

Word order variation with *decir* ‘say’ in spoken Peninsular Spanish: A case of residual V2?

Peter Herbeck

University of Vienna & University of Wuppertal
peter.herbeck@univie.ac.at

Pekka Posio

University of Helsinki
pekka.posio@helsinki.fi



Received: 17-12-2022

Accepted: 26-03-2022

Published: 01-09-2022

How to cite: Herbeck, Peter & Pekka Posio. 2022. Word order variation with *decir* ‘say’ in spoken Peninsular Spanish: A case of residual V2? In *Residual Verb Second in Romance*, eds. Silvio Cruschina, Antonio Fábregas & Christine Meklenborg Nilsen. Special issue of *Isogloss. Open Journal of Romance Linguistics* 8(3)/7, 1-33.
DOI: <https://doi.org/10.5565/rev/isogloss.196>

Abstract

In this paper, we investigate word order with the verb *decir* ‘say’ in two corpora of spoken Peninsular Spanish. This verb is particularly interesting because it has been observed to have a high frequency of postverbal subjects in previous studies. Beside features that influence optional subject-verb inversion, such as the type of subject, the type of complement and the type of sentence in which *decir* appears, we investigate whether factors related to residual V2 trigger the postverbal subject position in some contexts. It will be shown that fronted objects, but not preverbal adverbs, trigger subject-verb inversion in the speech data investigated here. Furthermore, the factor of subject type and complement type are relevant for explaining word order patterns. Lastly, we also discuss

the existence of semi-fixed, ‘constructionalized’ word order patterns with *decir* and how they can be accounted for in a framework that makes use of perspectival functional categories in the left periphery.

Keywords: word order, residual V2, subject-verb inversion, spoken Spanish, corpus study.

1. Introduction

In this paper, we present a quantitative and qualitative study of Spanish word order with one particular verb – *decir* ‘say’ – in two speech corpora. A detailed study of this verb is interesting because it is associated with particularly high frequencies of postverbal subjects at least in Peninsular Spanish (Posio & Pešková 2020) even though it does not fall into semantic verb classes that have been argued to favor subject-verb inversion, such as intransitive and unaccusative predicates. One of our main research questions is whether factors related to ‘residual V2’ might be responsible for the high frequency of subject inversion associated with this verb lexeme.

We combine generative and functional perspectives to account for the patterns of subject placement: on the one hand, some factors that might be classified as belonging to a residual V2 effect are deeply rooted in the syntactic structure, such as a preferred postverbal position of subjects with fronted objects (see §1.1 and §3.1.1). This pattern is best explained by means of assuming that fronting operations (of e.g. objects) to Spec,IP trigger postverbal subjects in Spanish (see e.g. Zubizarreta 1998). On the other hand, it can be observed that the postverbal position of subjects with *decir* ‘say’ is preferred in frequently occurring semi-fixed sequences that serve different discourse-pragmatic functions and show a certain degree of constructionalization (Rostila 2004, Wiemer & Bisang 2004, Traugott & Trousdale 2013), i.e., grammaticalization of constructions. Thus, we believe that in order to fully understand the patterns of subject-verb inversion with *decir*, it is beneficial to combine insights from formal, generative syntax and usage-based, functional linguistics.

The paper is structured as follows. In Section 1.1 we discuss previous studies on variable word order and residual V2 in Spanish from the generative perspective, and in Section 1.2 we present what is known from variable word order with *decir* from more usage-based studies. Section 2 introduces the research questions, data, and methodology. In Section 3 we present the results of our analysis, and the concluding Section 4 contains the discussion of the results and their consequences, as well as suggestions for future research.

1.1. Variable word order and residual V2 in Spanish from the generative perspective

It is well known that Spanish and other Romance languages allow for variable word order patterns. Thus, Spanish allows SVO, VSO as well as VOS order, as illustrated in examples (1)–(3) (see Contreras 1991, Zubizarreta 1998, Ordóñez 2000, Leonetti 2017, among many others):

- (1) *Juan leyó el periódico.* (SVO)
 Juan read.3SG the newspaper
 ‘Juan read the paper’

(2) *Leyó el periódico Juan.* (VOS)

(3) *Leyó Juan el periódico.* (VSO)

The property of variable word order has notoriously been linked to the possibility of having null subjects (see Chomsky 1981, Rizzi 1982, among others). However, not all constituent orders are equally present in actual language use, where the choice of a particular order has been related to various morpho-syntactic, semantic, and pragmatic factors. Thus, postverbal subjects are preferred if the agentivity of the subject is reduced (see Cuartero Sánchez 2007), with unaccusative and intransitive verbs as opposed to transitives (see Fernández Soriano 1993, Leonetti 2017), and with indefinite NP subjects compared to definite NPs and pronouns (see Fernández Soriano 1993, Leonetti 2017). Furthermore, information structure has been singled out as a particularly relevant factor for word order, especially the concepts of topic and focus (see Rizzi 1997, Zubizarreta 1998, Belletti 2004). In general, subject-verb inversion is favored with focal subjects, in comparison with sentences having a topic-comment sequential order.

In the literature SV(O) has been observed to be the unmarked word order in Spanish, being used in the widest range of contexts: It has been argued that preverbal subjects are in a (left peripheral) topic position in several studies (e.g. Alexiadou & Anagnostopoulou 1998 and related studies). In fact, SV orders can receive a topic-comment partition, in which the subject is the topic of the sentence. However, it has been argued that SV orders are also compatible with an all-focus reading (see Zagona 2002, Leonetti 2017 for discussion):¹

- (4) A: ¿*Qué pasó?*
 ‘What happened?’
 B: [_{FOCUS} *José fue a casa.*]
 ‘José went home.’ (Zagona 2002:209)

Furthermore, SV(O) orders are compatible with contrastive focus on the subject if it receives emphatic stress (cf. Zubizarreta 1998:125, Zagona 2002:210f.) .

With respect to VS, on the other hand, the topic interpretation of the subject is excluded, i.e. VS orders are typically used in narrow or wide focus contexts (cf. Leonetti 2017:893 and references). Thus, intransitive verbs sanction VS with wide focus in presentational contexts in Spanish (cf. Sheehan 2010, citing Suñer 1982; see Corr 2016 for further detailed discussion):

- (5) A: ¿*Qué pasó?*
 ‘What happened?’
 B: *Se rió Juan.*
 REFL laughed Juan
 ‘Juan laughed.’ (Sheehan 2010:241)

With transitive verbs, the alternation between (XP)VSO and VOS has also been argued to be related to information structure in that both SVO and VSO can be used in

¹ See Suñer (2003) and López (2009) for further arguments that preverbal subjects can, but need not be in a left-dislocated topic position in Spanish.

‘neutral’ contexts and receive a wide focus interpretation, while the VOS order correlates with narrow focus on the subject (see Zubizarreta 1998, Ordóñez 2000, Zagana 2002, among many others).

In formal terms, the possibility of VS orders in Romance *pro*-drop languages has been linked to the strength of agreement morphology and correlated EPP absorption (see Alexiadou & Anagnostopoulou 1998, Ordóñez & Treviño 1999, Barbosa 2009) or expletive *pro* (Chomsky 1981, Rizzi 1982). While in English an overt subject must satisfy the EPP of T in a spec/head relation, in languages with variable word order this requirement is satisfied by agreement morphology, verb movement or a null expletive in Spec,TP, making it possible for the subject to remain in the base-generated Spec,vP position or to raise to a preverbal (topic) position. In (6), we indicate this position as Spec,FP (functional projection), given that it has been assumed in the literature that preverbal subjects are not obligatorily (left-dislocated) topics (see Suñer 2003, López 2009 for discussion):

- (6) [FP (*Juan*) [TP T-Durmió [vP (*Juan*) *durmió* ...]]
 Juan slept Juan

The actual surface position of the subject would be governed by topic or focus features, being situated in the high, CP-related (Rizzi 1997) or low, vP-related (Belletti 2004) left periphery.

Even though information structure and, in particular, the notions of topic and focus have been argued to be decisive to explain the positioning of the subject in Spanish, there are reasons to believe that they are not the only decisive factors triggering SV or VS orders with the verb *decir* ‘say’: as we will see in the analysis and discussion sections, several examples of VS orders in authentic spoken corpus data cannot easily be explained by these two information structural notions, above all if pronominal elements are included into the analysis. Several examples have apparently redundant overt subjects, both in preverbal and postverbal position:

- (7) *Yo de... de hecho a León yo no he vuelto o sea... / yo dije que*
 I in in fact to León I not have returned that is I said that
yo a León... / se lo dije en serio a un juez [...]
 I to León him it said.1SG seriously to a judge
 ‘In fact, I haven’t returned to León, that is, I said that, to León, ... Seriously, I said it to a judge [...]

(PRESEEA; Madrid_H12-007)

- (8) *digo “oye tú” digo “el...” / se lo dije a mi sobrina digo yo*
 say.1SG listen you say.1SG the her it said.1SG to my niece say I
 / “el padre de... / G” / digo “es que es... / es de gente noble” porque [...]
 the father of G say.1SG is that is is of people noble because
 ‘(I) say “listen” (I) say “the...” (I) said it to my niece I say “G’s father” (I) say
 “the thing is that he belongs to high society” because [...]

(PRESEEA; Alcalá_M33-018)

The use of some instances of overt strong pronouns in 1SG seem to be related to perspectival notions such as subjectivity and epistemicity/evidentiality (Posio 2012,

2015, Herbeck 2021 and references) rather than to the notions of topic and focus. Furthermore, even though some subjective uses of strong pronouns might be related to a subtype of contrast (Herbeck 2021), the operationalization of contrast in the investigation of spoken corpus data has been shown to be particularly problematic (see Travis & Torres Cacoullos 2012 for discussion).

