Pulido, Manuel F. (in press). Remapping variable subject position in Spanish intransitives: A proposal for functionally defined categories in motion verbs*. Spanish in Context, 18.2.*

Abstract

In Spanish, a SVO language with variable word order, post-verbal subjects have been proposed to be favored for particular verb categories. For instance, based on agentivity, unaccusatives are proposed to favor VS as a whole. Motion verbs are regarded as unaccusatives generally favoring VS order. An alternative analysis is presented here, using data from two conversational corpora. Motion verbs are recategorized based on their predicted tendency to include adverbials in the sentence and compared with other unaccusatives. Motion verbs are divided according to their Deictic Function (Talmy 2000) into "come" verbs (i.e., "motion-toward-the-center", that is, the speaker), and "go" verbs. "Come" verbs do not often require target specification through an adverbial, whereas "go" verbs do. Adverbials were found to appear as post-verbal path specification in "go" verbs; due to weight factors, such specifiers favor pre-verbal subjects. Importantly, even when no modifier is present, trends persist, suggesting entrenchment of usage patterns.

Keywords: subject position, unaccusatives, weight factors

Remapping variable subject position in Spanish intransitives: A proposal for functionally defined categories in motion verbs.

1. Introduction

The variable position of subjects is a question of interest for linguists studying the effect of discourse on syntax, as well as for those exploring factors that affect the ordering of elements at the sentence and clause level. A large body of studies has revealed that the dichotomic distinction in S(ubject) V(erb) / V(erb) S(ubject) variation in Spanish, and in other languages with variable subject position, is motivated by a host of factors. Earlier research on subject position typically investigated the role of information structure (e.g., Givón 1983, Myhill 2005, Bentivoglio and Weber 1986), operationalized through quantifiable measures of topicality (referential distance, topic persistence), and based on cognitive and psychologically-motivated scales of referent giveness (i.e., accessibility, identifiability). More recent research has also explored other factors operating at the clause level, such as the effect of verb type (e.g., Mayoral Hernández and Chen 2006), the presence and weight of constituents in the clause (Mayoral Hernández 2010, Silva Corvalán 1982), the form and/or grammatical person of the subject (Posio 2012), the form and presence of verb objects (Ocampo 2014), as well as psychologically based factors such as pragmatic focus (Rivas 2008), or potential cross-linguistic influence (Benevento and Dietrich 2015), among others.

All of these factors have been shown to systematically influence variability in subject position. However, only in a few occasions has attention been given to how

these factors impact the language more generally, through conventionalization. An example of this is embodied in the Presentative Function (e.g., Bentivoglio and Weber 1986, Givón 1983, Naro and Votre 1999, inter alia), in which unaccusative (often motion) verbs are seen as associated with the introduction of new referents in discourse, by means of post-verbal subjects. In the literature on Spanish variable subjects, the Presentative Function accounts for the fact that VS order is strongly favored when subjects are first introduced in discourse, often via an unaccusative verb. But while the Presentative Function illustrates well how one same account can bring together discourse dynamics and syntactic variation, it is also a rare exception in the literature on subject variation. Defining variability based on similar functions tied to the discourse level is important for a more comprehensive account of word order variation, as it allows to go beyond the transient contingencies within a particular clause, and to make connections between discourse and factors that influence syntactic variability.

Under current classifications, motion verbs are typically characterized as unaccusative verbs that favor post-verbal subjects (e.g., Mayoral Hernández 2005, Mendikoetxea 1999). This paper investigates factors that may allow to characterize subclasses of motion verbs associated with different patterns of subject position. In particular, it will explore the impact of the presence and position of adverbials in the clause, which have been shown to influence the position of the subject (Mayoral Hernández 2010, Roggia 2018). More specifically, I consider whether particular communicative pressures may motivate the use of adverbials, in connection with the mentioned impact of adverbials on subject position. Based on the literature on Deictic Function (e.g., Talmy 2000), in which the speaker is taken as a center of refence in speech, motion verbs are classified as expressing either “motion-towards-the-center”

(e.g., come, appear) or "motion-from-the-center" (e.g., exit, go) (Talmy 2000). In this study I draw on this classification to explore a novel prediction: while "come" verbs should rarely require path specification via adverbials, with a referent being described as coming towards the speaker, increased rates of adverbials will be present in sentences with "go" verbs (to specify where a referent goes). Based on the literature, this expectation for different rates of use of adverbials across groups of verbs is predicted to be associated with different patterns of subject position.

In what follows, I briefly review evidence on how communicative routines at the discourse level may shape syntactic preferences at the clause level. I then describe how a connection can be explored between adverbials and verbs of motion, that may predict variable patterns of subject position. The following sections report on the criteria, analysis and results of corpus data from conversational Peninsular Spanish. Finally, I discuss the implications of applying some traditional verb taxonomies in variationist studies (e.g., the Unaccusative Hypothesis), and argue for an approach based on functionally-defined categories where possible.

1.1 A function-driven approach to the study of variability in Spanish subject position

Under the Preferred Argument Structure (e.g., Du Bois 2003), the Quantity and Role constraints describe generalizations grounded in cross-linguistic evidence that allow us to explain how conventionalized structure (i.e., syntax) arises from function (although see e.g., Haig and Schnell 2016). Based on cross-linguistic data, Du Bois suggests that there is (a) an "overall constraint on the quantity of new information that can be handled within a single processing unit"; and (b) "a predictable locus for the heaviest cognitive demands" –that is, a particular *locus of*

focus within the clause- through which users know where to "direct their limited attentional resources."

The way these two principles (i.e., cognitive constraints, and a designated place of pragmatic focus in linguistic structure) interact and lead to conventionalization in languages is often tied to function. In order to find trends of systematicity one should consider communicative routines present in discourse, given their potentially important implications in variation and in giving rise to functional-syntactic patterns. In the case of subject placement, the interplay between the use of adverbials, subject position and verb type is a scenario in which multiple factors potentially converge to have an effect on syntax.

Adverbials are often seen as peripheral elements in the sentence structure, and one might even wonder whether they may play a prominent role in discourse. But there is compelling evidence that adverbials have a non-negligible influence on subject position. Indeed, several studies indicate that the presence and position of adverbials has an influence on subject position in Spanish (Mayoral Hernández 2010, Roggia 2018, Silva Corvalán 1982). In a recent study (Roggia 2018), in which Mexican speakers answered questions following presentation of a story, Roggia compared the relative weight of factors, including the type of verb (along the Split Intransitivity Hierarchy), subject heaviness, definiteness, and location of the adverbial phrase. In responses to broad focus questions (What happened?), the location of the adverbial was the most powerful predictor of subject position.

If the presence and position of adverbials have an important influence on the ordering of the subject, it is relevant to identify what circumstances and/or groups of verbs may determine when and how adverbials are used. A previous study by Mayoral Hernández (2005) examined the association between adverbials and subject

ordering in different groups of verbs. However, rather than considering how adverbials may influence subject position, the conclusions in the study emphasized how SV/VS preferences, assumed to be the norm for some verbs, affect the ordering of other elements. He found that unaccusatives favored VS in a significantly higher proportion (33.8%) than intransitive, transitive and copulative verbs (12.5%). Given the fact that overt adverbials and subjects tend to be in complementary distribution (i.e., they tend not to be both pre-verbal or both post-verbal), the study concluded that the tendency for unaccusatives to have post-verbal subjects is a predictor for adverbials to be found to the left of the verb. However, the assumption that unaccusatives homogeneously favor VS, influencing the ordering of other elements (e.g., adverbials), appears to be somewhat circular. Are adverbials shifted because of a pre-existing tendency for unaccusatives (or any other specific type of verb) to have post-verbal subjects, or is the trend of VS influenced to some extent by the use and placement of adverbials?

