



de Rocafiguera, Núria & Bel, Aurora. 2025. The interplay between syntax and pragmatics on pronoun resolution in Catalan. *Glossa: a journal of general linguistics* 10(1). pp. 1–37. DOI: <https://doi.org/10.16995/glossa.16989>



The interplay between syntax and pragmatics on pronoun resolution in Catalan

Núria de Rocafiguera, Universitat Pompeu Fabra, ES, nuria.rocafiguera@upf.edu

Aurora Bel, Universitat Pompeu Fabra, ES, aurora.bel@upf.edu

The aim of the present study is to disentangle the role of syntactic and pragmatic factors on pronominal subject anaphora resolution in a null subject language, Catalan. Through the manipulation of information structure, we assess how and to what extent the syntactic function (subject vs. object), the information status (focus vs. topic), and the sequential or hierarchic position of the antecedents (first vs. second position) in the main clause guide the interpretation of ambiguous null and overt subject pronouns in the subordinate clause. The results of a two-alternative forced-choice task reveal a different sensitivity of null and overt pronouns to syntactic and pragmatic cues, in line with multifactorial and form-specific approaches, and no overriding factor is identified. Null pronouns appear highly sensitive to both the syntactic function and the information status of their antecedent, preferring both subjects and topics and not showing a specialization when these syntactic and pragmatic features are misaligned. The interpretation of overt pronouns, on the other hand, is mainly guided by the syntactic function and hierarchical position of the antecedent, preferring coreference with objects and second-mentioned, hierarchically lower positions. Different from null pronouns, overt pronouns appear less sensitive to the information status of their antecedent—showing that they can be interpreted as pragmatically ‘inert’.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2025 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction

As a null subject language, Catalan allows subjects to be omitted through null pronouns and overtly expressed using overt pronouns or lexical subjects. In terms of anaphora resolution (i.e., the process of identifying which expression a pronoun or anaphor refers to in a text), each pronoun has specific interpretive preferences. There is consensus to claim that null pronouns preferentially refer to more prominent antecedents than overt pronouns. These preferences result from a complex interplay of factors from different linguistic and cognitive domains.

In essence, the prominence of an antecedent for a (null or overt) subject pronoun is determined by a combination of linguistic and cognitive factors. These factors include the syntactic position of the antecedent (Carminati 2002), its information status (Kaiser 2011; de la Fuente 2015; Lozano 2016), thematic roles (Schumacher & Dangi & Uzun 2016), coherence relations—e.g., conveyed by causal, consequential, or concessive connectives (Stevenson et al. 2000; Kehler & Rohde 2013; Godoy & Weissheimer & Mafra 2018), implicit causality (Pyykkönen & Järvikivi 2010; Koornneef & Sanders 2013) or order of mention (Gernsbacher & Hargreaves 1988). However, there is no consensus on the extent to which this set of factors contributes to the resolution of null and overt subject pronouns, and it remains to be fully comprehended. Briefly, pronoun subject anaphora resolution remains a complex and multifaceted phenomenon.

While previous research has explored various factors influencing anaphora resolution, such as syntactic and pragmatic factors, a clear understanding of how these factors interact in null subject languages is still lacking. In the following sections, we will attempt to address this gap by synthesizing existing literature and examining through experimental data the interplay between syntactic and pragmatic factors in pronominal anaphora resolution in Catalan, a specific null subject language, which has received limited research attention.

1.1 Syntactic factors on pronominal subject anaphora resolution: the Position of Antecedent Hypothesis

Pronominal subject anaphora resolution has been shown to be highly sensitive to syntactic factors. A very strong preference of personal pronouns for subject antecedents has been widely attested in previous studies in non-null subject languages (Kaiser & Trueswell 2008 for Finnish; Arnold 2010; Kaiser 2011 for English; Bader & Portele 2019 for German). In null subject languages, capitalizing on the Position of Antecedent Hypothesis (PAH) proposed by Carminati (2002) for intra-sentential anaphora in Italian, null and overt pronouns have been claimed to display a division of labor in their preferential interpretation choices. This hypothesis predicts that, in referentially ambiguous contexts, null pronouns have a strong preference for antecedents in Spec,IP (i.e., subject antecedents), whereas overt pronouns show a clear bias for antecedents in lower syntactic positions (i.e., object antecedents). As indicated in (1) by the subindex, when ambiguity is at stake, the null pronoun (*pro*) typically prefers to retrieve the subject antecedent in

the preceding clause and the overt pronoun (*ell/ella*, ‘he/she’), the object antecedent. Although Carminati (2002) proposed the PAH for Italian, these patterns have also been attested for Catalan (Mayol & Clark 2010; Bel & García-Alcaraz 2018) and for other Romance and non-Romance null subject languages (e.g., Papadopoulou et al. 2015 in Greek; Rinke & Flores 2018 in Portuguese; Contemori & Di Domenico 2021 in Spanish; Wolna & Durlík & Wodniecka 2022 in Polish), both at intra- (1–3) and inter-sentential level (4).

- (1) a. La Laura_j va espantar la Maria_k quan *pro*_j va entrar a
 the Laura AUX-PST frighten the Maria when *pro* AUX-PST enter in
 l’ habitació. (Null pronoun)
 the room
- b. La Laura_j va espantar la Maria_k quan ella_k va entrar a
 the Laura AUX-PST frighten the Maria when she AUX-PST enter in
 l’ habitació. (Overt pronoun)
 the room
 ‘Laura scared Maria when she went into the room.’

Concerning anaphora resolution in Catalan, only the two aforementioned studies have previously experimentally examined null and overt pronoun interpretive biases in globally ambiguous contexts within the framework of the PAH. Bel & García-Alcaraz (2018) analyzed semantically ambiguous intra-sentential contexts using an acceptability judgement task and found Catalan to show well-defined PAH-like biases of both null and overt pronouns in subordinate-main clause order, as illustrated in (2), and only a clear bias of overt pronouns towards object antecedents in main-subordinate sequences, as shown in (3). Indeed, the order of clauses and the structural organization of linguistic information (as we will see in the next section) have been shown to introduce nuances in the validity of the PAH, particularly evident in subordinate-main sequences (see Chamorro 2018; de Rocafiguera & Bel 2022 for a discussion).

The second study mentioned on Catalan anaphora resolution, by Mayol & Clark (2010), assessed inter-sentential contexts using a two-alternative forced-choice task and a self-paced reading task. Although these are not the specific contexts for which the PAH was formulated, null pronouns were also found to prefer subject antecedents and overt pronouns to prefer object antecedents, as in (4) and in line with the PAH, expanding its predictions formulated for intra-sentential contexts.

- (2) *Intra-sentential anaphora in subordinate-main clause order*
 Mentre en Cèsar_j desmentia en Joaquim_k, *pro*_j/*ell*_k es va posar vermell.
 while the Cèsar disprove-PST the Joaquim *pro*/*he* REFL AUX-PST put red
 ‘While Cèsar disproved Joaquim, he turned red.’

(Example from Bel & García-Alcaraz 2018: 46)

(3) *Intra-sentential anaphora in main-subordinate clause order*

La Irene_j va saludar la Catalina_k quan *pro*_{j/k}/ella_k va entrar
 the Irene AUX-PST greet the Catalina when *pro*/she AUX-PST enter
 a la botiga.
 in the store.
 'Irene greeted Catalina when she entered the store'

(Example from Bel & García-Alcaraz 2018: 46)

(4) *Inter-sentential anaphora*

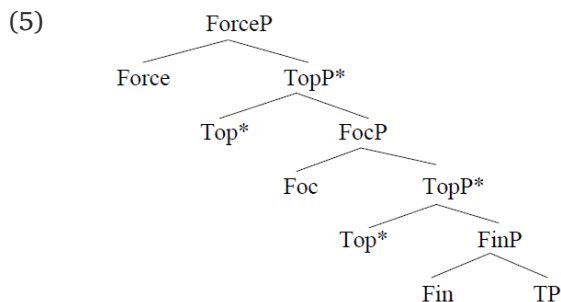
La Marta_j escrivia sovint a la Raquel_k. *pro*_j/Ella_k vivia als Estats Units.
 the Marta write-PST often to the Raquel *pro*/she live-PST in the United States
 'Marta wrote frequently to Raquel. She lived in the United States.'

(Example from Mayol & Clark 2010: 784)

Nevertheless, Carminati (2002) recognizes that pronoun interpretation may also be sensitive to discourse factors such as the information status of the antecedents, which is examined in the next section.

1.2 Pragmatic factors on pronominal subject anaphora resolution: the information status of the antecedent

Assuming Rizzi's (1997) articulated structure of the expanded left periphery shown in (5) (from Fábregas 2016: 4), Carminati (2002: 184) further hypothesizes that referents introduced in positions higher than Spec,IP, such as topic positions (i.e., Spec,TopP), i.e. the position that typically host topics such as clitic-left dislocations (CLLD) in Romance languages, may be similarly preferred for null pronouns as antecedents in Spec,IP. She speculates that topicalized referents would compete with subject antecedents without overriding the preference of null pronouns for antecedents in Spec,IP. Focalized referents (i.e., in Spec,FocP) such as fronted focus, on the other hand, would not compete with subject referents. She attributes this asymmetry between topicalized and focalized referents to discourse factors. Null pronouns typically retrieve old referents in discourse, associated with topicalized antecedents, whereas focalized referents often introduce new information into the discourse.



More recently, considering the different position of subjects and topics, Torregrossa & Andreou & Bongartz (2020) proposed a reformulation of the PAH, defining the relative prominence of a constituent based on its hierarchical height. Their proposal is grounded on Rizzi (2018)'s claim that null subjects in Italian are mainly sensitive to "aboutness", a property shared by both subject and topic positions, rather than to "subjecthood". More specifically, Torregrossa et al. (2020: 9) propose the principles in (6) to account for the interpretation of null pronouns.

- (6)
- a. A null subject is expected to have the referent of a prominent DP.
 - b. A DP is more prominent than another DP if the former is hierarchically higher than the latter.
 - c. Prominence of a DP depends on other factors beyond syntax (e.g., verb-type, coherence relations, discourse topicality, prosody, etc.).

In this proposal, the hierarchical position of an antecedent is granted special importance in terms of making an antecedent prominent, or more preferred, for subsequent null pronouns. As displayed in (6b), the preferred antecedent of a null pronoun will be the one appearing in a hierarchically higher position, typically occupied by subjects or topics. In contrast to Carminati (2002), the authors predict that null pronouns may not similarly prefer subject antecedents in Spec,IP and antecedents in Spec,TopP. Instead, the preference for the hierarchically higher position may prevail (i.e., Spec,TopP). In this regard, they suggest that "the greater the difference between constituents in terms of hierarchical height, the more evident this bias is [the bias of null pronouns toward the hierarchically higher antecedent]" (Torregrossa et al. 2020: 9). As such, the bias of null pronouns towards subject antecedents is argued to be more pronounced when the object occurs in situ than when it is, for instance, left-dislocated. When the object occurs in situ there is greater distance in terms of hierarchical height between Spec,IP (i.e., the subject) and the object constituent. However, these observations were not empirically tested in their study. Finally, in (6c), they acknowledge the need of adopting a multifactorial approach to account for the possibility for null subjects to corefer with object antecedents, given that pronominal preferences are not absolute. However, they lack sufficient evidence to define how, or to what extent, these multiple factors interact (Torregrossa et al. 2020: 22).

Further research and experimentation are needed to address this gap. In the current study, we will use clitic-left dislocations as a topicalization device and it-cleft constructions as a focalization device to alter the syntactic and pragmatic positions within the phrase structure to serve as potential loci for hosting antecedents for null and overt subject pronouns.

