

Information Structure and Syntax in Ibero-Romance: A Literature Review.

Leonora Sampaio ^a.

^a Faculty of Letras, Federal University of Minas Gerais, Belo Horizonte, Brazil, ldfsampaio@gmail.com.

Abstract. The relationship between information structure and syntax is a close one throughout many languages. This interface is particularly interesting to the Romance group, which is known for conveying information through grammatical structures, establishing a direct (albeit not singular) link between form and meaning. As the size of this group is large and its members vary depending on the criteria, this paper limited itself to the Ibero-Romance languages, mainly Spanish and Portuguese varieties. It aimed to gain a better knowledge of what has been discussed about Ibero-Romance languages on the interface between information structure and syntax in the past decade. In this sense, a thorough literature review was conducted in the academic-renowned databases *ScienceDirect* and *Scopus* during the months of January and February of 2024. Although the research was limited to the free-of-charge papers available in these databases, it acquired some perspective on what has been the findings and discussions for the scientific community of the chosen field of interest in the last 10 years. Ultimately, the collected articles discussed the matters of word order inversion, focus and topic structures, as well as specific positions within a sentence. These papers were not only concerned with parametrical variation within a single language but also with a cross-linguistic analysis.

Keywords. Ibero-Romance, Information Structure, Syntax, Literature Review.

1. Introduction

The relationship between information structure and syntax is a close one throughout many languages. This interface is particularly interesting to Romance languages, as this group often opts for syntactic means to convey different types of information. To convey focus structures, for example, these languages lack morphological markers and instead resort to the syntax and prosody field [1]. In a complementary way, studying information structure through the lens of Romance languages has proved helpful in solving many debates in the field [1], especially regarding the already mentioned marking of focus structures, which will be further explained below.

The Romance group, however, is a large one and its members vary depending on which criteria is chosen to define it. For these reasons, the present study chose to focus on the subgroup of Ibero-Romance languages, mainly Portuguese and Spanish, in a search to discover what has been discussed about these languages concerning the interface between information structure and syntax for the past 10 years.

With that in mind, one might wonder, then, how information structure can be expressed via syntactic

means, and for that, this study already has an answer: word order, i.e. the disposition of constituents inside a sentence or clause (S = subject; V = verb; O = object). By lacking morphological markers to determine the communicative intent of a word or expression [1], Ibero-Romance languages move and rearrange the order of their constituents to express meaning. Nevertheless, it isn't possible to affirm that there is a singular direct correlation between order and information. Indeed, the literature shows that a specific word order doesn't always convey the same information in different languages, nor in the varieties of the same language. In this sense, there is a canonical (or unmarked) word order in every language. While Classical Latin is known for displaying a predominantly SOV order and it is speculated that Romance languages disposed of a V2 stage in Medieval times, all Romance languages present an SVO order today [1]. Ergo, both Portuguese and Spanish display an unmarked SVO order today.

To better comprehend what is studied through the information structure-syntax interface, a brief explanation is needed. Let us simulate a communicative situation in Portuguese with Speaker A and Speaker B.

(1) *Speaker A spots an empty plate at the kitchen table:*

A: Quem comeu o último pedaço de bolo?
Who ate the last piece of cake?

B: A Maria comeu o último pedaço de bolo.
Mary ate the last piece of cake.

Under the present analysis, a few things can be said about Speaker B's answer, also known as a declarative sentence. On a syntactic level, this sentence has an SVO order, in which *Maria* (Mary) is the subject, *comeu* (ate) is the verb, and *o último pedaço de bolo* (the last piece of cake) is the object. Besides that, it can also be said that *o último pedaço de bolo* is a shared information between both speakers or rather it is the topic of which the sentence is about. *Maria*, on the other hand, is the new information that Speaker B shared with Speaker A. In information structure terms, *o último pedaço de bolo* is known as the topic of the sentence, whereas *Maria* is labeled as its focus or, more specifically, its information focus.

Topic and focus constitute a key notion of information structure, and they can affect the grammatical properties of the sentence on a syntactic level [2]. For its part, Cruschina says that topicality can be divided between the notions of aboutness topic (AT), which identifies what a sentence is about, and general topic (GT), which conveys old information in the communicative situation [2]. The syntactic mechanism used to mark an expression as the topic of a sentence is known as topicalization [2]. On it, a constituent is dislocated, usually to the left periphery of the sentence.

On the other hand, focus is generally classified as new information introduced in the discourse [1]. However, there isn't a singular focus structure and scholars don't necessarily agree on a classification. In a general sense, there is a division between the previously mentioned information focus, which Cruschina classifies as the information that indicates covert questions implicit in the context of communication; and contrastive focus, which presents a correction or denial of information in the discourse [1], as can be seen in (2):

(2) A: Quem comeu o último pedaço de bolo,
Maria ou João?