In the light of some of these findings, it seems that it might be possible to describe groups of verbs that favor VS based on their tendency to be complemented by adverbial clauses in discourse, rather than relying on groups of verbs predefined by syntactic-semantic attributes. While in many cases the use of adverbials may be quite unpredictable, some contexts may be described in which the use of adverbials is favored and predictably affects word order offer. The question, of course, is whether and why there is a theoretical motivation for certain verbs to be used in conjunction with adverbials more often than others. This paper takes a functional approach that tries to depart from a priori assumptions about subject position stemming directly from verb semantics. In such an approach, different groups of intransitive verbs (especially verbs of motion) can be more flexibly categorized based on their trend for

complementation through adverbial. The current study will focus on types of motion verbs, discussed in the following section, which are predicted to favor overt path specification via adverbials or not (e.g., Talmy 2000; Lewandowski 2007).

Exploring function-based groups of verbs may allow for idiosyncratic patterns to emerge, which may have been difficult to detect under previous verb taxonomies. That is, because previous variationist studies including intransitive verbs have considered categories such as unaccusatives or unergatives as cohesive units, it is possible that groups that unevenly favor SV or VS may have been conflated under a particular label. I will test the hypotheses that, first, certain verbs of motion may be grouped according to their functional need for path specification, as independent of other intransitive verbs (including other verbs of motion and unaccusatives); and that, secondly, this classification will allow for a better characterization of variability in subject position in Spanish in those verbs. The analysis reported below will examine these questions and present evidence in support of a usage-based approach taxonomy of motion verbs, often considered as a homogeneous group.

1.2 A functional approach for a classification of motion verbs

The focus of the present study will be on self-directed verbs of motion (e.g., *ir* 'go' *venir* 'come'*, llegar* 'arrive'*, bajar* 'descend'*,* etc.); these tend to be considered unaccusatives insofar as they do not present agentive, but rather object-like, features in the subject. In present-day Spanish, it has been proposed that the main direct syntactic implication derived from the status of motion verbs as unaccusatives is the

preference for post-verbal subjects[1] (for syntactic implications at earlier stages of Spanish, see Aranovich 2003).

There have been several proposals for taxonomies of motion verbs (e.g., Levin 1993, Levin and Rappaport Hovav 1992, Talmy 2000; for taxonomies specific to Spanish see, e.g., Morimoto 2001). There are some important differences in the classification of verbs that are considered part of this class, and there is also ample variability in their degree of specificity and the criteria considered, as will be discussed below. More important to the issue at hand, it is not always clear how some taxonomies of motion verbs would relate to the syntactic preferences in discourse. It appears that the most insightful approach to date comes from the proposal that motion verbs should be considered unaccusatives, based on the thematic roles of the subject (e.g., Mayoral Hernández 2005, 2010; Mendikoetxea 1999).

The notion that unaccusative verbs will favor VS order relies on the assumption that semantically defined categories will be directly mapped to syntactic preferences that affect the position of the subject. In the current study, rather than considering semantically defined categories as directly affecting the subject, I consider how the verb may have direct consequences for other elements in the clause that have an *indirect* effect on the subject. Inherently-directed motion verbs may provide one such scenario in which the focus may be on a constituent other than the subject or verb objects is found, namely, adverbials. Building on previous work on deixis in motion verbs, I propose that function-defined categories can be established. These inherently differ in their need for specification through the use of adverbials,

[1] While this seems to be the most accepted view, De Miguel (1999, 74-78) proposed that inherently-directed verbs be subdivided into two groups with different syntactic preferences. Inherently achieved location verbs (*subir* 'ascend', *bajar* 'descend', *entrar* 'enter', *salir* 'exit', *llegar* 'arrive) are described as being unaccusatives, while verbs of inherent path (*dirigirse* 'go towards', *moverse* 'move oneself', *regresar* 'return') are classified as unergative in this view.

which affect the ordering of elements in the sentence, and ultimately influence the position of the subject. Inherently-directed motion verbs (henceforth, "motion verbs") have been semantically divided into "go" and "come" verbs in some taxonomies (e.g., Talmy 2000, Lewandowski 2007). While "come" verbs do not need to explicitly code the path of motion, which is implicitly available, "go" verbs often accompanied by a directional complement in order to express path (i.e., an adverbial phrase). This is particularly relevant in the study of subject position, given that post-verbal adverbials exert a "repelling force" on the subject through clause-level weight effects, pushing the subject to a pre-verbal position, S+V+Adv (Mayoral Hernández 2005, 2010; Silva-Corvalán 1982). Consequently, "go" verbs would presumably be a case in which focus is routinely placed on a constituent of the clause (i.e., an adverbial) which triggers a particular subject position.

Motion verbs can be classified by considering a center of reference based on the speaker –or sometimes an interlocutor. Deictic motion verbs are classified as expressing either "motion-towards-the-center" (e.g. come, appear) or "motion-from-the-center" (e.g. exit, go) (Talmy 2000). A classification based on the notion of the Deictic Function allows to explore the predictions in the current study: that increased rates of modifiers will be present in sentences with "go" verbs, while the opposite pattern is expected for "come" verbs, in both cases influencing the position of the subject. This approach differs from other classifications concerned with verb semantics, in which deiction (critically opposing "come" and "go") is not present. For instance, in Levin (1993, 263-270), a category of "leave verbs" is proposed (e.g., *abandon, leave, desert*), but *come* and *go* are still part of the same category of inherently directed motion verb. They are likewise grouped together under the "arrive class" in Levin and Rappaport Hovav (1992, 252-253). Other taxonomies proposed

for Spanish verbs, such as Morimoto (2001, 82), distinguish inherent direction verbs according to their type of trajectory (trajectory *Hacia* 'Towards'; trajectory *De / A* 'From / To'), but "come" and "go" verbs are still conflated within the same type of trajectory (*De / A*). This classification also differs from De Miguel's (1999) division of motion verbs into unaccusatives and unergatives. Therefore, a categorization based on the deictic function allows for an approach absent in previous studies examining variation.

Based on the rationale described above, a novel aspect of the present proposal is that a verb can be more flexibly categorized: for instance, some tokens of the verb *llegar* 'arrive' may express motion towards-the-center while others from-the-center (examples from corpus data are discussed below). Rather than assuming that the semantics of the verb determine a SV / VS preference, the prediction is that the presence and position of adverbials will be associated with the described categories, and will in turn be predictive of subject position. Despite proposals for a gradience in unaccusatives (e.g., Sorace 2000, 2004), previous variationist studies tend to assume a uniform unaccusative category when examining SV/VS patterns across groups of verbs in Spanish (e.g., Mayoral Hernández 2005, 2010). This study will test the idea that motion verbs do not homogeneously favor VS, while testing an alternative approach that differs from the assumption that VS is determined by unaccusativity. The analysis presented below will show that such a functional categorization of verbs gives insight into how the interplay between verb semantics and adverbial usage patterns influence subject position.

