



ON WORD ORDER VARIATION AND INFORMATION STRUCTURE IN PENINSULAR SPANISH AND ITALIAN ‘WHY’-INTERROGATIVES¹

VARIAÇÃO DA ORDEM DE PALAVRA E DA ESTRUTURA DE INFORMAÇÃO NAS INTERROGATIVAS-Wh DO ESPANHOL PENINSULAR E DO ITALIANO

Svenja Schmid²
Klaus von Heusinger³
Georg A. Kaiser⁴

ABSTRACT

In this paper, we investigate the effect of information structure on word order in Italian and Peninsular Spanish ‘why’-interrogatives, and whether these two languages differ from each other. To this end, we conducted two empirical studies. In a parallel text corpus study, we compared the frequency of the word order patterns ‘why’SV and ‘why’VS, as well as the distribution of focal and non-focal subjects in the two languages. In order to get a deeper understanding of the impact of the information structural categories *focus* and *givenness* on word order in ‘why’-interrogatives, we conducted a forced-choice experiment. The results indicate that word order is affected by focus in Italian, while it is not determined by any information structural category in Peninsular Spanish. We show that Italian and Peninsular Spanish ‘why’-interrogatives differ from each other in two ways. First, non-focal subjects occur preverbally in Italian, while they occupy the postverbal position in Peninsular Spanish. Second, Italian reveals a lower level of optionality with respect to word order patterns. Even though we find a high preference for the postverbal position in Peninsular Spanish, we argue that this limitation is related to a higher flexibility regarding word order in Peninsular Spanish than in Italian which does not allow for ‘why’VSO in contrast to Peninsular Spanish. **Keywords:** word order; information structure; ‘why’-interrogatives.

¹ We would like to thank the audience of the workshop ‘Subject, Topic, and clause structure’, and in particular Sandra Quarezemin and her team for organizing the event and editing this volume and two anonymous reviewers for helpful comments and suggestions. Furthermore, we would like to thank Laura González López, María del Pilar Pérez Ocón, and Silvia Operato for their help in preparing the stimuli of our experiment and Anika Lloyd-Smith for proofreading. The research of the first and third author was supported by the German Research Foundation (DFG) as part research group FOR-2111 ‘Questions at the interfaces’ within the project P2 KA 1004/4-2 ‘The structure of *wh*-utterances and the interpretation of *wh*-words in Romance (and Germanic) languages’. The research of the second author was funded by the German Research Foundation (DFG) as part of the project ‘Composing events in Romance causative constructions and the semantics of causation’ (DO 2113/1-1) at the University of Cologne, Department for German Language and Literature I, Linguistics.

² Doctoral student, Department of Linguistics, Universität Konstanz, Germany. svenja.schmid@uni-konstanz.de

Orcid: <https://orcid.org/0000-0001-6524-8486>

³ Full professor, Faculty of Philosophy, Universität zu Köln, Germany. klaus.vonheusinger@uni-koeln.de. Orcid: <https://orcid.org/0000-0001-8108-3191>

⁴ Full professor, Department of Linguistics, Universität Konstanz, Germany. georg.kaiser@uni-konstanz.de

Orcid: <https://orcid.org/0000-0003-4336-910X>

RESUMO

Neste trabalho, investigamos o efeito da estrutura da informação na ordem de palavras em orações interrogativas com ‘porquê’ em italiano e espanhol peninsular e a questão se estas duas línguas diferem uma da outra. Para este fim, realizámos dois estudos empíricos. Num estudo de corpora paralelos, comparámos a frequência dos padrões ‘porquê’SV e ‘porquê’VS, bem como a distribuição de sujeitos não focais e focais nas duas línguas. A fim de obter uma compreensão mais profunda do impacto das categorias estruturais da informação *focus* e *givenness* na ordem de palavras em interrogativas com ‘porquê’, conduzimos um estudo de escolha forçada. Os resultados indicam que *focus* afeta a ordem de palavras interrogativas com ‘porquê’ em italiano, enquanto que a ordem das palavras em espanhol peninsular não é determinada por nenhuma categoria estrutural de informação. Mostramos que as interrogativas com ‘porquê’ em italiano e em espanhol peninsular diferem em duas maneiras. Primeiro, os sujeitos não focais ocorrem em posição pré-verbal em italiano, enquanto que ocupam a posição pós-verbal em espanhol peninsular. Segundo, o italiano revela um nível inferior de opcionalidade no que diz respeito aos padrões de ordem de palavras. Apesar de encontrarmos uma alta preferência pela posição pós-verbal no espanhol peninsular, supomos que esta limitação está relacionada a uma maior flexibilidade em relação à ordem das palavras em espanhol peninsular do que em italiano que não permite ‘porquê’VSO em contraste com o espanhol peninsular.

Palavras-Chave: ordem de palavras; estrutura da informação; interrogativas com ‘porquê’.

1. ON WORD ORDER IN ITALIAN AND PENINSULAR SPANISH WH-INTERROGATIVES

Like all Romance languages, Italian and (Peninsular) Spanish belong to the group of SVO languages (Gutiérrez-Bravo, 2002; Cardinaletti, 2004; Lahousse & Lamiroy, 2012). Besides the canonical SVO word order, they have all the typical typological characteristics of SVO languages, such as having prepositions (instead of postpositions), postnominal genitives (instead of prenominal ones), and auxiliary-verb sequences (instead of verb-auxiliary sequences). Deviations from SVO order are generally possible under certain pragmatic circumstances, although these are more frequent in Peninsular Spanish than in Italian (Leonetti, 2017). In some specific syntactic contexts, there are also deviations from SVO. This is the case in *wh*-interrogatives where Italian and Peninsular Spanish obligatorily exhibit subject-verb inversion (*wh*VS). The fronted *wh*-phrase is immediately followed by the finite verb, while the subject is expressed postverbally, as in (1) and (2), respectively (Suñer, 1994; Rizzi, 1996):

(1) a. Che cosa ha fatto *Gianni*?
what has done John

a'. *Che cosa *Gianni* ha fatto?
what John has done

b. Dove lavora *Maria*?
where works Mary

b'. *Dove *Maria* lavora?
where Mary works

(2) a. ¿Qué hizo *Juan*?
what did John

a'. *¿Qué *Juan* hizo?
what John did

b. ¿Dónde trabaja *María*?
where works Mary

b'. *¿Dónde *María* trabaja?
where Mary works

However, one major exception to this restriction is observed in combination with (bare) reason *wh*-words, in particular ‘why’ (*perché / por qué*), where both ‘why’VS and ‘why’SV order is possible (for Italian see Calabrese, 1982; Rizzi, 2001; 2006; Zipf & Quaglia, 2017, and for Peninsular Spanish Torrego, 1984; Ordóñez, 1998):⁵

- (3) a. *Perché balla Maria?*
 why dances Mary
 b. *Perché Maria balla?*
 why Mary dances
 ‘Why does Mary dance?’

