2023. *Proc Ling Soc Amer* 8(1). 5542. https://doi.org/10.3765/plsa.v8i1.5542.

¿Va primero el verbo? or *¿El sujeto va primero?*: Subject-verb order in Latin American Spanish

Lee-Ann Vidal Covas*

Abstract. This paper investigates subject-verb placement for *unaccusative* and *unergative* verbs in Spanish, focusing on syntactic, pragmatic, and sociolinguistic factors that predict placement. The study aims to answer three questions: (1) Does the unergative/unaccusative divide influence SV/VS order acceptability?, (2) What are the dialectal differences in subject placement acceptability in Spanish?, and (3) Does sentence context affect subject placement preference? The study collected data from sixty-nine Spanish speakers from the Caribbean, Chile, and Mexico, who provided 1656 acceptability ratings on sentences with different subject-verb orders. The findings indicate that both verb type and pragmatic conditions predict word preferability, with VS order preferred when the verb is unaccusative, and SV order preferred overall. The study adds to the literature by establishing the connection between argument structure and information structure and supporting the Unaccusative Hypothesis.

Keywords. Latin-American Spanish; subject-verb order; intransitive verbs; unaccusative verbs; unergative verbs; regional dialects; sociosyntax; Spanish language

1. Introduction. Spanish has been shown to have systematically constrained word order variation when it comes to subject and verb. Subject-Verb (SV) and Verb Subject (VS) constructions, such as the ones seen in example (1), have not only been shown to be acceptable by native speakers but are also produced in spontaneous speech by Spanish speakers on a regular basis.

(1) a. Catalina cantó. (SV)
Catalina sing-3SG.PST
'Catalina sang.'

b. Cantó Catalina. (VS)
sing-3SG.PST Catalina
'Catalina sang.'

This word order variability has garnered the attention of scholars from many areas of linguistics for quite some time. Although many have tried to understand and explain SV and VS word order variation in Spanish, accounts of said variation remain insufficient. These accounts have focused on different aspects of the subject placement phenomenon. Generative syntactic theorists (Chomsky 1981; Rizzi 1982) and variationist studies (Silva Corvalán 1982; Ortiz López 2009; Ortiz López & Citivello 2016; Raña-Risso & Barrera-Tobón 2018) alike have linked the availability of SV & VS variation to being a null subject language. Corr (2016:2) mentions, however, that while some instances of VS order "correlate with the null-subject parameter," being a null-subject subject language is not a pre-requisite for said order. Some studies have declared the canonical Spanish order as SVO, and departure from said order has been attributed to internal variables such as focus, topicalization, and constituent length (Ocampo 2009, 2010; Zagona 2002). Corr establishes that in a subset of VS constructions, referred to by syntacticians as *wide-focus verb*

* I am indebted to Daniel Erker and Neil Myler for their continuous and valuable support and advising. Special thanks to Paul Hagstrom and Chris Lee, as well as all the members of my Writing Circle. Funding for this study was granted by Boston University's Linguistics Department. All errors are my own. Lee-Ann Vidal Covas, Department of Linguistics, Boston University (vidall@bu.edu).

inversion, the information provided by the post-verbal subject and the verb is new, i.e., information that was not previously known by the listener (2016:1-2). The new information then creates a context where the sentence receives focus, referred to as "wide-focus." Bergen (1976:95) explores the possible reasons for variability in subject placement in interrogatives. He proposes that subject usage and placement in wh-questions in Spanish is related to emphasis: a question that appears with no subject has no emphasis, whereas a question with a subject can indicate slight or strong emphasis.

Studies have also investigated the effect of verb type on subject placement (Silva Corvalán 1982; Ortiz López 2009). Many studies have indeed investigated what factors motivate the variation that exists in Spanish when it comes to SV/VS order, yet questions still remain. For example, Ortiz López (2009) found that Caribbean Spanish speakers seem to exhibit a preferred SV word order, a finding that goes against what the author hypothesized was to be expected based on previous research. The main drive of this study was that much of the previous literature regarding this phenomenon presupposes that word order preferences come about as a result of one factor, whether it be only pragmatic condition or the verb type, and, in many of the studies, the possibility of there being a multivariate and probabilistic explanation is not explored.

The current study looks to add to the growing understanding of subject placement by assessing the acceptability (Corr 2016) and preference (Corr 2012; Ortiz López 2009) of SV/VS constructions among Spanish speakers in Latin America.[1] The idea is to determine whether the results found by Ortiz López (2009), Corr (2012, 2016) were indeed generalizable to other dialects of Spanish. The objective is to study subject placement preference in Spanish speakers, focusing on *unaccusative* and *unergative* verbs,[2] and analyzing the restrictions on subject placement using syntactic, pragmatic, and sociolinguistic factors. A focus of the study is to compare the findings to those of Ortiz López's 2009 on Caribbean Spanish, as well as Corr's 2012 & 2016 findings on European, Mexican, and Rio Platense Spanish. Another goal is to identify any potential dialectal differences in grammatical judgments on SV/VS order. This study endeavors to highlight potential predictors of SV/VS acceptability in order to create a clearer picture of which grammatically conditioned predictors need to be taken into account when designing future experiments.

1.1. RESEARCH QUESTIONS. I set out to answer the following research questions:

1. Does the unergative/unaccusative divide influence subject placement?
2. What are the dialectal differences, if any, regarding subject placement acceptability in Spanish?
3. Does sentence context affect subject placement preference?

This paper provides an overview of a sociolinguistic study that examines the shaping effects of linguistic and social factors on word order variability in a specific community. The study found that the type of verb was the key driving factor for word order acceptability and preferability, followed by the context of the sentence. Region was not found to be a predictor in VS order preferability. The paper concludes with a discussion of the results, framed by the three research questions.

[1] Acceptability refers to whether an utterance is accepted at all by the speakers, while preference refers to whether one order is preferred over another when different predictor variables are taken into account.

[2] Unaccusative verbs are those whose subject is the undergoer of the action. Unergative verbs are those whose subject is the doer of the action.