Further, if the Deictic Function can account for usage trends in verbs of motion, it is possible that the verb categories associated with that function may show signs of conventionalization. That is, not only should we ask whether certain verbs are

used in conjunction with adverbials, and if this has an effect on subject placement, but also whether stable patterns of use associated with the ordering of elements may have given rise to entrenchment even in the absence of the elements (the adverbial phrases) that motivate patterns of variation. In other words, a more general question this study addresses is how the higher level of discourse may both affect and transcend the domain of the clause, leading to emerging patterns of conventionalization in some verbs.

1.3 Hypotheses

Considering all of the above mentioned, in this paper I explore the validity of the following hypotheses for understanding syntactic patterns of subject position in a groups of intransitive verbs in Spanish:

(1) certain types of motion verbs (namely, "go" verbs) are associated with adverbials in focalized (i.e., post-verbal) position;

(2) the presence and position of adverbials in the sentence predicts subject position;

(3) patterns of subject position motivated by communicative functions (1) and weight factors (2) may remain in the absence of overt specifiers, providing evidence of conventionalization.

2. The present study

2.1 Data and envelope of variation

The data analyzed were obtained from two corpora: *Corpus Oral de Referencia del Español Contemporáneo (CORLEC),* (Marcos Marín 1992) and the barrio de Salamanca (Cestero et al. 2012) subcorpus of the *Proyecto para el Estudio Sociolingüístico del Español de España y América* (Moreno Fernández et al. 2001-present). The corpus sections analyzed consisted of conversational data collected from speakers in Madrid.

The analysis of written texts in the exploration of subject placement in Spanish seems to have become a general pattern, with only a few authors (e.g., Ocampo 2009, 2014, Rivas 2008, Alamillo 2009) having analyzed conversations. Data from conversational corpora provide the golden standard for variationist analysis, as they consist of spontaneous language use, free of instruction- or register-bound conventions and capture language produced in real time (Labov 1984). The present study is, to my knowledge, the only one in which only Peninsular Spanish is analyzed.

The analysis of the conversational data produced by Madrilean speakers in both corpora yielded 757 tokens of sentences containing unaccusative verbs: 344 tokens from 79 conversations in CORLEC, and 413 from 36 conversations in PRESEEA. Two anonymized speakers that were identified through the available corpus information as being from countries other than Spain were excluded.

2.2 Circumscribing the variable context

As described above, this study will test the idea that motion verbs do not homogeneously favor VS, contrary to the prediction that this would be the case if motion verbs are a homogeneous subclass of an unaccusative category. Therefore, the

analysis will also examine and consider these motion verbs against other unaccusatives, included here as a baseline of the common finding that verbs classified as unaccusatives tend to favor VS (e.g., Mayoral Hernández, 2005). This examination of pre- and post-verbal subject placement includes only utterances with overt subjects in intransitive verb clauses, either in third person singular or third person plural. This is due to the fact that the first grammatical person, as well as the second one, are anchored in the speakers and differ from the third person in their referentiality.

Only affirmative declarative sentences were included, as other types of sentences, such as interrogatives, may be affected by intonation patterns and structure (Brown and Rivas 2012, 23). Exclamatives work in a similar way to interrogatives, where SV tends to be predominant too. To exclude polarity effects, negative sentences were excluded as well as, interrogative and exclamative sentences.

An important point to note is that, as discussed below, in order to consider a variable context that allows for different verb types to be considered, all unaccusatives (and not just motion verbs) were included. This allows to actually compare (1) verbs that are associated within a specific category (e.g., verbs expressing deictive or non-deictive motion, or no motion); but also, independently (2) tokens of verbs that tend to appear with a complement and those that do not. While the hypotheses described above propose that some verbs (e.g., some verbs of motion) will be often complemented, the variable context allows to contrast these and other verbs. Consider Example 1 with *nacer* 'to be born', an unaccusative expressing change of state, but which might often appear with a post-verbal complement, and thus favor SV. The analysis below will examine these different factors.

Finally, a few specific verbs exhibiting an almost categorical distribution were excluded, namely *haber* 'there is/are' following Bentivoglio and Weber (1986)

(Example 2) and also the copulative verb *ser* 'to be', given the difficulty of objectively determining subjects and predicates (compare ex. 3 and ex. 3'). If (3) and (3') are analyzed differently, this is solely done on the assumption of the categoricality of SV with *ser*; if judged the same, there is no objective way to identify the subject. Either way, these were considered grounds for their exclusion.

(1)	*Una*	*de*	*mis*	*sobrinas*	*nació*	*en*	*Salamanca*
	One	of	my-PL	niece-PL	be born-3SG	in	Salamanca

'One of my nieces was born in Salamanca' (MADR_H21_020)

(2)	*Hay*	*una*	*cosa*	*muy*	*curiosa*
	There.is-3SG	one	thing	very	peculiar

'There's one interesting thing' (ACON001A p12)

(3)	*Esto*	*era*	*el*	*antiguo*	*campo*	*Vallehermoso*
	This	was-3SG	the	old	field	Vallehermoso

'This was the old "Vallerhemoso" field' (ACON006C p227)

(3')	*El*	*antiguo*	*campo*	*Vallhermoso*	*era*	*esto*
	The	old	field	Vallehermoso	was-3SG	this

'The old "Vallehermoso" field was this'

2.3 Linguistic Variables

2.3.1 Measuring contextual accessibility of the subject: Referential Distance

The Presentative Function is perhaps the most established finding about the role of information structure in Spanish word order (e.g., Bentivoglio and Weber 1986), also demonstrated in other Romance languages (e.g., Naro and Votre 1999). Referents mentioned for the first time, and in particular those the speaker intends to establish as topical in discourse, have a tendency to be presented post-verbally, while already topical subjects are SV (Bentivoglio Weber 1986, Givón 1983, Naro and Votre 1999, inter alia). The measure of accessibility of referents was provided through Referential Distance (RD) of the subject, i.e., distance since the last preceding mention (Myhill 2005), within the 12 previous clauses. Considering general trends of information flow, the presentative function is a functional instantiation of the dichotomic division of information into "old" vs. "new", or "theme" and "rheme" in the tradition of the Prague School: "Old" information, which is accessible and tends to be presented first (a.k.a. "rheme"), tends to precede new information. Based on previous studies of the presentative function, it is expected that subjects that have not been mentioned within the preceding 12 clauses will be re-introduced or mentioned for the first time as VS.

Tokens were coded as containing a subject that had been mentioned within the 12 preceding clauses or not. Referents that were not present within the 12 preceding clauses will be referred to as "first mentions," favoring VS order. In addition, given the cognitive constraints and the load placed on short-term memory during discourse (Hawkins 2003) it is expected that a shorter RD will correlate with a higher probability of a pre-verbal subject.