- (4) a. *¿Por qué baila María?*
 why dances Mary
 b. *¿Por qué María baila?*
 why Mary dances
 ‘Why does Mary dance?’

This variation shows that ‘why’-phrases behave in a different way than other *wh*-phrases, since they do not obligatorily trigger subject-verb inversion. Rizzi (2001) explains this difference by assuming that ‘why’ is base-generated in the left periphery, namely in the Specifier of the Interrogative Phrase (IntP), see (5a). By contrast, other *wh*-phrases are moved from a lower (IP internal) position to the Specifier of the Focus Phrase (FocP) (see (5b)) and freeze in this position. As a consequence, they cannot co-occur with another focal fronted element. In ‘why’-interrogatives, this position is available for another fronted focal element. This contrast is shown in (6) for Italian and (7) for Peninsular Spanish.

- (5) a. [ForceP [TopP* [IntP *perché/por qué* [TopP* [FocP [TopP* [ModP [TopP* [QembP [FinP [IP ...]]]]]]]]]]]]
 b. [ForceP [TopP* [IntP [TopP* [FocP *che cosa/qué*_i [TopP* [ModP [TopP* [QembP [FinP [IP ... *t_i]]]]]]]]]]]]] (cf. Rizzi & Bocci, 2017: 2179)*

- (6) a. *Perché questo_[+Foc] avremmo dovuto dirgli, non qualcos’altro?*
 why this have.COND.1PL should say=him not something else
 ‘Why should we have said THIS to him, not something else?’ (Rizzi 2001: 294)
 b. **A chi questo_[+Foc] dobbiamo dire, non qualcos’altro?*
 to whom this should.1PL say not something else
 ‘To whom should we say this, not something else?’

⁵ There are two further remarkable exceptions to the obligatory subject-verb inversion in Italian and (Peninsular) Spanish *wh*-interrogatives. First, complex or so-called d-linked *wh*-phrases allow for preverbal subjects in Italian and Spanish (Torrego, 1984; Contreras, 1999; Rizzi, 2006)

- (i) *In quale cassetto Gianni ha messo il libro?*
 in which drawer Gianni has put the book
 ‘In which drawer Gianni put the book?’
 (ii) *¿En qué época del año María sale de vacaciones?*
 in which season-of-the year Mary goes on holidays
 ‘In which season of the year, Mary goes on holiday?’

Second, in Caribbean Spanish, and some other Spanish dialects, preverbal subjects are also possible with different (argumental and non-argumental) *wh*-phrases (see among others Davis, 1971; Toribio, 1993; Ordóñez & Olarrea, 2006; Zimmermann, 2019).

- (7) a. ¿Por qué *esto*_[+Foc] habríamos debido decirle, y no otra cosa?
 why this have.COND.1PL should say=him and not other thing
 ‘Why should we have said THIS to him, not something else?’
 b. *¿A quién *esto*_[+Foc] deberíamos decir, y no otra cosa?
 to whom this should.1PL say and not other thing
 ‘To whom should we say this, and not something else?’

There is, so far, little empirical research on the question of what triggers the occurrence of pre- and postverbal subjects in Italian and Peninsular Spanish ‘why’-interrogatives. Bianchi, Bocci & Cruschina (2017) show by means of a forced-choice experiment that the optionality between (3a) and (3b) is due to differences in the information structure of the subject. Accordingly, the subject in (3a) is focal, while it is non-focal in (3b).

Leonetti (2017) relates these findings to the canonicity of word order, which is defined in terms of textual frequency and pragmatic neutrality (Herring, 2000; Kemmerer, 2012). In Italian, the canonical word order is SV. He shows that the optionality in (3) is in line with the distribution of subjects in Italian declaratives (e.g., Zubizarreta, 1998; Belletti, 2004; Bocci & Pozzan, 2014; Bianchi, Bocci & Cruschina, 2017; Leonetti, 2018). Preverbal subjects generally receive a non-focal interpretation, while postverbal subjects are normally focal and their position corresponds to the non-canonical word order in Italian, as shown in (8). In other words, in both Italian declaratives and ‘why’-interrogatives, the position of the subject depends on its information structure (Leonetti, 2017).

- (8) a. *Maria* ha vinto il posto. (canonical word order)
 Mary has won the position
 ‘Mary has won the position.’
 b. Ha vinto il posto *Maria*_[+Foc]. (non-canonical word order)
 has won the position Mary
 ‘MARY won the position. / Mary is the one who got the position.’

Leonetti (2017) makes the same observation for Peninsular Spanish declaratives. Assuming SV as the canonical word order for Peninsular Spanish, he shows that the preverbal subject in (9) is non-focal, while the subject is focal when it occurs in postverbal position (cf. Leonetti, 2018: 18-19).

- (9) a. *María* ha ganado la plaza. (canonical word order)
 Mary has won the post
 ‘Mary has won the position.’
 b. Ha ganado la plaza *María*_[+Foc]. (non-canonical word order)
 has won the post Mary
 ‘MARY won the position. / Mary is the one who got the position.’

However, it is unclear whether Peninsular Spanish ‘why’-interrogatives display the same distribution of focal and non-focal subjects. In this paper, we want to fill this gap by answering the following research questions:

1. Does information structure determine word order in Italian and Peninsular Spanish ‘why’-interrogatives?
2. Do Peninsular Spanish and Italian ‘why’-interrogatives differ from each other and, if yes, why?

The paper is organized as follows: Section 2 presents the results of the parallel text study and the forced-choice experiment, comparing Italian and Peninsular Spanish ‘why’-interrogatives. In Section 3, we discuss the results from both studies in light of the research questions, before concluding in Section 4.

2. EMPIRICAL STUDIES ON WORD ORDER VARIATION IN ITALIAN AND PENINSULAR SPANISH ‘WHY’-INTERROGATIVES

We conducted two empirical studies in order to investigate the effect of information structure on word order in Italian and Peninsular Spanish ‘why’-interrogatives. The first study is a parallel text corpus study comparing translations of the same text in both languages. The second study consists of a forced-choice experiment that allows for a deeper understanding of those information structural categories that appear with a very low frequency in the corpus.