Who ate the last piece of cake, Mary or John?

B: Maria comeu o último pedaço de bolo.

Mary ate the last piece of cake.

Although some scholars have suggested a change in the nomenclature of contrastive focus [1], the basic difference in focus structures has been shown here. With this brief description, we can advance in our paper to better understand what have been the discussions and findings regarding information structure and syntax in Ibero-Romance languages.

This paper was structured as follows: Section 1 briefly introduces the area of our investigation and its basic concepts; Section 2 presents the Methodology used to create our literature review; Section 3 exhibits the findings regarding word order; Section 4 touches on the discussions about topicality and focus structures; Section 5 illustrates

the controversy regarding a V2-system for Ibero-Romance; and Section 6 concludes our paper.

2. Methodology

The aim of this paper was to gain a better knowledge of what was being discussed about Ibero-Romance languages on the interface between information structure and syntax. In this sense, a thorough literature review was conducted in the academic-renowned databases *ScienceDirect* and *Scopus* during the months of January and February of 2024. For this, keywords related to the theme were combined and scrambled to generate different results, such as "information structure", "word order", "focus structures", "topicalization", "romance languages", "Spanish" and "Portuguese". Due to the limited access to free-of-charge papers on these platforms, *Google Scholar* and *SciELO Brazil* were also used as databases, although less thoroughly.

In total, 49 papers were found related to Romance languages and Information Structure. Consequently, a selection of filters was needed to continue our research. Initially, a 10-year gap was settled (that is, only papers published from 2014 to 2024 were considered); after reading the Abstract and Conclusion section of the remaining texts, articles that extrapolated the limits of the information structure-syntax interface and/or Ibero-Romance languages were also excluded. In the end, 10 papers were considered relevant to our research.

3. Word order: an overview

Kato and Martins draw a comparison of European and Brazilian Portuguese (EP and BP, respectively) regarding word order. The authors show that both varieties of Portuguese agree on the unmarked order SVO for simple declaratives and exhibit similar behavior for negative sentences, yes/no questions, and the negative concord phenomenon [3]. However, their differences start with clitic pronouns. In fact, EP displays a system that differs from Ibero-Romance languages in general as it: i. allows both enclisis and proclisis (unlike Portuguese, which possesses a system for generalized proclisis); ii. isn't governed by the opposition of finite/non-finite (unlike Catalan and Spanish); and iii. possess a proclisis trigger (which isn't useful for languages such as Portuguese, Spanish, and Catalan) [3].

In a general sense, EP is less restrictive than BP regarding word order. For instance, it has a higher acceptability of VS order, whereas BP imposes stronger restrictions on this inversion [3]. As an example, for *wh*-questions, EP accepts a VXS order for its answer (*X* = another constituent such as an object or locative), while BP manipulates its response into a cleft sentence [3]. To illustrate, we adapt example (1) to the appropriate answer an EP and a BP speaker would give to the question "Who ate the last piece of cake?":

(3) EP: *Comeu o último pedaço de bolo a Maria.*

Ate the last piece of cake Mary.

(4) BP: Foi a Maria que comeu o último pedaço de bolo.

It was Mary who ate the last piece of cake.

Kato and Martins show that an exception to the rule is when a locative is selected by the verb to appear at the beginning of the sentence [3]. The authors claim that the general rule is that VS is accepted both by EP and BP varieties “if some XP precedes the verb, even if this XP is a covert deictic expression” [3].

This is similar to the findings of Corr. The author defends that, in Ibero-Romance languages, wide-focus inversion is a type of systematic phenomena that necessarily involves a null-locative element [4]. Her paper researches wide-focus inversion in six varieties of Ibero-Romance languages: BP, EP, Asturian (Ast), European Spanish (ES), Mexican Spanish (MS), and River-Plater Spanish (RS). BP, in particular, supports her claim that wide-focus inversion is related to null-locatives and cannot be explained by a null-subject parameter as previous research indicated, for BP is considered only as a partial null-subject language and yet still presents this type of inversion [4].

Corr argues that “Ibero-Romance displays fine-grained, systematic inversion phenomena that vary according to variety and verb class” [4]. She is able to create a “scale” of inversion for Ibero-Romance which consists of four classes of verbs, classified according to three semantic features: path, location, and deixis. With that, it is possible to conclude that intra-familiar variation regarding verbs allowing SV/VS order persists in Ibero-Romance [4]. When both orders are available, speakers will have systematically varying degrees of “preference” depending on verb, variety, and pragmatic interpretation [4]. This “preference” is later explained in syntactic terms as the availability and further selection of null locatives in a language [4]. Among the findings, it is also stated that Portuguese is the most restrictive language regarding wide-inversion, and BP is considered its most restrictive variant [4], similar to what is claimed by Kato and Martins.