From a broader perspective within discourse pragmatics the notions of topicality and givenness have often been intuitively related to prominence (see Cowles & Walenski & Kluender 2007; Kaiser 2011; von Heusinger & Schumacher 2019). Topical or given information has been considered to be more prominent and thus more likely to be pronominalized than new information (Chafe 1976; Prince 1981; Givón 1983). In agreement with Torregrossa et

al. (2020)’s proposal (see above), most experimental studies demonstrate that antecedents in topic position tend to be good candidates for subject pronouns in reference assignment processes (e.g., Kaiser 2011 for English; Colonna & Schimke & Hemforth 2012 for German; de la Fuente 2015 for Spanish). In these studies, the preference for topic antecedents does not seem to override the preference for subjects, only to compete with it, as already hypothesized by Carminati (2002).

Apart from the pragmatic features carried by antecedents, null or overt subject pronouns themselves convey a discourse pragmatic meaning. In null subject languages, null pronouns have traditionally been assumed to be specialized in conveying topic maintenance (thus preferring to corefer with topical antecedents) and overt pronouns in conveying topic shift (thus preferring to corefer with non-topical antecedents) (e.g., Lozano 2009; Sorace et al. 2009; García-Alcaraz & Bel 2019). This has been mainly supported by the analysis of (semi)spontaneous production. However, as demonstrated in several corpora studies, null pronouns can also be felicitously used to express topic shift (e.g., Lubbers-Quesada & Blackwell 2009; Lozano 2016; García-Alcaraz & Bel 2019; Giannakou & Sitaridou 2022) (see also Leonetti 2021). Bel & Perera & Salas (2010) examined anaphoric third person subject pronouns in oral and written narratives in Catalan. Their findings showed that null pronouns mainly expressed topic maintenance and preferred to refer to antecedents in subject position, whereas the scarce number of overt pronouns did not show a clear coreference pattern.

Languages employ topicalization mechanisms that are crucial for our analysis. Clitic-left dislocation (CLLD), in particular, serves as a topicalization device (as opposed to focalization). At this point, it is relevant to cite Mayol (2010b) since it empirically examined the role of word order and topicalization via CLLD in the resolution patterns of both null and overt pronouns. As shown in the examples below (from Mayol 2010b: 129), she contrasted SVO unmarked sentences (7a) and OVS CLLD sentences (7b) in a two-alternative forced-choice task.

- (7) a. A: Què li va passar a la Marta?
 what CL-DAT AUX-PST happen to the Marta
 ‘What happened to Marta?’
- B: La Marta escrivia sovint a la Raquel. *pro*/Ella vivia als
 the Marta write-PST often to the Raquel *pro*/she live-PST in the
 Estats Units.
 United States
 ‘Marta wrote frequently to Raquel. (She) lived in the United States.’
- b. A: Què li va passar a la Raquel?
 what CL-DAT AUX-PST happen to the Raquel
 ‘What happened to Raquel?’

B: A la Raquel, l' escrivia sovint la Marta. *pro/Ella* vivia als
 to the Raquel OBJ write-PST often the Marta *pro/she* live-PST in the
 Estats Units.
 United States
 'To Raquel, Marta wrote (to her) frequently. (She) lived in the United States.'

Based on her results, Mayol (2010b) argues that null pronouns are sensitive to syntactic factors but not to the manipulation of information status or word order. Topicalizing an object antecedent via CLLD did not affect the preference of null pronouns for subject antecedents. On the other hand, overt pronouns showed sensitivity to both syntactic and pragmatic factors, preferring object antecedents only in the unmarked SVO condition (non-subject, non-topic constituents¹), and not showing a clear preference in the CLLD OVS condition. It is worth noting that Mayol (2010b)'s results on null pronouns do not support the predictions of Carminati (2002), who hypothesized that null pronouns should be sensitive to both subjecthood and topicality cues. However, de la Fuente (2015), who similarly analyzed the impact of topicalization via hanging topic left dislocations (HTLD) on the preferences of null pronouns in Spanish, found a significant increase in the preference of null pronouns for topicalized object antecedents.

Another linguistic device that can influence accessibility is focalization. In this context, clefting is a focalization mechanism that manipulates information structure. Research on anaphora resolution in marked information structures, such as it-clefted clauses, shows mixed results. Psycholinguistic research has found that linguistic focus increases the accessibility of focused entities (Foraker & McElree 2007; Káldi & Babarczy 2021). However, it is not always straightforward that focused entities are chosen as the preferred antecedent to interpret a subject pronoun (Blything et al. 2021). In fact, focusing it-cleft structures have also been found to make an antecedent less accessible for an anaphoric subject pronoun, leading to an "anti-focus effect" (Colonna et al. 2012; Colonna & Schimke & Hemforth 2015; de la Fuente 2015; Patterson & Esaulova & Felser 2017; Patterson & Felser 2020).

No previous studies have addressed the impact of it-cleft structures on anaphora resolution in Catalan, but de la Fuente (2015) examined the effects of both subject and object focalization via it-cleft structures in Spanish. His results revealed that null pronouns consistently showed a higher preference for the non-clefted antecedent, conveying presupposed, topic-like, information. Similar to Colonna et al. (2012; 2015) for French, de la Fuente (2015) concluded that focusing an antecedent in an intra-sentential (but not in an inter-sentential) context leads to an anti-focus effect. His design, however, did not include overt pronouns, and to our knowledge, there has not been any other research examining the effects of information status on the interpretive preferences of subject pronouns in a null subject language. The present study will expand the findings of

¹ Using Vallduví (1992)'s terms, Mayol (2010b) does not refer to *topics* but to *links*.

Mayol (2010b) in Catalan and de la Fuente (2015) in Spanish by exhaustively investigating the interpretation of both null and overt pronouns in both topicalization and focusing structures (clitic-left dislocation and it-clefts).

1.3 Sequential factors on pronominal subject anaphora resolution: the order of mention of the antecedents

A consequence of manipulating the information status of the antecedent using syntactically marked structures is that word order, or the sequential position of the antecedents, is also altered. In this regard, the initial position has been argued to be a privileged position for an antecedent to be picked up by a subsequent pronoun, leading to a first-mention advantage (Gernsbacher & Hargreaves 1988) attributed to general cognitive processes (see also Carreiras & Gernsbacher & Villa 1995). However, subsequent studies assessing this first-mention advantage, such as Fukumura & van Gompel (2015) in an eye-tracking study on English subject pronouns, have demonstrated that the preference of pronominal subjects for subject antecedents is stronger than the preference for first-mentioned antecedents (Cowles et al. 2007; Kaiser & Trueswell 2008; see also Bader & Portele 2019). On the other hand, Järvikivi et al. (2005) showed no first-mention advantage but claimed that both syntactic function and order of mention play a role in the interpretation of personal pronouns in Finnish, and d-pronouns in German were also claimed to be sensitive to order of mention, different from personal pronouns (Bader & Portele 2019).

In the present study, we will interpret a preference for first-mentioned antecedents as a preference for those in the highest structural position, as suggested by Carminati (2002) or Torregrossa et al. (2020). This proposal implies a prevalence of syntactic/structural factors over sequential factors, although they overlap. We will be studying how the information status of the antecedent interacts with its syntactic position, being aware that manipulating information structure to alter the pragmatic status of referring expressions also change their surface position.

1.4 Form-specific approaches to pronominal subject anaphora resolution

As already described, in previous studies neither syntactic nor pragmatic factors have been found to act as an overarching factor governing anaphora resolution. At the same time, it is not clear that they are similarly relevant. Kaiser & Trueswell (2008) propose a form-specific multiple-constraints approach to account for pronominal anaphora resolution, suggesting that these factors may be prioritized differently or carry different weights in shaping coreference preferences depending on the specific pronoun form involved. More specifically, Kaiser & Trueswell analyzed the interpretive preferences of Finnish personal and demonstrative subject pronouns and found the two forms to have different degrees of sensitivity to syntax and information structure: whereas personal pronouns were mainly sensitive to the syntactic function of the antecedent (preferring subjects), the preference of demonstrative pronouns were guided by a combination

of information status and syntactic function (preferring discourse-new and object referents). Similarly, Bader & Portele (2019) identified German personal pronouns to be mainly sensitive to syntactic cues and demonstrative pronouns to be sensitive to a combination of linear position and topicality. In the case of Catalan, Mayol (2010b) also suggested that null pronouns may have a simple preference for previous subjects and overt pronouns be sensitive to both syntactic and pragmatic factors (see also Fedele & Kaiser 2014 for the different weight of syntax, semantics and sentence boundaries in the interpretation of null and overt pronouns in Italian).

Still other studies on non-null subject languages have confronted factors from different linguistic domains, focusing solely on personal pronouns. In a study on English, Kaiser (2011) found subjecthood (and agentivity) to have particularly strong effects on pronoun interpretation, regardless of information structure. These pronouns were sensitive to both topicalization and focusing effects, which boosted the preference for topicalized and focused referents. However, these preferences did not overcome the preference for subjects. Similarly, Blything et al. (2021) found subjecthood/agentivity/first-mention cues to have a stronger role in the offline interpretation of English subject pronouns in focus position (via prosodic marking and it-clefts). On the other hand, Colonna et al. (2012) found that the preference of personal pronouns in German toward subject antecedents was only attested when the subject coincided with the sentence topic. When the object was topicalized, personal pronouns remained unbiased, indicating a similar weight of syntactic function and information structure. Finally, Schumacher & Roberts & Järvikivi (2017) argued that thematic role is a more powerful cue in pronoun resolution in German than grammatical function and linear position (personal pronouns preferring proto-agents and d-pronouns preferring proto-patients). However, these preferences guided by thematic role became weakened when these cues were misaligned with grammatical function and word order cues (i.e., proto-agents not being marked with nominative case, and proto-patients not being postverbal or second-mentioned).

In summary, these studies support a multifactorial and form-specific approach, indicating that there are multiple factors involved in the anaphoric interpretation of a particular pronominal form and that they may vary compared to another form. In the present study, we will consider two pronominal forms (null and overt pronouns) and their potentially different interactions with syntactic and pragmatic factors.

2 The present study

2.1 Main aim and rationale

The main interest of the present research is to elucidate how syntactic and pragmatic factors interact on pronominal anaphora resolution in null subject languages by using syntactically and pragmatically marked information structures. As a novelty, this approach thus puts together these different dimensions in a single study and it does so in a language that has received little

attention. Specifically, we aim at disentangling the contribution in the interpretation of null and overt pronouns of three factors: 1) the syntactic function of the antecedent (subject vs. object), 2) the information status of the antecedent (topical vs. non-topical), and 3) the hierarchical position of the antecedent (higher vs. lower position in the syntactic structure; also interpreted as first-mentioned vs. second-mentioned in alternative perspectives). These factors overlap in unmarked structures, but, crucially, they do not in marked CLLDs and it-clefts. With this purpose, we will test the Position of Antecedent Hypothesis (PAH; Carminati 2002) in non-canonical sentences where the information status of antecedents is manipulated, thereby altering both their linear and hierarchical positions. More precisely, we will contrast the interpretive preferences of anaphoric null and overt subject pronouns in the following four structures:

(8.1) *Unmarked structures*

La Laura va interrompre la Maria quan *pro/ella* va començar a
 the Laura AUX-PST interrupt the Maria when *pro/she* AUX-PST start to
 parlar.
 speak
 ‘Laura interrupted Maria when she started speaking.’

(8.2) *Clitic-left dislocation structures with topicalized object antecedents*

A la Maria la va interrompre la Laura quan *pro/ella* va
 to the Maria CL-OBJ AUX-PST interrupt the Laura when *pro/she* AUX-PST
 començar a parlar.
 start to speak
 ‘Maria, Laura interrupted her when she started speaking.’

(8.3) *It-cleft structures with focused subject antecedents*

Va ser la Laura qui va interrompre la Maria quan *pro/ella*
 AUX-PST be the Laura who AUX-PST interrupt the Maria when *pro/she*
 va començar a parlar.
 AUX-PST start to speak
 ‘It was Laura who Maria interrupted when she started speaking.’