Furthermore, several instances of VS with *decir* show a degree of fixation and it is the construction as a whole that fulfils a particular discourse function and not just the subject:

- (9) *le estoy comentando que la educación es importante / para la*
 you.DAT am commenting that the education is important for the
erradicación de esa de esa lacra ¿no? Digo yo ¿no? ¿eh?
 eradication of that of that evil no say.1SG I no eh
 ‘I’m saying that education is important to eradicate that evil, right? That’s what I say, right?’
 (PRESEEA; Madrid_H32-043)

Here, *digo yo* ‘(I) say’ is not used as a verb of saying but, rather, as a marker of intersubjectivity (cf. Hennemann 2016 for *creo yo* ‘I think’). Furthermore, several instances of DECIR + *yo*, such as *como digo yo* ‘as I say’ in (10) or *diría yo* ‘I would say’ in (11) are used parenthetically and they seem to trigger a shift towards a subjective evaluation:

- (10) *el trabajo que tengo me gusta / soy de la gente como digo yo*
 the work that have.1SG me like am of the people as say.1SG I
agraciada
 lucky
 ‘I like my job. I am one of the, as I say, ‘lucky’ people’
 (PRESEEA; Alcalá_H11-037)

- (11) *no sé es muy oscura es como no sé /*
 not know.1SG is very dark is like not know.1SG
un tugurio diría yo no sé no me gusta
 a shack would-say I not know.1SG not me like
 ‘I don’t know, it is very dark, it is like, I don’t know, a shack I would say, I don’t know. I don’t like it.’
 (PRESEEA; Alcalá_M12-023)

These structures are frequent in spoken discourse and, thus, they must be considered in a quantitative analysis of word order variation in spoken Spanish in order to offer a full picture of the phenomenon.

Lastly, it has been argued that some configurations of Peninsular Spanish do not only favor, but obligatorily require subject-verb inversion (Fernández Soriano 1993, Zubizarreta 1998, Goodall 2004, Holmberg 2015, and Leonetti 2017 for an overview). In these cases, subject-verb inversion is not triggered by focus on the subject. Some of these VS orders have been argued to be due to a “residual V2” effect (cf. Rizzi 1996, Holmberg 2015 and references). Note that the asterisks refer to the agrammaticality or unacceptability of the examples in Peninsular Spanish: for

example, (12) would be acceptable in Caribbean Spanish varieties where interrogative words do not necessarily trigger VS order (Gutiérrez-Bravo 2008, Zimmermann 2019). In the remainder of this paper, we will focus only on the Peninsular varieties of Spanish.

(12) *wh*-fronting:

- a. *¿*Qué Juan compró?*
what Juan bought.3SG
- b. ¿*Qué compró Juan?*
what bought.3SG Juan
'What did Juan buy?'

(Goodall 2004: 101)

(13) focus fronting:

- a. *CON MARÍA habló Pedro (y no con Marta).*
With María spoke.3SG Pedro (and not with Marta)
'Pedro spoke with María (and not with Marta).'
- b. **CON MARÍA Pedro habló (y no con Marta).* (Zubizarreta 1998: 103)

(14) resumptive preposing:

- Eso creo yo.* (Leonetti 2017: 909)
this think.1SG I
'This is what I think.'

(15) locative inversion:

- a. *Aquí llega el tren.*
Here arrives the train
'Here comes the train.'
- b. *Aquí, el tren llega.*
- c. **Aquí el tren llega.* (Leonetti 2017: 911)

(16) quotative inversion:

- a. *No es un enanito – rectifica el viejo.*
not is a dwarf corrects the old-man
'It's not a dwarf – corrects the old man.'
- b. **No es un enanito – el viejo rectifica.*
(Matos 2013, cited in Leonetti 2017: 896 [adapted & emphasis added])

The term 'residual V2' refers to the assumption that such word orders are a residue of the verb second order having been productive in older stages of (some) Romance languages. In fact, Medieval Spanish has been argued to have been a V2 language (see Fontana 1993, Pinto 2015 and references for discussion). Configurations such as (12)–(16) would thus be due to a residual effect of a phenomenon that can be observed in V2 languages like German, in which a preverbal XP (a fronted XP or an adverbial phrase) triggers the obligatory postverbal position of the subject:

- (17) a. *Gestern kam der Zug zu spät.*
 yesterday came.3SG the train too late
 ‘Yesterday, the train arrived late.’
 b. **Gestern der Zug kam zu spät.*

However, given that Medieval Spanish also sanctions V1, V3, V4, etc. orders, its status as a V2 language has been questioned (Elvira 2015, Batllori 2015; see also López Izquierdo & Castillo Lluch 2015 for discussion and references). In addition to the purported historical features, the term “residual V2” is also used to refer to the synchronic existence of apparent V2 in only a subset of specific configurations (such as those in (12) to (16)). In the present paper, we adopt this synchronic usage of the term.

From a generative perspective, two main proposals have been put forth to explain obligatory subject-verb inversion. The first approach deals with obligatory subject-verb inversion in *wh*-interrogatives in terms of V-to-T-to-C movement (cf. Rizzi & Roberts 1989, Rizzi 1996, among others). This way, only one position (Spec,CP) is available in interrogatives, which is occupied by the fronted *wh*-element:

- (18) [CP *Qué_j* C-*compró_i* [TP *Juan_k* T-*compró_i* [vP *Juan_k* *compró_i* ...*qué_j*]]]
 what bought.3SG Juan
 ‘What did Juan buy?’

V-to-C movement forces the subject to be in postverbal position either in Spec,TP or in Spec,vP. However, some problems for such an approach have been discussed in the literature (see Suñer 1994, among others). One problem is that it fails to provide an explanation for the existence of apparent V2 effects also in subordinate clauses in Spanish (Zubizarreta 1998, Barbosa 2001), which are absent from a V2 language like German:

- (19) a. **No sé qué cosa María comió.*
 not know.1SG what thing María ate.3SG
 ‘I don’t know what María ate.’
 b. **No sé a quién María invitó.*
 not know.1SG P who María invited.3SG
 ‘I don’t know who María invited.’ (Zubizarreta 1998: 105)

Furthermore, while V-to-C movement can, in principle, account for inversion in interrogatives, the approach is less straightforward for an explanation of other obligatory subject-verb inversion patterns, such as Focus Fronting, resumptive preposing, locative inversion and quotative inversion.

Another approach that has been proposed is that the verb does not move to C, but it moves to INFL just as in declarative clauses.² The difference to English is that not only subjects, but also *wh*-elements (Solà 1992, Goodall 1993) and any type of focal XP can be moved to Spec,TP (Zubizarreta 1998, Zagana 2002, among others). If we follow this approach, the EPP on T in Spanish is not obligatorily satisfied by nominative subjects, but the EPP is topic/focus-related (Zagana 2002, see also Gallego

² For a discussion of two approaches to obligatory subject-verb inversion in terms of V-to-I or V-to-C movement in Old Spanish, see e.g. Pinto (2015) and references.

2010 and Herbeck 2015 for discussion and references).³ Also within these approaches, it is an open issue whether Spec,TP is the only position available for preverbal XPs (i.e. CP is not projected) or whether CP is projected and Spec,CP is available under certain conditions:

- (20) ([_{CP} C) [_{TP} *Qué_j* T-*compró_i* [_{vP} *Juan* v-*compró_i* ... *qué_j*]]
 what bought.3SG Juan

In this paper, we will have a look at variable word order patterns with the verb *decir* ‘say’ in data from speech corpora in the light of apparent residual V2 properties. At the same time, we aim at describing and modeling the word order patterns found in actual language use and the connection of these patterns with different discourse functions of this particular verb. The latter point is particularly important with the verb *decir* ‘say’ because several word order patterns seem to correspond to specific discourse functions related to speaker/hearer interaction and (inter-)subjectivity, rather than to information structure and residual V2 effects.⁴ Such discourse-pragmatic patterns are typical of highly frequent verbs like *decir* (Posio 2015).

In a generative architecture, this can be encoded if, apart from TopP and FocP, ‘perspectival’ functional categories in the left periphery (such as an EvaluativePhrase and a Speech Act Phrase; Cinque 1999, Speas & Tenny 2003) can be targeted by subjects with certain verbs (see §4.3).

1.2. Variable word order in language use and the case of *decir* ‘say’

While in generative studies of subject expression and subject-verb inversion, a main focus has been on the formal conditions that sanction word order patterns (such as agreement morphology and properties related to the C-T spine), functional studies have focused on the specific discourse-pragmatic uses that subject expression and word order patterns serve. In some functional studies, the term “variable subject expression” is therefore preferred over the term of “pro-drop” or “null subject” language (see Posio 2012). Given the variable nature of subject expression in Spanish, any account of variation in the placement of subjects must also consider why the subject is expressed if it could have been omitted. While the present paper focuses only on sentences with the verb *decir* ‘say’ and expressed subjects, a brief overview of the factors affecting subject expression is thus in order.

Subject expression is governed by slightly different tendencies depending on the grammatical person and whether the subject is a pronoun or a noun phrase. In general, third-person referents need to be introduced into the discourse as lexical noun phrases before they can be referred to by pronouns and, as predicted by the Preferred Argument Structure model (DuBois 2003), in languages with SVO as their basic word order, discourse-new referents tend to be placed in the postverbal position as objects of transitive verbs or subjects of intransitive verbs, while discourse-old referents are more prone to occur in the preverbal subject position and as subjects of transitive verbs. In Spanish, third-person subjects may be omitted if they are accessible to the

³ In Gallego’s (2010) *phase sliding* approach, V-to-T movement has the effect of extending the vP phase level to TP, making Spec,TP a position with A’-properties.

⁴ See Herbeck (2021) for an account of the overt/covert alternation of subjects in Spanish which takes into account the notions of epistemicity, evidentiality and subjectivity with certain verb forms, such as (*yo*) *creo que* ‘(I) think that’.

addressee, i.e., their referent can be identified in the context, but they are usually expressed if they do not fulfill this criterion.