2. Previous research.

2.1. GENERAL RESEARCH ON WORD ORDER. Corr (2012) and (2016) explore SV/VS word order in six Ibero-Romance varieties: Asturian, Brazilian Portuguese, European Portuguese, European Spanish, Mexican Spanish, and River-Plate Spanish. They focus on what they call *wide-focus verb inversion* (explained above), having a type of locative element that can occupy different positions to the left of the verb, which is variably available in the languages, thus explaining the variation. According to the author, many languages with access to this null locative element have a positive setting for the null subject parameter; however, being a null subject language is not a prerequisite for having said locative element (Corr 2016:2). Gupton (2010) collected quantitative and qualitative data regarding the syntax-information structure interface in Galician. Focusing on agentive transitive verbs, she examined the acceptability of multiple word orders using seven pragmatic conditions. Her results suggested a marked preference for SVO order, which is markedly different from studies using similar contexts for Spanish, such as Zubizarreta (1998) and Casielles-Suárez (2004). Gupton (2010), in reference to findings indicating that SV order in Spanish occurs because said subjects are canonical or left-peripheral, mentions that the generalizability of these results comes into question given that regional variation has been found, something that could help explain why there is so much diversity in the findings.

Ciconte (2018) investigated post-verbal subjects in old Italo-Romance, concluding that variable word order is pragmatically motivated and that for the subject to occur in its obligatorily postverbal position in sentence-focus structures with unaccusative verbs, "the subject referent must carry no presupposition and must be low in agentivity, nonidentifiable and non-specific, whether formally indefinite or definite" (2018:149). Establishing modern Romance word order as SVO, the author adds that "sentence-focus structures with unaccusatives exhibit VS order in classical and late Latin and modern Italo-Romance, suggesting that patient/undergoer subjects are invariably postverbal over time" (Ciconte 2018:148). Lobo & Martins (2017) studied constituent order in Romance languages and what motivates and licenses these orders. According to the authors, subjects in SV orders are usually interpreted as topics when VS is an available option. They also state that VS orders have a common feature in that their subjects are usually not topicalized. In contrast, in root subjects with transitive verbs, subject focalization derives a VOS order, while thetic sentences and non-degree exclamatives derive a VSO order (Lobo & Martins 2017:32).[3]

2.2. WORD ORDER IN SPANISH. Sitaridou (2012) studied word-order phenomena in Old French, Old Spanish, Old Portuguese, and Old Occitan from a comparative perspective to establish if Old Romance languages were all V2, a claim that has been previously made. Several findings regarding the felicitousness of certain word orders in modern Spanish exist. Casielles-Suárez (2004) ascertains that SVO word order in Spanish can be felicitous when answering questions where nothing is presupposed, as well as those where the subject is presupposed. She, however, makes no claims about whether other word orders such as VSO or VOS are felicitous and how, or even if, they would compare in acceptability to SVO. Zubizarreta, on the other hand, did ascertain that SVO is not a felicitous order in response to a subject-narrow focus question, but VOS is (1998:123). She also mentions that SVO and VSO orders are compatible when nothing is pre-

[3] Thetic sentences are those which describe the situation without expressing any type of topicalization. Non-degree exclamatives are those which do not include a wh-operator and "comment upon a fact (or state of affairs) and express the speaker's emotive attitude towards its unexpectedness" (Lobo & Martins 2017:47).

supposed. Ortiz López (2009) studied word order in Spanish with intransitive verbs. The Unaccusative Hypothesis (UH), developed by Perlmutter & Postal (1974), proposed that intransitive verbs can be divided into two groups: unaccusative and unergative. The basic idea behind this division is that unergative subjects seem to behave like transitive subjects. In contrast, unaccusative subjects behave like transitive objects [see Perlmutter (1978) for a full explanation of the theory]. The terms themselves are by courtesy of Geoffrey K. Pullum.[4] Using examples (2) and (3) below, Ortiz López (2009:80) explains that, while the type of verb can affect the word order, in an utterance with no focus, where both word orders are grammatically possible, one of them is pragmatically infelicitous. He hypothesized that, in cases where the subject is not the only new piece of information, argument structure would be the determinant of the word order, i.e., unergatives would prefer SV and unaccusatives would prefer VS.

(2) ¿Qué ocurr-ió anoche en el programa de don Francisco?

What happened in Don Francisco's program last night?

a. Vino el presidente Hugo Chávez. (VS)
come.PST.3SG the president Hugo Chavez
'President Hugo Chavez came.'

b. *El presidente Hugo Chávez vino. (SV)
the president Hugo Chavez come.PST.3SG
'President Hugo Chavez came.'

Ortiz López (2009:81), *glossing & translations are mine*

(3) ¿Qué pas-ó anoche en el programa de don Francisco?

What happened in Don Francisco's program last night?

a. *Bail-ó toda la noche el presidente Hugo Chávez. (VS)
dance-PST.3SG all the night the president Hugo Chavez
'President Hugo Chavez danced all night.'

b. El presidente Hugo Chávez bail-ó toda la noche. (SV)
the president Hugo Chavez dance-PST.3SG all the night
'President Hugo Chavez danced all night.'

Ortiz López (2009:81), *glossing & translations are mine*

However, Ortiz López (2009) adds that when focus is added to the utterance, as in (4) and (5) below, said focus should affect the word order in favor of the SV order. In other words, when the subject is the only new information, as in (4) and (5), then SV order is preferred.

(4) ¿Quién vin-o anoche al programa de don Francisco?

Who came to Don Francisco's program last night?

a. *Vino el presidente Hugo Chávez. (VS)
come.PST.3SG the president Hugo Chavez
'President Hugo Chavez came.'

[4] Read Pullum (1988) for an amusing lesson on the origins of UH and the history of miscitations it has endured.

b. El presidente Hugo Chávez vino. (SV)
the president Hugo Chavez come.PST.3SG
'President Hugo Chavez came.'

Ortiz López (2009:81), *glossing & translations are mine*

(5) ¿Quién bail-ó anoche en el programa de don Francisco?

Who danced in Don Francisco's program last night?

a. *Bail-ó toda la noche el presidente Hugo Chávez. (VS)
dance-PST.3SG all the night the president Hugo Chavez
'President Hugo Chavez danced all night.'

b. El presidente Hugo Chávez bail-ó toda la noche. (SV)
the president Hugo Chavez dance-PST.3SG all the night
'President Hugo Chavez danced all night.'