For his part, Martín presents a study that distances itself from word order but presents expressive data about Spanish syntax under a generative approach. Alongside Cruschina, he agrees that the default (or unmarked) word order in Spanish is SVO, although with exceptions [5]. However, his work focuses on the left-periphery of Spanish and the controversy regarding the nature of the syntactic position of the Spanish preverbal subject.

In his paper, Martín relates position to discourse functions. To him, verbs may reach a different position (FocP or TopicP) depending on what type of feature it bears, and their position may be even higher than generally accepted in Spanish [5]. Furthermore, it is stated that null subjects and dative pronouns have a bound relationship and that it’s possible that accusative clitic pronouns have both a bound and conferential relationship [5]. In this regard, Martín affirms that there is an asymmetry between morpho-syntactic features of

subject/dative objects and accusative objects, but not with their preverbal position [5]. He also defends that all preverbal positions are related to the Complementizer field [5]. Besides that, he also shows that Spec-of-FocP is one of the positions where Spanish subjects may appear preverbally [5]. In this sense, it is stated that, although there is a canonical subject position in Spanish, other predicative layers are present in the preverbal position (one of them consisting of the position adjacent to the verb) [5]. As his final conclusion, Martín affirms that different preverbal positions serve as landing or base-generation sites for preverbal subjects [5]. For the author, this analysis further elucidates why OVS is not a possible word order for thetic sentences [5].

4. Strategies to convey information

4.1 On topicality

In what concerns topicality and, more specifically, topicalization in Ibero-Romance languages, de Andrade carried out a study about the emergence of topicalization in European Portuguese in which it is shown that this variety of Portuguese has taken a different path regarding left-dislocation in comparison to the Ibero-Romance sub-group [6]. The author defends that, for other Romance languages, the transition from medieval into their modern stages has led to the loss of V2 Topicalization (V2T) in exchange for Clitic Left Dislocation (CLLD) [6]. As strategies for marking topicality, Modern European Portuguese showcases CLLD and Hanging Topic Left Dislocation (HTLD) but also remnants of V2T [6]. In his paper, de Andrade tests out three scenarios: one in which this change happened in two stages (V2T to CLLD; CLLD to TOP); another in which the transformation from CLLD into TOP happens due to information-structural shifts; and another in which V2T changes into TOP due to syntactic changes [6].

At first, the author attests that late stages of Classical Portuguese had no preference for the use of CLLD, diverging from other Romance languages and therefore denying the first scenario [6]. In sequence, the second scenario is contradicted by pragmatic criteria, which show a “stability of the correlation between discourse function and syntactic form during time” [6]. Thus, the third hypothesis is confirmed as the last remaining possibility is the change from V2T into TOP and (Modern) Topicalization is mainly understood as the reanalysis of V2 Topicalization [6]. Additionally, as it was proved that EP diverges from his group of languages in the matter of TOP, the author raises the question of why this construction is not found in other Romance languages that opt for the CLLD strategy [6].

4.2 On focus structures

In what concerns focus structures regarding Portuguese, Kato and Martins have shown that, without clear context, Brazilian Portuguese presents

an ambiguity between informational or contrastive readings [3]. In European Portuguese, however, this ambiguity is solved by prosody: when stress falls on the preverbal subject, the reading is of contrastivity [3]. The authors also state that a VS order is obligatory for focus movement in EP when either a negative element is part of the focus or the fronted constituents include a focus marker (like *only* or *even*) [3].

For Spanish, Jiménez-Fernández has researched the microvariation of subject position and focus structures in two of its varieties: Standard (SS) and Southern-Peninsular (SPS) Spanish. Initially, he attempts to draw a typology of focus, which “has been shown to be best understood as a bundle of discourse features which describe the distinct types” [7]. The author classifies focus into four sub-types: information (IF), contrastive (CF), mirative (MF), and quantifier fronting (QF) [7]. Besides the first two which are known to us, mirative focus is explained as a type of focus that presents unexpected information, whereas quantifier fronting presents information that also carries the feature of polarity [7].

Another novelty of Jiménez-Fernández work is the perspective of resumptive proposing (RP) as an aboutness topic instead of a type of focus fronting [7], as it is usually regarded in the literature. The author claims that, in all varieties of Spanish studied, resumptive proposing showcased a different type of verb movement in comparison with the already mentioned focus structures [7].