First and second person subjects are, in principle, always omissible since their referent is present in the communicative situation. However, studies have shown that subject expression tends to be more common precisely in first and second person singular, which suggests that subject expression in these persons serves other functions than referent tracking. In particular, postverbal placement of pronominal subjects is often associated with special pragmatic functions and may be related to constructionalized or (semi)formulaic sequences (see Posio & Rosemeyer 2021 for an overview). Although the most frequent and canonical constituent order in Spanish is SVO, postverbal subjects are not particularly infrequent either, representing ca. 20–40% of all subjects, depending on the dialect, medium and discourse type (see Bentivoglio & Weber 1987, Meyer-Hermann 1990, Bentivoglio 2003). Pronominal subjects typically occur much more rarely in the postverbal position than lexical subjects, and the verbs that favor postverbal subject placement are typically cognitive and speech act verbs (in particular, *decir* ‘say’; Posio 2012) as opposed to unaccusative and existential verbs that favor postverbal placement in the case of lexical subjects (Fernández Soriano 1993).

There is ample evidence to the effect that frequency of occurrence affects the collocational and morphosyntactic patterns of lexical items, particularly when high token frequency is combined with low type frequency (Bybee & Thompson 1997). For instance, frequently occurring verb forms have been shown to display patterns of subject expression different from the tendencies found with less frequent items (Erker & Guy 2012). Posio (2015) further suggests that patterns of subject expression and position are not affected by frequency of use alone, but rather by the combination of frequency and discourse-pragmatic uses of certain very frequent lexical verbs such as *decir* ‘say’, *creer* ‘think’ or *saber* ‘know’, all characterized by high token frequency and low type frequency, whereas other frequently used verbs such as auxiliaries and copulas do not display such specific patterns.

According to Posio & Pešková (2020), while the verb *decir* ‘say’ is associated with a lower-than-average rate of expressed subjects in Peninsular Spanish, a higher-than-average rate of these expressed subjects are postverbal. These authors report a subject expression rate of 23% for this verb in their Peninsular Spanish corpus, and 39% of pronominal subjects and 63% of lexical subjects of *decir* in their data are postverbal (Posio & Pešková 2020: 402).⁵ In Peninsular Spanish, postverbal subject placement with *decir* is often associated with the discourse-pragmatic functions of this verb in introducing both indirect and direct reported speech (example (21)). It also occurs with particular pragmatic uses like the (inter)subjective *digo yo* that serves as a hedge limiting the speaker’s responsibility or access to the information, or as a marker of subjective stance (example (23)). Different forms of the verb *decir* without expressed subjects have also grammaticalized into pragmatic markers like the corrective *digo* (example (24)) and the quotative particles *digo* ‘I say’ and *dice* ‘s/he says’ that are used in Peninsular Spanish to indicate the beginning of reported speech

⁵ This contrasts strikingly with their Argentinian data, where the rate of expressed subjects is slightly higher (27%) but only 12% of pronominal and 16% of lexical subjects of *decir* are postverbal. These results point at dialect-specific constraints in the expression and placement of subjects that may override general tendencies governing subject placement, related with, e.g., information structure and accessibility of the subject referent.

even when the verb *decir* or another speech act verb already occurs in the sentence (see the use of *digo* in example (22); Posio & Pešková 2020).

a. Quotative use of *decir* + postverbal lexical subject with indirect reported speech:

- (21) *Me ha dicho mi hermana que tenemos que ir a finales de mayo.*
 To-me has said my sister that (we)have to go.INF at end of
 may
 ‘My sister has said to me that we have to go at the end of may.’
 (CORLEC, Posio & Pešková 2020: 399)

b. Quotative use of *decir* + postverbal pronominal subject with direct reported speech:

- (22) *Dice: “bueno, pues na”. Le dije yo, digo: “Oye, mañana te podría yo localizar en algún sitio”*
 says well well nothing to-him said I say.1SG listen.IMPER
 tomorrow you.ACC could I locate.INF in some place
 ‘He says: “well, nothing then”. I said to him: “Look, tomorrow I could meet you at some place.”’
 (CORLEC, Posio & Pešková 2020: 405)

c. (Inter)subjective use of *digo yo*

- (23) *Será que ahora colecciona bastones. Digo yo, no, no tengo ni idea, vamos.*
 be.FUT that now collect.3SG sticks say.1SG I not not have.1SG
 nor idea go.1PL
 ‘Maybe now he collects walking sticks. Or that’s what I think, I don’t, I don’t have an idea, you know.’
 (CORLEC, Posio 2012: 173)

d. Corrective use of *digo*

- (24) *—No, ese color no me gusta.*
 no that color not to-me please.3SG
—¿Larga? Digo, ¿blanca?
 long say.1SG white
 ‘—No, I don’t like that color.
 —Long? I mean, white?’
 (CORLEC, Posio & Pešková 2020: 399)

As argued by Posio & Pešková (2020), many of such pragmatic uses of *decir* ‘say’ in speech show different grades of fixation towards formulaic sequences or even

quotative particles.⁶ This can be observed very clearly in the case of (23), where 1st person singular *digo yo* ‘I say’ with a postverbal subject pronoun correlates with an intersubjective use (see Posio 2012) and an epistemic, rather than communicative meaning, similarly to the same word order with the cognitive verb *creer* ‘think’ (i.e. *creo yo* ‘(I) think’).

In Section 4.3, we will discuss how the existence of semi-fixed word order patterns with the verb *decir* ‘say’ could be encoded in a generative framework making use of perspectival functional projections in the left periphery above TP (Cinque 1999, Speas & Tenny 2003, Cruschina 2015, Herbeck 2021).

2. Research questions, data and methodology

2.1. Research questions

Our study attempts to find out what factors govern subject-verb inversion with *decir* ‘say’. We consider the properties of the subject (type of nominal phrase), the subtype of *decir* ((non-)quotative uses, etc.) which is closely related to the type of complement the verb takes (e.g. direct quotes, subordinate clauses, etc.), as well as patterns related to residual V2, such as sentence type and the presence of preverbal XPs.

Given the findings in previous literature (see §1.1), we would expect that sentences with preverbal XPs favor subject-verb inversion. However, it is unclear whether all types of preverbal phrases (e.g., preverbal adverbials vs. fronted objects) have the same effect on word order patterns. We furthermore expect subject-verb inversion to be more frequent with lexical subjects than with personal pronouns, given the obligatory definiteness of the latter.

With respect to sentence type, it has been suggested that subordinate clauses trigger certain subject-verb inversion patterns more easily than main clauses (see Gutiérrez-Bravo 2005 for a discussion of relative clauses). As for complement type, it has been observed that, in Peninsular Spanish, *decir* as a verb introducing direct reported speech is associated with particularly high frequencies of postverbal subjects (see Posio & Pešková 2020). However, some studies posit that the ‘size’ or ‘weight’ of the subject and object also influence word order. Thus, ‘large’ or ‘heavy’ elements tend to be placed after the verb, be them subjects or clausal complements, and the presence of a ‘heavy’ object such as a clausal complement after the verb may in turn trigger the preverbal placement of the subject (Silva-Corvalán 1982, Cuartero Sánchez 2007). Gabriel (2007) furthermore argues that the presence of object clitics before the verb favors the postverbal position of subjects so that this factor should be controlled for.

In the study presented in the following sections, we intend to examine the interaction of these factors in order to account for subject-verb inversion patterns with the verb *decir*.

⁶ This is particularly salient in the case of the corrective discourse particle *digo* (cf. Martín Zorraquino & Portolés 1999), which should be excluded from studies focusing on variable expression and placement of subjects, as it does not present such variation.

2.2. Methodology

2.2.1. The data: CORLEC and PRESEEA

Our data come from two corpora of spoken Peninsular Spanish. Most of the data were extracted from the PRESEEA (*Proyecto sociolingüístico para el estudio del español de España y América*) corpus of spoken Spanish (PRESEEA 2014–), more precisely from the subcorpora representing speakers from Alcalá de Henares (ca. 155,000 words) and Madrid (ca. 215,000 words). This selection was complemented by spoken data from the subcorpora consisting of “interviews” and “conversations” of the CORLEC (*Corpus oral de referencia del español contemporáneo*; Marcos Marín 1992) and contains ca. 69,000 words. PRESEEA is a corpus of transcribed semi-structured sociolinguistic interviews. This means that the data represent a very specific type of interaction with highly asymmetric discourse roles: an interviewer who asks questions and provides feedback, and an informant who is expected to answer the questions and elaborate on the topics chosen by the interviewer but is not expected to ask questions back. The CORLEC data, on the other hand, is more varied in that it contains some naturally occurring conversational exchanges between two or more speakers, but it also comprises interviews from radio and television that resemble more closely the sociolinguistic interviews of the PRESEEA corpus in terms of interactional characteristics. The nature of our data naturally limits the kind of generalizations that can be made, for instance, regarding spontaneous, colloquial conversations. In Section 4.2, we will discuss some properties of the different data types in the two corpora that might affect subject-verb inversion. Furthermore, our conclusions can only be limited to the varieties of Peninsular Spanish as represented by our data.

Our PRESEEA data contain the turns of the informants as well as of the interviewers. An important point is that the structure of the data made it necessary to restrict our analysis of subject-verb inversion to declaratives. Given that most of the interrogatives were produced in the less frequent and shorter turns of the interviewers in the sociolinguistic interviews of PRESEEA, they do not provide enough data for a quantitative analysis (see Table 1):

Table 1. Number of analyzed sentences containing *decir* ‘say’ plus an overt subject

sentence type	number
main declarative	333
subordinate	152
interrogative	35
total	520

Given the low number of interrogatives in the corpus sample, we restrict the quantitative analysis to declarative main and subordinate clauses. However, since *wh*-interrogatives have been singled out as one main context of residual V2 in the literature on Peninsular Spanish, we will nevertheless offer a qualitative discussion of these sentences in Section 4.

2.2.2. Variables and coding

Our study investigates the dependent variable of ‘subject position’ (preverbal vs. postverbal) with respect to the independent variables of (i) preverbal XP type, (ii)

subject type, (iii) sentence type, (iv) complement type, and (v) presence of indirect object (IO) clitics.