2.1 Parallel text corpus

2.1.1. Data collection and annotation

As the source of data, we used the Italian and Peninsular Spanish translations of detective novels and other stories published in ‘The Complete Sherlock Holmes’, as well as several scriptures from the Old and New Testament (Old Testament: Gen. - 2 Kings, New Testament: Matt. - Acts). In total, our corpus includes 860 interrogatives introduced by the *wh*-element ‘why’ (*perché* and *por qué*). Out of these ‘why’-interrogatives, 208 were extracted from the Italian (Doyle, 2015) and 216 of the Peninsular Spanish translation of ‘The Complete Sherlock Holmes’ (Doyle, 2018), 197 ‘why’-interrogatives were extracted from a modern Italian (Nuova Riveduta, 2006 (NR)), and 239 from a modern Peninsular Spanish Bible translation (Reina Valera, 1995 (RV)). For the present study, we used a subset of 196 ‘why’-interrogatives with overt subjects (excluding 664 sentences with null subjects). These interrogatives were annotated manually with respect to the following syntactic features, discourse categories and other variables:

- **word order**: relative position of subject, verb, and *wh*-phrase, as well as the realization of the subject (overt vs. null) (abbreviated as ‘why’V, ‘why’VS, ‘why’SV)
- **question type**: matrix vs. embedded interrogatives
- **language**: Italian vs. Peninsular Spanish
- **discourse categories of the subject**:
 - **focus** ([+Foc] vs. [-Foc]): set of alternatives that are relevant for the interpretation of the context in the sense of Krifka (2008: 247) and Heidinger (2018: 46).
 - **information status** ([+Given] vs. [-Given]): new vs. given information in the sense of Krifka (2008: 262) and Heidinger (2018: 46).

For the annotation of the discourse categories of the subject, we considered the preceding context and searched for a possible antecedent and for relevant alternatives of the subject. An example of how we annotated these discourse categories is provided in (10). In this example, an antecedent for the subject in (10a) and (10b), *Mrs. Barrymore*, is mentioned in the previous context. Thus, the subject is characterized by [+Given]. Since the context does not contain any relevant alternatives to the subject, it is classified as [-Foc]:

(10) Context: And yet he lied as he said it, for it chanced that after breakfast I met *Mrs. Barrymore* in the long corridor with the sun full upon her face. She was a large, impassive, heavy-featured woman with a stern set expression of mouth. But her tell-tale eyes were red and glanced at me from between swollen lids. It was she, then, who wept in the night, and if she did so her husband must know it. Yet he had taken the obvious risk of discovery in declaring that it was not so. Why had he done this?

a. Italian:

E perché *quella donna* aveva pianto così amaramente?
and why this woman has cried so bitterly (Doyle, 2015: 424)

b. Peninsular Spanish:

Y ¿por qué lloraba *su mujer* tan amargamente?
and why cried his wife so bitterly (Doyle, 2018: 246)
'And why did this woman cry so bitterly?'

We asked a second annotator to annotate about 40% of our data in order to check the original annotation. For focus, the inter-annotator agreement is 97.3% (Cohen's $\kappa = 0.73$ (substantial)) and for information status 100% (Cohen's $\kappa = 1.0$ (perfect)).

2.1.2 Results (parallel text corpus)

In this section, we first present the results for textual frequency which serve as a starting point and approximation for determining the effect of information structure on word order. Afterwards, we consider the distribution of 'why'SV and 'why'VS with respect to the discourse properties of the subject. Finally, we present a statistical analysis including the effects of the discourse properties, language, and text type on word order.

Textual frequency:

For textual frequency, the results are summarized in Table 1 and Figure 1, showing the absolute numbers and percentages for pre- and postverbal subjects in Italian and Peninsular Spanish 'why'-interrogatives. We find a crucial difference between both languages with respect to the frequency of the word order patterns. In Italian, preverbal subjects occur more frequently than postverbal ones, while Peninsular Spanish displays the opposite pattern.

	'why'VS	'why'SV	total ⁶
Italian (<i>perché</i>)	18 22.0%	64 78.0%	82 100%
Peninsular Spanish (<i>por qué</i>)	91 79.8%	23 20.2%	114 100%

Table 1: 'why'VS and 'why'SV order in 'why'-interrogatives (absolute numbers and percentages) in the parallel text corpus.

⁶ The divergent (total) number of overt subjects in Italian and Peninsular Spanish mostly results from the differences between Italian *lei* ('you' polite) and Peninsular Spanish *usted* ('you' polite). The use of overt *lei* in Italian is more restricted and requires either a narrow focus reading or a topic reading, while Peninsular Spanish *usted* can occur without any focal or contrastive value (Leonetti & Escandell-Vidal 2020).

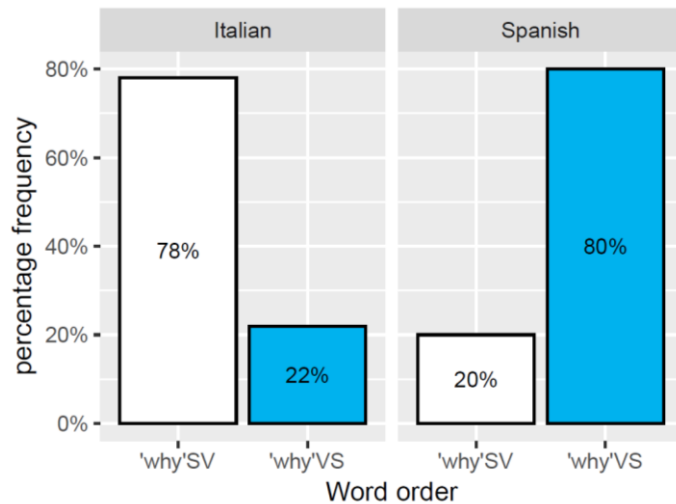


Figure 1: Percentage of preverbal subjects ('why'SV) and postverbal subjects ('why'VS) in Italian (left) and Peninsular Spanish 'why'-interrogatives (right).

Discourse categories:

The results for the discourse categories are presented in Table 2, illustrating the absolute number of all combinations. Due to the low number of subjects marked by focus [+Foc] and/or information status [-Given], we cannot consider all four combinations of discourse categories separately and we merely differentiate between non-focal and focal subjects, as shown in Figure 2. For non-focal subjects, we observe a clear difference between the two languages. While in Italian 'why'-interrogatives, non-focal subjects are more frequent in preverbal position, they predominantly occur postverbally in Peninsular Spanish 'why'-interrogatives. Focal subjects, on the other hand, occur more frequently in postverbal position in both languages.

		[+Given] [-Foc]	[-Given] [+Foc]	[+Given] [+Foc]	[-Given] [-Foc]
Italian	'why'(x)SV	53	1	5	5
	'why'(x)VS	1	0	14	3
Peninsular Spanish	'why'(x)SV	9	1	7	6
	'why'(x)VS	72	1	15	3

Table 2: Word order in 'why'-interrogatives across information structure conditions (absolute numbers) in the parallel text corpus.