Ultimately, Jiménez-Fernández carries on to show the microparametric variation displayed in different Spanish varieties. Diverging from other varieties, SPS doesn't require the verb to be next to the focus constituent in focus fronting, nor does it reserve its left periphery to contrastive focus, also accepting information focus in this position [7]. Additionally, in an experiment with native speakers in which the grammaticality of fronted focus constructions was evaluated, it was proved that SPS speakers regarded preverbal subject constructions with CF, MF, and QF as fully grammatical, while being less receptive to preverbal subjects in RP [7].

5. The V2 controversy

To better understand the controversy surrounding Ibero-Romance languages and a V2-system, we turn to Cruschina and Sailor, who, in their quest to settle the terminological issue regarding residual V2, help elucidate this matter.

In their paper, the authors claim that an idealized V2-system is characterized by the movement of the finite verb to a higher clausal position (either in main clauses, in asymmetric V2 languages, or in all finite clauses, in symmetric V2 languages) [8]. This system is attested in all Germanic languages [8], which prompts multiple cross-linguistic comparisons between this group and that of Romance.

In what concerns Romance languages, there has been an ongoing debate regarding the possibility of

an early V2 system. The hypothesis is that V2 served as a transitional stage between the general SOV order of Classical Latin and the standard SVO of Modern Romance, a theory that has precedents across Romance but is especially disputed for Old Ibero-Romance languages [8].

In this sense, the debacle regarding Ibero-Romance's V2 nature can be understood through clitic placement: one of the arguments for an early stage of V2 has to do with the presence of enclisis with finite verbs in Western Peninsular Ibero-Romance languages, such as Portuguese, Galician, and Asturian [8]. However, it is possible to argue against this “proof” by claiming that these languages only preserve a part of V2 syntax, i.e. verb movement past the pronoun, but not the movement of a constituent to its specifier position; aside from that, it is well known that enclisis isn't dependable of a V2 order [8].

Amid this controversy, Wolfe's work explores the status of V2 for Old Spanish, which was challenged based on the criticism regarding Old Portuguese. The author not only analyzes the possibility of Spanish as a V2-system, but also as a symmetrical one. In his paper, he finds “compelling evidence in favour of analysing the system instantiated in the text as V2”, although it is “shown that a symmetrical analysis is not motivated by the data” [9]. In other words, he finds that Old Spanish is, in fact, a V2 system, although not a symmetrical one. As evidence, he showcases the placement of the finite verb in matrix clauses, similar to proven V2 languages such as Germanic and Rhaeto-Romance [9]. Additionally, he finds that, in the samples collected, verb-second allows a wide range of phrasal constituents in the first position of the clause, substantiating the claim of a V2 system [9]. Regarding the presence of V1 and V3 orders, Wolfe affirms that they are not expressive and in fact “allow us to refine exactly what kind of v2 system is instantiated within the text” [9], improving the definition of a V2-system. In the end, Old Spanish is classified as an asymmetrical V2 language, alike with other Romance languages [9].

For their part, de Andrade and Galves' work focuses on the nature of V2 in Old and Classical Portuguese (OP and CIP, respectively). More precisely, they argue that contrast is a relevant feature for the description of the V2 status of these languages [10]. In their research, they find that OP and CIP share similar features regarding a KP projection but differ regarding other grammatical aspects, such as the usage of demonstratives and boundedness [10]. Ultimately, OP is considered a V2 symmetrical language and CIP is an asymmetrical one [10]. The authors also argue that, in the change from CIP to Medieval Portuguese, the KP projection was lost, which implied the loss of V2 [10].

6. Discussion and Conclusion

The interface between information structure and syntax in Ibero-Romance languages is a prolific field of study, as this review of literature has shown.

Among the current topics that have interested the academic community in the past decade, we can mention word order inversion, focus and topic structures, as well as the specific position of the left-periphery of the sentence. In the Ibero-Romance languages, these topics have been studied from a singular perspective to a cross-linguistic analysis contemplating multiple languages within this subgroup. What this paper has also demonstrated is that this is by no means a closed research field. In fact, several questions arise from the presented research.

As the literature has shown, European Portuguese seems to stand out as a particular variation in the Portuguese language and even within the Ibero-Romance subgroup [3] [6]. Brazilian Portuguese, however, doesn't seem to share the same divergences [3], sometimes approaching other Romance Languages, other times presenting its singularities. It makes us wonder, then, if Brazilian Portuguese is following the path of other Ibero-Romance languages, or if it is creating its particularities instead.