Ortiz López (2009:81), *glossing & translations are mine*

His results went against his predictions; he found that SV order was favored with unaccusatives 71% of the time, contra the Unaccusative Hypothesis (*UH*) (Ortiz López 2009:84). While VS was favored more when it appeared with unaccusatives, Ortiz-López indicated that he did not have an explanation for this as neither the argument structure (verb type) nor information structure (+/-focus) proved to have a significant effect on word order preferability (Ortiz López 2009:84-86).[5] Although participants appeared to prefer SV order in focused sentences, the analysis revealed no significant differences. There was also no regional difference between the three Caribbean dialects: Cuban, Dominican, and Puerto Rican. These results led Ortiz-López to conclude that both types of intransitive verbs patterned the same (contra *UH*) and that SV was the unmarked order in Spanish.

Erker et al. (2017) studied subject placement in finite clauses. They used the first 100 finite verb tokens from fourteen participants' interviews, seven newcomers, and seven New York-raised Cubans. This yielded 1,400 finite verb tokens, half from Cuban newcomers and half from Cuban participants raised in New York (NYR). Their research questions related to whether Cuban newcomers choose to place subjects before or after verbs and if said preference was something the New York-raised Cubans did similarly or if a difference between the two groups existed. They explored social factors such as proficiency in English and Spanish to determine whether "the differences of frequency of use and proficiency between the generations in the two groups and between the patterns of subject placement in the two languages were associated with some degree of intergenerational change in the variable placement of subjects of Spanish finite verbs" (Erker et al. 2017:63). After considering the variables that motivate SV/VS placement in Spanish, they found two main differences between the two groups: (1) the New York-raised Cubans were sensitive to a predictive variable that the newcomers ignored (whether the clause was main or subordinate), and (2) the generations differed in terms of order of preference for the predictor variables. Newcomers preferred the type of finite verb as their top predictor, while the New York-raised Cubans were concerned with the type of subject (i.e., whether the subject was a personal

[5] If example (2b) were to be reduced to "Chávez vino" (Chavez came), then the utterance would be more likely to become more acceptable to speakers. For the current study, the heaviness of the Noun Phrase (NP) was taken into consideration, and "light" lexical NPs to reduce possible noise in the data.

pronoun, a lexical noun phrase, a demonstrative pronoun, a quantifier, or a clause). The results showed that, for both generations, SV order was more likely to occur when pronominal subjects appear with a human referent, in declarative sentences and when they occurred with either a copulative or occurrence verb (Erker et al. 2017:74). On the other hand, post-verbal subjects were more probable when the subject was a lexical noun phrase, a quantifier, or a clause appearing with a non-human referent in interrogative sentences and when they occurred with either an experiential or presentative verb (Erker et al. 2017:74). Ortiz López (2009) did not control for NP size and, unlike Erker et al. (2017), he did not look at type of subject, which could help explain why his results appeared to be unexplainable. Given the differences observed between the newcomers and the New York-raised Cubans, Erker et al.'s 2017 findings suggest that Subject/Verb order should be added to the list of features regularly employed as diagnostics for investigating outcomes of contact. This study aims to add to previous research by not just adding to the field of sociolinguistic variation but by looking at it through the lens of sociosyntax and applying syntactic ideas and methods to variation patterns.

3. Speakers, data, & methods. The present study takes an experimental approach, which is not only helpful in avoiding research bias that can come from intuition-based theoretical linguistic research (Gupton 2014:109), it can also aid in gathering information on possible dialectal variation. The survey was created and hosted on Qualtrics Research Suite (2005) and administered through Prolific (2014). Inspired by Ortiz López (2009), Corr (2012) and Gupton (2010), informants were asked to provide acceptability judgments on SV/VS constructions in Spanish.

3.1. SPEAKERS. A total of 69 participants were recruited through Prolific. The 69 speakers in the study vary along several social dimensions. They represent three regions: Andean, Central, and Caribbean and four countries: Chile (25), Cuba (10), Mexico (25), and Puerto Rico (9). All were born in one of the countries, and either still live there or emigrated to the United States. They range from 18 to 70 years of age. While the vast majority (61 speakers) have at least some university level education, nine do not. Of these, eight were educated through high school, while the other was educated through elementary school. Fifty-nine participants consider Spanish to be their L1, while ten consider both Spanish and English their L1s. All reported using Spanish at home and in general daily or frequently, while thirteen reported never using it at work.

3.2. SURVEY. The experimental part of the survey consisted of a quantitative data-gathering task hosted on Qualtrics and administered through an anonymous link participants could access after being pre-screened on Prolific. Participation was entirely voluntary and anonymous. Participants were compensated for their time through Prolific. The survey consisted of a total of 24 day-to-day scenarios. Inspired by both Ortiz López (2009) and Gupton (2010), the scenarios were followed by a question in one of three pragmatic conditions: wide-focus, subject-narrow focus, and subject-old information (explained below). Following Ortiz López's 2009 methodology, two lexically identical sentences with two different syntactic orders (SV/VS) were given as possible answers, which participants were to rate using a Likert scale. Half of the scenarios were filler items, while the other half were target items. Taking into consideration the dependent variable (Word Order preference) and the two independent variables: (i) Verb Type (with two values) and (i) Pragmatic Condition (with three values), the survey was designed using a 2x3 factorial approach. The task had 12 target items, with 1/3 of the items devoted to each Pragmatic Condition. Half of the verbs were unaccusative for all items, and half were unergative, a la Ortiz López

(2009). Questions were randomized on Qualtrics to reduce potential bias.

3.3. PREDICTOR VARIABLES. As mentioned above, the study was designed with two independent variables in mind: *verb type* and *pragmatic condition*. In this section, the two linguistic predictor variables are defined and explained.

3.3.1. VERB TYPE. All verbs in the study are intransitive verbs, meaning that they are verbs that have only one argument. The two types of intransitive verbs that used in the experiment are *unergative* and *unaccusative*. *Unaccusative* verbs are those whose subject is the undergoer of the action (as in example 6a), whereas, in *unergative* verbs, the subject is the doer of the action (see example 6b).

(6) a. Las flores florecieron.
The flowers bloom.PST.3PL
'The flowers bloomed.'

b. Casandra caminó.
Casandra walk.PST.3SG
'Casandra walked.'