Discourse markers and fillers (*vaya* 'well,' etc.) were not counted as intervening clauses. Example (4), with 1 intervening clause, illustrates how RD was coded:

(4) *Entonces, si yo dijera todo lo que **ellos** me han dicho,*

[previous mention**]**

*<u>se armaría</u>, **ellos** mismos me lo han dicho así*

[Target**]**

Then, if I say-SUBJ-PAST all it that **they**-M me have-3SG said,

REFL <u>kick.up</u>-COND, **they** selves me-IO it-DO have-3SG said like.this

'Then, if I were to say everything **they** told me, <u>there would be chaos</u>, **they** told me so.' (ACON009A p67)

2.3.2 Definiteness of the subject

Definiteness of the subject is known to strongly influence word order, especially in verbs in the unaccusative class (Lozano and Mendikoetxea 2011). A stronger constraint affects indefinite bare subjects (e.g., *mujeres* 'women'), which in Spanish cannot take a pre-verbal position; if a noun is to appear pre-verbally, it is preceded by a determiner (e.g., *unas mujeres* 'some women'). Definiteness was operationalized as a binary variable, with tokens coded as having either definite or indefinite subjects.

2.3.3 Other constituents within the sentence: Adverbial expressions

As stated above, it is expected that other constituents present in the clause will affect the position of the subject. Previous studies have investigated the influence of adverbials on the reorganization of elements in the clause, with tendency to avoid the co-occurrence of a subject and an adverbial both simultaneous pre- or post-verbally (Silva-Corvalán 1992, Mayoral Hernández 2005). Here adverbials of place, time and manner were coded for presence and position before (Example 5) or after (6) the verb.

Cases in which two or more modifiers were simultaneously present in different positions with respect to the verb (Example 7) were coded separately; due to their low frequency (20 tokens) they were not included in the analysis. Tokens where a specifier was present between the subject and the verb had also very low counts (23 tokens) and were not included. The adverbial *entonces* was not included as an adverbial, as it may be used as either a filler or a discourse marker.

(5) *y entonces al día siguiente llega* ***la poli***
and then to.the day following **arrive**-3SG **the police**-3SG
'and then the following day the police arrives' (CCON018B p68)

(6) *Lo que pasa que* ***los de Erasmus*** *se quedan seis meses.*
It-N that happen-3SG that **those of Erasmus**
REFL remain-3PL six months
'What happens is that the Erasmus people stay for six months' (ACON006C p159)

(7) *Mi* ***dama*** *mañana* ***se va*** *para Lanzarote con todo el equipo televisivo*
My lady tomorrow **go**-REFL.3SG towards Lanzarote with all the team televisive
'My lady is leaving tomorrow for Lanzarote with the whole TV crew' (BCON022B p79)

2.3.4 Verb type

As stated above, verb type has been found to be a factor consistently influencing subject position in Spanish. While most Spanish verbs are in agreement with the general tendency to follow the subject, unaccusative verbs have been claimed to favor VS order.

However, as described in the introduction, it was hypothesized that a finer-grained analysis based on verb semantics might reveal interesting patterns in discourse. As indicated above, all instances of intransitive verbs are considered as part of the variable context, which also includes all unaccusatives. While the inclusion criteria for unaccusatives (which encompass motion verbs) follows Mendikoetxea (1999), the classification criteria for subgroups of verbs was based on the Deictic Function. Following the distinction discussed in the introduction, unaccusatives were divided into three groups: non-deictic motion/no motion verbs, "go" verbs, and "come" verbs.

To operationalize this approach with the speaker/interlocutor as the center of reference, verbs were coded based on the modifier information expressed in the sentence and on contextual information. It should be noted that, while in Spanish instances of *venir* 'come' are unequivocally identified as "toward-the-center" or *inbound,* other motion verbs such as *llegar* 'arrive', *entrar* 'enter', *salir* 'exit', etc., may be classified as either "go" or "come" verbs, based on adverbial specification and/or contextual information. Thus, *llegar* 'arrive' in (Example 8) is a "motion-towards-the-center" or inbound verb, whereas in (9) it is outbound. Additionally, some Spanish motion verbs may be non-deictic in some instances of use (Lewandowski 2007), as is the case of *llegar* in (10), which expresses motion but is not oriented with reference to a center or origin. A third, broader category, termed

"non-deictic", included all other verbs that either did not express motion to- or from-the-center, or expressed no motion at all. Evidently, a considerable number of intransitives, including unaccusatives (verbs of existence, etc.) do not express motion. These also included all intransitive verbs that were not the "go" or "come" types. Consequently, verbs were coded as "come" verbs, "go" verbs, or "non-deictic". This coding scheme allows to test the hypothesis described above. To remind the reader, the hypothesis regarding "come" and "go" verbs is that the former group will present particularly low rates of adverbial complements (given that the center of reference is already implicit), while the opposite should be true for "go" verbs, requiring patch specification. This hypothesis can be tested relative to other verbs (non-deictic) that serve as a baseline for comparison.

(8) ***Llegaron dos amigos*** *de la hermana de Tato* *cuando eh cuando n/ nos íbamos nosotros*

Arrived-3PL **two friends** of the sister of Tato, when uh when u/ us leave-PST-1PL we

'Two friends of Tatos' sister arrived when we were about to leave'

(ACON006D p87)

(9) *Te imaginas que* ***El Quijote*** ***llega*** *al mar*

You imagine-2SG that **the Quixote** **arrive**-3SG to sea

'Imagine that Quixote gets to the sea' (CCON018B p224)

(10) *porque* ***llega*** *antes* ***el calor*** *abajo*

because **arrive**-3SG before **the heat** below

'because the lower floor gets heated first' (CCON031A p67)

3. Results and Discussion

To test the hypotheses outlined above, a Generalized Linear Mixed-Effects Regression was performed on the data using the glmer function in the lme4 package (Bates et al. 2015) in R (version 3.3.2, R Core Team 2016). The analysis started with the maximal converging model, which included all variables as fixed effects and by-subject and by-verb random intercepts, as well as random slopes for adverbial presence/position. A backward selection procedure followed, in which one by one, the variable that least contributed to the model was tentatively removed. Each resulting model was compared to an identical model containing the variable by performing ANOVAs and using the resulting p-values and AIC values. Variables that did not significantly improve the model were removed. The selected model is presented below[2].

```
glmer (VS ~ SubjectRD + AdvPosit + Verb.class + Definiteness +
            (1 + AdvPosit | verb) + (1 | subject),
                    data = plots.labels, family = binomial,
```

[2] Given the fact that data came from two different corpora, collected in different decades, each token was also identified as part of CORLEC or PRESEEA. This extralinguistic variable made no significant contribution to the model and was therefore removed.

control = glmerControl (optimizer = "bobyqa",

optCtrl = list (maxfun=2e5)))

Table 1 shows the output of the generalized mixed-effects logistic regression analysis. The reference level for the intercept were tokens with no adverbials and definite "old" subjects in the "Come" verb category. In what follows, I will discuss the significant results in connection with the hypotheses and predictions.

Table 1. Output of generalized linear mixed-effects model for post-verbal subjects

	Estimate	**SE**	***z***	***p|t|***
Post-verbal Subject				
Intercept (ref. level 'Come')	1.809	0.441	4.106	<0.001
SubjectRDold	-0.376	0.235	-1.602	0.102
AdvPositPost-verbal	**-2.189**	**0.369**	**-5.927**	**<0.0001**
AdvPositPre-verbal	7.335	5.355	1.37	0.171
Verb.class non-deictic	-0.645	0.404	-1.599	0.11
Verb.class 'Go' verb	**-1.029**	**0.476**	**-2.16**	**<0.05**
Indefinite	**1.133**	**0.331**	**3.422**	**<0.001**

3.1 Communicative functions and verb modifiers

The view tested in the present study is that discourse functions and information structure may have a direct influence on sentence level elements (i.e., presence of

adverbials), which would then indirectly influence the ordering of the subject relative to the verb. Therefore, the analysis must first examine the distributions of adverbials across verb groups, before arguments about clause weight effects on subject position can be put forward.