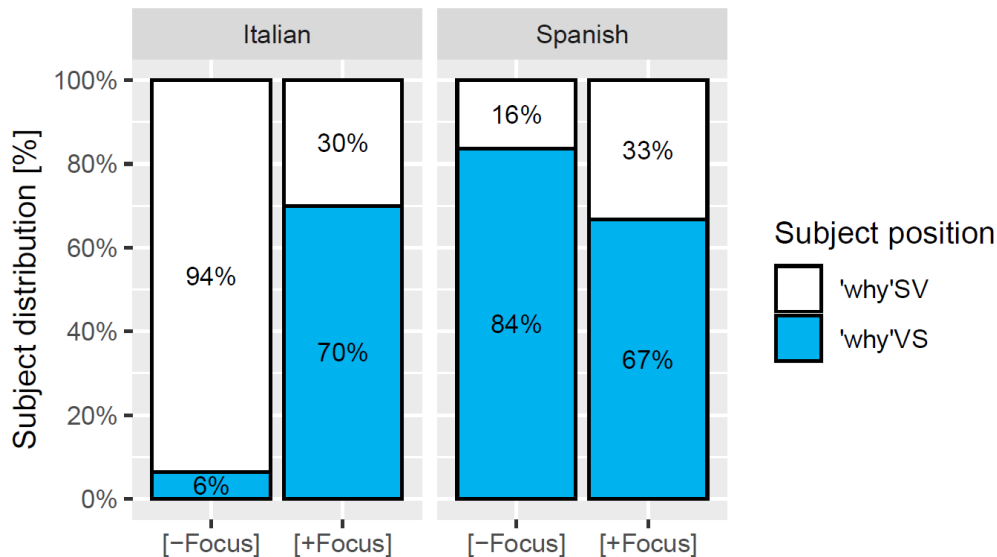


Figure 2: Percentage frequency of non-focal and focal subjects in the preverbal ('why'SV) and postverbal ('why'VS) subject position in Peninsular Spanish and Italian 'why'-interrogatives.

For the statistical analyses, we ran logistic mixed-effects regression models in R, using the packages lme4 (Bates, Maechler, Bolker, & Walkers, 2015). We defined *word order* as dependent variable and we considered *language* (Italian vs. Peninsular Spanish), *focus* ([+Foc] and [-Foc]), and text type as independent factors. The findings confirm the trends suggested by the descriptive results: The word order pattern depends on language ($\beta = 5.48$, $SE = 0.81$, $z = 6.74$, $p < .001$). In Italian, subjects occur more frequently in preverbal position, while they are more frequent in Peninsular Spanish in postverbal position. Focus significantly affects word order in Italian 'why'-interrogatives ($\beta = 4.50$, $SE = 1.10$, $z = 4.08$, $p < .001$). In Peninsular Spanish, by contrast, focus does not have an effect on word order in 'why'-interrogatives and the postverbal position is more frequent regardless of the focus interpretation of the subject ($\beta = -1.00$, $SE = 0.520$, $z = -1.19$, $p > .05$). A further noteworthy result concerns the text types. We do not find any significant difference between Sherlock Holmes translations and Bible translations ($\beta = -0.57$, $SE = 0.41$, $z = -1.39$, $p > .1$).

2.2 Forced-choice experiment

2.2.1 Methodology

We further ran a web-based forced-choice experiment on the platform *SoSci Survey* (Leiner, 2014) due to the small number of marked subjects in the corpus data that did not allow for a differentiated analysis of the discourse categories focus and information status. The aim of the experiment is therefore to investigate whether subjects marked by focus or new information affect word order in Peninsular Spanish and Italian 'why'-interrogatives.

2.2.2 Materials

The task was based on a 2x2 factorial design and consisted of 24 items, resulting in four conditions for each item and in a total of 96 experimental stimuli. The experimental stimuli were manipulated with respect to the dependent variable *word order*, considering 'why'VS and 'why'SV, and three independent variables. These independent

variables are the same categories that we annotated in the corpus study. Accordingly, two independent variables focus on the information structure of the subject, namely *information status*, with two levels: [+Given] and [-Given], and *focus*, distinguishing between [+Foc] and [-Foc]. Both factors were manipulated within items and within participants. The third independent variable considers the *languages* Peninsular Spanish and Italian varying between items and participants. All 24 items were introduced by a brief context which led to the respective information structural reading of the subject, followed by two interrogatives, varying in word order. Right after the interrogatives, an answer was added to underline the information structure of the subject. To keep constant as many other factors as possible, we used common Peninsular Spanish and Italian proper names for the subject and intransitive verbs.

Additionally, we created 24 fillers including pairs of declarative sentences that varied also in the position of one constituent (cf. Bianchi, Bocci & Cruschina, 2017). We used 16 grammatical and 8 ungrammatical sentences in order to control for the participants' attention during the experiment.

We created four lists so that each participant saw only one of the four conditions per item, and six sentences per condition during the experiment. The test sentences were presented along with the same number of fillers. In sum, each participant saw a total of 48 experimental stimuli.

2.2.3 Participants

A total of 372 participants, who were recruited via Facebook, took part in the experiment. Of these, 157 were Italian monolinguals (mean age = 32.3, range = 20-63, 111 female, 46 male) who were born in Italy and lived in Italy at the time of completion. The remaining 215 participants were native speakers of Spanish from Spain (mean age = 34.3, range = 20-78, 169 female, 45 male, 1 other). Further, 48 of the Peninsular Spanish participants were also native speakers of one of the co-official languages in Spain: Two were speakers of Basque, 13 of Galician and 33 of Catalan or Valencian.

Before the data analysis, we excluded three Italian and three Peninsular Spanish participants based on their performance on the control items. They performed unexpectedly on three or more items.

2.2.4 Procedure

The experiment consisted of two parts. In the first part, the participants read a short introduction followed by three practice items. Two of the practice items were constructed identically to the test stimuli, and one similarly to the fillers. For each stimulus, participants read a brief context, the two possible 'why'-interrogatives, and a short answer. They were then asked to specify which interrogative sounded more natural to them. The experimental stimuli were presented in a pseudo-randomized order combined with the filler sentences. The second part of the experiment consisted of a background questionnaire with sociolinguistic information on the participants. The total duration of the experiment was around 20 minutes.