3.3.2. PRAGMATIC CONDITIONS. As mentioned, the questions that followed each scenario were designed using three possible Pragmatic Conditions:

Wide-focus Wide-focus relates to situations where there is no presupposition; all the information is new and unknown to the listener (Corr 2016:1-2). Wide-focus utterances would answer the question: *'what happened?'*. An example of a Wide-focus scenario can be seen in (7):

(7) Alexandra está haciendo la tarea. De repente se escucha una conmoción en el patio. Su compañera de cuarto, Adriana, sale alarmada de su habitación, y le pregunta: *'¿Qué pasa?'* Alexandra anuncia:

Alexandra's doing her homework. Suddenly there is a commotion in the backyard. Her roommate, Adriana, comes out of her room alarmed, and asks: 'What's going on?' *Alexandra replies:*

a. Está gritando Daniela angustiosamente. (VS)
be.PRS.3SG scream.PROG Daniela agonizingly
'Daniela is screaming agonizingly.'

b. Daniela está gritando angustiosamente. (SV).
Daniela be.PRS.3SG scream.PROG agonizingly
'Daniela is screaming agonizingly.'

Subject narrow-focus Subject Narrow Focus relates to situations where the subject is the only piece of information not known to the listener. These situations would ask the question *'Who did X?'*, as in example (8):

(8) Ashley tiene una compañera de cuarto que sufre de migrañas. Una amiga de Ashley viene a visitarla a su casa y Ashley le dice que tienen que hablar en voz baja porque una de sus compañeras de cuarto sufre de migrañas y está durmiendo. Su amiga le pregunta: *'¿Quién está sufriendo de una migraña?'* Ashley contesta:

Ashley has a roommate who suffers from migraines. A friend of Ashley's comes to visit her at her house and Ashley tells her that they have to speak quietly because one of her roommates suffers from migraines and is sleeping. Her friend asks, 'Who is suffering from a migraine?' *Ashley replies:*

a. Sufre de migraña Marta. (VS)
suffer.PRS.3SG from migraines Martha
'Martha suffers from migraines.'

b. Marta sufre de migraña. (SV)
Martha suffer.PRS.3SG from migraines
'Martha suffers from migraines.'

Subject as old information Subject as old information refers to situations where the subject has already been given within the discourse. These situations would be driven by the question: *'What did X do?'*. See example (9) below:

(9) Camila, Isabela y Olivia están de vacaciones en Canadá en invierno. Luego de pasar el día haciendo varias actividades separadas, deciden cocinar juntas pero Camila llega más tarde que las otras dos y se va a bañar. Isabela le pregunta a Olivia: *'¿Qué hizo Camila hoy?'* Olivia dice:

Camila, Isabela, and Olivia are on vacation in Canada in winter. After spending the day doing various activities, they decide to cook together but Camila arrives later than the other two and goes to bathe. Isabela asks Olivia: 'What did Camila do today?' Olivia says:

a. Patinó Camila todo el día. (VS)
skate.PST.3SG Camila all the day
'Camila skated all day.'

b. Camila patinó todo el día. (SV)
Camila skate.PST.3SG all the day
'Camila skated all day.'

If speakers prefer an SV order regardless of any pragmatic or discourse restrictions, they will pattern with those found in Caribbean speakers by Ortiz López (2009). However, if, in scenarios with neutral text, speakers prefer a VS order in sentences with unaccusative verbs and an SV in sentences with unergative verbs, then they would pattern with findings of other groups of speakers. In a scenario where the subject is focused, the SV order would be expected for both types of verbs.

The statistical results of the survey will appear in the following section. These will give an overview of the preferences regarding word order for all three pragmatic conditions and both unergative and unaccusative verbs.

4. Results. The 69 questionnaires yielded a total of 1656 acceptability ratings, with 828 belonging to SV order, and 828 to VS order. The results of this study are presented in four parts. First, I walk you through the distribution of the data without any predictor variables. Then, focusing on VS order, I give a quick summary of what the univariate analysis highlights were. Third, I

explore a multivariate analysis. Lastly, I discuss interactions between some of the predictor variables. All visuals and statistics were created using R (2021).[6] Results are presented using diverging stacked bar charts; the further an entire bar is to the left on the x-axis, the less participants liked it for the order in question and thus used the unfavorable values. On the other hand, the further to right an entire bar is, the more highly it was rated by participants, thus more preferable.

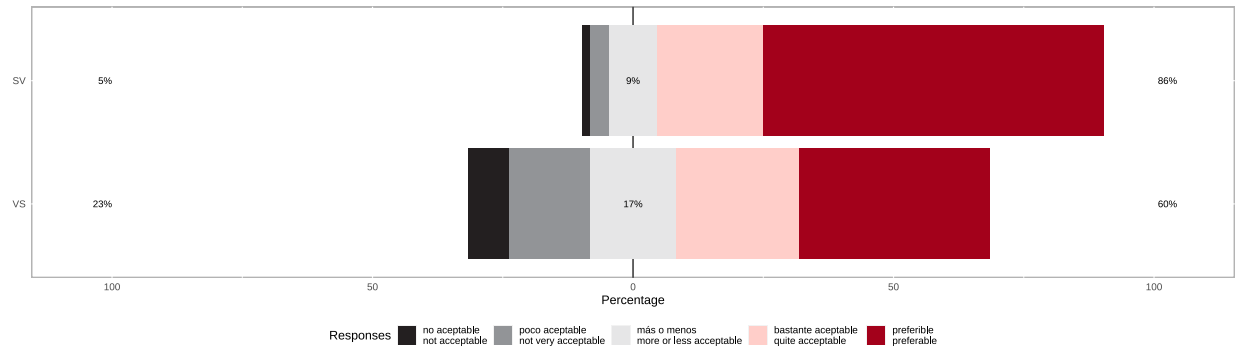

Figure 1. Word order ratings

4.1. VARIATION IN WORD ORDER PREFERABILITY RATINGS. Figure 1 shows that, overall, SV order is highly preferable (86%) compared to VS order which is rated as preferable only 60% of the time. SV order is only dispreferred 5% of the time, whereas VS is rated negatively 23% of the time. T-test results are significant, $t(1469.1) = 14.189$, $p < .001$, with SV order being rated more positively on average ($M = 4.44$, $SD = 0.91$) than VS order ($M = 3.66$, $SD = 1.32$). It should be noted that the word values used in the experiment were converted to numbers (1-5). In other words, the categorical values were transformed into a continuous numeric variable, as this allowed me to report the means and interpret the results in a simpler and more informative manner. These results confirm that the data is consistent with what we know to be true of Spanish, although speakers do prefer SV order, they are also accepting of VS order. Based on these results, and given that we know SV order is usually accepted, the next steps are to analyze what factors affect the favorability of VS order. In other words, what predictor variables affect when VS is rated favorably or disfavorably.