(8.4) *It-cleft structures with focused object antecedents*

Va ser a la Maria a qui va interrompre la Laura quan *pro/ella*
 AUX-PST be to the Maria to who AUX-PST interrupt the Laura when *pro/she*
 va començar a parlar.
 AUX-PST start to speak
 ‘It was Maria whom Laura interrupted when she started speaking.’

As already mentioned, the notions of subject, topic, and higher position in the syntactical hierarchy (or initial position) overlap in unmarked structures (8.1). Clitic-left dislocation (CLLD)

and it-cleft structures will allow us to disentangle the role of these syntactic and pragmatic factors in the process of solving pronominal anaphoric dependencies, as detailed below.

Firstly, in CLLD structures (8.2), object antecedents are topicalized and appear first-mentioned (OVS word order), in the hierarchically highest position of the syntactic structure (Spec,TopP). Subject antecedents convey focal (or non-topical) information² and appear second-mentioned, in a postverbal position, and in a hierarchically lower position than objects (i.e., either in Spec,IP or in situ in a VP-internal position)³.

Secondly, in subject it-cleft structures (8.3), subject antecedents are focused and appear first-mentioned (SVO word order), in the hierarchically highest position of the syntactic structure (Spec,FocP). Object antecedents convey presupposed (topical) information⁴ and appear second-mentioned, in a hierarchically lower position than subjects.

Finally, in object it-cleft structures (8.4), object antecedents are focused and appear first-mentioned (OVS word order), in the hierarchically highest position in the syntactic structure (Spec,FocP). Subject antecedents convey presupposed (topical) information⁵ and appear second-mentioned, in a postverbal position, and in a hierarchically lower position than objects.

As reviewed in the introduction, recent proposals, like those by Torregrossa et al. (2020), suggest a need to reevaluate the framework grounded on the influential PAH reconsidering the relative prominence of antecedents in terms of hierarchical height. While previous research has highlighted the significance of information status and word order in anaphora resolution, there are discrepancies regarding the importance of syntactic function versus information structure and sequential word order. We identify gaps in the literature concerning the interaction between these levels, particularly in cases where these cues are misaligned. The current study aims to explore how these factors interact in non-canonical sentence structures. By analyzing experimental data, manipulating the information status of the antecedents, and analyzing their linear and hierarchical positions, we aim to elucidate the mechanisms governing pronominal anaphora resolution in Catalan. By doing so, we seek to contribute to a deeper understanding of language interpretation mechanisms in null subject languages and inform future research in this area.

² According to some authors, postverbal subject antecedents in CLLD structures are syntactically marked as conveying focus (e.g., Vallduví 1992; see also Domínguez 2018).

³ Depending on the analysis, a postverbal subject appears in Spec,IP/Spec,TP (cf. Olarrea 1998) or remains in its VP-internal position (cf. Gutiérrez-Bravo 2003). In any case, it appears in a postverbal position.

⁴ Given that presupposed elements express given information, they can be considered to convey topical information, as opposed to “new” or focal information.

⁵ See the previous note.

2.2 Research questions and hypotheses

The research questions and hypotheses that will be tested in the present study are the following:

- RQ1. Does the syntactic function of the antecedent affect the interpretation of null and overt subject pronouns in unmarked sentences in Catalan?

In the light of the studies that have previously addressed these contexts (Mayol & Clark 2010; Bel & García-Alcaraz 2018), we expect to attest the biases predicted by the Position of the Antecedent Hypothesis (PAH; Carminati 2002): null pronouns should preferably corefer with subject antecedents (topical, in first position) and overt pronouns with object antecedents (focal, in second position). Given that consistent evidence has been found in these previous studies on Catalan, we expect to replicate previous findings. However, these previously attested robust biases could be affected by two different factors. First, although previous research has shown that pronominal biases are more pronounced in subordinate-main sentences, we opted for main-subordinate sentences to ensure consistency across the four analyzed structures. Opting for the subordinate-main order would result in ungrammatical structures in the case of it-cleft sentences, as they would not fit within that sequence (see section 3.2 below). Second, we consistently used the conjunction *quan* ('when') to introduce the second subordinate clause, minimizing variability caused by the connector (de Rocafiguera & Bel 2022; Martín-Villena 2023).

- RQ2. How and to what extent do syntactic and pragmatic factors interact in determining null and overt subject pronoun resolution in marked information structures?

As already stated, we do not expect the interpretive preferences of null and overt subject pronouns to be guided by an only factor and different pronominal forms could be differently sensitive to each factor (i.e., syntactic function, information status or hierarchical/linear position; see Kaiser & Trueswell 2008; Bader & Portele 2019; among others). However, there could be an outranking factor having a higher strength in defining the interpretive patterns of null and overt subject pronouns. In this regard, three hypotheses should be put forward:

- (i) The first hypothesis predicts that syntactic factors will outrank pragmatic factors. It is compatible with two different outcomes: (a) if only the syntactic function matters, null pronouns will prefer subject antecedents and overt pronouns object antecedents regardless of the structure of the sentence; (b) if only the syntactic function matters, null pronouns will prefer antecedents in higher syntactic positions and overt pronouns in lower positions, regardless of the specific syntactic function of the antecedent. Within Carminati (2002)'s framework, null pronouns will preferably be interpreted as corefering with antecedents in Spec,IP. In the light of Torregrossa et al. (2020), the preferred position for the antecedent of a null pronoun will be the hierarchically highest position in the syntactic structure (i.e., Spec,IP in unmarked contexts, Spec,TopP in

CLLDs, and Spec,FocP in it-clefts). Complementarily, overt pronouns should prefer antecedents in lower structural positions across contexts. In line with Mayol (2010b), only null pronouns would be mainly sensitive to syntactic factors, systematically preferring subject antecedents.

- (ii) The second hypothesis predicts that pragmatic factors will outrank syntactic factors, and the information status of the antecedent will be crucial for pronoun resolution patterns (e.g., de la Fuente 2015; Papadopoulou et al. 2015). As has been generally assumed across null subject languages (e.g., Lozano 2016; García-Alcaraz & Bel 2019), we expect null pronouns to be more prone to convey topic maintenance and overt pronouns to be more prone to convey topic shift. Another possibility could be to attest a general anti-focus effect affecting it-cleft structures and driving subject pronouns to systematically reject focused antecedents in intra-sentential contexts as the ones in our items (Colonna et al. 2012; 2015; de la Fuente 2015; Patterson et al. 2017).
- (iii) The third hypothesis predicts that there will not be a prevalent factor when syntactic and pragmatic cues are misaligned. As also hypothesized by Carminati (2002), when topicality and subjecthood are not aligned, antecedents in topic position (Spec,TopP) may compete with antecedents in subject position (Spec,IP). In this case, the bias of null pronouns towards subject antecedents would be weakened—although not overridden—by the preference for topical antecedents. If overt pronouns convey a complementary function to that of null pronouns, we expect their bias towards object antecedents to be weakened when the object conveys focal information, which would be in line with Mayol (2010b)’s results on CLLD contexts.

3 Method

3.1 Participants

The participants in the present study are 53 Catalan-dominant speakers (mean age: 22.26, SD: 3.81, range: 18–35) that were born and raised in Catalonia, and that were studying a university-level degree or had already graduated from university. Since there are no monolingual speakers in Catalonia with Catalan as their first and only language, we aimed to be rigorous by selecting participants with the highest degree of exposure and mastery of Catalan (referred to as Catalan-dominant speakers). Catalan speakers are also exposed to Spanish in the society and end up acquiring both languages. To identify the language dominance profile of the participants, they completed a background questionnaire based on the Bilingual Language Profile (BLP; Birdsong & Gertken & Amengual 2012)⁶ before participating in the experiment.

⁶ Some minimal adaptations were made to three questions in the BLP to make them more appropriate to the bilingual context in Catalonia: the one referring to the language of schooling in the language history module, and the statements in the language attitudes module referring to nativeness. This slightly adapted version of the BLP can be found in Appendix A.

All participants declared that they began acquiring Catalan at the same time (50.94%) or earlier (49.06%) than Spanish, most of them were raised in monolingual Catalan-speaking families (73.58%) and they all declared that they use Catalan to talk to their parents and/or siblings. Regarding language use, they reported using Catalan most of their time during the week (mean: 88.34%; SD: 8.21) and showed a much more limited use of Spanish (mean: 13.30%; SD: 6.59) and other languages (mean: 1.83%; SD: 3.47). Concerning self-rated language proficiency, most of them rated it as being 'very good' in all skills of both languages, although they perceived their oral production abilities to be slightly higher in Catalan than in Spanish (mean out of 6 in Catalan: 5.91; SD: 0.35; mean out of 6 in Spanish: 4.92; SD: 0.87). None of them rated their abilities in Catalan to be lower than in Spanish. Regarding the global dominance scores provided by the BLP questionnaire, which allows for placing participants within a numerical continuum of language dominance, the participants in the present study received a mean score of -82.59 (SD: 16.01; range: -121.24 to -56.76). The values of the global dominance score in the BLP range from -218 to 218 and are obtained from the subtraction of the score of one language from the score of the other one. In the present study, negative values are indicative of Catalan dominance, and positive values indicative of Spanish dominance. A score close to zero would suggest balanced bilingualism.

3.2 The experimental task

3.2.1 The nature of the task

The task completed by participants was an offline written combined two-alternative forced-choice plus slider task (Figure 1). The participants were presented with sentences containing an ambiguous subject pronoun and were asked to interpret it by choosing between two possible antecedents in a continuous visual scale with no visible numerical values (participants just relied on spatial reasoning). They could choose any point in the scale, included the middle point, which would express ambiguity.

La Laura va esperar la Maria quan va arribar a l'estació.
 ¿Qui va arribar a l'estació?

La Maria
La Laura




Figure 1: Example screen of an item of the task.

The main reason to use this task and a continuous scale was to uncover nuances in interpretive biases that could remain imperceptible in a categorical dichotomous decision (Sorace & Keller 2005; Schütze & Sprouse 2014; Sprouse & Almeida 2017; Langsford et al. 2018). In addition to obtaining a referential choice, we aimed at obtaining a value reporting the strength of this choice (Marty & Chemla & Sprouse 2020). However, as will be shown in Section 3.4, it does not seem that we succeeded in capturing the intended nuances.

3.2.2 Design and materials

Two within-subject independent variables were manipulated in a 4×2 factorial design: 1) information structure of the main clause (unmarked structure, object topicalization via clitic left dislocation, subject focalization via it-cleft, or object focalization via it-cleft), and 2) type of pronoun (null or overt). An example for each condition is shown in **Table 1**. Overall, the task included 48 critical items ($k = 6$) and 72 non-critical items. The experimental items were lexically matched in all their versions (i.e., they used exactly the same words and only differed in the information structure of the main clause and the subject pronoun form in the subordinate clause). The complete list of experimental items in the task, along with the instructions given to the participants and a description of the non-critical items, can be found in Appendix B.

The properties of the two potential antecedents in each context could be summarized as shown in **Table 2**.

The experimental items were globally ambiguous sentences that consisted of two clauses, as shown in (9), an example for the baseline condition. The first clause was a main clause in which two characters of the same gender—feminine for half of the items, masculine for the other half—were introduced using proper names in subject and object position. The second clause was a temporal subordinate clause containing an ambiguous pronoun (null or overt) in subject position, followed by a verb and a prepositional complement. Depending on the condition, the information structure of the main clause and the type of subject pronoun in the subordinate clause were manipulated. Each sentence was followed by a question asking for an interpretation of the ambiguous subject pronoun (null or overt). Participants had to answer by choosing between the subject (e.g., *Laura* in (9)) or the object antecedent (e.g., *Maria* in (9)) in the main clause.

