure of distance from the previous mention) and importance (with the cataphoric measure of persistence in subsequent mentions). Applying these measures to variable subject position, we have seen that pragmatic functions may reside in discourse-oriented constructions, as in existential-presentative constructions with an indefinite postverbal NP subject for new and topical referents. Greater attention to constructions as units of analysis may allow pragmatic effects to be more readily discerned and interpreted.

Quantitative assessments of contexts of occurrence—the distribution tendencies of linguistic structures according to features of the environment—allow us to identify usage patterns and the associations between particular structures and discourse-pragmatic

functions. We can then evaluate candidate pragmatic functions of a structure with respect to the contexts in which it appears, but also with respect to the variants with which it competes.

Appendix: Transcription Conventions

(Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993)

Carriage return	new Intonation Unit
.	final intonation contour
,	continuing intonation contour
?	appeal intonation contour
--	truncated intonation contour
-	truncated word
=	lengthened syllable
..	short pause (0.5 secs)
...	medium pause (0.5–0.7 secs)
...()	timed pause (longer than 0.7 secs)
@	one syllable of laughter
<@ @>	speech produced while laughing
!word	speech produced with notably high pitch

References

- Alarcos Llorach, E. (1972). El artículo en español. In E. Alarcos Llorach (Ed.), *Estudios de gramática funcional del español* (2nd ed., pp. 166–177). Madrid: Gredos.
- Alonso, A. (1951). Estilística y gramática del artículo en español. In A. Alonso (Ed.), *Estudios lingüísticos, temas españoles* (pp. 151–194). Madrid: Gredos.
- Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schillperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: John Benjamins.
- Ashby, W. J., & Bentivoglio, P. (1993). Preferred argument structure in spoken French and Spanish. *Language Variation and Change*, 5(1), 61–76.
- 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. (1983). Topic continuity and discontinuity in discourse: A study of spoken Latin-American Spanish. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-linguistic study* (pp. 255–311). Amsterdam: John Benjamins.
- Bentivoglio, P. (1993). Full NPs in spoken Spanish: A discourse profile. In W. J. Ashby, M. Mithun, G. Perissinotto, & E. Raposo (Eds.), *Linguistic perspectives on the Romance languages* (pp. 212–224). Amsterdam and Philadelphia: John Benjamins.
- Bentivoglio, P., & Sedano, M. (2007). The light subject constraint in spoken Spanish. *Lodz Papers in Pragmatics*, 3, 195–205.
- Bentivoglio, P., & Weber, E. G. (1986). A functional approach to subject word order in spoken Spanish. In O. A. Jaeggli & C. Silva-Corvalán (Eds.), *Studies in Romance linguistics: Selected papers of the 14th linguistics symposium on Romance languages* (pp. 23–40). Riverston: Foris.
- Bolinger, D. (1954). English prosodic stress and Spanish word order. *Hispania*, 37(2), 152–156.
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form*. Philadelphia: John Benjamins.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Cameron, R. (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.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.
- Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago, IL: University of Chicago Press.
- 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(2), 141–150.
- Delancey, S. (1981). An interpretation of split ergativity and related patterns. *Language*, 57(3), 626–657.
- Delbecque, N. (1988). On subject position in Spanish: A variable rule analysis of the constraints at the level of the subject NP and of the VP. *Literary and Linguistic Computing*, 3, 85–200.

- Du Bois, J. W. (1980). Beyond definiteness: The trace of identity in discourse. In W. Chafe (Ed.), *The pear stories: Cognitive, cultural and linguistic aspects of narrative production* (pp. 203–274). Norwood, NJ: Ablex.
- 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.
- Dumont, J. (2016). *Third person references: Forms and functions in two spoken genres of Spanish*. Amsterdam and Philadelphia: John Benjamins.
- Givón, T. (1983a). 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.
- Givón, T. (Ed.). (1983b). *Topic continuity in discourse: A quantitative cross-linguistic study*. Amsterdam: John Benjamins.
- Givón, T. (1988). The pragmatics of word order: Predictability, importance and attention. In M. Hammond, E. Moravcsik, & J. Wirth (Eds.), *Studies in syntactic theory* (pp. 243–384). Amsterdam: John Benjamins.
- Givón, T. (2001). *Syntax: An introduction* (Vol. 1, 2nd ed.). Amsterdam: John Benjamins.
- Goldberg, A. E. (2013). Constructionist approaches. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 15–31). Oxford: Oxford University Press.
- Grundy, P. (2000). *Doing pragmatics*. London: Arnold.
- Haiman, J., & Munro, P. (1983). Introduction. In J. Haiman & P. Munro (Eds.), *Switch reference and universal grammar: Proceedings of a symposium on switch reference and universal grammar* (pp. ix–xv). Amsterdam: John Benjamins.
- Hopper, P. J., & Thompson, S. A. (1984). The discourse basis for lexical categories in universal grammar. *Language*, 60(4), 703–752.
- Hopper, P. J., & Thompson, S. A. (1993). Language universals, discourse pragmatics, and semantics. *Language Sciences*, 15(4), 357–376.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Basil Blackwell.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–53). Englewood Cliffs, NJ: Prentice Hall.
- LaCasse, D. (2018). *The subjunctive in New Mexican Spanish: Maintenance in the face of language contact*. PhD thesis, Department of Spanish, Italian and Portuguese, Pennsylvania State University.