4.2. UNIVARIATE ANALYSIS SUMMARY. A univariate analysis returned significant results for both linguistic factors, yet showed that *Region*, was not a significant variable for either word order. The univariate analysis showed that, while VS is generally less preferred than SV order overall, as discussed earlier and visualized in Figure 1, it is rated more highly when it appears with *unaccusative* verbs, and in *wide-focus* contexts. An example that illustrates the conditions under which VS is rated most favorably can be seen in (10), where there is no presupposition, except for the knowledge that something happened:

(10) Carolina, Julio y Olivia están sentados en un restaurante para desayunar. La mesera trae la comida y se va. Olivia hace gestos para llamarle la atención a la mesera y Carolina le pregunta: *¿Qué pasa?'* Olivia contesta:

[6] Packages used in *R* to wrangle data, visualize and run stats: likert (Bryer & Speerschneider 2016) lme4 (Bates et al. 2015), stargazer (Hlavac 2018), kableExtra (Zhu 2021), dplyr (Wickham et al. 2021), ggpubr (Kassambara 2020), ggplot2 (Wickham 2016), and tidyr (Wickham 2021)

Carolina, Julio, and Olivia are sitting in a restaurant for breakfast. The waitress brings the food and leaves. Olivia gestures to get the waitress's attention and Carolina asks: 'What's going on?' Olivia replies:

a. Falta el café.
miss.PRS.3SG the coffee
'The coffee is missing.'

4.3. MULTIVARIATE ANALYSIS. I now proceed with the multivariate analysis, which unfolds in two steps. First, a mixed effects model combining linguistic and social factors is built and commented on for the VS word order. This is followed by a conditional inference (CI) tree being constructed for both word orders and commented on. Given that *Region* was not found significant, it was left out of the CI. Finally, as mentioned before, the categorical values in the experiment were transformed into a continuous numeric variable to conduct the analyses. This allowed me to interpret the results in a more straightforward and informative manner.

4.3.1. FIXED AND MIXED EFFECTS LINEAR REGRESSIONS - VS ORDER. Using the function lmer() in the *lme4* package, a mixed effects model,[7] shown in Table 2, was run with the dependent variable as a function of the linguistic and social variables combined, all treated as fixed effects. It additionally includes **participant** and **answer** (test-item) as random effects. When the random effects are taken into account, approximately 63% of the variance is accounted for. The reference level in the intercept for the model is a sentence in a *wide-focus* context, with an *unergative* verb. Two variables appear to be statistically significant. When the subject is presupposed (i.e., *subject old information*), it is associated with a lower rating of VS order compared to the reference level of *wide-focus* context, where nothing is presupposed. Similarly, when the verb is *unaccusative*, it is associated with a higher rate for VS orders when compared to the reference level of *unergative*. Given the results of the univariate analysis, it is not surprising that no significance for *Region* was found by the models.

4.3.2. MODELING WITH CI TREES. The conditional inference (CI) trees used for this study are non-parametric models that "select variables in an unbiased way" and use a partitioning algorithm not affected by overfitting (Hothorn et al. 2006:670).[8] CI trees can provide insight into the data, even though they cannot be used in lieu of a regression analysis. These are designed to group the dependent variable into maximally homogeneous groups. Figure 2 shows both word orders, where the continuous transformation of the ratings mentioned above is treated as a categorical variable. As is expected, the first partition encountered is the word order, followed by *verb type*, confirming again that the most critical factor in word order preferability in these data is, in fact, the type of verb. We can see that *pragmatic condition* is an essential factor predicting

[7] Various models were run in the exploration of the data. I used the function lm() in the base package in *R* to model the variation in the dependent variable as a function of (1) the study's linguistic variables, (2) the social variable, and (3) the linguistic and social variables combined. In the fixed effects models, the first model accounts for approximately 40% of the variance, the second for less than 1% of the variance, and the combined model accounts for approximately 40% of the variance.

[8] Packages used: rpart (Therneau & Atkinson 2019), partykit (Hothorn & Zeileis 2015), and ctree (Hothorn et al. 2006).

	VS word order rating (1-5 scale)			
Predictors	*Estimate*	*SE*	*t-value*	*p-value*
Intercept	3.23	0.29	11.23	**<0.001**
Pragmatic Condition: Subject Narrow-Focus	-0.34	0.32	-1.04	0.33
Pragmatic Condition: Subject Old Information	-0.76	0.32	-2.34	**0.05**
Verb Type: Unaccusative	1.51	0.27	5.71	**<0.001**
Region: Caribbean	0.00	0.17	0.01	0.99
Region: Central	0.11	0.16	0.66	0.51
Random Effects				
σ^2	0.67			
$\tau_{00\ \text{participant\_id}}$	0.27			
$\tau_{00\ \text{answer}}$	0.20			
ICC	0.41			
$N_{\ \text{participant\_id}}$	69			
$N_{\ \text{answer}}$	12			
Observations	828			
Marginal R^2 / Conditional R^2	0.371 / 0.631			

Table 2. Mixed effects regression model predicting vs word order acceptability

preferability only for the type of verb that is less preferred in a given word order. For VS order, *unergative* verbs divide into two new nodes where *SOI* (*subject old information*) is in one of the nodes and *SNF* (*subject-narrow focus*) and *WF* (*wide-focus*) are together in the other. This is because, in VS order, sentences with unergative verbs in *SOI* contexts are rated more negatively when compared to both *SNF* and *WF*, where the ratings are more evenly distributed amongst all possible options. For SV order, unaccusatives, the verb type rated least favorably, also splits into two nodes. In this case, *SNF* and *SOI* pattern together as they are rated more favorably than *WF*, where the ratings are more distributed. This will be discussed more in the following section, which addresses the topic of interactions.

4.4. INTERACTIONS. Taking into account the results of multivariate analyses, we now turn to look at the possible interactions between *type of verb* and *pragmatic condition*. Given that *Region* was not a significant factor in the univariate or multivariate analyses with the response variable, word order preferability, it was not be included in the analysis of interactions.