With respect to (i) preverbal XP type, we examine whether the presence of a preverbal (non-clitic) phrase has any effect on subject-verb inversion, as would be predicted by (residual) V2 effects. Within this category, we differentiate between adverbials (i.e., adverbs, PPs, and adverbial clauses), discourse markers, fronted XPs, and configurations in which no phrasal XP precedes the verb.⁷ If more than one XP precedes the verb, we consider only the constituent closest to the verb. Thus, in example (25), the preverbal XP *pues* ‘well’ was annotated as “DM” (= discourse marker), while in the example (26), it was annotated as “adverbial”, given that it is the adverb *siempre* ‘always’ that appears closest to the verb, not the preceding DM *hombre*, literally ‘man’:

- (25) *y entonces el arquitecto de Villalbilla pues dijo* “[...]”
 and then the architect of Villalbilla well said.3SG
 ‘And then the architect of Villalbilla, well, said to me “[...]”’
 (PRESEEA; Alcalá_H33-051)

- (26) *hombre mis amigas siempre me dicen “¿si te*
 man my friends always me say.3PL if you.DAT
pasara algo?”
 happened something
 ‘Well, my friends always say to me: “And what if something happened to you?”’
 (PRESSEA-Madrid_H32-043)

The following demonstrates an example with a fronted XP, here the demonstrative pronoun *eso* ‘that’ – a case of resumptive pronoun fronting (Leonetti 2017 and references):

- (27) *¡pues eso digo yo / qué suerte tuve!*
 well that say.1SG I what luck had.1SG
 ‘Well that’s what I said, I had so much luck!’
 (PRESEEA-Madrid_M33-054)

As for (ii) subject type, we differentiate between lexical NPs (including indefinite NPs such as *una madre* ‘a mother’, definite NPs such as *el médico* ‘the doctor’, and proper names such as *Juan*) and pronouns (including personal pronouns like *yo* ‘I’ and indefinite ones, e.g., *uno* ‘one’, *alguien* ‘somebody’, *nadie* ‘nobody’). Turning to (iii) sentence type, we make a distinction with respect to whether *decir* occurs in a main clause (including coordinate clauses) or a subordinate clause.

In the category of (iv) complement type, we annotated whether the complement of *decir* is a direct quote, a subordinate clause, an NP, a clitic, a relative pronoun /clause, or whether *decir* appeared without a complement. This category is crucial because the different uses of *decir* in each context (direct reported speech marker, main verb, intransitive uses, etc.; see §1.2) largely depends on the type of complement it

⁷ Negation and complementizers were not counted as preverbal XPs, given their non-phrasal character.

takes. It is important to note that the category “no complement” in its majority contained parenthetical uses like *como digo yo* and *como yo digo* ‘like I say’ (see (28) and discussion in §1.1).

(28) a. direct reported speech:

Porque me... al irme, me dice Alberto, dice: “[...]”
because me at-the go.INF-me me says Alberto says (CORLEC)
‘Because to me... when I was leaving, Alberto says to me, he says: “[...]”’

b. indirect reported speech (subordinate clause):

entonces yo eso ...yo les dije que me podían llamar de tú [...]
then I this I them said that me could call of you
‘Then I this... I said to them that they could call me of your [...]’
(PRESEEA- Madrid_H23-033)

c. NP complement:

pues yo te voy a decir una cosa
well I you go to say.INF a thing
‘Well, I’m going to tell you something’
(PRESEEA-Alcalá_M21-047)

d. clitic complement:

espero a que ellos me lo digan
wait.1SG to that they me it say-3PL.SBJV
‘I’ll wait until they say it to me’
(PRESEEA-Madrid_M11-004)

e. relative pronoun/clause (including pseudo-clefts):

[...] y lo que tú dices [...]
and what you say
‘And what you say [...]’
(PRESEEA- Alcalá_M21-047)

f. no complement:

ni madrugo ni trasnocho como digo yo
neither get-up-early.1SG nor stay-up-all-night.1SG like say I
‘I’m neither an early bird nor a night owl, like I use to say.’
(PRESEEA- Alcalá_H21-043)

Lastly, while the presence of a direct object (DO) clitic is integrated into the category of complement type, (v) indirect object (IO) clitics were annotated as a separate factor because they can occur both with and without direct objects in the same clause.

2.2.3. Statistical analysis

Our study analyzes whether the predictors of subject type (pronoun vs. lexical), sentence type (main vs. subordinate), complement type (none, relative pronoun/clause, clitic, NP, indirect or direct reported speech), preverbal XPs (no XP, adverbial, discourse marker, fronted XP) and the presence or absence of an IO clitic have an effect on subject-verb inversion with the verb *decir* ‘say’. All variables are categorical, and the dependent variable of subject position is binary: preverbal vs. postverbal. We

apply a binomial logistic regression model in R (R Core Team 2018). The reference category for the dependent variable of subject position is set to ‘preverbal’. For the independent variables, the most frequent type was taken as reference category: for complement type ‘direct reported speech’, for sentence type ‘main clause’, for subject type ‘pronoun’, for preverbal material ‘none’, and for IO clitics ‘none’. Throughout Section 3, we indicate the relevant coefficients (in logits) and *p*-values for the analysis of subject-verb inversion.

3. Results

In this section, we present the main results of the effect of the variables described in Section 2.2.2 on subject-verb inversion with the verb *decir* ‘say’. First, we will deal with the most crucial factor for residual V2 effects: the presence or absence of a preverbal XP. Thereafter, we turn to the factors of subject type, sentence type, complement type and the presence or absence of IO clitics.

3.1 Preverbal material

Table 2 presents the main results of preverbal and postverbal subjects according to different types of preverbal XPs or absence thereof:

Table 2. Subject-verb inversion according to the presence of a preverbal XP

preverbal XP	subject position			% -post
	preverbal	postverbal	all	
no XP	174	148	322	46%
adverbial	63	23	86	27%
DM	54	9	63	14%
fronted XP	3	11	14	79%
total	294	191	485	39%

As can be seen in Table 2, subjects are most frequently postverbal with the verb *decir* ‘say’ if a fronted XP appears before the verb ($11/14 = 79\%$), as in examples (29) and (30) that both have the pronoun *eso* ‘that’ as their fronted direct object:

(29) *eso te iba a decir yo*
 that to-you went.1SG to say.INF I
 ‘That’s what I was going to say to you’ (PRESEEA-Alcalá_M32-035)

(30) *bueno, eso dicen los medios de comunicación*
 well that say.3PL the media
 ‘Well, that’s what the media says’ (PRESEEA-Madrid_H32-043)

According to the statistical analysis outlined in Section 2.2.3, the probability of a postverbal subject significantly increases with preverbal fronted XPs (coefficients in logits: 1.757; $p = 0.015$; reference category: ‘no preverbal material’). With adverbial elements and discourse markers, on the contrary, the probability of a postverbal subject decreases when compared to configurations without preverbal material. Thus, the

subject is postverbal in 23/86 cases (= 27%) with a preverbal adverbial phrase. The difference to configurations without preverbal material is significant, the probability of a postverbal subject being lower if an adverbial phrase precedes the verb (coefficient in logits: -0.804, $p < 0.01$; reference category = ‘no preverbal XP’). In the case of preverbal discourse markers, we have the lowest frequency of postverbal subjects: 9/63 = 14% (coefficient in logits: -1.578; $p < 0.001$; reference category = ‘no preverbal XP’).

The results of our study thus show that in the spoken peninsular Spanish data examined here, neither preverbal adverbial phrases nor discourse markers occurring in the preverbal position favor the VS order, if compared with configurations without preverbal phrases. Importantly, those configurations without any preverbal XP have a high frequency of postverbal subjects (148/322 = 46%). Fronted preverbal objects favor subject-verb inversion to the highest degree (11/14 = 79%). Note furthermore that the only three cases of preverbal subjects co-occurring with fronted XPs were preverbal IOs:

- (31) *No; yo yo, a Lola Gaos le a... le diría que [...]*
 no I I to Lola Gaos her to her would-say.1SG that
 ‘No, I, I, to Lola Gaos I... I would say to her that [...]’ (CORLEC)

If we split the category of preverbal subjects into adjacent and non-adjacent ones, it can be observed that preverbal adverbial phrases trigger dislocation, rather than inversion, of the subject (see Table 3). The category of ‘adjacent’ preverbal subjects includes all cases in which the subject appeared adjacent to the verb or only a clitic element (and negation) intervened between them.⁸ Non-adjacent preverbal subjects are those in which a non-clitic constituent intervened between the verb and the subject.

Table 3. Type of preverbal material and position of subjects

preverbal XP	preverbal (adjacent)	preverbal (non-adjacent)	postverbal	all
adverbial	21 (24%)	42 (49%)	23 (27%)	86
DM	43 (68%)	11 (17%)	9 (14%)	63
fronted XP	1 (7%)	2 (14%)	11 (79%)	14

Preverbal discourse markers, in contrast to adverbial elements, are not associated with high frequencies of either postverbal or non-adjacent preverbal subjects.

In the next section, we will have a look at the results with respect to the type of subject.

3.2. Type of subject

The following Table 4 presents the results of subject-verb inversion with respect to the type of subject:

⁸ We use the terms ‘adjacent’ and ‘non-adjacent’ instead of ‘dislocated’ and ‘non-dislocated’ because of the potential ambiguity of preverbal ‘adjacent’ subject with respect to their dislocation status.

Table 4. Subject-verb inversion and type of subject

	preverbal	postverbal	total
pronoun	202 (70%)	87 (30%)	289
lexical	92 (47%)	104 (53%)	196
all	294 (61%)	191 (39%)	485

Lexical DP subjects are postverbal more frequently (53%) than pronominal subjects (30%). According to the statistical analysis outlined in Section 2.2.3, this difference is statistically significant, the probability of a postverbal subject increasing with lexical subjects (coefficient in logits: 1.301, $p < 0.001$; reference category = ‘pronoun’).

3.3. Sentence type, complement type, and preverbal IO clitics

The following Table 5 shows the frequencies of preverbal and postverbal subjects according to sentence type:

Table 5. Subject-verb inversion and sentence type

	preverbal	postverbal	total
main clause	218 (65%)	115 (35%)	333
subordinate	76 (50%)	76 (50%)	152
all	294 (61%)	191 (39%)	485

As can be seen, the frequency of postverbal subjects is higher in subordinate than in main clauses (50% vs. 35%, respectively), but the difference is not statistically significant ($p = 0.46$).