- (9) La Laura va esperar la Maria quan *pro/ella* va arribar a
 the Laura AUX-PST wait the Maria when pro/she AUX-PST arrive to
 l' estació. ¿Qui va arribar l' estació?
 the station who AUX-PST arrive the station
 'Laura waited for Maria when she arrived at the station. Who arrived at the station?'

Information Structure	Pronoun	Example
Unmarked structure	Null	La Laura va esperar la Maria quan va arribar a l'estació. 'Laura waited for Maria when (she) arrived at the station.'
	Overt	La Laura va esperar la Maria quan ella va arribar a l'estació. 'Laura waited for Maria when she arrived at the station.'
Dislocated object (CLLD)	Null	A la Maria la va esperar la Laura quan va arribar a l'estació. 'Maria, Laura waited for her when (she) arrived at the station.'
	Overt	A la Maria la va esperar la Laura quan ella va arribar a l'estació. 'Maria, Laura waited for her when she arrived at the station.'
Focused subject (subject cleft)	Null	Va ser la Laura qui va esperar la Maria quan va arribar a l'estació. 'It was Laura who waited for Maria when (she) arrived at the station.'
	Overt	Va ser la Laura qui va esperar la Maria quan ella va arribar a l'estació. 'It was Laura who waited for Maria when she arrived at the station.'
Focused object (object cleft)	Null	Va ser a la Maria a qui va esperar la Laura quan va arribar a l'estació. 'It was Maria whom Laura waited for when (she) arrived at the station.'
	Overt	Va ser a la Maria a qui va esperar la Laura quan ella va arribar a l'estació. 'It was Maria whom Laura waited for when she arrived at the station.'
Question:		Qui va arribar a l'estació? 'Who arrived at the station?'
Possible answers:		Laura / Maria

Table 1: Experimental items per condition.

Information Structure	Antecedent
Unmarked SVO structure	Subject: + topical, + high/first-mentioned Object: -topical, -high/first-mentioned
Dislocated object (CLLD)	Subject: -topical, -high/first-mentioned Object: + topical, + high/first-mentioned
Focused subject (subject cleft)	Subject: -topical, + high/first-mentioned Object: + topical, -high/first-mentioned
Focused object (object cleft)	Subject: + topical, -high/first-mentioned Object: -topical, + high/first-mentioned

Table 2: Properties of subject and object antecedents in each information structure.

Given that subordinate clauses lack the topic and focus positions in CP, the only suitable order to test the interpretive preferences of null and overt pronouns in syntactically marked information structures is the main–subordinate clause order. This is the reason why we did not use a subordinate–main clause order, which is the context that would be more favorable to make the preferences of the PAH arise (Carminati 2002; Chamorro 2018; de Rocafiguera & Bel 2022).

We considered important to have globally ambiguous sentences to be able to observe the bare referential biases that subject pronouns may inherently have. Ambiguous sentences provide the optimal context for identifying these inherent biases in their purest form: in the absence of other disambiguating cues, the weighted factors emerge clearly. To ensure the ambiguity of the pronominal subjects in the items, we controlled for semantico-pragmatic factors by using verbs with neutral implicit causality biases toward the subject or the object and temporal subordinate clauses. Based on Goikoetxea & Pascual & Acha (2008), who analyzed the implicit causality biases of 100 interpersonal verbs in Spanish, sixteen verbs with a bias to the subject or the object higher than 37.5% or lower than 62.5% were chosen so that the selected verbs remained within a neutral range of neutrality (Järvikivi et al. 2005; García-Alcaraz 2015; Järvikivi & van Gompel & Hyönä 2017).⁷ Each of these verbs was used three times in the main clause of the experimental sentences, in different contexts for each item (the verb in the subordinate clause was always different).

Regarding the conjunction that introduced the adjunct subordinate clause, it was a temporal connective, to avoid the underlying coherence relations signaled by other types of connectives and following previous studies (e.g., Carminati 2002; Tsimpli et al. 2004; de la Fuente 2015; García-Alcaraz 2015; Torregrossa et al. 2020). All the experimental sentences used the same connective—*quan* ('when')—to avoid introducing variability (see Martín-Villena 2023) and to maintain the verbs' grammatical aspect homogeneous across items. They were all in past tense and perfective aspect, both the verbs in the main clause and in the subordinate clause.

At this point, we would like to anticipate some limitations of the design of the present study. First, we would like to acknowledge the potential lack of statistical power given the number of participants and the complexity of the experimental design (Brysbaert 2021). The limitations in the data collection that conditioned the choice of the sample size were related to time, financial resources and the constraints of being part of a PhD thesis, as well as to the lack of an a-priori power analysis (Lakens 2022).

Second, we would like to bring up the fact that we will use non-canonical sentences out of context in our experimental design. We have chosen CLLD and clefting structures because they inherently signal topicality and, hence, testing these structures without context will allow us to examine their intrinsic contribution to topicality. To ensure methodological consistency, we will evaluate both

⁷ We assumed that the effects of implicit causality are similar across languages (Hartshorne et al., 2013), particularly in closely related languages such as Catalan and Spanish.

canonical and non-canonical constructions under similar decontextualized conditions, which will support valid cross-structural comparisons. We acknowledge that examining topicality in these conditions may not fully capture the dynamics of natural discourse, which will be a limitation of our study. Nevertheless, this approach will provide valuable baseline insights into how syntactically (non-)marked topicality influences subject pronoun interpretation, offering a foundation for further exploration in more context-rich settings and complementing broader discourse findings.

3.2.3 Procedure

The task was implemented as a Qualtrics Survey questionnaire and took around 40 minutes to complete. The 120 items were counterbalanced and presented in a pseudorandomized order across eight presentation lists constructed following a Latin Square design. No more than two experimental items were presented in succession (never in the same condition or containing the same neutral implicit causality verb). Only one item was presented at a time, and it was not possible to go back to the previous screen. The number of words referring to subject and object antecedents that appeared in the right and the left side of the slider bar were also counterbalanced.

A break in the middle of the task was included to allow participants to rest if needed. Given that this study is part of a broader project that aimed at assessing bilingualism effects in anaphora resolution and making crosslinguistic comparisons between Catalan and Spanish, all participants also completed an equivalent task in Spanish in a separate session (a minimum of 7 days and a maximum of 1 month separated the two sessions, half of the participants began with the Catalan version and the other half with the Spanish one). Most participants completed the tasks in a controlled lab environment, except from those that we did not have the time to collect before the lockdown during the Covid-19 pandemia ($N = 21$).

3.3 Data preparation

One participant was excluded because s/he consistently and systematically considered all the pronouns across conditions as ambiguous (and did not make an interpretive choice, as indicated in the instructions of the task). No items were excluded.⁸ The Qualtrics Survey interface measured

⁸ Although no items were removed, following an item analysis prior to statistical analyses, we observed a slightly different behavior of the three items using the verb *suportar* ('to stand'). Different from the other verbs, in which the subject antecedent was mapped to an agent-like thematic role and the object antecedent to a patient-like thematic role, *suportar* was the only verb that took an experiencer-stimulus structure (following the terminology in Goikoetxea et al., 2008). We ran supplementary analyses excluding these three items to see whether they affected the general results. However, no substantial differences were attested with the results reported in this paper. In any case, we think it is worth mentioning this observation given that we are not aware of any study that has contrasted the role of agentivity on anaphora resolution in Catalan or in other null subject languages. Schumacher et al. (2017) signaled this factor as a better predictor than subjecthood when investigating the resolution of personal and demonstrative pronouns in German. Further studies could contrast the contribution of subjecthood and agentivity on the resolution of null and overt pronouns.

the time that participants took to give a response for each item in the task, which was used to clean the data. All items that had been answered faster than 3 seconds were excluded, since we considered that this amount of time was not enough to carefully read the sentences and provide an answer accordingly. Overall, the removed observations represented a 3.64% of the initial dataset.

At this point, reference is needed to the responses obtained in the combined two-alternative forced choice plus slider task. As shown in **Figure 2**, participants gave answers that were mainly at the edges of the continuous scale: they expressed an interpretive choice, but the task did not seem to capture a gradation in the strength or certainty of these choices. Given the distribution of the data, the answers were finally reduced to a binary choice (i.e., as subject or object choices), as would have been in a traditional two-alternative forced choice task.⁹

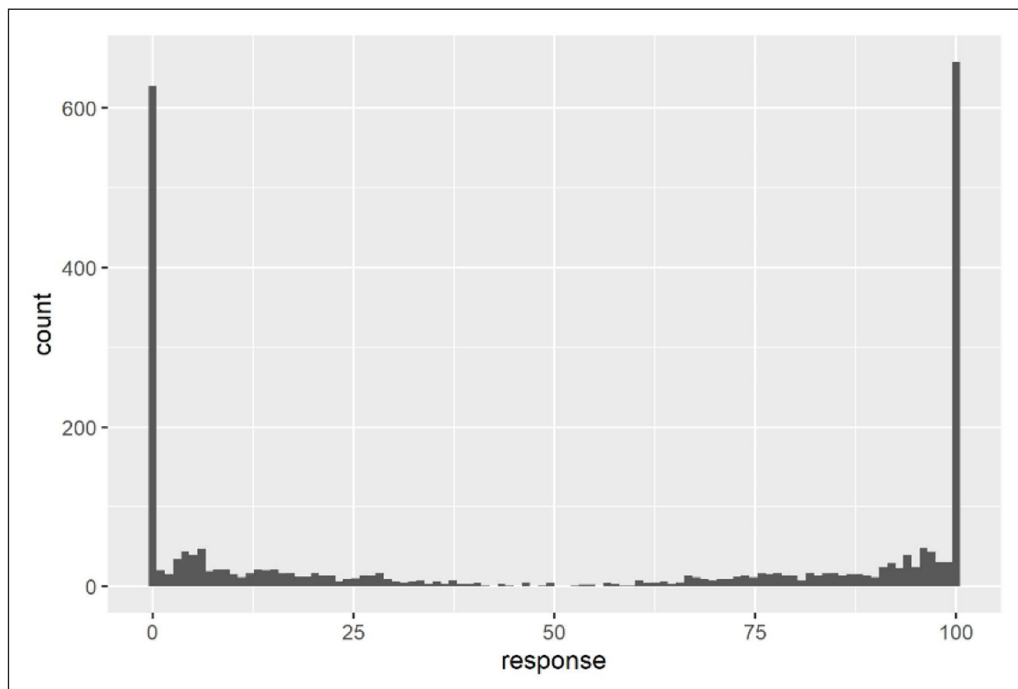


Figure 2: Histogram of the responses showing the preference toward subject or object antecedents in a 0-to-100 visual analogue scale.

3.4 Data analysis

To analyze the results, a mixed-effects logistic regression model was run in R (v. 4.0.5; R Core Team 2021) using the *glmer* function of the *lme4* package (v. 1.1.27; Bates et al. 2015) and the *emmeans* package to obtain pairwise contrasts (v. 1.7.0; Lenth 2021). The dependent variable was

⁹ We analyze the responses as binary, but we also include supplementary analyses with the original responses.

the choice to interpret an ambiguous pronoun as referring back to either a subject or an object. The mixed-effects logistic regression model included Pronoun (null, overt) and Information structure (unmarked structures, topicalized objects via CLLD, focused subjects via it-clefts, and focused objects via it-clefts) as fixed effects, as well as their interaction. As random effects, varying intercepts for participants and items were added to the model and a by-participant varying slope for the effect of Pronoun. Additional random slopes were tested but they led to estimation problems within the models. The final model had the maximal random structure that converged and was the model offering the best fit (we performed a maximum likelihood ratio comparison between the models that converged). The model's total explanatory power was moderate (conditional $R^2 = 0.23$), and the part related to the fixed effects alone (marginal R^2) was 0.08. No multi-collinearity issues were detected (highest VIF = 1.001). This model had a C-index of concordance of 0.76. The dataset and the code used to run the analyses in R are available in the Open Science Framework repository. The coefficient table of the mixed-effects logistic regression model and the table of pairwise contrasts can be found in Appendix C. Supplementary analysis of the responses in their original scale using an ordered beta regression can be found in Appendix D.