2.2.5 Results (forced-choice experiment)

The results of the forced-choice experiment are summarized in Figure 3, which illustrates the distribution of preverbal and postverbal subjects in Italian and Peninsular Spanish across the four information structure conditions ([+Foc][-Given], [-Foc][-Given],

[+Foc][+Given], [-Foc] [+Given]). In Italian, the participants preferred postverbal over preverbal subjects in the case of contexts that lead to a focus interpretation of the subject. In contexts that favor a [-Foc] interpretation, there is a clear preference for preverbal over postverbal subjects. When the subject was already introduced in the context, there is a slight tendency towards the preverbal subject position as compared to the [-Given] conditions. In Peninsular Spanish, the postverbal subject position was clearly preferred over the preverbal subject position across information structure conditions.

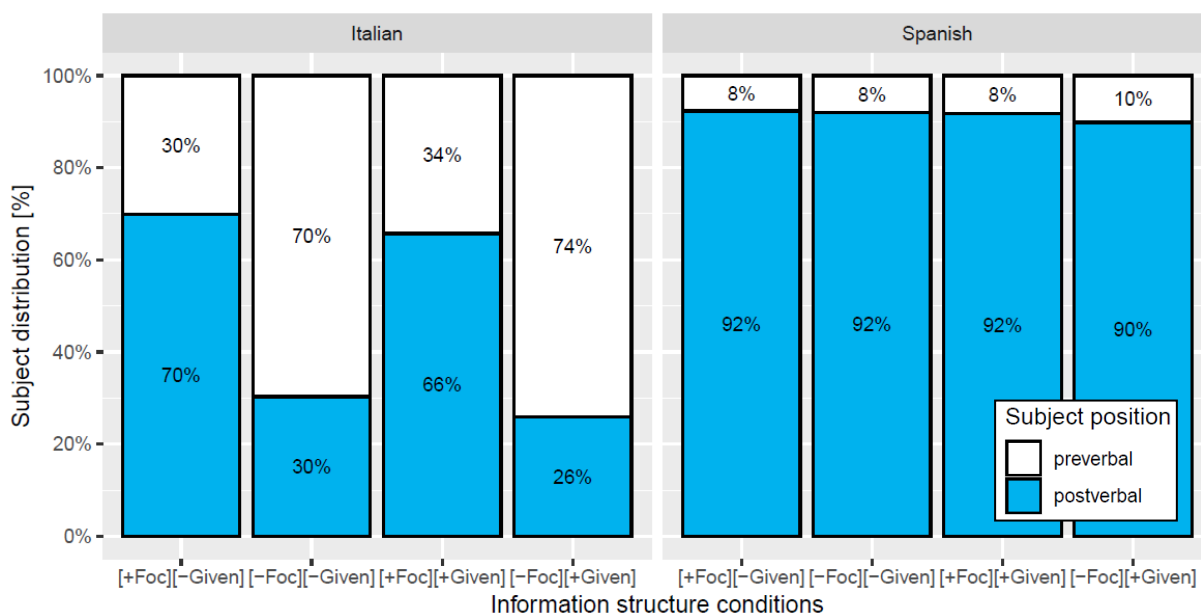


Figure 3: Distribution of pre- and postverbal subjects in Italian and Peninsular Spanish across information structure conditions in the forced-choice experiment.

For the statistical analyses, we ran logistic mixed-effects regression models in R. We defined *word order* as dependent variable, including two levels: the preverbal and the postverbal subject position. We specified three different independent variables with two levels each. The first model took *language* (Italian vs. Peninsular Spanish) into account as independent factor. In the second and the third model, we separately considered *information status* ([+Given] and [-Given]), and *focus* ([+Foc] and [-Foc]) for each language, including main effects and interaction terms. For each model, we further added *participant* and *item* as random effects.

The results show a clear difference between the two languages ($\beta = -2.80$, $SE = 0.07$, $z = -39.05$, $p < .001$). In Peninsular Spanish, we find an overall preference for postverbal over preverbal subjects, while we do not observe a general preference in Italian.

In the following analyses, we consider whether the information structure of the subject determines word order in ‘why’-interrogatives. As anticipated, Italian subjects that are marked by focus are preferred in the postverbal position. By contrast, preverbal subjects are significantly more likely in contexts with a [-Foc]-interpretation ($\beta = 2.01$, $SE = 0.12$, $z = 17.51$, $p < .001$). Surprisingly, we do not observe an effect of focus in Peninsular Spanish ($\beta = -0.06$, $SE = 0.16$, $z = -0.35$, $p = 0.72$). Regardless of whether the subject qualifies as focus or not, Peninsular Spanish speakers prefer the postverbal subject position. For information status, we find a small effect in Italian ‘why’-interrogatives ($\beta = 0.23$, $SE = 0.11$, $z = 1.99$, $p < .05$). The preference for the preverbal subjects increases slightly if the subject is already introduced in the context ([+Given]). For Peninsular Spanish, we observe the same effect for information status ($\beta = 0.31$,

$SE = 0.16, z = 1.97, p < .05$). However, a closer look at the individual items for Peninsular Spanish shows that this effect is only driven by two items. If we exclude these two items, the effect disappears ($\beta = 0.18, SE = 0.16, z = 1.09, p = .28$). In Italian, by contrast, the effect of information status can be found over a large majority of the items.

3. DISCUSSION

The first research question was concerned with whether the information structure of the subject affects word order in Italian and Peninsular Spanish ‘why’-interrogatives. For Italian, the results of our corpus study and of our forced-choice experiment confirm Bianchi, Bocci & Cruschina’s (2017) findings. We identified focus as the discourse category that has a major impact on word order in ‘why’-interrogatives. Non-focal subjects occur most frequently in preverbal position, while focal subjects normally appear postverbally, as illustrated in (11):

- (11) a. Perché Maria_[-Foc] balla?
 why Mary dances
 ‘Why does Mary dance?’
 b. Perché balla Maria_[+Foc]?
 why dances Mary
 ‘Why does MARY dance?’

Information status only shows a minor effect on word order in Italian ‘why’-interrogatives. If the subject is already introduced in the context, the preference for the preverbal position increases to a slight extent.

For Peninsular Spanish, the results of our two studies (parallel text corpus study and forced-choice experiment) show that non-focal subjects mostly occur in postverbal position ((12a)). In our corpus, 90% of these subjects are in this position, which is also the highly preferred position in our forced-choice experiment. For focal subjects, both studies show that the postverbal position is also clearly preferred over the preverbal one ((12b)), but the corpus study reveals a higher percentage of preverbal subjects in comparison to the forced-choice experiment.

- (12) a. ¿Por qué baila María_[-Foc]?
 why dances Mary
 ‘Why does Mary dance?’
 b. ¿Por qué baila María_[+Foc]?
 why dances Mary
 ‘Why does MARY/Mary dance?’