Looking at the effect of complement type, the highest frequency of postverbal subjects arises in those sentences where *decir* does not take any complement (58% postverbal subjects). The second most frequent case of postverbal subject placement are those sentences where *decir* occurs with a DO clitic complement (= 55% postverbal subjects).

Table 6. Subject-verb inversion according to complement type of *decir*

type of complement	pre	post	total	%-post
no complement	26	36	62	58%
DO clitic	10	12	22	55%
relative pronoun	21	18	39	46%
NP	14	10	24	42%
direct reported speech	145	80	225	36%
indirect reported speech	78	35	113	31%
all	294	191	485	39%

However, departing from the reference category “direct reported speech” (i.e., *decir* used as a direct reported speech marker), the only significant type of complement for subject-verb inversion is the category of ‘no complement’ (coefficient in logits: 0.908, $p < 0.01$). This means that, even though other complement types (NP, DO clitics, and relative pronouns) co-occur with higher frequencies of postverbal subjects than the most frequent complement type of “direct reported speech”, this tendency observed in the data would have to be tested against a larger data set in future studies.

It should be stressed that the category “no complement” also comprises constructions with *como* ‘like’ of the following type (32):

- (32) *el trabajo que tengo me gusta / soy de la gente como digo yo*
 the job that have.1SG me pleases am of the people like say I
agraciada
 lucky
 ‘The job that I have... I like it. I’m one of those, like I say, lucky people.’
 (PRESEEA, Alcalá_H11-037)

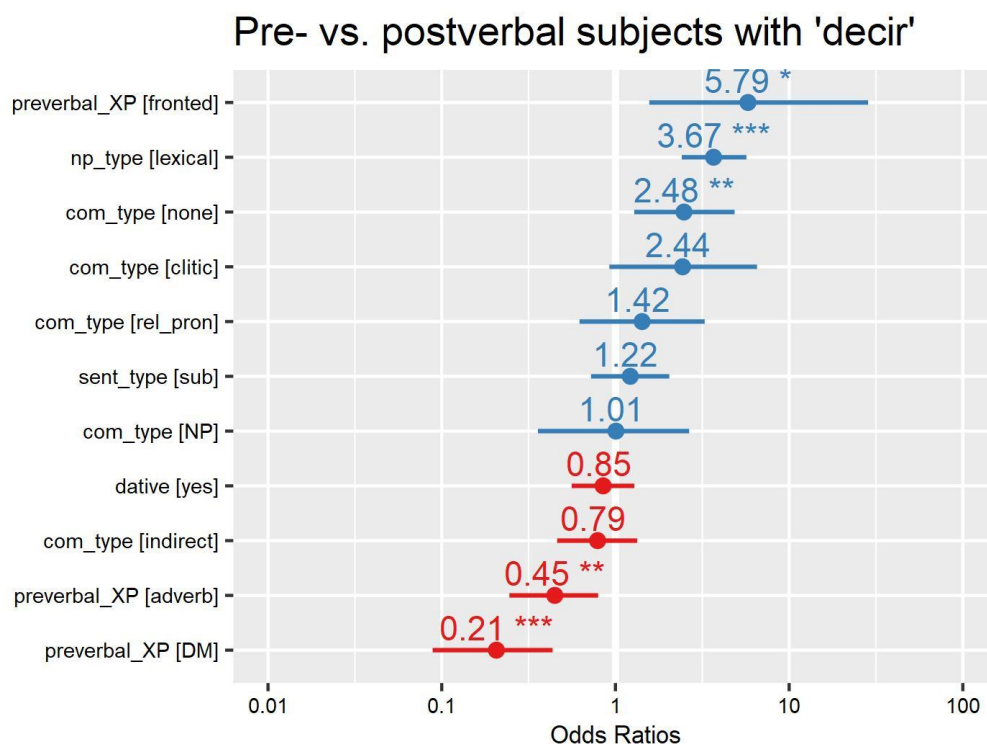
The semi-formulaic constructions *como* DECIR X ‘as X say(s)’ and *eso* DECIR X ‘that’s what X say(s)’ (see examples (29) and (30)) are thus frequent contexts of occurrence for postverbal subjects with the verb *decir*.

Turning to the presence or absence of a preverbal IO clitic, we do not observe any significant effect on subject-verb inversion in our data ($p = 0.44$). The frequencies of postverbal subjects are very similar, comparing clauses containing an IO clitic (= 37% postverbal subjects) and lacking it (= 41% postverbal subjects).

3.5. Interim summary

Figure 1 depicts the influence of each factor on subject-verb inversion with *decir* ‘say’ according to the statistical model outlined in Section 2.2.3.⁹

Figure 1. Preverbal and postverbal subjects with *decir*



⁹ The figure has been generated by means of the sjPlot package (Lüdtke 2021).

As Figure 1 shows, subject-verb inversion is significantly favored if an XP is fronted to preverbal position. On the contrary, preverbal subjects are significantly favored with adverbial elements (adverbs, PPs, and adverbial clauses) and discourse markers appearing before the verb. Furthermore, postverbal subjects are significantly favored if *decir* does not take any complement and if the subject is a lexical DP (in comparison to a pronoun).

4. Discussion and theoretical consequences

4.1. Residual V2 with *decir* and the preverbal field

One of our main research questions was to find out whether the occurrence of preverbal, phrasal material affects subject-verb inversion and, thus, whether high frequencies of postverbal subjects with the verb *decir*, noticed in previous literature, could be related to a residual V2 effect. Our results do not show a general preference of VS orders with a preverbal XP (see Table 2 and Figure 1). On the contrary, preverbal adverbial elements trigger peripheral preverbal subjects rather than subject-verb inversion and discourse markers favor preverbal adjacent subjects. The only type of preverbal material associated with frequent VS order are fronted XPs. This indicates an important difference with respect to canonical V2 languages like German, where adverbs as well as fronted XPs trigger the VS order.

As mentioned earlier, the data we analyzed do not allow carrying out a quantitative analysis of subject-verb inversion in *wh*-interrogatives with the verb *decir* – one potential context for ‘residual V2’: interrogatives were mainly produced by the interviewers of the sociolinguistic interviews and constitute only 35 of 520 sentences in total. Out of these, 24 were classified as *wh*-interrogatives. However, an interesting pattern can be observed in these sentences (see Table 7).

Table 7. Subject position in *wh*-interrogatives with *decir*

	preverbal (adjacent)	preverbal (non-adjacent)	postverbal	total
<i>wh</i> -interrogatives	0 (0%)	9 (37.5 %)	15 (62.5%)	24

As Table 7 shows, preverbal adjacent subjects are absent with *wh*-interrogatives in our data of Peninsular Spanish, as is expected if we are dealing with a residual V2 effect. However, it is interesting to observe that speakers use the strategy of left-dislocation or clause external location of the nominative pronoun/DP to a high extent (= 37,5 % of the sentences):

- (33) E: *¿pero tú qué dirías?*
 but you what would-say.2SG
 ‘E: But you, what would you say?’ (PRESEEA Madrid_M32-047)
- (34) y *¿tu madre qué dijo?*
 and your mother what said.3SG
 ‘And your mother, what did she say?’ (PRESEEA Madrid_M13-018)

- (35) *y al final dice: “Bueno, ¿y éste qué dice? [...]”.*
 and at-the end says well and this-one what say.3SG
 ‘And at the end, s/he says: “Well, and what does this one say? [...]”’
 (CORLEC)

Judging from these data, it seems to be the case that, even though the low preverbal subject position (Spec,IP) cannot be targeted in *wh*-interrogatives, a usual strategy in speech data is situating the subject before the *wh*-element in a high topic position:

- (36) [_{TopP} Tú [_{IP} qué T-dirías [_{VP} tú ~~v-dirías~~ [... qué]]]]
 you what would-say

In fact, it seems to be the case that the peripheral nominative pronoun or DP appears above all in shifting contexts (for example, at the beginning of a new turn) and shifting topics have been argued to be high in the structure (see Frascarelli & Hinterhölzl 2007). However, given that the low number of data points does not allow definite conclusions, more research is necessary to determine the exact syntactic position of these nominative pronouns/DPs, either in the CP-internal left periphery or in a clause external position.

Thus, the most plausible analysis of sentence structure for *decir* ‘say’ in spoken Spanish is an analysis in which focus fronted constituents and fronted resumptive pronouns can target the low (IP-related) position for preverbal subjects (see Zubizarreta 1998), making this position unavailable for subjects in *wh*-interrogatives. However, speakers make use of a high position before the *wh*-element:¹⁰

- (37) *wh*-interrogative:
 Y [_{TopP} (tú) [_{IP} qué dices [_{VP} (tú) ~~dices~~ ...]]]
 and you what say.2SG you

- (38) declarative:
 [_{TopP} (Tú) [_{IP} (tú) dices [_{VP} (tú) ~~dices~~ ...]]]
 you you say.2SG

A similar reasoning applies to the high frequency of postverbal subjects with fronted XPs: These cases of resumptive pronoun fronting confirm the observation in the literature that they trigger postverbal subjects: in our data, all cases of the fronted resumptive pronoun *eso* trigger a postverbal subject, indicating that the two elements compete for the same (low) preverbal subject position:

- (39) [_{IP} *Eso* I-digo [_{VP} yo ~~v-digo~~ [_{VP} digo *eso*]]]
 that say.1SG I

¹⁰ In the literature, it has been argued that preverbal subjects can be located either in a low IP-related or high C-related topic position (see e.g. López 2009 for discussion). However, in the case of *wh*-interrogatives, more research is necessary whether preverbal nominative pronouns and DPs above the *wh* elements in shifting and turn taking contexts corresponds to a CP-internal or a CP-external position.