- Lapesa, R. (1975). El sustantivo sin actualizador en español. *Boletín de la Comisión Permanente de Academias*, XXI, 50–67.
- 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.
- 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 and New York: Walter de Gruyter.
- Naro, A. J., & Votre, S. J. (1999). Discourse motivations for linguistic regularities: Verb/subject order in spoken Brazilian Portuguese. *Probus*, 11(1), 75–100.
- Ocampo, F. (1995). The word order of two-constituent constructions in spoken Spanish. In P. Downing & M. Noonan (Eds.), *Word order in discourse* (pp. 425–447). Amsterdam and Philadelphia: John Benjamins.
- Ocampo, F. (2003). The expression of topic in spoken Spanish: An empirical study. In R. Nuñez-Cedeño, L. López, & R. Cameron (Eds.), *A Romance perspective in language knowledge and use: Selected papers from the 31st linguistic symposium on Romance languages (LSRL), Chicago, 19–22 April 2000* (pp. 195–208). Amsterdam and Philadelphia: John Benjamins.
- 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 and Philadelphia: John Benjamins.
- Prince, E. F. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.
- Rivas, J. (2013). Variable subject position in main and subordinate clauses in Spanish: A usage-based approach. *Moenia*, 19, 97–113.
- Sankoff, D., Dion, N., Brandts, A., Alvo, M., Balasch, S. & Adams, Jackie. (2015). Comparing variables in different corpora with context-based model-free variant probabilities. In R. Torres Cacoullos, N. Dion & A. Lapiere (Eds.), *Linguistic variation: Confronting fact and theory* (pp. 335–346). New York: Routledge.
- Sedano, M., & Bentivoglio, P. (2014). Uso del sujeto y del objeto directo en el habla de once ciudades hispanohablantes. In A. Enrique-Arias, M. Gutiérrez, A. Landa, & F. Ocampo (Eds.), *Perspectives in the study of Spanish language variation: Papers in honor of Carmen Silva-Corvalán* (pp. 137–176). Santiago de Compostela: Universidad de Santiago de Compostela, Editorial Verba.

- 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.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Sorace, A. (2004). Native language attrition and developmental instability at the syntax—discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition*, 7(2), 143–145.
- Thompson, S. A. (1997). Discourse motivations for the core-oblique distinction as a language universal. In A. Kamio (Ed.), *Directions in functional linguistics* (pp. 59–82). Amsterdam: John Benjamins.
- Thompson, S. A., & Hopper, P. J. (2001). Transitivity, clause structure, and argument structure: Evidence from conversation. In J. Bybee & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 27–59). Amsterdam and Philadelphia: John Benjamins.
- Torres Cacoullous, R., & Aaron, J. E. (2003). Bare English-origin nouns in Spanish: Rates, constraints and discourse functions. *Language Variation and Change*, 15(3), 289–328.
- Torres Cacoullous, R., & Travis, C. E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Travis, C. E. (2005). *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin and New York: Mouton de Gruyter.
- Travis, C. E., & Torres Cacoullous, R. (2018). Discovering structure: Person and accessibility. In N. L. Shin & D. Erker (Eds.), *Questioning theoretical primitives in linguistic inquiry (Papers in honor of Ricardo Otheguy)* (pp. 67–90). Amsterdam and Philadelphia: John Benjamins.
- Vázquez Rozas, V. (2006). Animación, accesibilidad y estructura argumental preferida. In M. Sedano, A. Bolívar, & M. Shiro (Eds.), *Haciendo lingüística: homenaje a Paola Bentivoglio* (pp. 393–409). Caracas: Comisión de Estudios de Postgrado-Facultad de Humanidades y Educación Universidad Central de Venezuela.

1 Corpus of Conversational Colombian Spanish (cf. Travis, 2005, pp. 9–25).

2 That syntactic roles differ in distribution according to discourse referential uses and accessible referents has been proposed to motivate the cross-linguistic core-oblique distinction, a canonical example of the role of pragmatics in shaping linguistic structure (Thompson, 1997, pp. 73–75).

3 Subjects account for the vast majority of mentions, for both distance (this section) and persistence (Section 3.2.3): 85% of mentions of the same referent in the immediately preceding clause occur as a subject (858/1007, Torres Cacoullos & Travis, 2018, p. 86), and 85% of referents that are mentioned at least once in the subsequent three clauses are mentioned as a subject (61/410).

4 Excluded are instances where precise distance cannot be determined, for example because of unclear speech (n = 160).

5 For example, in split-ergative marking, the split is most commonly drawn between speech act participants vs. others (e.g., Delancey, 1981, p. 628); switch-reference marking may apply only to the third person (e.g., Haiman & Munro, 1983, p. xi); and zero-marked agreement is more likely on the third person (e.g., Bybee, 1985, p. 53).