4 Results

Table 3 summarizes the proportion of subject interpretations of null and overt pronouns in the four analyzed contexts (unmarked structures, topicalized objects via CLLD, focused subjects via it-cleft, and focused objects via it-cleft). Complementary and clear-cut interpretations of null and overt pronouns seem to be attested only in canonical unmarked sentences, with 62% of null pronouns referring back to subject antecedents and 72% of overt pronouns to object antecedents. However, when the information structure of the antecedents' clause is manipulated, the division of labor between null and overt pronouns seems to be attenuated. This is especially the case of object topicalization and object focusing structures, which present in both cases an OVS word order.

	Unmarked	Topicalized object	Focused subject	Focused object
Null	.62 (.48)	.48 (.50)	.53 (.50)	.65 (.48)
Overt	.28 (.44)	.44 (.50)	.34 (.47)	.54 (.50)

Table 3: Proportion of subject interpretations, by information structure and type of pronoun (SD).

Several likelihood ratio tests were also performed comparing nested models to evaluate the contribution of the predictors and their interactions to the overall fit of the model: both Pronoun and Information structure show a significant main effect in the interpretive choice of

the ambiguous pronoun (respectively: $\chi^2(1) = 24.993, p < .001$; $\chi^2(3) = 46.882, p < .001$), and the two-way Pronoun \times Information structure interaction also makes a significant contribution ($\chi^2(3) = 39.572, p < .001$).

The estimates of the model are graphically illustrated in **Figure 3**, which shows the predicted probability of interpreting null and overt subject pronouns as coreferring with subject antecedents in each of the information structures under analysis.¹⁰ It can be easily observed that null pronouns only show a significant bias towards subject antecedents in canonical unmarked structures and in object clefts, two structures in which the subject antecedent conveys topical information. On the other hand, overt pronouns only show a significant bias towards object antecedents in canonical unmarked structures and in subject clefts, two structures that maintain the SVO word order and in which the object antecedent occupies a hierarchically lower position in the syntactic structure.

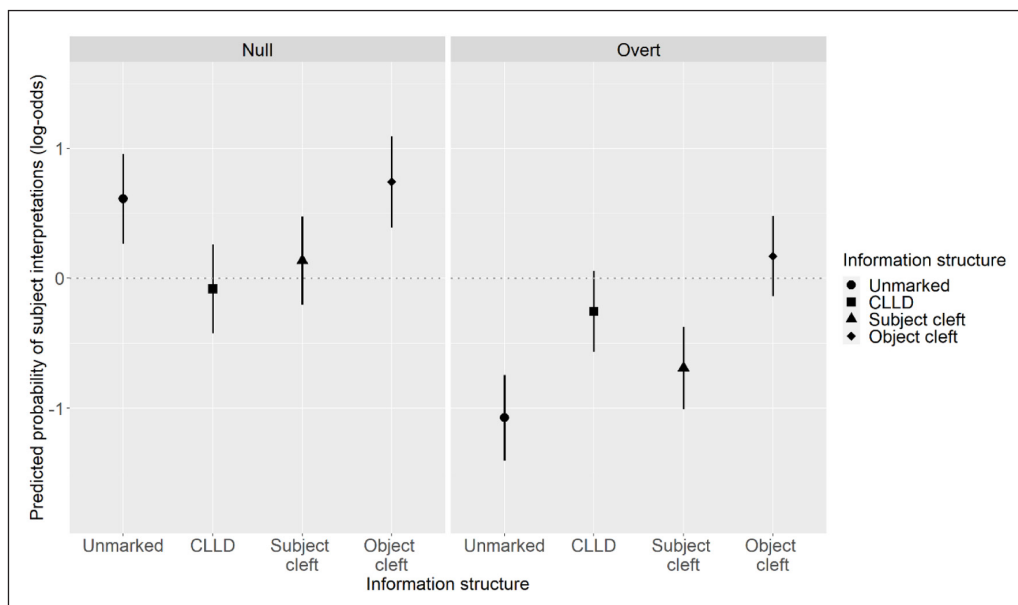


Figure 3: Predicted subject interpretations of null and overt pronouns by information structure (± 95 CI).

In canonical sentences, the results show PAH-like patterns (Carminati 2002): null pronouns are preferably interpreted as coreferring with subject antecedents ($\beta = 0.611, p < .001$), and

¹⁰ Note that results are presented in log odds. A value of 0 corresponds to chance level (or a 50% of probabilities). If the CI bar illustrated in the graph crosses the line representing chance level, it means that the mean bias represented in the graph (with a point shape) is not reliable or significant. If the predicted probability of subject interpretations for a pronoun is a positive value, it means that this the preferred interpretation of the pronoun is to corefer with subject antecedents. Alternatively, if it is a negative value, its preferred interpretation is to corefer with object antecedents.

overt pronouns as coreferring with object antecedents ($\beta = -1.074, p < .001$). In marked information structures, however, these intrinsic clear-cut preferences that null and overt pronouns show in unmarked contexts may disfigure, as observed in **Figure 3**. In CLLDs, neither null nor overt pronouns show any significant preferences in their interpretations and they both remain unbiased (null pronouns: $\beta = -0.082, p = .636$; overt pronouns: $\beta = -0.255, p = .106$). In subject clefts, null pronouns also remain unbiased ($\beta = 0.135, p = .436$), but overt pronouns do show a significant preference for object antecedents ($\beta = -0.692, p < .001$). Finally, in object clefts, null pronouns show a significant bias towards subject antecedents ($\beta = 0.741, p < .001$), but the interpretative preferences of overt pronouns are not different from chance ($\beta = 0.168, p = .285$). These results are summarized in the following examples, which illustrate the resolution patterns of null (10) and overt (11) pronouns in each of the contexts under analysis:

(10) *Resolution patterns of null pronouns*

La **Laura**_j va espantar la Maria quan **pro**_j va entrar a l'habitació. (*Unmarked*)

A la Maria_k la va espantar la Laura_j quan **pro**_{j/k} va entrar a l'habitació. (*CLLD*)

Va ser la Laura_j qui va espantar la Maria_k quan **pro**_{j/k} va entrar a l'habitació. (*Subject cleft*)

Va ser a la Maria_k a qui va espantar la **Laura**_j quan **pro**_j va entrar a l'habitació. (*Object cleft*)

(11) *Resolution patterns of overt pronouns*

La Laura_j va espantar la **Maria**_k quan **ella**_k va entrar a l'habitació. (*Unmarked*)

A la Maria_k la va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació. (*CLLD*)

Va ser la Laura_j qui va espantar la **Maria**_k quan **ella**_k va entrar a l'habitació. (*Subject cleft*)

Va ser a la Maria_k a qui va espantar la Laura_j quan **ella**_{j/k} va entrar a l'habitació. (*Object cleft*)

The pairwise contrasts in the interaction Pronoun \times Information structure from the perspective of Information structure confirm the different impact of marked information structures on resolution patterns depending on the form of the pronoun. Compared to canonical sentences, null pronouns show a decreased preference for subject antecedents in marked structures like CLLDs and subject clefts (respectively: $\beta = 0.693, p < .001$; $\beta = 0.476, p = .032$). This suggests a (dis)preference for focal (i.e., non-topical) subject antecedents, regardless of whether antecedents occupy or not the hierarchically highest position in the syntactic structure. The subject bias of null pronouns is similar in unmarked and object cleft conditions ($\beta = -0.130, p = .885$), where the subject antecedent maintains a topical status.

Interestingly, when comparing the two OVS structures (CLLD and object clefts), null pronouns exhibit a stronger preference for subject antecedents in object clefts (when they do not convey focus) compared to CLLDs (when they convey focus) ($\beta = -0.823, p < .001$). The comparison between subject and object clefts also shows a lower preference of null pronouns for focal subjects ($\beta = -0.606, p = .003$).

In contrast, overt pronouns show a decreased preference for object antecedents in CLLDs and in object clefts (respectively: $\beta = -0.818, p < .001$; $\beta = -1.242, p < .001$). In these two structures, the object antecedent appears first-mentioned, in a hierarchically higher syntactic position in the phrase structure than the subject antecedent. In subject clefts, the object bias remains unchanged compared to canonical sentences ($\beta = -0.382, p = .149$). Overt pronouns also show a lower preference for object interpretations in CLLDs and in object clefts compared to subject clefts, indicating an increased preference for subjects in lower postverbal positions (respectively: $\beta = 0.436, p = .056$; $\beta = -0.860, p < .001$).

When comparing the two OVS structures, a lower preference of overt pronouns for object interpretations is attested in CLLDs compared to object clefts, although it does not reach significance ($\beta = -0.424, p = .059$).

All the significant differences between conditions reported along this were also confirmed as reliable differences in the supplementary analysis using an ordered beta regression (see Appendix E). Additionally, the lower preference of overt pronouns for object interpretations in object clefts, compared to CLLDs, and the higher preference of overt pronouns for object interpretations in object clefts, again compared to CLLDs, were also found to be reliable in the supplementary analysis (even though they did not reach significance in the reported results). Thus, the preference of overt pronouns for object antecedents would be similar in unmarked sentences and subject clefts (clear and well-defined), lower in CLLDs and even lower in object clefts.

Overall, while null pronouns seem to prefer subject and topical antecedents, regardless of their structural position, overt pronouns favor object antecedents when they appear in lower structural position, regardless of pragmatic features.

5 Discussion

Firstly, referring to unmarked canonical sentences in Catalan (RQ1), null and overt pronouns have been shown to display a clear division of labor as predicted by the Position of Antecedent Hypothesis (Carminati 2002). Whereas null pronouns tend to be interpreted as coreferring with antecedents in subject (or Spec,IP) position, overt pronouns are prone to be resolved toward antecedents in object or non-subject (or non-Spec,IP) position. In line with Bel & García-Alcaraz (2018) and Mayol & Clark (2010), these findings confirm that null and overt pronouns in Catalan show clear and well-defined interpretive PAH-like biases.

It is interesting to notice that, in contrast to Bel & García-Alcaraz (2018), who did not attest a clear bias of null pronouns in main-subordinate contexts (only in subordinate-main sequences), we found a clear bias of null pronouns toward subject antecedents in main-subordinate contexts.

This difference could be related to a task effect, which could make clause order effects remain imperceptible if a forced-choice task is used, as in the present study (see de Rocafiguera & Bel 2022). Indeed, in acceptability tasks such as the one utilized by Bel & García-Alcaraz (2018), participants have the freedom to adjust their judgments and are not compelled to select a particular antecedent. This flexibility in judgment can occasionally mitigate preferential biases. In Spanish, other studies using forced-choice tasks have shown this clear biases of null pronouns also in main-subordinate contexts (e.g., de la Fuente 2015; García-Alcaraz 2015; Contemori & Di Domenico 2021; Martín-Villena 2023), which is not attested when acceptability judgements are used. Forced-choice tasks may increase statistical power to detect differences between conditions and may not capture subtle nuances to the same degree as acceptability judgments (Schütze & Sprouse 2014).

Secondly, referring to the interplay between syntactic and pragmatic cues on these interpretive biases of null and overt pronouns in canonical sentences (RQ2), our results confirm that the syntactic function of the antecedent is not the only factor to guide anaphora resolution. The preference of null and overt pronouns towards subject and object antecedents, respectively, was not always maintained when the information status of subject and object antecedents was manipulated using syntactically marked information structures. This means that information structure impacts pronoun resolution and provides further information on the interpretive properties of both null and overt subject pronouns. The impact of the three sets of factors being studied on the interpretation of null and overt pronouns is summarized in **Table 4**.¹¹

	Null pronouns	Overt pronouns
Syntactic function	subject > object	subject < object
Information status	topical > focal	topical = focal
Structural position	higher = lower position	higher < lower position

Table 4: Summary of features influencing the antecedent preference of null and overt subject pronouns.