4.2. Factors beyond residual V2

We have discussed that predictors other than those related to residual V2 have an effect on inversion patterns (see §1.1. and §1.2). With respect to subject type, pronouns favor the preverbal position if compared with lexical subjects. This seems to indicate that information structure and the size of the subject (see Givón 1983, Levinson 1987, Gundel et al. 1993, DuBois 2003) is a strong factor for triggering postverbal subjects: elements encoding given information (such as pronouns) tend to be more frequently preverbal than constituents that have a stronger tendency to encode new information.

However, a further factor might be the ‘size’ of the relevant subject: ‘smaller’ elements like pronouns tend to appear preverbally more frequently than ‘larger’ or ‘heavier’ elements such as full NPs and clauses. This might also explain why subjects are less frequently postverbal when *decir* takes a clausal complement (if compared with *decir* taking clitic complements or no complement)..

Let us consider, in this context, a comparison between definite and indefinite lexical subjects with respect to subject position:

Table 8. Subject-verb inversion and the definiteness of subjects

NP-type	pre	post	total	%-post
indefinite pronoun	8	2	10	20%
personal pronoun	194	85	279	30%
indefinite NP	15	7	22	32%
definite NP	65	73	138	53%
proper name	12	24	36	67%
all	294	191	485	39%

If subjects encoding new information triggered the postverbal position more frequently than subjects encoding given information, we would expect entities encoding new information such as indefinite NPs to occur more frequently in the postverbal position than definite ones while definite NPs and proper names would occur more frequently in the preverbal position. However, the pattern observed in Table 8 is rather the opposite: In our data, 39% of all explicit subjects with *decir* ‘say’ appear in postverbal position. Definite NPs and proper names, i.e., elements typically referring to old or given information, are the ones most frequently placed after the verb (53% and 67% postverbal subjects, respectively). Indefinite NPs appear only in 7/22 (=32%) of the cases in postverbal position. The preference of pronouns (both definite and indefinite) to occur preverbally, if compared with full NPs, might thus be attributed to their smaller ‘size’ or lower ‘weight’ rather than to a different status with respect to givenness.

These findings are in line with the ones presented by Posio and Pešková (2020) from Peninsular Spanish: in their data, 37% (39/95) of the pronominal subjects and as much as 63% (33/52) of the NP subjects of the verb *decir* were postverbal (however, they did not distinguish between definite and indefinite NPs and pronouns or proper and common names). In Section 4.3, we will come back to the issue of why some indefinite elements (both pronouns and NPs) show a higher frequency to occur in preverbal position if compared with definite ones in our data.

Three further things should be noted when evaluating the findings in Table 8. First, the frequency of postverbal subjects with *decir* is rather high if we compare it to

the rates of postverbal position of subjects in general. Since the percentages of postverbal subjects reported in the literature vary between 20% and 40%, depending on the medium (spoken vs. written language) and the variety of Spanish in question (see, e.g., Meyer-Hermann 1990: 75, Bentivoglio 2003), it seems evident that there the verb *decir* presents a strong lexical tendency towards postverbal subjects, in particular proper name and definite NP subjects. Second, it seems that the high rate of postverbal subjects associated with this verb is dialect-specific. While Posio and Pešková's (2020) observation of a high frequency of postverbal subjects of *decir* from Peninsular Spanish are in line with our findings, they did not find a similar tendency in their Argentinian data, where only 12% (12/102) of pronominal subjects and 16% (9/58) of NP subjects were postverbal. However, since our data come exclusively from Peninsular Spanish, we will have to leave the question of dialectal differences aside. Third, our data seem to show that discourse genre affects word order patterns with *decir*. This can be observed if we divide our data according to the two corpora from which they were drawn, PRESEEA (sociolinguistic interviews) and CORLEC (conversations and media). Table 9 shows a comparison of the occurrence of postverbal subjects according to the variable 'complement type' in these two sets of data.

Table 9. Subject position according to complement type in the two corpora PRESEEA and CORLEC

Corpus PRESEEA (Madrid + Alcalá)

complement type	preverbal	postverbal	total	%-postverbal
direct reported speech	115	46	161	29%
indirect reported speech	65	24	89	27%
NP	7	5	12	42%
clitic	6	8	14	57%
relative pronoun	14	13	27	48%
none	17	27	44	61%
all	224	123	347	35%

Corpus CORLEC

comp-type	preverbal	postverbal	total	%-postverbal
direct reported speech	30	34	64	53%
indirect reported speech	13	11	24	46%
NP	7	5	12	42%
clitic	4	4	8	50%
relative pronoun	7	5	12	42%
none	9	9	18	50%
all	70	68	138	49%

As can be seen in Table 9, the frequency of postverbal subjects when the complement is reported speech (both direct and indirect) is much higher in the CORLEC than in the PRESEEA data. This might indicate that subject-verb inversion is influenced by the degree of interaction in the data. The sections of CORLEC included in the analysis contain spontaneous, unplanned conversations between speakers who do not have such

strictly institutionalized discourse roles as the informants and interviewers in the PRESEEA data, and they seem to exploit the full communicative potential of different *decir* constructions to a greater extent than speakers in the PRESEEA data. This is also reflected in the higher frequency of direct reported speech in CORLEC (9.27 occurrences per 10,000 words) than in PRESEEA (4.35 occurrences per 10,000 words).

4.3. Subject-verb inversion, semi-fixed sequences and left peripheral categories

The fact that definite NPs and especially proper names are particularly likely to occur in the postverbal position suggests that postverbal subject placement with *decir* is related to constructions where this verb typically occurs, namely semi-fixed parentheticals, intersubjective uses, and some (but not all) of the quotative constructions used to introduce reported speech. Arguably, such constructions present their own, seemingly idiosyncratic word orders that are not affected by regular information structural factors to the same extent as other contexts. For instance, in example (40), the proper name *Elena* has been mentioned three times in the part of the preceding conversation included in the corpus and thus, it does not introduce a new referent as such. The information of the embedded clause represents given information and is present in the immediately preceding discourse. What is newly introduced by the speaker in (40), however, is the source of information: *dice Elena* indicates that the assertion *explotaban en el aire* ‘they exploded in the air’ relies on reported evidence. Some types of word order patterns and subject-verb inversion with *decir* ‘say’ might thus be related to evidentiality (see Rooryck 2001, Aikhenvald 2018 and references for discussion of this concept) in that introduction of *dice* ‘say’ plus a postverbal DP indicates that the information is not based on personal experience by the speaker, but on a report by another referent:

- (40) *¿a qué sí? Explotaban en el aire.*
 that yes exploded.3PL in the air
Dice Elena que explotan en el aire, los mosquitos.
 says Elena that explode.3PL in the air the mosquitoes
 ‘Elena says that they explode in the air, the mosquitoes.’ (CORLEC)

Consider, in this context, example (41). Here, the postverbal DP subject *Carmen* is neither newly introduced nor bears narrow focus. Furthermore, the verb *decir* ‘say’ plus the postverbal DP is immediately followed by a repetition of the quotative particle *dice* (cf. also Posio & Pešková 2020):

- (41) *Y ésta me dice, me dice la Carmen, dice: “Pues*
 and this-one to-me says to-me says the Carmen says well
hija, no sé qué decirte”.
 daughter not know.1SG what say.INF.to-you
 ‘And this one says to me, Carmen says to me, she says: “Look girl, I don’t know what to say to you.”’ (CORLEC)

Here, it could be argued that main prominence is on the stretch of reported speech (Posio & Pešková 2020) and *decir* plus the DP introduces the source of information on which this stretch is based. The high frequency of postverbal subjects that are not narrowly focused

might thus be related to the functional nature of *decir* ‘say’ as indicating the type and source of information that the complement clause encodes.

In a generative architecture, the particular behavior of *decir* ‘say’ with respect to subject-verb inversion in reported speech contexts could be encoded in a system which makes use of a functional category encoding Speech Act in discourse sensitive projections above TP (Cinque 1999, Speas & Tenny 2003, Speas 2004; see Herbeck 2021 for an account of overt and covert pronouns with cognitive verbs in this spirit). According to these authors, perspectival factors relating to epistemicity, evidentiality, subjectivity and speech acts have a designated position in the left periphery of the clause:

(42) [_{SAP} SA [_{EvalP} Eval [_{EvidP} Evid [_{EpistP} Epist [_{TP} T ...

If *decir* ‘say’ is used as a marker of reported speech, it can be argued that the verb is moved into the functional category encoding speech act:

(43) [_{SAP} SA-*Dice* [_{IP} *Elena* I-*dice* [_{VP} ~~*Elena*~~ v-*dice* ... [_{CP} *que* ...

This might be further motivated by the observation that *decir* ‘say’ shows various degrees of grammaticalization and/or pragmaticalization and is used in several Romance languages either as a mere direct speech marker or, in some Spanish and Italian varieties, *decir* plus *que* has evolved towards an evidentiality marker (Sp. *dizque*, It. *dice che*; Travis 2006, Cruschina & Remberger 2008, Cruschina 2015, Demonte & Fernández Soriano 2013 and 2022). If the functional SAP phrase is in a high position within the extended IP (see Cinque 1999) or above CP (Speas & Tenny 2003), it is expected that (non-focal) postverbal subjects are frequent when *decir* ‘say’ adopts more functional meanings, such as a reported speech marker or a marker of intersubjectivity (see example (9)), and is therefore located in the high SAP projection.¹¹

This approach would crucially link the preferred postverbal position of subjects with *decir* to the functional properties of the construction where it occurs (i.e., in contexts where the speaker may want to foreground the contents of the stretch of reported speech and, at the same time, introduce the source of information).

This does not, however, explain why indefinite pronouns are frequently preverbal. Note that pronouns annotated as ‘indefinite’ in the preverbal position also include the frequently used generic pronoun *uno* ‘one’ and quantifiers like *nadie* ‘nobody’. It has been argued in the literature (e.g. Barbosa 2001 and related work) that non-referential NPs and QPs occupy a different preverbal position than referential and topical subjects. Thus, the data in Table 8 above might also be evidence for the assumption that certain non-referential expressions can occupy a preverbal position (Spec,TP) lower than referential preverbal subjects (Spec,FP in the vein of Uriagereka 1995 or Spec,TopP):

¹¹ As argued in Herbeck (2021) in the context of the cognitive verbs *creer* ‘think’ and *saber* ‘know’, a high degree of pragmaticalization might lead to loss of movement and direct merger of the verb into the perspectival functional category (in the vein of Roberts & Roussou’s 2003 loss of movement). This would yield a configuration in which *decir* is directly generated in SAP, as Cruschina (2015) argues for Italian *dice che* and Sicilian *dicica* ‘say+COMP’.