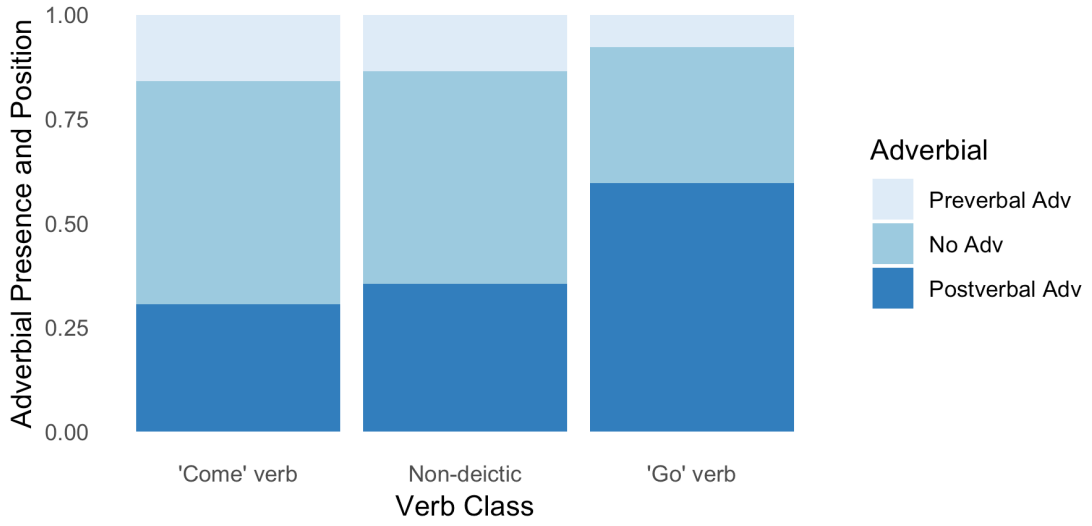
Above it was proposed that "go" verbs (or motion from-the-center verbs) do not *per se* convey path related information; it is only through either periphral information provided elsewhere in discourse or via explicit complementation that path specification can become evident to the interlocutor. Consistent with this, the data in Table 2 show that 68% of "go" verb tokens are used in conjunction with adverbials. On the opposite end, "come" verbs, which take the speaker or interlocutor as the center of reference for motion, should disfavor the presence of adverbials. The data show that this is indeed the case, as 54% of "come" tokens lack an adverbial. This proportion is similar to that of tokens that fall in neither category do not express deictic motion, which also disfavor the presence of adverbials. The patterns are in line with the hypothesis that tokens coded as "go" verbs will favor the presence of adverbials after the verb, whereas "come" verbs and non-deictic tokens tend not to co-occur with adverbials.

Table 2. Token counts and percentages for the presence and position of adverbials across verb categories.

Adverbial	**"Come" verb**		**non-deictic**		**"Go" verb**		**Total**	
No Adv.	54%	98	51%	204	33%	42	48%	344
Post-verbal Adv.	31%	56	36%	142	60%	77	39%	275
Pre-verbal Adv.	16%	29	14%	54	8%	10	13%	93

It was also expected that the information modifying the verb would be in pragmatic focus, appearing post-verbally. Figure 1 below presents the rates of presence and position of adverbials with the different verb classes coded. Across the different categories, it becomes apparent that adverbials do in fact occur predominantly to the right of the verb. This preference is preserved across the verb categories despite notable differences in the degree of presence of specifiers, with 60% of adverbials in "go" verbs being post-verbal, relative to only 31% of adverbials in "come" verbs.

Figure 1. Presence and position of complements in each verb category.

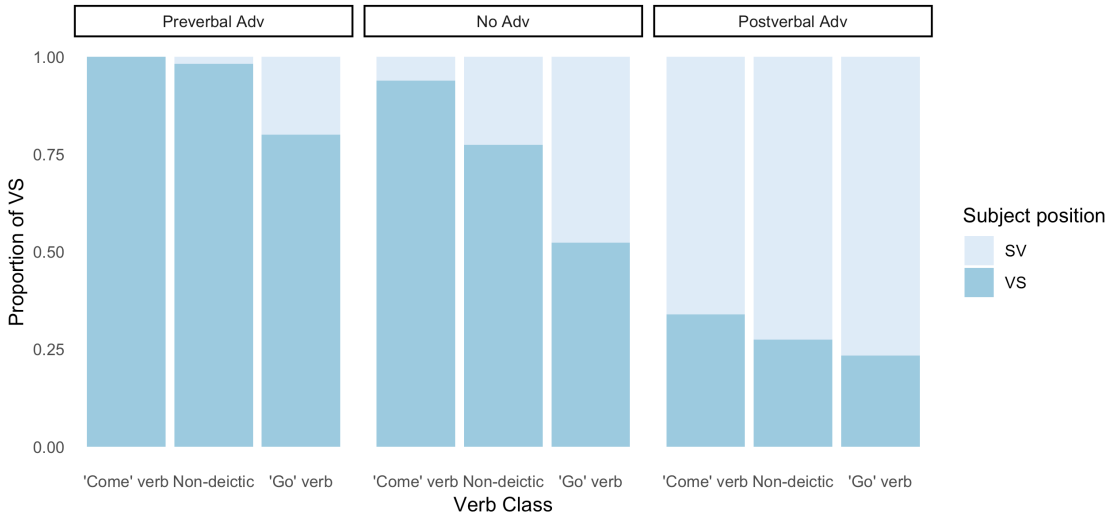


3.2 Specification, weight factors and their effect on subject position

The finding of different patterns of complementation in the verb categories examined here is of importance insofar as the presence of adverbials can be shown to have an influence on subject position. As discussed, previous studies have

demonstrated that weight factors play a role in how linguistic material is arranged around the verb in a sentence. Weight effects are evidence of the tendency for a balance in how elements in the same sentence are displayed pre- and post-verbally. Post-verbal arguments favor a subject in pre-verbal position, and the opposite is true for pre-verbal arguments. The results of the statistical analysis replicate the result in previous studies that the presence of post-verbal modifiers is a highly significant factor affecting subject placement, accounting for a significant decrease in the probability of post-verbal subjects (*b:* -2.19*;* SE: 0.37; $p<0.0001$). As shown in Figure 2, the effect is robust across the different verb categories. Rates of post-verbal subjects are consistently disfavored if adverbials are found after the verb, while the reverse is true when adverbials are absent or placed before the verb. “Go” verbs show the highest rates of post-verbal adverbials, and lower rates of VS than the other categories.