On the one hand, null pronouns only show a clear subject bias when the subject antecedent conveys non-focal (i.e., topical) information, regardless of the hierarchical or surface position of the antecedents. On the other, overt pronouns only show a clear object bias when the object

¹¹ As mentioned above (see 1.3), the structural position can be understood in relation to the order of mention, or linear position. From this perspective, null pronouns do not differentiate between first and second mentions, treating them equivalently. In contrast, overt pronouns typically favor the second mention over the first. Schematically, this can be represented as:

- Null pronouns: first = second mention
- Overt pronouns: first < second mention

antecedent appears in a non-initial position (i.e., in its canonical VP-internal postverbal and hierarchically lower position), regardless of the information status of the antecedents. Hence, we can posit a generalization from our findings: null pronouns are highly sensitive to the syntactic function and the information status of their antecedents, while overt pronouns are primarily influenced by the syntactic function and the hierarchical position (or linear position according to other perspectives) of their antecedents.

The different sensitivity of null and overt pronouns to the analyzed cues supports a form-specific multiple constraint approach: different pronominal forms are affected by different—and multiple—factors (in line with Kaiser & Trueswell 2008; Bader & Portele 2019). The fact that pronominal interpretive preferences are determined by a combination of factors confirms that there is not a single cue that can account for the interpretive biases of null and overt pronouns. These findings, derived from both informatively marked and unmarked sentences in isolation, offer valuable insights that go beyond the scope of this study, particularly in broader discourses concerning the factors influencing the interpretive biases of null pronouns specifically. The interplay between syntactic prominence, as indicated by subjecthood, and pragmatic salience, associated with topicality, creates an optimal setting for null pronouns to resolve anaphorically to subject and topical antecedents, which inherently possess discursive greater salience and accessibility. As a result, null pronouns may show a preference for referring back to topical antecedents, capitalizing on their heightened salience and accessibility within discourse contexts. Of course, considering discourse prominence opens alternative explanations: it is possible that subject antecedents, by virtue of being prominent discourse entities, are more likely to attract anaphoric reference from null pronouns. Similarly, topical elements, being established topics within discourse, may also exhibit heightened salience and accessibility, making them preferred candidates for anaphoric reference by null pronouns. Hence, the preference for subject and topical antecedents observed in our study may partly stem from their prominence in discourse. Should this be confirmed, it would ultimately validate the features highlighted by our research.

Regarding the already well-known need to account for multiple factors to explain pronominal anaphora resolution, a relevant finding in our data is that when conflicting cues were encountered, pronouns were interpreted at chance. This means that when subject antecedents conveyed focal information and object antecedents, topical information, null pronouns were interpreted as unbiased. Similarly, when object antecedents did not appear *in situ* but in the left periphery, in an initial surface position, overt pronouns were interpreted as unbiased. If the relevant cues are misaligned, pronominal preferences become flexible and indeterminate. This finding confirms that pronoun resolution is governed by multiple constraints that cannot be ranked, i.e., they have similar weights. Hence, there does not appear to be a more determining cue that can outrank the others in the final interpretation of either null or overt pronouns (among the syntactic and pragmatic factors under analysis). However, our data reveal that null pronouns are clearly

sensitive to the interaction between syntax and pragmatics, whereas overt pronouns are not so sensitive to pragmatics, or the information status of its plausible antecedents, and syntactic/structural constraints seem to outrank pragmatic constraints.

This asymmetry in the interpretation of null and overt pronouns also implies that null and overt pronouns may not be in complementary distribution, as has sometimes been assumed across null subject languages. In line with our results, in completely ambiguous contexts, null pronouns may prefer to be interpreted as expressing topic continuity and may also be the preferred form to express topic continuity (Lozano 2016; García-Alcaraz & Bel 2019; Martín-Villena & Lozano 2020 for Spanish). However, overt pronouns may not be as specialized on expressing topic shift. Instead, and in contrast to null pronouns, they appear to be governed mainly by syntactic and structural constraints in our data, and not so much by discourse constraints. It should be noticed that overt subject pronouns are scarcely used in production and corpus studies have shown that the preferred form to express topic shift is generally through lexical subjects (Lozano 2009; 2016; see, for data in Catalan, Ribera 2008; Bel et al. 2010; see, for data in Spanish, Giannakou & Sitaridou 2020; Quesada 2021; Lozano et al. 2023). It is possible that, in pragmatic terms, null pronouns are in complementary distribution with lexical subjects, instead of overt pronouns, as suggested by Lozano et al. (2023) and Torregrossa et al. (2020). In this regard, it would be interesting to further explore the interpretive preferences of demonstrative pronouns and to what extent do they diverge from overt pronouns (see Giannakou 2018 for results and discussion on demonstrative pronouns in Chilean Spanish).

Based on our experimental data, we have observed that the interpretation of null pronouns is indeed multifaceted and aligns well with the multifactorial approach. Given these findings, it is relevant to consider Torregrossa et al.'s (2020) proposal. Despite recognizing the need for a multifactorial approach to account for pronoun resolution, these authors give special importance to the hierarchic position of an antecedent in the syntactic structure for the interpretation of null pronouns, based on Rizzi (2018). More specifically, they claim that an antecedent will be more prominent (and thus more preferred) for a null pronoun if it is hierarchically higher than the competing antecedent. In relation to the PAH, this claim implies that null pronouns will prefer antecedents that occupy a higher position in the phrase structure rather than antecedents in Spec,IP, regardless of their syntactic function (subject vs. object). The present research demonstrates that an antecedent that appears higher in the syntactic configuration is not necessarily more preferred for a null pronoun than an antecedent in a lower position. In other words, the interpretation of null pronouns does not seem to be especially determined by the hierarchic position of the antecedent. Antecedents in Spec,IP (i.e., subjects)¹² are preferred over

¹² Considering that the postverbal subject antecedent in CLLD and object clefts occupies a Spec,IP position, assuming Olarrea's (2002) analysis.

antecedents in the highest configurational position in the interpretation of null pronouns (e.g., in object clefts, the subject antecedent is preferred over the object antecedent, which appears higher in the syntactic structure).

When the object antecedent appears in a higher Spec,TopP position than the subject (i.e., in CLLDs), it does become more preferred for null pronouns—blurring their subject bias. However, in object clefts (e.g., *Va ser a la Maria a qui va espantar la Laura, quan pro_j va entrar a l'habitació*), null pronouns show a clear-cut bias toward (postverbal) subject antecedents. Despite being in a hierarchically higher position (Spec,FocP), object antecedents do not appear to have any feature that makes them compete with subject antecedents when it comes to the interpretation of null pronouns. This asymmetry in the preference of null pronouns for objects in Spec,FocP and in Spec,TopP reveals that topicality is a more determining cue than structural height (overlapping with surface position), as summarized in **Table 4** above. According to our data, the antecedent in Spec,IP (i.e., the subject) would be more preferred for null pronouns than the antecedent in the highest position in the syntactic structure. Our study also provides specific evidence on the weighting of subjecthood: at least in ambiguous contexts, subjecthood is as relevant as topicality.

There are still some results regarding overt pronouns that deserve further examination. We have found that they may not be governed by pragmatic features, or not as much as previously assumed in the literature. Overt pronouns do not show a preference towards focal or non-topical antecedents, they show no specialization on topic shift in the examined contexts and, as predicted by the PAH (Carminati 2002), overt pronouns only show a clear preference to corefer with the antecedent in the lowest position in the syntactic structure, provided that it is not a subject (i.e., it appears in Spec,IP). Filiaci (2011) attested a different interpretation of overt pronouns in Italian and in Spanish—only showing a clear bias toward object antecedents in Italian—and referred to Cardinaletti & Starke's (1999: 219) crosslinguistic typology of deficient forms to explain these differences. Unlike the Italian overt pronouns *lui/lei* and considering previous research showing a lack of strong bias for overt pronouns in Spanish, the Spanish overt pronouns *él/ella* could be analyzed as weak or structurally deficient elements, similar to *egli/ella* in Italian (Cardinaletti & Starke 1999). This would explain why overt pronouns in Spanish would not be so restricted to corefer with non-topical object antecedents and would more easily allow coreferences with topical or subject antecedents. As for Catalan, based on the uneven results for non-marked constructions in our experimental data, it could be suggested that Catalan overt pronouns may behave similarly to the “weak” overt pronouns in Spanish, as described by Filiaci (2011). In a similar vein, Licerias & Alba de la Fuente (2015) propose that Spanish might have two kinds of overt subject pronouns: 1) weak overt pronouns that would behave like a free phonetically realized counterpart of null pronouns and would convey the same pragmatic features, and 2) strong pronouns, conveying pragmatic features such as focus. According to our data, we cannot say that overt pronouns convey the same pragmatic features as null pronouns, but the overt

pronouns included in our items were not interpreted as strong pronouns conveying focus, neither (see Rigau 1989; Mayol 2010a; Herbeck 2018 for discussions on the relation between overt subject pronouns and the expression of focus in Romance languages such as Catalan). They are somehow weak pronouns, devoid of pragmatic features, but clearly differ from null pronouns. A term introduced by Vallduví (1994: 13) may become useful to describe the behavior of overt pronouns in the assessed contexts. He refers to the dichotomy between strong and weak pronouns and defines weak pronouns to be “inert as far as information packaging is concerned” (see also Vallduví & Engdahl 1996: 476). Although he relates weak pronouns to null or clitic pronouns in Catalan, his proposal could well be expanded to overt pronouns. In the sentences that we assessed in the present study, they do not appear to show a special sensitivity to topic and focus features. Overt pronouns could also be interpreted as weak pronouns and, hence, pragmatically ‘inert’. This is not far from what Liceras & Alba de la Fuente (2015) termed weak overt pronouns, with no content. Of course, this fact does not exclude the existence of strong overt pronouns, that could clearly convey emphasis, contrast or focus features in marked contexts.

6 Conclusion and future directions

Overall, our findings demonstrate the need for a form-specific and multiple-constraint approach to account for pronominal anaphora resolution in a null subject language like Catalan (in line with Kaiser & Trueswell 2008, among others). These multiple factors that jointly affect pronominal anaphora resolution, respectively, have similar weights and cannot be ranked: when the relevant multiple cues are misaligned, neither null nor overt pronouns show well-defined biases. We have demonstrated that syntactic, structural and pragmatic features intervene to a different extent in shaping the interpretation of null and overt subject pronouns. On the one hand, the interpretation of null pronouns is guided by the interaction between syntactic and pragmatic factors (i.e., the syntactic function and the information status of the antecedent): null pronouns prefer both subject and topical antecedents. On the other hand, the interpretation of overt pronouns is guided by the interaction between syntactic and structural factors (i.e., the syntactic function of the antecedent and word order, or its hierarchical position in the syntactic configuration): overt pronouns prefer object and postverbal (second-mentioned or hierarchically lower) antecedents. We have interpreted these findings as showing 1) that null and overt pronouns are not in complementary distribution in null subject languages (at least in Catalan), but differently constrained by a combination of factors, 2) that subjecthood and topicality—syntactic and pragmatic features—are similarly relevant in guiding the interpretation of null pronouns (in line with Carminati 2002’s predictions), and 3) that overt pronouns appear less sensitive to pragmatic features and can be interpreted as pragmatically ‘inert’ weak pronouns (Vallduví 1994).

In terms of future directions, we highlight the interest in replicating and extending the present study to other null subject languages. The reported findings have already been replicated in Spanish using the same design, yielding similar results (see de Rocafiguera 2023). Microvariation in anaphora resolution within null subject languages in canonical sentences has already been attested when comparing languages such as Catalan and Spanish (Bel & García-Alcaraz 2018; de Rocafiguera 2023), Spanish and Italian (Filiaci & Sorace & Carreiras 2014; Contemori & Di Domenico 2021) or Spanish, Greek and Italian (Leonetti-Escandell & Torregrossa 2024). From these studies, Spanish and Greek have been shown to allow more flexibility in the interpretation of overt pronouns than Catalan or Italian. However, de Rocafiguera (2023) identified similar patterns between Catalan and Spanish when comparing the effects of marked information structures. A cross-linguistic comparison between Catalan and Italian has not yet been conducted, nor have comparisons including Portuguese or other Romance null subject languages. Such comparisons will require further exploration to deepen our understanding of microvariation patterns.