Note, however, that this strong preference for the postverbal subject position over the preverbal one does not mean that preverbal subjects with a focus interpretation are unacceptable in Peninsular Spanish ‘why’-interrogatives. The results of our corpus study clearly reveal that there is variation with respect to the subject position. Evidence for this variation is presented in Kaiser, von Heusinger & Schmid (2019) who provide data showing that focal subjects are acceptable in both the preverbal and the postverbal position:

- (13) a. ¿Por qué prepara esta profesora el material para la clase de
 why prepares this teacher the material for the lesson of
 Matemáticas?
 Mathematics
 ‘Why does THIS TEACHER prepare the lesson material for the Math class?’
 Mean rating: 5.2 (on a Likert scale from 1 (‘unnatural’) to 7 (‘unnatural’))
- b. ¿Por qué esta profesora prepara el material para la clase de
 why this teacher prepares the material for the lesson of
 Matemáticas?
 Mathematics
 ‘Why does THIS TEACHER prepare the lesson material for the Math class?’
 Mean rating: 5.5 (on a 7-point Likert scale)

We assume that one possible reason for the higher percentage of preverbal subjects in the corpus study than in the forced-choice experiment in Peninsular Spanish could be the test design of the latter. In order to get statistically reliable and robust data, we were forced to restrict the number of variables in the forced-choice experiment and we therefore only presented sentences containing intransitive verbs without any further adjuncts. In our corpus study, however, we also encounter ‘why’-interrogatives with transitive verbs or intransitives with adjuncts, which might affect the subject position and generally increase the number of possible positions for the subject to appear in.

Before addressing the possible impact of verb type on word order, let us first discuss our second research question, namely whether Italian and Peninsular Spanish ‘why’-interrogatives differ from each other. When comparing the frequency of the word order patterns in the corpus study, we find that ‘why’SV is by far the most frequent pattern in Italian, while ‘why’VS occurs most frequently in Peninsular Spanish. Both our corpus study and our forced-choice experiment reveal that these word order patterns occur with non-focal subjects. Put differently, non-focal subjects appear in different position in these two languages. In Italian, they occur preverbally (see (11a)), whereas they occupy the postverbal position in Peninsular Spanish (see (12a)).

Comparing the position of non-focal subjects in ‘why’-interrogatives with the canonical subject position in neutral declaratives, we found that the position of non-focal subjects in Italian ‘why’-interrogatives coincides with their position in declaratives (compare (14a) and (14b)). Furthermore, focal subjects occur postverbally in Italian ‘why’-interrogatives, and therefore match the pattern of declaratives with non-canonical word order (compare (14c) and (14d)).

- (14) a. Perché *questo mastino* non era legato questa notte? (Doyle, 2015: 459)
 why this hound not was tied this night
 ‘Why was this hound not tied up last night?’
- b. *Holmes* si segnò l’ indirizzo. (Doyle, 2015: 16)
 Holmes himself wrote down the address
 ‘Holmes wrote down the address.’
- c. E perché dovrebbe farlo *il signor McFarlane*_[+Foc]?
 and why should make=this the Mister McFarlane
 ‘And why should Mr. McFarlane do this?’ (Doyle, 2015: 494)
- d. L’ho scritto *io*_[+Foc], per farla venire qui. (Doyle, 2015: 516)
 it have written I for make=her come here
 ‘I wrote it to make her come here.’

Note that Leonetti (2018) argues that word order in Italian ‘why’-interrogatives and declaratives can be explained by the same effect. More specifically, he argues that if the language allows an alternation between preverbal and postverbal subjects, postverbal subjects have a focus interpretation because of a correlation between optionality of inversion and focus. Given this correlation, we would expect to make the same observation in Peninsular Spanish. However, according to the results of our studies, this assumption does not seem to hold for Peninsular Spanish ‘why’-interrogatives, in which non-focal subjects occur in postverbal position while occurring preverbally in neutral declaratives:

- (15) a. ¿Por qué tendría que estar suelto *ese animal* [...] esta noche?
 why have.COND.3SG to be loose this animal this night
 ‘Why should this animal be on the loose tonight?’ (Doyle, 2018: 312)
- b. *Holmes* tomó nota de la dirección. (Doyle, 2018: 34)
 Holmes took note of the address
 ‘Holmes took note of the address.’

In sum, the word order in Italian ‘why’-interrogatives corresponds to that of declaratives. In both clause types, two subject positions are available, but only the postverbal position is related to a focus interpretation. In Peninsular Spanish, this correlation does not exist. Non-focal subjects occur in postverbal position in ‘why’-interrogatives, while they are preverbal in neutral declaratives. We assume that this difference between both languages is related to crosslinguistic differences in word order restrictions. Recall that both languages display SVO as canonical word order, as shown in Section 1. Apart from this canonical SVO order, (Peninsular) Spanish allows all types of word order except SOV (i.e. SVO, VOS, OSV, VSO and OVS) (Ordóñez, 1998; Zubizarreta, 1998; Gutiérrez-Bravo, 2002; 2008; Lahousse & Lamiroy, 2012; Olarrea, 2012; Leonetti, 2017). Italian, however, is more restrictive in this sense. In addition to SOV, it also lacks VSO (Belletti, 2004; Cardinaletti, 2004; Lahousse & Lamiroy, 2012; Leonetti, 2017).⁷ This difference is illustrated in the following example:

- (16) Italian:
 a. *Ha letto *Gianni* il giornale. (Cardinaletti, 2004: 118)
 has read John the newspaper
 ‘John read the newspaper.’
- Peninsular Spanish:
 b. Ayer ganó *Juan* la lotería. (Cardinaletti, 2004: 118)
 Yesterday won John the lottery
 ‘Yesterday John won the lottery.’

⁷ Note that this applies to all sentences without marginalization, right-dislocation or an intervening pause preceding the subject (Belletti, 2004). An exception to this general lack of VSO in Italian is found in cases in which the subject is pronominal (Belletti, 2004: 31):

- (i) Di quel cassetto ho *io* le chiavi.
 of that drawer have I the keys.
 ‘I have the keys for that drawer.’

Given that in Italian declaratives VSO is generally excluded, we suppose the same restriction for Italian ‘why’-interrogatives.⁸ As a consequence, in Italian ‘why’-interrogatives containing an object, subjects can either occur immediately in front of the finite verb (‘why’SVO) or at the right edge of the clause in these sentences (‘why’VoS):

- (17) a. *Perché il mio signore* perseguita il suo servo? (NR, 2006: 1 Sam 26:18)
 why the my lord pursue the his servant
 ‘Why is my lord pursuing his servant?’
 b. *E perché* dovrebbe farlo il signor McFarlane? (Doyle, 2015: 494)
 and why should make=it the Mister McFarlane
 ‘And why should Mr. McFarlane do it?’