Another question that remains regards the relationship between Ibero-Romance and V2. It seems that, in the past 10 years, researchers have developed a tentative agreement about certain aspects of this debate, such as the presence of a transitional V2 stage in Medieval Romance and vestiges of a V2 order in Ibero-Romance syntax. Notwithstanding, the controversies haven't been fully solved. As Wolfe has pointed out, a better understanding of the V2 phenomena is needed [9]. In this sense, a cohesive cross-linguistic account has to be executed to account for the idiosyncrasies of the V2 system. More restrictively, the varieties of Medieval Ibero-Romance have to be further researched, as scholars have yet to agree on the nature of its verb position.

Furthermore, recent findings regarding specific Ibero-Romance languages urge for the reconsideration of previously cemented notions in the literature, such as V2 topicalization in EP [6], which calls for new research to contemplate this phenomenon in other Ibero-Romance (and even other Romance) languages; and the status of Resumptive Proposing in Spanish [7], which calls for further researches to put this claim to test under a cross-linguistic perspective. These are just a few of the gaps still open in the chosen field of interest.

7. Acknowledgement

The author would like to thank the INCBAC Institute and the Unigou Remote Program for providing the Scientific Training to further her understanding of academic research which allowed her to write this article. The author would also like to acknowledge the Brazilian National Council for Scientific and Technological Development, CNPq, for funding her undergraduate research, and her supervisor, Aroldo Leal de Andrade.

8. References

- [1] Cruschina, S. Focus and Focus Structures in the Romance Language. In: Aronoff, M, editor. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press; 2022. p. 1-24. Available from: <https://helda.helsinki.fi/server/api/core/bitstreams/bfcd838b-ed7f-40e1-991f-a81b31e4b60f/content>.
- [2] Cruschina, S. Topicalization in the Romance Languages. In: Aronoff, M., editor. *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press; 2021. p. 1-24. Available from: <https://helda.helsinki.fi/server/api/core/bitstreams/71b94f32-d1e2-4ed1-aa2d-fac4d136493c/content>.
- [3] Kato, MA, Martins, AM. The Main Varieties of Portuguese: an overview on word order. In: Wetzels, L., Menuzzi, S., Costa, J., editors. *The Handbook of Portuguese Linguistics*. Hoboken (NJ): Wiley-Blackwell; 2016. p. 15-40. Available from: <https://core.ac.uk/download/147885142.pdf>.
- [4] Corr, A. Wide-focus subject-verb inversion in Ibero-Romance: A locative account. *Glossa* [Internet]. 2016 Jun 29 [cited 2024 Mar 10];1(11):1-33. Available from: <https://www.glossa-journal.org/article/id/4822/>. doi: <https://doi.org/10.5334/gjgl.85>.
- [5] Martín, J. Spanish preverbal subjects and the cartography of the Spanish left periphery. 2022 Mar 4 [cited 2024 Mar 10];270:103197-103233. Available from: <https://www.sciencedirect.com/science/article/abs/pii/S0024384121001698?via%3Dihub>. doi: <https://doi.org/10.1016/j.lingua.2021.103197>.
- [6] de Andrade, AL. On the Emergence of Topicalization in Modern European Portuguese: A Study at the Syntax-Information Structure Interface. *Estudos Linguísticos* [Internet]. 2015. [cited 2024 Mar 10];11:13-34. Available from: <https://encr.pw/wrNlq>.
- [7] Jiménez-Fernández, ÁL. Towards a typology of focus: Subject position and microvariation at the discourse-syntax interface. *Ampersand* [Internet]. 2015 [cited 2024 Mar 10];2:49-60. Available from: <https://www.sciencedirect.com/science/article/pii/S2215039015000077>. doi: <https://doi.org/10.1016/j.amper.2015.03.001>.
- [8] Cruschina, S., Sailor, C. What is "residual verb second"? And what does Romance have to do with it? *Isogloss* [Internet]. 2022 Aug 31 [cited 2024 Mar 10];8(3):1-28. Available from: <https://revistes.uab.cat/isogloss/article/view/v8-n3-cruschina-sailor>. doi: <https://doi.org/10.5565/rev/isogloss.235>.

[9] Wolfe, S. The nature of Old Spanish verb second reconsidered. *Lingua* [Internet]. 2015 [cited 2024 Mar 10];164A:132-155. Available from: <https://www.sciencedirect.com/science/article/pii/S0024384115001126>. doi: <https://doi.org/10.1016/j.lingua.2015.06.007>.

[10] de Andrade, A., Galves, C. Contrast and word order: A case study on the history of Portuguese. *Glossa* [Internet]. 2019 Sep 11 [cited 2024 Mar 10];4(107):1-29. Available from: <https://www.glossa-journal.org/article/id/5214/>.