Data availability and supplementary files

All data associated with this study (language background questionnaire, experimental materials, dataset and analysis script, supplementary analyses) is openly available, according to the FAIR principles (Findable, Accessible, Interoperable, Reusable) at <https://doi.org/10.17605/OSF.IO/XRCB3>.

Ethics and consent

Ethical approval was received from the Institutional Committee for Ethical Review of Projects in the Universitat Pompeu Fabra (CIREP-UPF; Reference code 0032). All participants gave informed written consent regarding the data use and received a stipend for their participation in the study.

Funding information

This research was supported by a predoctoral grant (2018FI_B_00959) from the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) as well as the FFI2016-75082-P project awarded by the Spanish Ministry of Innovation and Science and the Spanish Ministry of Universities.

Acknowledgements

We are thankful to the anonymous reviewers and to the audience at *Jornades de lingüística catalana a Viena 2022* for their valuable comments and suggestions. We would also like to thank those who participated in the study and those who helped us recruit participants.

Competing interests

The authors have no competing interests to declare.

References

- Arnold, Jennifer E. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass* 4(4). 187–203. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00193.x>
- Bader, Markus & Portele, Yvonne. 2019. The interpretation of German personal pronouns and d-pronouns. *Zeitschrift für Sprachwissenschaft* 38(2). 155–190. DOI: <https://doi.org/10.1515/zfs-2019-2002>
- Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Bel, Aurora & García-Alcaraz, Estela. 2018. Pronoun interpretation and processing in Catalan and Spanish bilingual and monolingual speakers. In Cuza, Alejandro & Guijarro-Fuentes, Pedro (eds.), *Language Acquisition and Contact in the Iberian Peninsula*, 37–62. De Gruyter. DOI: <https://doi.org/10.1515/9781501509988-003>

- Bel, Aurora & Perera, Joan & Salas, Naymé. 2010. Anaphoric devices in written and spoken narrative discourse: Data from Catalan. *Written Language & Literacy* 13(2). 236–259. DOI: <https://doi.org/10.1075/wll.13.2.03bel>
- Birdsong, David & Gertken, L. M. & Amengual, Mark. 2012. Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. (<https://sites.la.utexas.edu/bilingual/>)
- Blything, Liam P. & Järvikivi, Juhani & Toth, Abigail G. & Arnhold, Anja. 2021. The influence of focus marking on pronoun resolution in dialogue context. *Frontiers in Psychology* 12. 684639. DOI: <https://doi.org/10.3389/fpsyg.2021.684639>
- Brysaert, Marc. 2021. Power considerations in bilingualism research: Time to step up our game. *Bilingualism: Language and Cognition* 24(5). 813–818. DOI: <https://doi.org/10.1017/S1366728920000437>
- Cardinaletti, Anna & Starke, Michal. 1999. The typology of structural deficiency: on the three grammatical classes. *Working Papers in Linguistics* 4. 41–109.
- Carminati, Maria Nella. 2002. *The processing of Italian subject pronouns*. University of Massachusetts Amherst dissertation.
- Carreiras, Manuel & Gernsbacher, Mortonann & Villa, Victor. 1995. The advantage of first mention in Spanish. *Psychonomic Bulletin & Review* 2(1). 124–129. DOI: <https://doi.org/10.3758/BF03214418>
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, Charles N. (ed.), *Subject and Topic*, 25–55. New York: Academic Press. Retrieved from <https://typeset.io/papers/giveness-contrastiveness-definiteness-subjects-topics-and-1uhgi6etzk>
- Chamorro, Gloria. 2018. Offline interpretation of subject pronouns by native speakers of Spanish. *Glossa: A Journal of General Linguistics* 3(1). DOI: <https://doi.org/10.5334/gjgl.256>
- Colonna, Saveria & Schimke, Sarah & Hemforth, Barbara. 2012. Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. *Linguistics* 50(5). 991–1013. DOI: <https://doi.org/10.1515/ling-2012-0031>
- Colonna, Saveria & Schimke, Sarah & Hemforth, Barbara. 2015. Different effects of focus in intra- and inter-sentential pronoun resolution in German. *Language, Cognition and Neuroscience* 30(10). 1306–1325. DOI: <https://doi.org/10.1080/23273798.2015.1066510>
- Contemori, Carla & Di Domenico, Elisa. 2021. Microvariation in the division of labor between null- and overt-subject pronouns: the case of Italian and Spanish. *Applied Psycholinguistics* 42(4). 997–1028. DOI: <https://doi.org/10.1017/S0142716421000199>
- Cowles, H. Wind & Walenski, Matthew & Kluender, Robert. 2007. Linguistic and cognitive prominence in anaphor resolution: topic, contrastive focus and pronouns. *Topoi* 26(1). 3–18. DOI: <https://doi.org/10.1007/s11245-006-9004-6>
- de la Fuente, Israel. 2015. *Putting pronoun resolution in context: The role of syntax, semantics, and pragmatics in pronoun interpretation*. Paris: Université Paris Diderot dissertation. Retrieved from <https://tel.archives-ouvertes.fr/tel-01535977>. DOI: <https://doi.org/10.1515/9783110464108-003>
- de Rocafiguera, Núria. 2023. *Information structure and language dominance in subject pronoun resolution in Catalan-Spanish bilingualism*. Barcelona: Universitat Pompeu Fabra dissertation. Retrieved from <https://www.tdx.cat/handle/10803/687754>

- de Rocafiguera, Núria & Bel, Aurora. 2022. On the impact of clause order on pronoun resolution: evidence from Spanish. *Folia Linguistica* 56(1). 1–24. DOI: <https://doi.org/10.1515/flin-2021-2002>
- Domínguez, Laura. 2018. Information Structure. In Geeslin, Kimberly L. (ed.), *The Cambridge Handbook of Spanish Linguistics*, 372–391. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781316779194.018>
- Fábregas, Antonio. 2016. Information structure and its syntactic manifestation in Spanish: facts and proposals. *Borealis – An International Journal of Hispanic Linguistics* 5(2). 1–109. DOI: <https://doi.org/10.7557/1.5.2.3850>
- Fedele, Emily & Kaiser, Elsi. 2014. Resolving null and overt pronouns in Italian: An experimental investigation of syntax-semantics interactions. In Brown, Christopher & Gu, Qianping & Loos, Cornelia & Mielens, Jason & Neveu, Grace (eds.), *Proceedings of the 15th Texas Linguistic Society*, 53–72. University of Texas at Austin.
- Filiaci, Francesca. 2011. *Anaphoric preferences of null and overt subjects in Italian and Spanish: A cross-linguistic comparison*. Edinburgh: The University of Edinburgh dissertation.
- Filiaci, Francesca & Sorace, Antonella & Carreiras, Manuel. 2014. Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language, Cognition and Neuroscience* 29(7). 825–843. DOI: <https://doi.org/10.1080/01690965.2013.801502>
- Foraker, Stephani & McElree, Brian. 2007. The role of prominence in pronoun resolution: Active versus passive representations. *Journal of Memory and Language* 56(3). 357–383. DOI: <https://doi.org/10.1016/j.jml.2006.07.004>
- Fukumura, Kumiko & van Gompel, Roger P. G. 2015. Effects of order of mention and grammatical role on anaphor resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(2). 501–525. DOI: <https://doi.org/10.1037/xlm000041>
- García-Alcaraz, Estela. 2015. *Comprensión y producción de los pronombres nulos y explícitos de tercera persona en posición de sujeto en la adquisición temprana del español L2*. Barcelona: Universitat Pompeu Fabra dissertation. Retrieved from <http://www.tdx.cat/handle/10803/311971>
- García-Alcaraz, Estela & Bel, Aurora. 2019. Does empirical data from bilingual and native Spanish corpora meet linguistic theory? The role of discourse context in variation of subject expression. *Applied Linguistics Review* 10(4). 491–515. DOI: <https://doi.org/10.1515/applirev-2017-0101>
- Gernsbacher, Morton Ann & Hargreaves, David J. 1988. Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language* 27(6). 699–717. DOI: [https://doi.org/10.1016/0749-596X\(88\)90016-2](https://doi.org/10.1016/0749-596X(88)90016-2)
- Giannakou, Aretousa. 2018. *Spanish and Greek subjects in contact: Greek as a heritage language in Chile*. Cambridge: University of Cambridge dissertation. DOI: <https://doi.org/10.17863/CAM.30352>
- Giannakou, Aretousa & Sitaridou, Ioanna. 2020. Microparametric variation in the syntax of Spanish and Greek pronominal subjects. *Glossa: A Journal of General Linguistics* 5(1). 1–42. DOI: <https://doi.org/10.5334/gjgl.960>
- Giannakou, Aretousa & Sitaridou, Ioanna. 2022. (In)felicitous use of subjects in Greek and Spanish in monolingual and contact settings. DOI: <https://doi.org/10.16995/GLOSSA.5812>

- Givón, Thomas. 1983. Topic continuity in discourse: An introduction. In Givón, Thomas (ed.), *Topic continuity in discourse: A quantitative cross-language study*, 1–42. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.3>
- Godoy, Mahayana C. & Weissheimer, Janaina & Mafra, Matheus Araújo. 2018. When grammar meets pragmatics: Subject preference and coherence relations in Brazilian Portuguese pronoun interpretation. *Journal of Portuguese Linguistics* 17(1). DOI: <https://doi.org/10.5334/jpl.197>
- Goikoetxea, Edurne & Pascual, Gema & Acha, Joana. 2008. Normative study of the implicit causality of 100 interpersonal verbs in Spanish. *Behavior Research Methods* 40(3). 760–772. DOI: <https://doi.org/10.3758/BRM.40.3.760>
- Gutiérrez-Bravo, Rodrigo. 2003. Subject inversion in Spanish relative clauses. In Geerts, Twan & van Ginneken, Ivo & Jacobs, Haike (eds.), *Romance Languages and Linguistic Theory*, 16. Amsterdam: John Benjamins.
- Hartshorne, Joshua K. & Sudo, Yasutada & Uruwashii, Miki. 2013. Are implicit causality pronoun resolution biases consistent across languages and cultures? *Experimental Psychology* 60(3). 179–196. DOI: <https://doi.org/10.1027/1618-3169/a000187>
- Herbeck, Peter. 2018. Deriving Null, Strong and Emphatic Pronouns in Romance Pro-Drop Languages. In Patel-Grosz, Pritty & Grosz, Patrick Georg & Zobel, Sarah (eds.), *Pronouns in embedded contexts at the syntax-semantics interface*, 171–213. Cham: Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-56706-8_6
- Järvikivi, Juhani & van Gompel, Roger P. G. & Hyönä, Jukka. 2017. The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 46(3). 525–550. DOI: <https://doi.org/10.1007/s10936-016-9451-1>
- Järvikivi, Juhani & van Gompel, Roger P. G. & Hyönä, Jukka & Bertram, Raymond. 2005. Ambiguous pronoun resolution: Contrasting the first-mention and subject-preference accounts. *Psychological Science* 16(4). 260–264. DOI: <https://doi.org/10.1111/j.0956-7976.2005.01525.x>
- Kaiser, Elsi. 2011. Focusing on pronouns: Consequences of subjecthood, pronominalisation, and contrastive focus. *Language and Cognitive Processes* 26(10). 1625–1666. DOI: <https://doi.org/10.1080/01690965.2010.523082>
- Kaiser, Elsi & Trueswell, John C. 2008. Interpreting pronouns and demonstratives in Finnish: Evidence for a form-specific approach to reference resolution. *Language and Cognitive Processes* 23(5). 709–748. DOI: <https://doi.org/10.1080/01690960701771220>
- Káldi, Tamás & Babarczy, Anna. 2021. Linguistic focus guides attention during the encoding and refreshing of Working Memory content. *Journal of Memory and Language* 116. 104187. DOI: <https://doi.org/10.1016/j.jml.2020.104187>
- Kehler, Andrew & Rohde, Hannah. 2013. A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. *Theoretical Linguistics* 39(1–2). 1–37. DOI: <https://doi.org/10.1515/tl-2013-0001>
- Koornneef, Arnout W. & Sanders, Ted J. M. 2013. Establishing coherence relations in discourse: The influence of implicit causality and connectives on pronoun resolution. *Language and Cognitive Processes* 28(8). 1169–1206. DOI: <https://doi.org/10.1080/01690965.2012.699076>