4.4.1. VERB TYPE - PRAGMATIC CONDITION. Figure 3 shows an interesting asymmetry that appears in contexts in which there is no information being focused (*wide-focus*): SV is preferred with *unergatives*, and VS is preferred with *unaccusatives*. It should be emphasized that, when it comes to *unergative* verbs, the dispreference for VS is stronger when the subject is old info, amplifying the effect of people's general dislike of VS order in sentences with *unergative* verbs. While in *unaccusatives*, participants rate VS more favorably than in *unergatives*, we can see that when focus is added to a sentence, there is some pressure from the information structure that seems to counteract the argument structure preference we see with the wide-focused sentences. We also see that the preference for VS order in *unaccusative* verbs is restricted to *wide-focus* con-

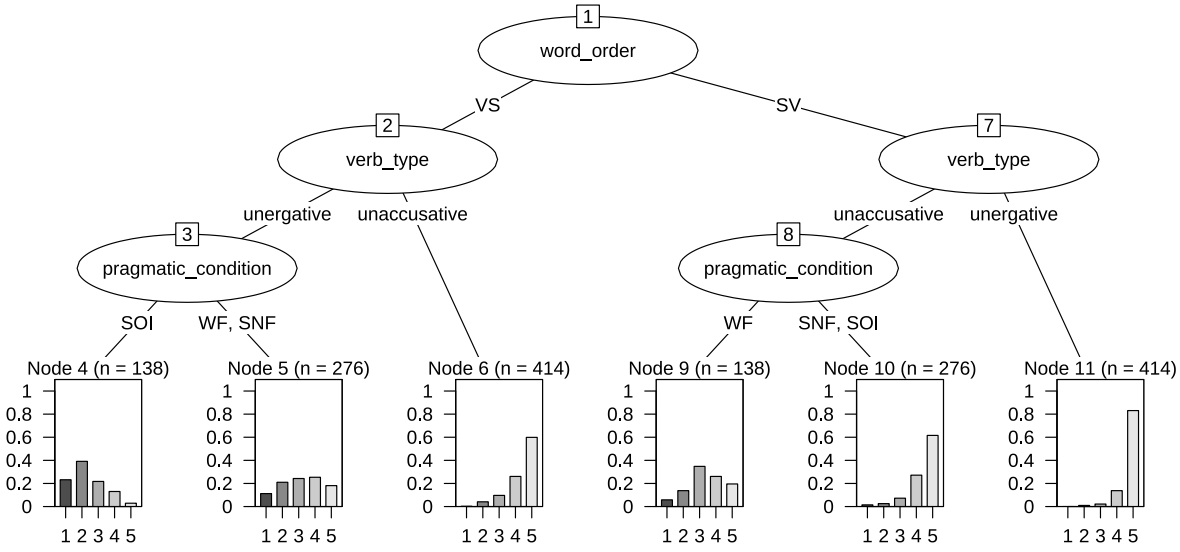

Figure 2. CI tree for word order preference

texts. While we see negative ratings only 1% with *wide-focus*, the negative ratings rise to 6% and 7% when the subject is not known (*subject narrow-focus*) and when the subject is the only information that is presupposed (*subject old info*). The regression results in Table 4, which have as a reference level a VS order sentence with an *unergative* verb, and in a *subject narrow-focus* context, corroborate what the rows related to VS order in Figure 3 show: *that the interactions between verb and pragmatic condition are indeed significant for VS order*. The conditions and verbs were re-leved to see if other combinations were significant, and the results consistently showed that the interactions are indeed significant. Table 3 shows examples of what these different contexts would look like. The examples, although shortened, come from the experiment itself.

5. Discussion. The primary objectives of this study were to determine whether the type of intransitive verb and/or information structure influenced SV/VS word order acceptability and whether there were any dialectal differences in these preferences. Although, as expected, SV order was the preferred order overall, the question remained as to what factors influenced favorable ratings for VS order. Whilst no regional differences were found for VS order, the linguistic factors proved to be illuminating.

Contra Ortiz López (2009), verb type significantly predicts word order preferability. Participants overwhelmingly preferred SV order (97%) with unergative verbs, while VS order was rated higher with unaccusative verbs, an order preference highlighted by Bosque Muñoz & Gutiérrez-Rexach (2009). This is unsurprising, given that it patterns with the expected divide via the Unaccusative Hypothesis (Perlmutter & Postal 1974). It bears mentioning that, to reduce possible noise in the data and to allow for a better understanding of the data, no extra information was included in our SV/VS structures; all sentences were relatively short, and yet the preferences still appeared.

In terms of the influence of information structure, the three pragmatic conditions (*wide-focus (WF)*, *subject-narrow focus (SNF)*, and *subject old information (SOI*) surveyed were found to be