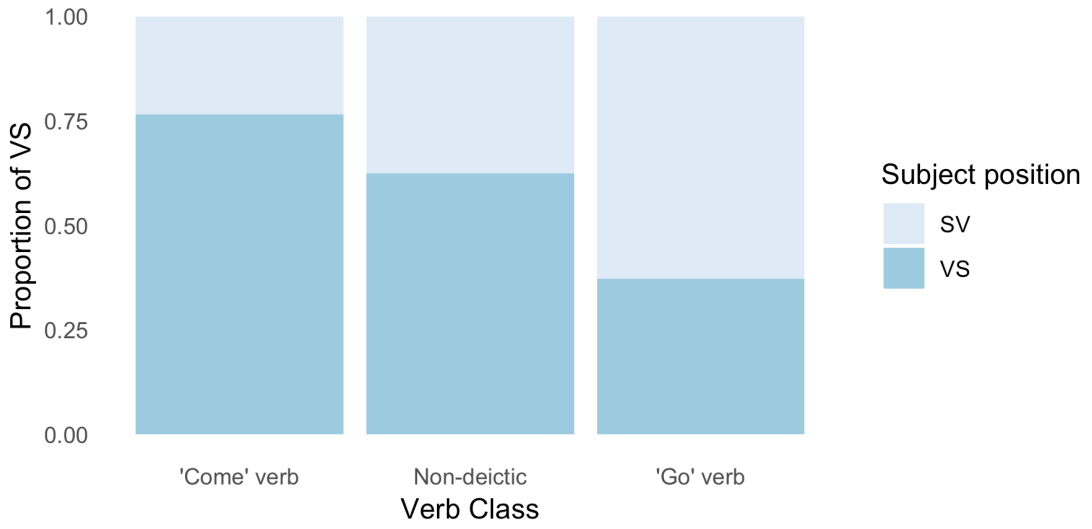
Figure 2. Rates of post-verbal subjects as a function of adverbial presence and position across verb categories.



Yet when it comes to their position in the sentence, do subjects influence adverbials and modifiers, or do the latter influence subjects? It is important to note that views anchored exclusively in a syntactic-semantic approach of the verb would argue for a trend in unaccusative subjects to be post-verbal across the board; if this holds true, adverbials would tend to occur before the verb. However, this is not always the case. Cases in which the subject is pre-verbal due to the influence of the presence of adverbials cannot be adequately accounted for without considering function and usage trends. Regarding adverbial placement, Figure 1 illustrates that the different verb classes included in the analysis do not appear to form a homogeneous group. Rather, the proportions of adverbials after the verb suggest a gradient scale, with "go" on one end and "come" verbs on the other. If post-verbal subjects, a syntactic pattern often associated with unaccusativity for Spanish, are found not to be homogeneously present across verbs in this category, this would have important consequences for verb categorization in variationist studies. More importantly, a preference for particular patterns in subject position may be accounted for in a usage-based approach as motivated by discourse functions and their interaction with clause weight effects on syntactic structure.

Based on the weight effects described above, we would predict that greater rates of post-verbal modifiers should be matched by a preference for pre-verbal subjects. Therefore, a tendency for post-verbal subjects in unaccusatives might be found to be favored only in verb classes with low rates of post-verbal modifiers, rather than across the board. That is, rates of post-verbal subjects should be inversely correlated to rates of post-verbal adverbials across verb categories, reversing the trends in Figure 1. This is precisely what Figure 3 shows.

Figure 3. Rates of VS across different verb categories.



3.3 Evidence of conventionalization of frequent usage patterns

These patterns provide initial support for a more nuanced explanation of variable subject position in Spanish motion verbs than has been provided to date. By re-grouping tokens based on their discourse functionality and allowing for the same verb to be categorized according to the Deictic function, the analysis trascends the rigid classification of verbs as lexical types reduced to one intrinsically motivated parameter. In other words, rather than being defined by *the verb type used*, subject position is influenced by *how a given verb token is used,* as determined by discourse.

So far, the data suggest that this alternative way of categorizing corpus data predicts when an adverbial expression will be found more often in a post-verbal position, and therefore influence subject position rates. However, that by itself is no evidence of conventionalization of usage. To investigate signs of entrenchment, we must turn to hypothesis (3): patterns of subject position motivated by communicative

functions (1) and weight factors (2) might remain in the absence of overt complementation.

As described above, the presence and position of adverbials was a highly significant variable in predicting subject position. Now, can these categories predict rates of subject position even in the absence of adverbials? The results of the mixed-effects logistic regression suggest that this is the case. The analysis revealed that "come" and "go" verb categories have significantly different rates of VS. As predicted, the category of "go" verbs (such as *ir* 'go', *salir* 'exit', *llegar* 'arrive') significantly reduces the likelihood of VS (*b:* -1.03*;* SE: 0.48; $p<0.05$). That is, even after accounting for the effect of presence and position of adverbials in the model, "come" and "go" verbs still are predictive of significantly different rates of post-verbal subject position. The patterns are further illustrated in Table 3, which presents the rates of VS by verb category considering the presence and position of modifiers. As Table 3 shows, the clearest contrast across verb categories is found when comparing tokens in which no adverbial is present relative to when it is found to the right of the verb (top shaded cells). The overall rates shown at the bottom are an indicator of the effect each category has on post-verbal subject position. The stark contrast between the "come" and "go" verb categories confirm the predicted patterns.

Table 3. Rates of VS by Verb Category as a Function of Adverbial Position

	Non-deictic		**'Come' verb**		**'Go' verb**	
No Adverbial	77%	204	94%	98	55%	42
Post-verbal Adverbial	27%	142	34%	56	23%	77
Pre-verbal Adverbial	98%	54	100%	29	80%	10
	63%	**400**	**77%**	**183**	**37%**	**129**

One might wonder about the weight that the different lexical types have in the effects observed in these categories. Given their high overall frequency cross-

linguistically, *ir* ‘go’ and *venir* ‘come’ would be expected to account for a large part of the tokens in their respective categories. The effects in “go” and “come” verbs are in fact mostly driven by these central members, with *ir* ‘go’ accounting for 76% of “go” verbs and *venir* ‘come’ for 65% of “come” verbs. A more detailed description of the lexical types each categories is provided in Table 4 below. Verbs representing less than 1% of the data within a given category were aggregated under “other.”

Table 4. List of verb types listed under each category.

Category		Verb		Percentage of category
"Come" verbs	N=183	*venir*	‘come’	64.5%
		llegar	‘arrive’	12.6%
		salir	‘exit’	6%
		entrar	‘enter’	3.8%
		ir	‘go’	3.3%
		aparecer	‘appear’	2.7%
		acercarse	‘approach’	1.6%
		meterse	‘get in’	1.6%
		volver	‘return’	1.1%
		other		3.5%
"Go" verbs	N=129	*ir*	‘go’	76%
		salir	‘exit’	14.8%
		llegar	‘arrive’	5.4%
		other		3.9%
Non-deictic	N=400	*pasar*	‘pass’	7.25%
		cambiar	‘change’	6.5%
		ir	‘go’	6.5%
		quedar	‘remain’	6.5%
		salir	‘exit’	6.5%
		morir	‘die’	4.75%
		estar	‘be’/‘stay'	4%
		caer	‘fall’	3.75%
		acabar	‘end’/‘end up’	3.5%
		nacer	‘be born’	3.5%
		llegar	‘arrive’	3%
		parecer	‘seem’	3%
		casarse	‘get married’	3%
		terminar	‘end’	2.5%
		entrar	‘enter’	2%

subir	'go up'	2%
pasar	'pass'	1.75%
desaparecer	'disappear'	1.5%
ocurrir	'happen'	1.5%
crecer	'grow'	1.25%
meterse	'get in'	1.25%
ponerse	'place onself'/'become'	1.25%
venir	'come'	1.25%
bajar	'go down'	1%
juntarse	'join'	1%
parar	'stop'	1%
quedarse	'stay'	1%
other		17.25%

The question is, then, whether the observed patterns are simply a result of the high number of tokens from these central types, or if the preferences for SV/VS are a common trait shared by the other verbs in these categories. Figure 4A shows the proportion of VS in these categories when their respective central lexical types are included (Go verbs: 37%, N=129; Come verbs: 77%, N=183) , while Figure 4B shows the patterns after *ir* 'go' and *venir* 'come' are removed (Go verbs: 35%, N=31; Come verbs: 83%, N=65). The data suggest that the same trends persist across a wider range of types even when the main contributors are removed.