Peninsular Spanish ‘why’-interrogatives reveal an even higher level of optionality regarding word order patterns. We observe besides the two subject positions available in Italian (in (18a) and (18c)) an additional postverbal subject position adjacent to the verb (VSO), see (18b):

- (18) a. *¿Por qué tus discípulos* quebrantan la tradición de los ancianos?
 why your disciples break the tradition of the elders
 ‘Why do your disciples break the tradition of the elders?’
 (RV, 1995: Matt. 15:2)
 b. *¿Por qué llenaba este individuo* la casa, en una ocasión así,
 why filled this person the house in a occasion like that
de fuertes olores? (Doyle, 2018: 1479)
 of strong odors
 ‘Why should this man at such a time be filling his house with strong odors?’
 c. *¿Por qué hizo esto Jehová* a esta tierra? (RV, 1995: Deut 29:24)
 why did this lord to this land
 ‘Why has the Lord done this to this land?’

This difference between Italian ‘why’-interrogatives on the one hand, and Peninsular Spanish ‘why’-interrogatives on the other raises the question of how this variation can be explained. We argue that this variation is related to differences in word order restrictions in declarative clauses. Following Belletti (2004), we assume that Peninsular Spanish is less restricted with respect to the positioning of the subject, since it has either a further subject position located higher in the structure than postverbal subjects in Italian, or it exhibits an additional option for case marking which does not exist in Italian.⁹ A further question that arises is how the higher flexibility with respect to word

⁸ Interestingly, we found evidence for this assumption in our corpus, considering the asymmetry between pronominal subjects and nominal subjects, described in footnote 7. Our corpus provides one example for ‘why’VS_{pronominal}O (see (i)) with a right-dislocated object, while it does not offer any example for ‘why’VS_{nominal}O in Italian, although this is a common word order pattern in Peninsular Spanish ‘why’-interrogatives (see (ii)).

(i) *Perché non le dici anche tu* le preghiere? (Doyle, 2015: 35)
 why not them say also you the wishes
 ‘Why don’t you give them some wishes yourself?’

(ii) *¿Por qué se llevó Cadogan West* los planos? (Doyle, 2018: 1222)
 why himself took Cadogan West the plans
 ‘Why did Cadogan West take the plans?’

⁹ An alternative explanation for this variation could be that Italian and Peninsular Spanish differ with respect to their canonical word order patterns. While SVO is generally considered to be the canonical word order in Italian, the determination of the canonical word order is more controversial in Peninsular Spanish.

order patterns and intonational focus marking in Peninsular Spanish ‘why’-interrogatives are related to each other (see Bosque & Gutiérrez-Rexach, 2009: 681, Olarrea, 2012, and Vanrell, 2013, for a discussion on intonational focus marking in Peninsular Spanish declaratives). We will leave the answers to these questions for further research.

4. CONCLUSION

In this paper, we examined the question of whether information structure affects word order in Italian and Peninsular Spanish ‘why’-interrogatives and whether there are differences between the two languages. To answer these questions, we conducted two empirical studies, a parallel text corpus study and a forced-choice experiment. Our studies show that Italian and Peninsular Spanish ‘why’-interrogatives differ from each other in two ways. First, non-focal subjects occur preverbally in Italian, while they occupy the postverbal position in Peninsular Spanish. Second, VSO does not occur in Italian ‘why’-interrogatives, while it does so in Peninsular Spanish. We assume, in accordance with Belletti, that in comparison with Italian, Peninsular Spanish has an additional higher subject position or an additional option for case marking.

Our results further show that word order in Italian ‘why’-interrogatives is determined by information structure. In contrast, information structure does not seem to affect word order in Peninsular Spanish ‘why’-interrogatives. Note that this result for Peninsular Spanish ‘why’-interrogatives might only apply to intransitive verbs since we did not include transitive verbs in our forced-choice experiment. Therefore, the focus distribution in Peninsular Spanish ‘why’-interrogatives with transitive verbs remains unclear. Future studies should address this issue by investigating the interplay of word order and information structure in this context. We further believe that these studies should take intonation into account as a potential contributing factor.

REFERENCES

PARALLEL TEXT TRANSLATIONS

- DOYLE, A. C. 2015. *Tutto Sherlock Holmes*. Roma: Newton Compton.
DOYLE, A. C. 2018. *Sherlock Holmes. La colección completa*. Stockholm: Atoz Classics.
Biblia Reina-Valera. Sociedades Bíblica Unidas, 1995.
<<https://www.biblegateway.com/versions/Reina-Valera-1995-RVR1995-Biblia/>>
Nuova Riveduta. Società Biblica di Ginevra, 2006.
<<https://www.biblegateway.com/versions/Nuova-Riveduta-2006-NR2006/>>

LINGUISTIC REFERENCES

- BATES, D., MAECHLER, M., BOLKER, B. & WALKERS, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1-48.
BELLETTI, A. 2004. Aspects of the low IP area. In: L. Rizzi (ed.). *The Structure of CP and IP. The Cartography of Syntactic Structures. Volume 2*. Oxford: Oxford University Press, 16-51.
BELLETTI, A. (ed.) 2004. *Structures and Beyond. The Cartography of Syntactic Structures. Volume 3*. Oxford: Oxford University Press.

In contrast to Leonetti (2017) and others, some researchers question whether Peninsular Spanish should be analyzed as having a basis SVO order by arguing that in specific contexts, such as in sentences with presentational focus, VS order is the unmarked one (Zagona, 2002: 48–50). In other words, the frequency of VS order in Peninsular Spanish ‘why’-interrogatives might be higher because the postverbal subject position is unmarked.