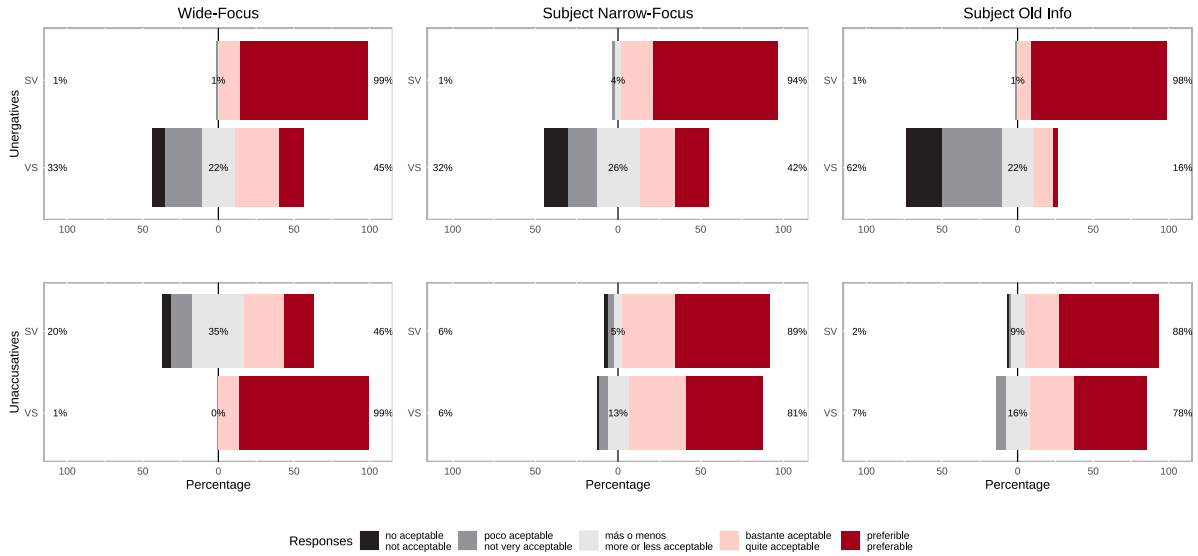

Figure 3. Word order ratings for verb types by pragmatic conditions

Pragmatic Condition	Context	SV	VS	Translation
Unergatives				
WF	Jorge pregunta: '¿Catalina, que pasó?' Catalina contesta:	Valentina bostezó.	Bostezó Valentina.	Jorge asks: 'What happened?' Catherine replies: 'Valentina yawned.'
SNF	Paula le pregunta a Elsa: '¿Quién se sonrío primero?' Elsa contesta:	Victor se río.	Se río Victor.	Paula asks Elsa: 'Who smiled first?' Elsa replies: 'Victor laughed.'
SOI	Isabela le pregunta a Olivia: ¿Qué hizo Camila hoy?' Olivia:	Camila patinó todo el día.	Patinó Camila todo el día.	Isabela asks Olivia: 'What did Camila do today?' Olivia: 'Camila skated all day.'
Unaccusatives				
WF	Emilio pregunta: '¿Ah? ¿Qué pasó?' Franchesca:	Una planta creció en el carro.	Creció una planta en el carro.	Emilio asks: 'Oh? What happened?' Franchesca: 'A plant grew in the car.'
SNF	La maestra pregunta: '¿Quién desapareció?' Anastasia contesta:	Florencia despareció.	Desapareció Florencia.	The teacher asks: 'Who disappeared?' Anastasia replies: 'Florence disappeared.'
SOI	Luna pregunta: '¿Qué pasa con el agua?' Cristina contesta:	El agua está hirviendo.	Está hirviendo el agua.	Luna asks: 'What is wrong with the water?' Cristina replies: 'The water is boiling.'

Table 3. Examples of the pragmatic conditions

significant predictors. When the only information presupposed is the subject (*SOI*), participants preferred SV order while rating VS order as preferable only 47% of the time. With no presupposition (*WF*), both orders were equally accepted, with 72% positive ratings. Although SV order was clearly preferred when the subject was the only information not presupposed (*SNF*) with a 92% preferability rating, VS order was not totally dispreferred, as it was rated 67% of the time favorably and disfavored 19% of the time. This is similar to what Lobo & Martins (2017) stated for Romance, where subjects in SV orders are usually interpreted as topics when VS is an available option. However, it should be noted that SV is still the overall preferred Spanish order. In

Predictors	*Estimate*	*SE*	*t-value*	*p-value*
Intercept	3.16	0.087	36.41	**<0.001**
Verb Type: Unaccusative	1.05	0.123	8.56	**<0.001**
Pragmatic Condition: Wide-Focus	0.04	0.123	0.35	0.723
Pragmatic Condition: Subject Old Information	-0.83	0.123	-6.73	**<0.001**
Verb Type: Unaccusative | Pragmatic Condition: WF	0.59	0.174	3.38	**<0.001**
Verb Type: Unaccusative | Pragmatic Condition: SOI	0.80	0.174	4.64	**<0.001**

Table 4. Interaction between Verb type & pragmatic conditions - VS word order rating

contrast, Ortiz López (2009) found no effect of information structure, while Corr (2012:12) mentions that, when asking her participants to explain their preferences, they explained that the question that is used in wide-focus contexts (*¿qué?/what?*) makes VS more acceptable.

Table 4 and Figure 3 present the most exciting find: the interaction of verb type and pragmatic condition. When information structure puts pressure on the argument structure preferences, we see some changes in the SV order preferability of unergatives. In these verbs, when the subject is presupposed, VS order is dispreferred 62% of the time. When there is no focus, VS is rated negatively 33% of the time and only positively rated 45% of the time. Similarly, when the subject is the only thing *not* presupposed in the sentence, VS is dispreferred 32% of the time and only rated 42% of the time positively. Similarly, we can observe the influence of the pragmatic conditions on the VS order preferability of unaccusatives. When the subject is the only thing not known in a sentence with an unaccusative verb, VS order is rated 81% of the time favorably while being rated unfavorably only 6% of the time. Analogously, when the subject is presupposed, VS order is rated 78% of the time positively and disfavored only 7% of the time.

Corr (2012) mentions that the systematic observation observed in her data should be explained by semantic differences in the verbs and the effects of PP structure. However, the data in the current study point to both argument structure and information structure working together to condition acceptability to both SV and VS word order. While, yes, overall, SV order seems to be the unmarked word order in Spanish, this doesn't mean that VS order is undesirable. The data shows that both syntactic and pragmatic factors constrain the preferability of these word orders.

6. Conclusion. This study endeavored to establish the potential predictors of SV/VS acceptability to highlight them for future studies and create a clearer picture of what would need to be considered when designing future experiments. Some studies looked at the effect of pragmatic conditions in transitive verbs only (Gupton 2010 for Galician). In contrast, others looked at only one or two pragmatic conditions and found diverging results (Ortiz López 2009; Corr 2012, 2016). The present study is the first to investigate the current combination of predictor variables while also controlling for subject properties, clause type, and subject size. Of the three predictors examined, verb type and pragmatic condition work together to condition word order acceptability. This study adds to the literature by establishing the connection between argument structure and information structure, as well as asserting that the Unaccusative Hypothesis is indeed borne out in the data.

Future studies could control for intonation by including audio with each context and possible responses as immediate responses to the orthographic representations might be different if heard

rather than read, thus effects of intonation are possible.

While I strove to include various dialects of Spanish in Latin America, this also meant that a wide net was cast. In the end, four countries and three Regions were represented, allowing for comparison between the dialects and to test whether what had been previously found for specific varieties regarding this variable was generalizable. The respondents of this study did not significantly diverge. While many exciting results have been explored, this study only scratches the surface regarding SV/VS order and allows future researchers to continue expanding this work.

References

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. https://doi.org/10.48550/arXiv.1406.5823.

Bergen, John J. 1976. The explored and unexplored facets of questions such as 'Qué tú tienes?'. *Hispania* 59(1). 93. https://doi.org/10.2307/339380.