Figure 4A. Proportion of VS in "Come" and "Go" verbs when *ir* 'go' and *venir* 'come' are included

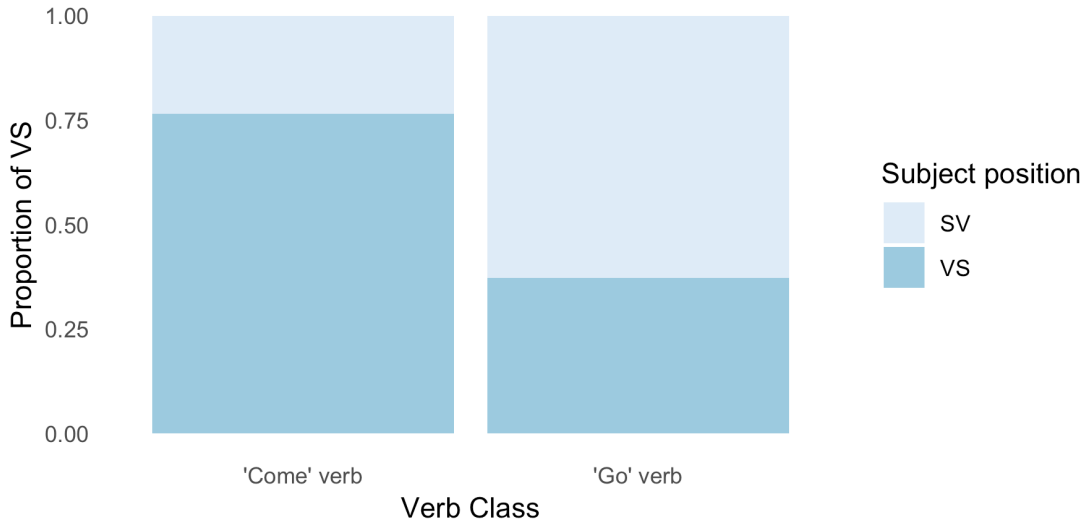
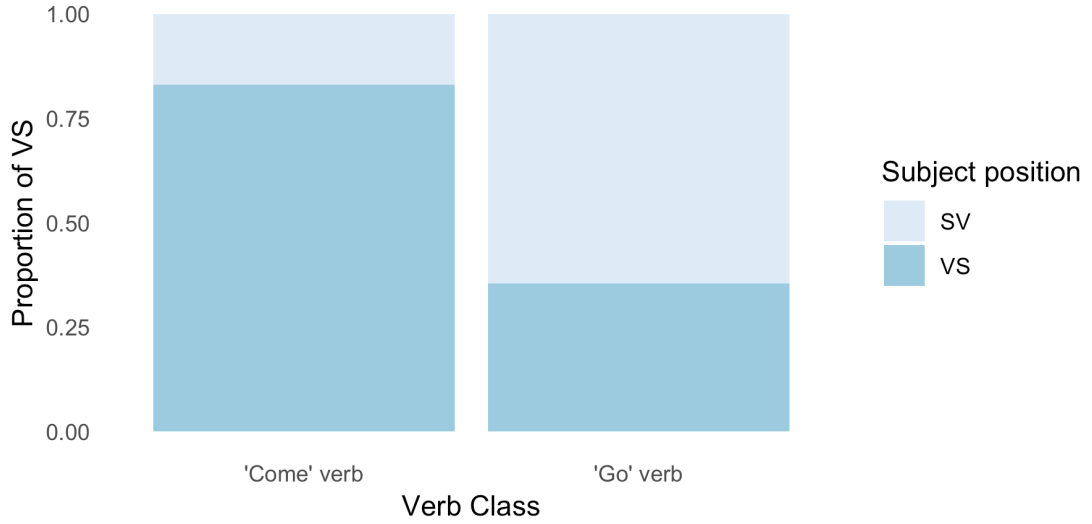


Figure 4B. Proportion of VS in "Come" and "Go" verbs after *ir* 'go' and *venir* 'come' are excluded



4. Conclusion

This study presented new evidence suggesting that patterns of variable subject position in Spanish motion verbs are largely influenced by function and weight

factors, by which the tendency to use adverbials is predictive of trends of subject position. The novel findings here link these two factors to putative verb class effects. The data examined, collected from two conversational corpora of Peninsular Spanish, casts doubt on the common view in variationist studies that Spanish verbs that are considered unaccusative should be expected to favor VS order.

The criteria for verb categorization were rooted in a functional perspective, such that motion verbs were divided according to the Deictive Function (Talmy 2000), which classifies verbs as expressing motion-from-the-center (i.e., away from the speaker[s] or their interlocutor[s]), also known as "go" verbs; or motion-toward-the-center, or "come" verbs. The proposed application of communicative functions to lexical types allowed for a flexible categorization of even the same verb, e.g. *llegar* 'arrive' as a "go" or a "come" verb based on its discourse functionality.

The analysis revealed that verbs which do not intrinsically define the direction of motion and require path specification (i.e., "go" verbs) tend to cooccur with adverbials in a post-verbal position. In the line of previous studies, the presence of a modifier to the right of the verb tended to be in complementary distribution with the subject, which tends to appear pre-verbally in those cases. "Come" verbs were found to have lower rates of modifiers and a stronger preference for VS than non-deictic verbs. But importantly, "go" verbs followed the opposite pattern, showing significantly higher rates of SV order both in connection with the presence of post-verbal modifiers, but also when no adverbial was present. Since the idiosyncratic categories described here tend to be conflated under a unitary unaccusative category in previous studies, the current approach offers a new way to classify certain verbs, providing a more fine-grained characterization of word order patterns. More broadly, this study encourages a functional approach for verb categorization.

In summary, the data show that the categories proposed, grounded in the Deictic Function, are by themselves reliable predictors of subject position. More importantly, the patterns observed in "come" and "go" verbs go in different directions, with the latter running counter to the hypothesis traditionally associating unaccusatives with VS. Thus, it offers an account that does not purely rely on intrinsic lexico-semantic features dictating syntactic patterns, but one that explains how lexico-semantic traits interact with discourse and usage to give rise to the observable patterns of variation in subject position. The data also give evidence of entrenchment in these function-based verb categories, even in the absence of the adverbial modifiers that influence patterns of variable subject position. These results encourage a revision of the view of unaccusativity as a blanket explanation for increased rates of VS in Spanish, and bring fresh insights into the issue of variable subject position.

References

Aranovich, R. 2003. "The Semantics of Auxiliary Selection in Old Spanish." *Studies in Language: International Journal Sponsored by the Foundation "Foundations of Language", 27*(1), 1.