- (44) y [ya [TP *nadie* te dice nada]]
 and yet nobody to-you say.3SG nothing
 ‘And then nobody says anything to you anymore’
 (ex. from PRESEEA; Alcalá_H31-050)

At least in the case of quantified and impersonal pronouns, it might be that the availability of an additional preverbal position favors a higher frequency of preverbal subjects. This issue deserves further research. In the next section, some additional perspectives for future research will be outlined.

5. Issues for future research

The present study demonstrates that one factor that has been claimed to be a residual V2 effect in the literature favors subject-verb inversion with the verb *decir* ‘say’: if an object (in almost all cases in our data, the resumptive pronoun *eso* ‘that’) is fronted, the subject of the verb is postverbal. However, residual V2 effects could not fully be tested in interrogatives given their scarcity in our sample. Hence, even though our data do not contain any adjacent preverbal subjects in *wh*-interrogatives, a different data set would have to be studied to confirm this tendency. Furthermore, apart from VS orders, another construction has the NP referring to the person to whom the reported speech is attributed, i.e. the dictioner, in a position higher than the *wh*-element (cf. (33)–(35)). This construction seems to be related to topic shift and turn taking. It would be fruitful to further investigate the exact syntactic position of the dictioner (a high position within the extended CP or a clause-external one) and the exact contextual factors that correlate with this word order by means of a larger data set.

Lastly, our data does not allow a study of the phenomenon of quotative inversion. Our data contained only one example where the quotative verb is located after the quote, and as expected there is quotative inversion, although the quotative also includes the complementizer *que*:

- (45) *El oportunismo es de derechas, que decía Lenin.*
 the opportunism is of rightwing that said Lenin
 ‘Opportunism is typical of right-wing supporters, like Lenin used to say.’
 (CORLEC)

The absence of preverbal quotes, and thus quotative inversion, in the data might be conditioned by the fact that our data consists of spoken language, in which quotative *digo* is preferably used before the introduction of direct reported speech.

6. Conclusion

In this paper, we have investigated the interaction of various morpho-syntactic and configurational factors in triggering subject-verb inversion patterns with the verb *decir* ‘say’ in two speech corpora of Spanish.

Our main results show that fronted objects significantly favor subject-verb inversion, which could be related to a residual V2 effect (resumptive pronoun fronting). With respect to *wh*-interrogatives, our data show a tendency towards blocking preverbal adjacent subjects, but these findings are based on a limited number of examples. However, apart from this finding, an interesting configuration awaiting future research is the existence of preverbal ‘dictioners’ appearing before the *wh*-element. Other factors of residual V2, such as preverbal adverbial elements (including adverbs, adverbial clauses and PPs), were shown to disfavor subject-verb inversion. In fact, left dislocation seems to be a more frequent strategy than subject-verb inversion with most adverbs.

The study of complement type and subject type shows that some factors not related to residual V2 also have a strong effect on subject-verb inversion patterns: first, lexical DPs are more frequently postverbal than pronominal ones, which receives a natural explanation if definiteness, information structure, and the ‘weight’ of subjects influence word order, as is standard in generative and functional studies. On the other hand, indefinite subjects do not trigger the postverbal position more frequently than expected. This is related to the fact that these involve in several instances non-referential or impersonal subjects (e.g., *uno* ‘one’, *nadie* ‘nobody’) which have a strong tendency to appear in the preverbal, non-dislocated subject position (see Barbosa 2001 for differences between dislocated topical subjects and preverbal non-referential ones). Second, postverbal subjects are favored if *decir* ‘say’ does not take any complement. On one hand, this might indicate that intransitive uses of the verb favor inversion. On the other hand, the category “no complement” included several instances of semi-fixed parentheticals such as *como digo yo* ‘as I say’ (or, more generally, *como DECIR X*). This shows that several word order patterns with the highly frequent verb *decir* ‘say’ have grades of constructionalization with specific discourse functions. While these factors are often not integrated in generative studies of variable word order, we have proposed that a theoretical solution within this framework might be the integration of perspectival functional categories related to speech act and epistemicity/evidentiality above TP (Cinque 1999, Speas & Tenny 2003) into the study of word order phenomena. In the case of functional uses of *decir* ‘say’ as a quotative particle, the verb might be located in the highest SAP phrase of the syntactic structure with the consequence that postverbal subjects appear in the unmarked subject position Spec,IP.

An account of variable word order with *decir* ‘say’ in spoken Spanish must consequently take into account morpho-syntactic, configurational, functional as well as constructional factors in order to be fully explanatory. Depending on the concrete function of *decir* ‘say’, it is not only the TP and CP area, but furthermore, the functional projection SAP (Cinque 1999, Speas & Tenny 2003) that is relevant for deriving word order patterns.

Acknowledgments

We would like to thank Suvi Tynnilä for her work as a research assistant for the project and for carrying out some of the coding. We would furthermore like to thank the organizers and participants of the workshop *Residual Verb Second in Romance*

(*Reverse1*) and two anonymous reviewers for helpful comments. All errors are our own.

References

- Aikhenvald, Alexandra Y. 2018. Evidentiality: The framework. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, 1–43. Oxford: Oxford University Press. <http://doi.org/10.1093/oxfordhb/9780198759515.013.1>
- Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parametrizing AGR: Word order, V-movement and EPP-checking. *Natural Language & Linguistic Theory* 16: 491–539. <https://doi.org/10.1023/A:1006090432389>
- Barbosa, Pilar. 2001. On inversion in *wh*-questions in Romance. In A. Hulk & J.-Y. Pollock (eds.), *Subject inversion in Romance and the theory of Universal Grammar*, 20–59. Oxford: Oxford University Press.
- Barbosa, Pilar. 2009. Two kinds of subject *pro*. *Studia linguistica* 63: 2–58. <https://doi.org/10.1111/j.1467-9582.2008.01153.x>
- Batlloori, Montserrat. 2015. Análisis comparativo de las construcciones de dislocación a la derecha y su incidencia en el orden de palabras del español, catalán y portugués medievales. In M. López Izquierdo & M. Castillo Lluch (eds.), *El orden de palabras en la historia del español y otras lenguas iberorromances*, 83–110. Madrid: Visor Libros.
- Belletti, Adriana. 2004. Aspects of the low IP area. In Luigi Rizzi (ed.), *The structure of CP and IP. The cartography of syntactic structures 2*, 16–51. Oxford: Oxford University Press.
- Bentivoglio, Paola. 2003. Orden de palabras en español: un análisis sintáctico-semántico-pragmático del sujeto. *Lexis*, XXVII(1-2): 235–260.
- Bentivoglio, Paola & Weber, Elizabeth G. (1987). “A Functional Approach to Subject Word Order in Spoken Spanish.” In O. Jaeggli & C. Silva-Corvalán (eds.), *Studies in Romance Linguistics: Selected Papers of the Fourteenth Linguistic Symposium on Romance Languages*. 23–40. Dordrecht: Foris.
- Castillo Lluch, Mónica & Marta López Izquierdo. 2015. El orden de palabras en la historia del español y en otras lenguas iberorromances : introducción. In: M. López Izquierdo & M. Castillo Lluch (eds.), *Orden de palabras en la historia del español y otras lenguas iberorromances*, 7–26. Madrid: Visor Libros.
- Chomsky, Noam. 1981. *Lectures of government and binding - the Pisa lectures*. Dordrecht: Foris Publications.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.

Contreras, Heles. 1991. On the position of subjects. In S. Rothstein (ed.), *Perspectives on phrase structure: Heads and licensing*, 63-79. San Diego: Academic Press.

Corr, A. 2016. Wide-focus subject-verb inversion in Ibero-Romance: A locative account. *Glossa: a journal of general linguistics* 1(1), 11.
<https://doi.org/10.5334/gjgl.85>

Cruschina, Silvio. 2015. The expression of evidentiality and epistemicity: Cases of grammaticalization in Italian and Sicilian. *Probus* 27: 1-31.
<https://doi.org/10.1515/probus-2013-0006>

Cruschina, Silvio & Eva-Maria Remberger. 2008. Hearsay and reported speech: Evidentiality in Romance. *Rivista di Grammatica Generativa* 33: 95-116.

Cuartero Sánchez, Juan Manuel. 2007. Sobre la naturaleza y el alcance de los factores que influyen en el orden de palabras (con referencia a las oraciones declarativas en español actual). *Lingüística española actual* 29: 101-118.

Demonte, Violeta & Olga Fernández-Soriano. 2013. Evidentials *dizque* and *que* in Spanish. Grammaticalization, parameters and the (fine) structure of Comp. *Revista de estudos linguísticos da Universidade do Porto* 8: 211-34.

Demonte, Violeta & Fernández-Soriano, Olga. 2022. A multidimensional analysis of the Spanish reportative epistemic evidential *dizque*. *Lingua* 266: 1-29.
<https://doi.org/10.1016/j.lingua.2021.103168>

DuBois, JohnW. 2003. Argument structure: Grammar in use. In J. DuBois, L. E. Kumpf, & W. J. Ashby (eds.), *Preferred argument structure: Grammar as architecture for function*, 11-60. Amsterdam: John Benjamins.

Elvira, Javier. 2015. ¿V2 en español antiguo? In M. López Izquierdo & M. Castillo Lluch (eds.), *El orden de palabras en la historia del español y otras lenguas iberorromances*, 27-47. Madrid: Visor Libros.

Erker, Daniel & Guy, Gregory R. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3): 526-557. <https://doi.org/10.1353/lan.2012.0050>

Fernández Soriano, Olga. 1993. Sobre el orden de palabras en español. *Dicenda. Cuadernos de filología hispánica* 11: 113-152.