- BIANCHI, V., BOCCI, G. & CRUSCHINA, S. 2017. Two types of subject inversion in Italian *wh*-questions. *Revue roumaine de linguistique* 62, 233-252.
- BOCCI, G. & POZZAN, L. 2014. Some questions (and some experimental answers) about Italian subjects: Subject positions in main and embedded questions in L1 and attrition. In: C. Contemore & L. D. Pozzo (eds.), *Inquiries into Linguistic Theory and Language Acquisition. Papers Offered to Adriana Belletti*. Siena: CISCL Press, 28-44.
- BOSQUE, I. & GUTIÉRREZ-REXACH, J. 2009. *Fundamentos de sintaxis formal*. Madrid: Akal.
- CALABRESE, A. 1982. Alcune ipotesi sulla struttura informazionale della frase in italiano e sul suo rapporto con la struttura fonologica. *Rivista di grammatica generativa* 7, 3-78.
- CARDINALETTI, A. 2004. Toward a cartography of subject positions. In: L. Rizzi (ed.). *The Structure of CP and IP. The Cartography of Syntactic Structures. Volume 2*. Oxford: Oxford University Press, 115-165.
- CONTRERAS, H. 1999. Relaciones entre las construcciones interrogativas, exclamativas y relativas. In: I. Bosque & V. Demonte (eds.), *Gramática descriptiva de la lengua española. Vol. 2: Las construcciones sintácticas fundamentales. Relaciones temporales, aspectuales y modales*. Madrid: Espasa Calpe, 1931-1962.
- DAVIS, J. C. 1971. Tú, ¿Qué tú tienes? *Hispania* 54, 331-333.
- GUTIÉRREZ-BRAVO, R. 2002. *Structural Markedness and Syntactic Structure: A Study of Word Order and the Left Periphery in Mexican Spanish*. Santa Cruz: University of California. Ph.D. Dissertation.
- GUTIÉRREZ-BRAVO, R. 2008. Identificación de los tópicos y los focos. *Nueva revista de filología española* 2, 363-401.
- HEIDINGER, S. 2018. *Sekundäre Prädikation und Informationsstruktur: Fokus und Informationsstatus bei spanischen Depiktiven*. Berlin: Lang.
- HERRING, S. C. 2000. Poeticallity and Word Order in Old Tamil. In: S. C. Herring, P. v. Reenen & L. Schøsler (eds.), *Textual Parameters in Older Languages*. Amsterdam: Benjamins, 197-236.
- KAISER, G. A., VON HEUSINGER, K. & SCHMID, S. 2019. Word order variation in Spanish and Italian interrogatives. The role of the subject in 'why'-interrogatives. In: N. Pomino (ed.), *Proceedings of the IX Nereus International Workshop: "Morphosyntactic and Semantic Aspects of the DP in Romance and Beyond"*. Konstanz: Fachbereich Linguistik, Universität Konstanz (= Arbeitspapier 131), 69-90.
- KEMMERER, D. 2012. The Cross-Linguistic Prevalence of SOV and SVO Word Orders Reflects the Sequential and Hierarchical Representation of Action in Broca's Area. *Language and Linguistics Compass* 6, 50-66.
- KRIFKA, M. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55, 243-276.
- LAHOUSSE, K. & LAMIROY, B. 2012. Word order in French, Spanish and Italian: a grammaticalization account. *Folia Linguistica* 46, 387-415.
- LEINER, D. J. 2014. SoSci Survey (Version 2.5.00-i) [Computer software]. Available at <http://www.sosicisurvey.com>.
- LEONETTI, M. 2017. Basic constituent orders. In: A. Dufter & E. Stark (eds.), *Manual of Romance Morphosyntax and Syntax*. Berlin: De Gruyter, 887-932.
- LEONETTI, M. 2018. Two types of postverbal subject. *Italian Journal of Linguistics* 30, 11-36.
- LEONETTI, M., & ESCANDELL-VIDAL, V. 2020. Strong pronouns as postverbal subjects in Spanish and Italian. In M. Georgiagentis, G. Giannouloupoulou, M. Koliopoulou & A. Tsokoglou (eds.), *Contrastive Studies in Morphology and Syntax*. London: Bloomsbury Academic, 210-224.
- OLARREA, A. 2012. Word order and information structure. In: J. I. Hualde, A. Olarrea & E. O'Rourke (eds.), *The Handbook of Hispanic Linguistics*. Malden: Blackwell, 603-628.
- ORDÓÑEZ, F. 1998. The inversion construction in interrogatives in Spanish and Catalan. In: J. Lema & E. Treviño (eds.), *Theoretical Analyses on Romance Languages. Selected Papers from the 26th Linguistic Symposium on Romance Languages (LSRL XXVI), Mexico City, 28-30 March 1996*. Amsterdam: Benjamins, 329-350.
- ORDÓÑEZ, F. & OLARREA, A. 2006. Microvariation in Carribean/non Carribean Spanish interrogatives. *Probus* 18, 59-97.
- RIZZI, L. 1996. Residual verb second and the *wh*-criterion. In: A. Belletti & L. Rizzi (eds.), *Parameters and Functional Heads. Essays in Comparative Syntax*. New York: Oxford University Press, 63-90.
- RIZZI, L. 2001. On the position 'Int(errogative)' in the left periphery of the clause. In: G. Cinque & G. Salvi (eds.), *Current Studies in Italian Syntax. Essays offered to Lorenzo Renzi*. Amsterdam: Elsevier, 287-296.
- RIZZI, L. 2006. Selective residual V-2 in Italian interrogatives. In: P. Brandt and E. Fuß (eds.), *Form, Structure and Grammar. A Festschrift Presented to Günther Grewendorf on Occasion of His 60th Birthday*. Berlin: Akademie Verlag, 229-241.

- RIZZI, L. & BOCCI, G. 2017. Left periphery of the clause. Primarily illustrated for Italian. In: M. Everaert & H.C.v. Riemsdijk (eds.), *The Wiley Blackwell Companion to Syntax. Volume IV. Im–Mi*. Hoboken: Wiley Blackwell, 2nd edition, 2171-2200.
- SUÑER, M. 1994. V-movement and the licensing of argumental *wh*-phrases in Spanish. *Natural Language and Linguistic Theory* 12, 335-372.
- TORIBIO, J.A. 1993. *Parametric variation in the licensing of nominals*. Ph.D. Dissertation, Cornell University.
- TORREGO, E. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 115, 103-129.
- VANRELL, M. del Mar. 2013. Variation at the interfaces in Ibero-Romance. Catalan and Spanish prosody and word order. *Catalan Journal of Linguistics* 12, 253-282.
- ZIPF, J. & QUAGLIA S. 2017. Asymmetries in Italian matrix *wh*-questions: Word order and information structure. In: M. Butt & T.H. King (eds.), *Proceedings of the LFG'17 Conference. University of Konstanz*. Stanford: CSLI Publications, 387-405.
- ZAGONA, K. 2002. *The Syntax of Spanish*. Cambridge: Cambridge University Press.
- ZUBIZARRETA, M. L. 1998. *Prosody, Focus and Word Order*. Cambridge: The MIT Press.
- ZIMMERMANN, M. 2019. Reconsidering the syntax of interrogatives in Caribbean Spanish, with special reference to Dominican Spanish. *Isogloss* 5, 1-36.

Recebido: 16/9/2020
Aceito: 17/7/2021
Publicado: 28/9/2021