- Lakens, Daniël. 2022. Sample Size Justification. *Collabra: Psychology* 8(1). 33267. DOI: <https://doi.org/10.1525/collabra.33267>
- Langsford, Steven & Perfors, Amy & Hendrickson, Andrew T. & Kennedy, Lauren A. & Navarro, Danielle J. 2018. Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa: A Journal of General Linguistics* 3(1). DOI: <https://doi.org/10.5334/gjgl.396>
- Lenth, Russell V. 2021. emmeans: Estimated Marginal Means, aka Least-Squares Means (Version R package version 1.7.0). Retrieved from <https://CRAN.R-project.org/package=emmeans>
- Leonetti, Manuel. 2021. Topics and the interpretation of referential null subjects. In Gianollo, Chiara & von Heusinger, Klaus & Napoli, Maria (eds.), *Determiners and quantifiers: functions, variation and change*, 94–129. Leiden: Brill. DOI: https://doi.org/10.1163/9789004473324_005
- Leonetti-Escandell, Victoria & Torregrossa, Jacopo. 2024. The interpretation of null and overt subject pronouns in Spanish compared to Greek and Italian. *Glossa: A Journal of General Linguistics* 9(1). DOI: <https://doi.org/10.16995/glossa.9725>
- Liceras, Juana M. & Alba de la Fuente, Anahí. 2015. Typological proximity in L2 acquisition: The Spanish non-native grammar of French speakers. In Judy, Tiffany & Perpiñán, Silvia (eds.), *The acquisition of Spanish in understudied language pairings*, 329–358. Amsterdam and New York: John Benjamins. DOI: <https://doi.org/10.1075/ihll.3.13lic>
- Lozano, Cristóbal. 2009. Pronominal deficits at the interface: new data from the CEDEL2 corpus. In Bretones Callejas, Carmen M. (ed.), *Understanding language and mind: Applied linguistics now*, 213–227. Almería: Universidad de Almería.
- Lozano, Cristóbal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface: advanced English learners of Spanish in the CEDEL2 corpus. In Alonso Ramos, Margarita (ed.), *Spanish Learner Corpus Research: State of the Art and Perspectives*, 236–265. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/scl.78.09loz>
- Lozano, Cristóbal & Quesada, Teresa & Papadopoulou, Despina & Charatzidis, Andreas. 2023. What do corpus data reveal about anaphora resolution? Spanish vs. Greek and the Type of Topic Hypothesis. *Glossa: A Journal of General Linguistics* 8(1). DOI: <https://doi.org/10.16995/glossa.9883>
- Lubbers-Quesada, Margaret & Blackwell, Sarah E. 2009. The L2 acquisition of null and overt Spanish subject pronouns: A pragmatic approach. In Collentine, Joseph (ed.), *Selected proceedings of the 11th Hispanic Linguistics Symposium*, 117–130. Somerville, MA: Cascadilla Proceedings Project.
- Martín-Villena, Fernando. 2023. *L1 morphosyntactic attrition at the early stages: evidence from production, interpretation, and processing of subject referring expressions in L1 Spanish-L2 English instructed and immersed bilinguals*. Granada: Universidad de Granada dissertation. Retrieved from <https://digibug.ugr.es/handle/10481/81920>
- Martín-Villena, Fernando & Lozano, Cristóbal. 2020. Anaphora resolution in topic continuity: Evidence from L1 English–L2 Spanish data in the CEDEL2 corpus. In Ryan, Jonathon & Crosthwaite, Peter (eds.), *Referring in a second language: Studies on reference to person in a multilingual world*, 119–141. Routledge. DOI: <https://doi.org/10.4324/9780429263972-7>

- Marty, Paul & Chemla, Emmanuel & Sprouse, Jon. 2020. The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: A Journal of General Linguistics* (2016–2021) 5(1). 72. DOI: <https://doi.org/10.5334/gjgl.980>
- Mayol, Laia. 2010a. Contrastive pronouns in null-subject Romance languages. *Lingua* 120(10). 2497–2514. DOI: <https://doi.org/10.1016/j.lingua.2010.04.009>
- Mayol, Laia. 2010b. Refining salience and the Position of Antecedent Hypothesis: a study of Catalan pronouns. *University of Pennsylvania Working Papers in Linguistics* 16(1). Retrieved from <https://repository.upenn.edu/pwpl/vol16/iss1/15>
- Mayol, Laia & Clark, Robin. 2010. Pronouns in Catalan: Games of partial information and the use of linguistic resources. *Journal of Pragmatics* 42(3). 781–799. DOI: <https://doi.org/10.1016/j.pragma.2009.07.004>
- Olarrea, Antxon. 1998. On the Position of Subjects in Spanish. *Anuario Del Seminario De Filología Vasca "Julio De Urquijo"* 32(1). 47–108.
- Papadopoulou, Despina & Peristeri, Eleni & Plemenou, Evagelia & Marinis, Theodoros & Tsimpli, Ianthi. 2015. Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua* 155. 98–120. DOI: <https://doi.org/10.1016/j.lingua.2014.09.006>
- Patterson, Clare & Esaulova, Yulia & Felser, Claudia. 2017. The impact of focus on pronoun resolution in native and non-native sentence comprehension. *Second Language Research* 33(4). 403–429. DOI: <https://doi.org/10.1177/0267658317697786>
- Patterson, Clare & Felser, Claudia. 2020. Cleft focus and antecedent accessibility: The emergence of the anti-focus effect. In Holler, Anke & Suckow, Katja & de la Fuente, Israel (eds.), *Information Structuring in Discourse*, Vol. 40. 56–85. Brill. DOI: https://doi.org/10.1163/9789004436725_004
- Prince, Ellen. 1981. Towards a taxonomy of given/new information. In Cole, Peter (ed.), *Radical Pragmatics*, 223–255. New York: Academic Press.
- Pyykkönen, Pirta & Järvikivi, Juhani. 2010. Activation and persistence of implicit causality information in spoken language comprehension. *Experimental Psychology* 57. 5–16. DOI: <https://doi.org/10.1027/1618-3169/a000002>
- Quesada, Teresa. 2021. *Studies on Anaphora Resolution in L1 Spanish – L2 English & L1 English – L2 Spanish adult learners: combining corpus and experimental methods*. Granada: Universidad de Granada dissertation. Retrieved from <https://digibug.ugr.es/handle/10481/72052>
- R Core Team. 2021. R: A language and environment for statistical computing (Version 4.0.5). Vienna, Austria: R Foundation for Statistical Computing.
- Ribera, Josep E. 2008. *La cohesió léxica en seqüències narratives*. Valencia: Universitat de València dissertation. Retrieved from <http://www.tesisenred.net/TDX-0917109-143808/>
- Rigau, Gemma. 1989. Connexity established by emphatic pronouns. In Conte, Maria-Elisabeth & Petöfi, János S. & Sözer, Emel (eds.), *Text and Discourse Connectedness*. John Benjamins Publishing. DOI: <https://doi.org/10.1075/slcs.16.17rig>
- Rinke, Esther & Flores, Cristina. 2018. Another look at the interpretation of overt and null pronominal subjects in bilingual language acquisition: Heritage Portuguese in contact with German and Spanish. *Glossa: A Journal of General Linguistics* 3(1). DOI: <https://doi.org/10.5334/gjgl.535>

- Rizzi, Luigi. 1997. The fine structure of the left periphery. In Haegeman, Liliane (ed.), *Elements of Grammar: Handbook in Generative Syntax*, 281–337. Dordrecht: Springer Netherlands. DOI: https://doi.org/10.1007/978-94-011-5420-8_7
- Rizzi, Luigi. 2018. Subjects, topics and the interpretation of pro. In Petrosino, Roberto & Cerrone, Pietro & Hulst, Harry van der (eds.), *From Sounds to Structures*. De Gruyter. DOI: <https://doi.org/10.1515/9781501506734-019>
- Schumacher, Petra B. & Dangl, Manuel & Uzun, Elyesa. 2016. Thematic role as prominence cue during pronoun resolution in German. In Holler, Anke & Suckow, Katja (eds.), *Empirical perspectives on anaphora resolution*, 213–240. De Gruyter. DOI: <https://doi.org/10.1515/9783110464108-011>
- Schumacher, Petra B. & Roberts, Leah & Järvikivi, Juhani. 2017. Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua* 185. 25–41. DOI: <https://doi.org/10.1016/j.lingua.2016.07.004>
- Schütze, Carson T. & Sprouse, Jon. 2014. Judgment data. In Podesva, Robert J. & Sharma, Devyani (eds.), *Research Methods in Linguistics*, 27–50. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139013734.004>
- Sorace, Antonella & Keller, Frank. 2005. Gradience in linguistic data. *Lingua* 115(11). 1497–1524. DOI: <https://doi.org/10.1016/j.lingua.2004.07.002>
- Sorace, Antonella & Serratrice, Ludovica & Filiaci, Francesca & Baldo, Michela. 2009. Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua* 119(3). 460–477. DOI: <https://doi.org/10.1016/j.lingua.2008.09.008>
- Sprouse, Jon & Almeida, Diogo. 2017. Design sensitivity and statistical power in acceptability judgment experiments. *Glossa: A Journal of General Linguistics* 2(1). DOI: <https://doi.org/10.5334/gjgl.236>
- Stevenson, Rosemary & Knott, Alistair & Oberlander, Jon & McDonald, Sharon. 2000. Interpreting pronouns and connectives: Interactions among focusing, thematic roles and coherence relations. *Language and Cognitive Processes* 15(3). 225–262. DOI: <https://doi.org/10.1080/016909600386048>
- Torregrossa, Jacopo & Andreou, Maria & Bongartz, Christiane M. 2020. Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa: A Journal of General Linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.1011>
- Tsimpli, Ianthi & Sorace, Antonella & Heycock, Caroline & Filiaci, Francesca. 2004. First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism* 8(3). 257–277. DOI: <https://doi.org/10.1177/13670069040080030601>
- Vallduví, Enric. 1992. *The informational component*. New York: Garland.
- Vallduví, Enric. 1994. The dynamics of information packaging. In Engdahl, Elisabet (ed.), *Integrating information structure into constraint-based and categorial approaches*, 1–26. Amsterdam: Institute for Logic, Language and Computation.
- Vallduví, Enric & Engdahl, Elisabet. 1996. The linguistic realization of information packaging 34(3). 459–520. DOI: <https://doi.org/10.1515/ling.1996.34.3.459>

von Heusinger, Klaus & Schumacher, Petra B. 2019. Discourse prominence: Definition and application. *Journal of Pragmatics* 154. 117–127. DOI: <https://doi.org/10.1016/j.pragma.2019.07.025>

Wolna, Agata & Durlik, Joanna & Wodniecka, Zofia. 2022. Pronominal anaphora resolution in Polish: Investigating online sentence interpretation using eye-tracking. *PLOS ONE* 17(1). DOI: <https://doi.org/10.1371/journal.pone.0262459>