Bosque Muñoz, Ignacio & Javier Gutiérrez-Rexach. 2009. *Fundamentos de sintaxis formal*. Madrid: Ediciones Akal.

Bryer, Jason & Kimberly Speerschneider. 2016. Likert: Analysis and visualization Likert items. Software package. https://cran.r-project.org/package=likert.

Casielles-Suárez, Eugenia. 2004. *The syntax-information structure interface: Evidence from Spanish and English*. New York: Routledge.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Ciconte, Francesco Maria. 2018. Postverbal subjects in Old Italo-Romance. *Italian Journal of Linguistics* 30(2). 127–158. https://hdl.handle.net/11564/801107.

Corr, Alice. 2012. *Subject inversion in Ibero-Romance*. Cambridge: University of Cambridge dissertation.

Corr, Alice. 2016. Wide-focus subject-verb inversion in Ibero-Romance: A locative account. *Glossa: A Journal of General Linguistics* 1(11). 1–33. https://doi.org/10.5334/gjgl.85.

Erker, Daniel, Eduardo Ho-Fernández, Ricardo Otheguy & Naomi Shin. 2017. Continuity and change in Spanish among Cubans in New York: A study of subject placement with finite verbs. In Alejandro Cuza (ed.), *Cuban Spanish dialectology: Variation, contact, and change*, 61–80. Washington, DC: Georgetown University Press.

Gupton, Timothy Michael. 2010. *The syntax-information structure interface: Subjects and clausal word order in Galician*. Iowa City: University of Iowa dissertation. https://doi.org/10.17077/etd.e8w28qrf.

Gupton, Timothy Michael. 2014. *The syntax-information structure interface: Clausal word order and the left periphery in Galician*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9781614512059.

Hlavac, Marek. 2018. Stargazer: Well-formatted regression and summary statistics tables. Software package. https://cran.r-project.org/package=stargazer.

Hothorn, Torsten, Kurt Hornik & Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15(3). 651–674. https://doi.org/10.1198/106186006X133933.

Hothorn, Torsten & Achim Zeileis. 2015. Partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16(118). 3905–3909. https://jmlr.org/papers/v16/hothorn15a.html.

Kassambara, Alboukadel. 2020. Ggpubr: 'Ggplot2' based publication ready plots. Software package. https://cran.r-project.org/package=ggpubr.
Lobo, Maria & Ana Maria Martins. 2017. 2 Subjects. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of romance morphosyntax and syntax*, 27–88. Berlin: De Gruyter.
Ocampo, Francisco. 2009. El orden de palabras en el español hablado. La construcción sujeto verbo objeto directo. In Montserrat Veyrat Rigat & Enrique Serra Alegre (eds.), *La lingüística como reto epistemológico y como acción social: Estudios dedicados al Profesor Ángel López García con ocasión de su sexagésimo aniversario*, 501–511. Madrid: Arco Libros.
Ocampo, Francisco. 2010. The place of conversational data in Spanish syntax: Topic, focus, and word order. *Studies in Hispanic & Lusophone Linguistics* 3(2). https://doi.org/10.1515/shll-2010-1086.
Ortiz López, Luis A. 2009. El español del Caribe: Orden de palabras a la luz de la interfaz léxico-sintáctica y sintáctico-pragmática. *Revista Internacional de Lingüística Iberoamericana* 7(2(14)). 75–93. https://www.jstor.org/stable/41678402.
Ortiz López, Luis A. & Ashley Dauphinais Citivello. 2016. Microvariation in the null subject parameter: Word order in Cuban Spanish. In Alejandro Cuza, Lori Czerwionka & Daniel J. Olson (eds.), *Hispanic Linguistics Symposium 14,* 281–300. Amsterdam: John Benjamins.
Perlmutter, David & Paul Postal. 1974. Lectures on Relational Grammar. Presented at Linguistic Society of America Linguistic Institute. University of Massachusetts, Amherst.
Perlmutter, David M. 1978. Impersonal passives and the Unaccusative Hypothesis. *Berkeley Linguistics Society (BLS)* 4. 157–190. https://doi.org/10.3765/bls.v4i0.2198.
Prolific. 2014. Prolific Online Research Platform. https://www.prolific.com.
Pullum, Geoffrey K. 1988. Citation etiquette beyond Thunderdome. *Natural Language & Linguistic Theory* 6(4). 579–588. https://www.jstor.org/stable/4047595.
Qualtrics. 2005. Qualtrics Research Suite. Software package. https://www.bu.edu/tech/services/cccs/desktop/distribution/qualtrics/.
R, Core Team. 2021. R: A language and environment for statistical computing. Software package. https://www.r-project.org/.
Raña-Risso, Rocío & Carolina Barrera-Tobón. 2018. On the relationship between subject placement and overt pronouns in the Spanish of New York City bilinguals. *Journal of Language Contact* 11(2). 324–347. https://doi.org/10.1163/19552629-01102007.
Rizzi, Luigi. 1982. *Issues in Italian syntax*. Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110883718.
Silva Corvalán, Carmen. 1982. Subject expression and placement in Mexican-American Spanish. In Jon Amastae & Lucia Elías-Olivares (eds.), *Spanish in the United States: Sociolin- guistic aspects*, 93–120. Cambridge: Cambridge University Press.
Sitaridou, Ioanna. 2012. A comparative study of word order in Old Romance. *Folia Lingüística* 46(2). 553–604. https://doi.org/10.1515/flin.2012.019.
Therneau, Terry & Beth Atkinson. 2019. Rpart: Recursive partitioning and regression trees. Software package. https://cran.r-project.org/package=rpart.
Wickham, Hadley. 2016. Ggplot2: Elegant graphics for data analysis. Software package. https://ggplot2.tidyverse.org.
Wickham, Hadley. 2021. Tidyr: Tidy messy data. Software package. https://cran.r-project.org/package=tidyr.
Wickham, Hadley, Romain François, Lionel Henry & Kirill Müller. 2021. Dplyr: A grammar of data manipulation. Software package. https://cran.r-project.org/package=dplyr.

Zagona, Karen. 2002. *The syntax of Spanish*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511613234.
Zhu, Hao. 2021. KableExtra: Construct complex table with 'kable' and pipe syntax. Software package. https://cran.r-project.org/package=kableExtra.
Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: MIT Press.