Alamillo, A. R. 2009. "Cross-Dialectal Variation in Propositional Anaphora: Null Objects and Propositional Lo in Mexican and Peninsular Spanish." *Language Variation and Change*, *21*(03), 381.

Bates, Douglas, Martin Maechler, Ben Bolker and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software,* 67(1), 1-48.<doi:10.18637/jss.v067.i01>.

Benevento, N. M., and A. J. Dietrich. 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. doi:10.1177/1367006913516038

Bentivoglio, Paola and Elizabeth G. Weber. 1986. "A Functional Approach to Subject Word Order in Spoken Spanish." In *Studies in Romance Linguistics: Selected papers of the fourteenth Linguistics Symposium on Romance Languages,* ed. by O. Jaeggli and C. Silva-Corvalán, 23-40. Riverton, USA: Foris.

Brown, Esther L., and Javier Rivas. 2012. "Grammatical relation Probability: How Usage Patterns Shape Analogy*." Language Variation and Change 24.*3:317-341.

Cestero, A. M., I. Molina and F. Paredes. 2012. *La Lengua Hablada en Madrid. Corpus PRESEEA-MADRID (Distrito de Salamanca). I. Hablantes de Instrucción Superior*. Alcalá de Henares: Universidad de Alcalá.

De Miguel, E. (1999). "El Aspecto Léxico." In I. Bosque & V. Demonte (Eds.) *Gramática Descriptiva de la Lengua Española*. Espasa Calpe, pp. 2987-3060.

Du Bois, John W. 2003. "Discourse and Grammar." In *The new psychology of language: Cognitive and Functional Approaches to Language Structure,* ed. by M. Tomasello, vol. 2, 47-88. Mahwah, NJ: Lawrence Erlbaum Associates.

Givón, Talmy. 1983. "Topic Continuity in Discourse: An Introduction." In *Topic Continuity In Discourse: A Quantitative Cross-Linguistic Study,* ed. by T. Givón. 1-41. Amsterdam: John Benjamins.

Haig, G., and S. Schnell. (2016). "The Discourse Basis of Ergativity Revisited." *Language, 92*(3), 591-618. doi:10.1353/lan.2016.0049

Hawkins, J. A. 2003. "Efficiency and Complexity in Grammars: Three General Principles." *The Nature of Explanation in Linguistic Theory*, 121–152

Labov, W. 1984. "Field Methods of the Project on Linguistic Change and Variation."

In *Language in Use: Readings in Sociolinguistics*, ed. by J. Baugh and J. Sherzer, 28–53. Englewood Cliffs, NJ: Prentice-Hall.

Levin, B. 1993. *English Verb Classes and Alternations. A Preliminary Investigation.* Chicago: Chicago University Press.

Levin, B., and M. R. Rappaport Hovav. 1992. "The Lexical Semantics of Verbs of Motion: The Perspective from Unaccusativity." In I. M. Roca (Ed.), *Thematic Structure: Its Role in Grammar*, Berlin & New York: Foris, 247-269.

Lewandowski, Wojciech. 2007. "Toward a Comparative Analysis of Coming and Going Verbs in Spanish, German, and Polish." Universitat Autonoma de Barcelona. M.A. Thesis.

Lozano, C., and A. Mendikoetxea. 2010. Interface Conditions on Post-Verbal Subjects: A Corpus Study of L2 English. *Bilingualism: Language and Cognition*, *13*(4), 475-497.

Marcos Marín, F. (Dir.) 1992. *Corpus Oral de Referencia de la Lengua Española CORLEC*. http://www.lllf.uam.es/ESP/Corlec.html (accessed August 2017).

Mayoral Hernández and Lindsey Chen. 2006. "Testing for Verb Types in Spanish: An Application of SAS to Corpus Linguistics." Proceedings of 14th Western Users of SAS Software

Mayoral Hernández, R. 2005. "A Typological Approach to the Ordering of Adverbials: Weight, Argumenthood and EPP." *Anuario del Seminario de Filología Vasca "Julio Urquijo"*, 39(2), 141-59

Mayoral Hernández, R. 2010. "The Locative Alternation: On the Symmetry between Verbal and Prepositional Locative Paradigms." *Probus: International Journal of Latin and Romance Linguistics, 22*(2), 211. doi:10.1515/prbs.2010.008

Mendikoetxea, Amaya. 1999. "Construcciones Inacusativas y Pasivas." In *Gramática*

Descriptiva de la Lengua Española, ed. by Ignacio Bosque and Violeta Demonte, vol. II, 1575-1629. Madrid: Espasa.

Morimoto, Y. 2001. *Los Verbos de Movimiento*. Madrid: Visor Libros.

Myhill, John. 2005. "Quantitative Methods of Discourse Analysis." In R. Köhler, G. Altmann, and R. G. Piotrowski (eds.), *Quantitative linguistics. An international handbook*. Berlin: de Gruyer. 471-498.

Naro, Anthony J. and Sebastião J. Votre. 1999. "Discourse Motivations for Linguistic Regularities: Verb/Subject Order in Spoken Brazilian Portuguese." Probus 11.1. 75-100.

Ocampo, F. A. 2009. "El Orden de Palabras en el Español Hablado. La Construcción Sujeto Verbo Objeto Directo: Estudios Dedicados al Profesor Ángel López García." In M. V. Rigat, and E. S. Alegre (Eds.), *La Lingüística como Reto Epistemológico y como Acción Social: Estudios Dedicados al Profesor Ángel López García.* (Vol. 1, pp. 501-511). Madrid: Arco Libros.

Ocampo, F. A. 2014. "El Orden de Palabras en Cláusulas Subordinadas Relativas con un Sujeto y un Verbo Transitivo." In Enrique-Arias et al. (eds.), *Perspectives in the Study of Spanish Language Variation*, Universidade de Santiago de Compostela

Posio, P. 2012. "The Functions of Post-Verbal Pronominal Subjects in Spoken Peninsular Spanish and European Portuguese." *Studies in Hispanic and Lusophone Linguistics*, *5*(1), 149–190.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/

Rivas, Javier. 2008. “La Posición del Sujeto en las Construcciones Monoactanciales del Español: Una Aproximación Funcional.” *Hispania*, 91 (4). 897-912.

Roggia, A. B. 2018. “An Investigation of Unaccusativity and Word Order in Mexican Spanish.” *Spanish in Context*, *15*(1), 77-102. doi:10.1075/sic.00004.rog

Silva Corvalán, Carmen. 1982. “Subject Expression and Placement in Mexican-American Spanish.” In J. Amastae and L. Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 93-120. New York: Cambridge University Press.

Sorace, Antonella. 2000. “Gradients in Auxiliary Selection with Intransitive Verbs.” *Language 76*. 859-890.

Sorace, Antonella. 2004. “Gradience at the Lexicon-Syntax Interface: Evidence from Auxiliary Selection and Implications for Unaccusativity.” In *The Unaccusativity Puzzle: Explorations of the Syntax-Lexicon Interface*, ed. by Alexiadou, A., Anagnostopoulou, E., and Everaert, M. 243-268. New York, NY: Oxford University Press.

Talmy, L. 2000. *Typology and Process in Concept Structuring. Toward a Cognitive Semantics*, vol. 2. Cambridge, London: The MIT Press.