Fontana, Josep M. 1993. *Phrase structure and the syntax of clitics in the history of Spanish*. Ph.D. dissertation, University of Pennsylvania.

Frascarelli, Mara & Roland Hinterhölzl. 2007. Types of topics in German and Italian. In Kerstin Schwabe & Susanne Winkler (eds.), *On information structure, meaning and form*, 87–116. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.100.07fra>

Gabriel, Cristoph. 2007. *Fokus im Spannungsfeld von Phonologie und Syntax Eine Studie zum Spanischen*. Berlin: Mouton de Gruyter.

Gallego, Ángel. 2010. *Phase theory*. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.152>

Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse*, 1–42. Amsterdam: John Benjamins. <http://doi.org/10.1075/tsl.3.01giv>

Goodall, Grant. 1993. Spec of IP and Spec of CP in Spanish *wh*-questions. In W.J. Ashby, M. Mithun, G. Perissinotto & E. Raposo (eds.), *Linguistic Perspectives on the Romance Languages: Selected Papers from the XXI Linguistic Symposium on Romance Languages*, 199–209. Amsterdam: John Benjamins.

Goodall, Grant. 2004. On the syntax and processing of *wh*-questions in Spanish. In Benjamin Schmeiser, Vineeta Chand, Ann Kelleher & Angelo Rodriguez (eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, 101–114. Somerville, MA: Cascadilla Press.

Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69 (2), 274–307. <https://doi.org/10.2307/416535>

Gutiérrez-Bravo, Rodrigo. 2005. Subject Inversion in Spanish relative clauses. In T. Geerts, I. van Ginneken & H. Jacobs (eds.), *Romance Languages and Linguistic Theory 2003*, 150–166. Amsterdam/Philadelphia: John Benjamins.

Gutiérrez-Bravo, Rodrigo. 2008. Topicalization and preverbal subjects in Spanish *wh*-interrogatives. In Joyce Bruhn de Garavito and Elena Valenzuela (eds.), *Selected Proceedings of the 10th Hispanic Linguistics Symposium*, 225–236. Somerville: Cascadilla Proceedings Project.

Herbeck, Peter. 2015. *Unifying Pro-Drop and Control – The Derivation of Spanish (Null) Subjects*. Doctoral dissertation, University of Salzburg. <https://resolver.obvsg.at/urn:nbn:at:at-ubs:1-15365>

Herbeck, Peter. 2021. Perspectival factors and *pro*-drop: A corpus study of speaker/addressee pronouns with *creer* 'think/ believe' and *saber* 'know' in spoken Spanish. *Glossa: a journal of general linguistics* 6(1). 1–34. <https://doi.org/10.16995/glossa.5873>

- Hennemann, Anja. 2016. A cognitive-constructionist approach to Spanish *creo* \emptyset and *creo yo* '[I] think'. *Folia Linguistica* 50(2). 449–474. <http://doi.org/10.1515/flin-2016-0017>
- Holmberg, Ander. 2015. Verb Second. In: Tibor Kiss & Artemis Alexiadou (eds.), *Syntax - Theory and Analysis. An International Handbook* (Volume 1), 342–382. Berlin: de Gruyter.
- Leonetti, Manuel. 2017. Basic constituent orders. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance Morphosyntax and Syntax*, 887–932. Berlin: De Gruyter. <https://doi.org/10.1515/9783110377088-024>
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora: A partial pragmatic reduction of Binding and Control phenomena. *Journal of Linguistics* 23. 379–434. <http://doi.org/10.1017/S0022226700011324>
- López, Luis. 2009. *A derivational syntax for information structure*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199557400.001.0001>
- Lüdtke, Daniel. 2021. sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.10.
- Marcos Marín, Francisco. 1992. *Corpus oral de referencia del español contemporáneo*. Universidad Autónoma de Madrid.
- Martín Zorraquino, María Antonia & José Lázaro Portolés. 1999. Los marcadores del discurso. In I. Bosque & V. Demonte (eds.), *Gramática descriptiva de la lengua española*, 4051–4214. Madrid: Espasa Calpe.
- Matos, Gabriela. 2013. Quotative Inversion in Peninsular Portuguese and Spanish, and in English. *Catalan Journal of Linguistics* 12: 111–130. <https://doi.org/10.5565/rev/catjl.86>
- Meyer-Hermann, Reinhard. 1990. Sobre algunas condiciones pragmáticas de la posición del sujeto en español. *ELUA - Estudios de Lingüística de la Universidad de Alicante* 6: 73–88. <http://dx.doi.org/10.14198/ELUA1990.6.04>
- Ordóñez, Francisco. 2000. *The Clausal Structure of Spanish: A Comparative Perspective*. New York: Garland.
- Ordóñez, Francisco & Esthela Treviño. 1999. Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua* 107: 39–68. [https://doi.org/10.1016/S0024-3841\(98\)00020-5](https://doi.org/10.1016/S0024-3841(98)00020-5)
- Pinto, Carlos Felipe. 2015. Algunas observaciones sobre el efecto V2 en el español antiguo. In: M. López Izquierdo & M. Castillo Lluch (eds.), *Orden de palabras en la historia del español y otras lenguas iberorromances*, 49–81. Madrid: Visor Libros.

Posio, Pekka. 2012. The functions of postverbal pronominal subjects in spoken Peninsular Spanish and European Portuguese. *Studies in Hispanic and Lusophone Linguistics* 5(1): 149–190. <https://doi.org/10.1515/shll-2012-1122>

Posio, Pekka. 2015. Subject pronouns in formulaic sequences: Evidence from Peninsular Spanish. In A. M. Carvalho, R. Orozco & N. Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal perspective*. Washington: Georgetown University Press.

Posio, Pekka & Andrea Pešková. 2020. *Le dije yo, digo*: Construccionalización de los introductores cuotativos con el verbo *decir* en español peninsular y argentino. *Spanish in Context* 17: 319–414.

Posio, Pekka & Malte Rosemeyer. 2021. Pragmatics and word order. In A. Koike & J. César Félix-Brasdefer (eds.), *The Routledge Handbook of Spanish Pragmatics - Foundations and Interfaces*, 73–90. Abingdon: Routledge. <https://doi.org/10.4324/9780429455643>

PRESEEA. 2014-. *Corpus del proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. <http://preseea.linguas.net> [Consulted: June 2019 & April 2021].

R Core Team. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Foris Publications.

Rizzi, Luigi. 1996. Residual Verb Second and the Wh-Criterion. In A. Beletti and L. Rizzi (eds.), *Parameters and Functional Heads*, 63–90. Oxford University Press, New York.

Rizzi, Luigi. 1997. The Fine Structure of the Left Periphery. In L. Haegeman (ed.), *Elements of Grammar: Handbook in Generative Syntax*, 281–337. Dordrecht: Springer.

Rizzi, Luigi & Ian Roberts. 1989. Complex Inversion in French. *Probus* 1: 1–30. <https://doi.org/10.1515/prbs.1989.1.1.1>

Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486326>

Rooryck, Johan. 2001. Evidentiality, Part I. *Glott International* 5(4): 125–133.

Rostila, Jouni. 2004. Lexicalization as a way to grammaticalization. In F. Karlsson (ed.), *Proceedings of the 20th Scandinavian conference of linguistics*, 1–13. Tampere: University of Tampere.

- Sheehan, Michelle. 2010. 'Free' Inversion in Romance and the Null Subject Parameter. In T. Biberauer, A. Holmberg, I. Roberts & M. Sheehan, *Parametric Variation: Null Subjects in Minimalist Theory*, 231-262. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511770784.007>
- Silva-Corvalán, Carmen. 1982. Subject expression and placement in Mexican American Spanish. In J. Amastae & L. Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*, 93–120. Cambridge/New York: Cambridge University Press.
- Solà, Jaume. 1992. *Agreement and Subjects*. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Speas, Margaret. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114: 255-7026. [https://doi.org/10.1016/S0024-3841\(03\)00030-5](https://doi.org/10.1016/S0024-3841(03)00030-5)
- Speas, Peggy & Carol Tenny. 2003. Configurational properties of point of view roles. In Anna Maria Di Sciullo (ed.), *Asymmetry in grammar* 1, 315–344. Amsterdam: John Benjamins. <https://doi.org/10.1075/la.57.15spe>
- Suñer, Margarita. 1982. *Syntax and semantics of Spanish presentational sentence-types*. Washington DC: Georgetown University Press.
- Suñer, Margarita. 1994. V-Movement and the licensing of argumental *wh*-phrases in Spanish. *Natural Language & Linguistic Theory* 12(2): 335-372. <https://doi.org/10.1007/BF00993148>
- Suñer, Margarita. 2003. The lexical preverbal subject in a Romance null subject language: Where are thou? In Rafael Núñez-Cedeño, Luis López & Richard Cameron (eds.), *A Romance perspective on language knowledge and use: Selected papers from the 31st LSRL*, 341–357. Amsterdam: John Benjamins. <http://doi.org/10.1075/cilt.238.25sun>
- Traugott, Elizabeth Closs & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Travis, Catherine E. 2006. *Dizque*: a Colombian evidentiality strategy. *Linguistics* 44(6): 1269–1297. <https://doi.org/10.1515/LING.2006.041>
- Travis, Catherine E. & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711–748. <http://doi.org/10.1515/cog-2012-0022>
- Uriagereka, Juan. 1995. An F position in Western Romance. In Katalin É. Kiss (ed.), *Discourse configurational languages*, 153–175. Oxford: Oxford University Press.

Wiemer, Björn & Walter Bisang. 2004. What makes Grammaticalization? An appraisal of its components and its fringes. In W. Bisang, N.P. Himmelmann & B. Wiemer, *What makes Grammaticalization?* Berlin: Mouton de Gruyter.

Zagona, Karen T. 2002. *The syntax of Spanish*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511613234>

Zimmermann, Michael. 2019. Reconsidering the syntax of interrogatives in Caribbean Spanish. With especial reference to Dominican Spanish. *Isogloss* 5/2: 1-36. <https://doi.org/10.5565/rev/isogloss.57>

Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, Mass: MIT Press.