



Perspectival factors and *pro*-drop: A corpus study of speaker/addressee pronouns with *creer* 'think/believe' and *saber* 'know' in spoken Spanish

PETER HERBECK 

RESEARCH


Open Library of Humanities

Abstract

This paper examines overt and covert speaker/addressee pronouns with the cognitive verbs *creer* 'think/believe' and *saber* 'know' in a corpus of spoken peninsular Spanish – the Madrid and Alcalá samples of PRESEEA (2014–) – with a focus on 1st person singular (*yo*) *creo que* '(I) think that'. Departing from the observation made in the literature that overt pronouns are highly frequent with *creer* and that topic shift cannot account for all of them, it will be argued that perspectival factors related to evidentiality/epistemicity and subjectivity influence overt pronoun realization. A corpus study was conducted to investigate whether (i) [person] and [polarity] and (ii) the type of complement affect overt pronoun realization with the cognitive verbs *creer* and *saber*. The results indicate that the type of belief expressed in the embedded clause should be taken into account, as well as person and polarity. The ultimate trigger for phonetic realization of speaker/addressee pronouns will be argued to be the notion of contrast: cognitive verbs whose embedded complement encodes evaluations and non-visual, abstract information have high frequencies of overt pronoun realization because these contexts favor the evoking of alternative perspective holders. Overt pronouns will be analyzed as the result of a [+contrast] feature which is assigned to the specifier of a functional category encoding perspective in the split IP.

CORRESPONDING AUTHOR:

Peter Herbeck

University of Vienna, AT

peter.herbeck@univie.ac.at

KEYWORDS:

pro-drop; strong pronouns; cognitive verbs; contrast; syntax of spoken Spanish; perspective

TO CITE THIS ARTICLE:

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): 122. 1–34. DOI: <https://doi.org/10.16995/glossa.5873>

The null subject property has been one of the most thoroughly studied phenomena in generative theory. In syntactic studies within the principles & parameters framework (Chomsky 1981), the formal possibility of having phonetically empty subjects – their “licensing” – was kept apart from their “identification” (Rizzi 1986). An explanation of what governs the overt realization of subject pronouns within a null subject language in actual language use was often linked to emphasis or contrast (e.g. Luján 1999).

Another strand of research focuses on the relation between null subjects and (non-)continuous topic-linking. According to Givón’s (1983: 18) *scale of phonological size*, there is a preference to use less (phonetic and/or structural) material for continuous/accessible topics while more material is used to encode non-continuous/inaccessible ones. Frascarelli (2007; 2018) provides a formal syntactic framework for the relevance of topic-continuity/shift, relying on the notion of topic chain and the syntactic operation of *Agree* that holds with an Aboutness (Shift) Topic in the left periphery (see section 2.1).

However, several corpus studies of spoken varieties of Spanish show that further factors must be taken into account, such as verb type and person specification (see Enríquez 1984; Davidson 1996; Aijón Oliva & Serrano 2010; Posio 2011; Travis & Torres Cacoullós 2012; Erker & Guy 2012; Adli 2019, among others). Epistemic verbs like *creer* ‘believe/think’ and *pensar* ‘think’ are frequently used with overt pronouns (see section 2.2). With respect to [person], it has been observed that (continuous) topic-linking, while it can be applied to 3rd person, is less straightforward in the account of 1st and 2nd person deictic pronouns (see Frascarelli 2007; 2018; Adli 2019). One hypothesis that can be found in the literature is that perspectival notions such as (inter-)subjectivity and/or epistemicity/evidentiality (Aijón Oliva & Serrano 2010; Posio 2011; 2013; 2014; Hennemann 2012; 2016; Grajales Alzate 2016) influence pronoun use with certain verbs.

One goal of this paper is to obtain a deeper understanding of the nature of high frequency overt pronouns with cognitive verbs in Spanish and to explore how they can be analyzed when perspectival notions in spoken language data are examined. A corpus study of the Madrid and Alcalá samples in PRESEEA (2014-),¹ was conducted to investigate null and overt 1st and 2nd person singular pronouns with two cognitive verbs – *creer* ‘believe/think’ and *saber* ‘know’. These verbs were chosen because they offer a testing ground for the role that subjectivity and epistemicity/evidentiality play for overt subject licensing. It will be argued that strong pronouns with certain verbs are best analyzed as *perspectival markers* (section 5). If subjectivity and epistemicity/evidentiality influence overt pronoun realization, one expectation is that the type of belief or knowledge that is expressed in the embedded clause might play a role. A classification of embedded complements will be proposed based on whether they encode visually/audibly perceivable information or not. The results indicate that complements expressing evaluations and abstract (non-visual) information trigger high frequencies of overt matrix subjects with *(yo) creo* ‘(I) think’ and *(yo) sé* ‘(I) know’.

Overt pronoun frequencies will be argued to be highest with verbs like *creer* because they can be used to express personal opinions (Aijón Oliva & Serrano 2010; Posio 2014) and, relatedly, they leave unspecified the degree of subjective truth probability (Lewis 1976; Davis et al. 2007) of the embedded proposition. Thus, [contrast] assignment to the pronoun triggers potential alternative perspectives, which can serve as a means of adapting the truth probability of the embedded information. The verb *saber* implies a higher degree of truth probability of the embedded claim by means of its lexical semantics; overt pronoun realization is hence expected to be lower.²

It will be argued that a form of “weak contrast” (Mayol 2010) is assigned to the Spec of a functional category encoding perspectival notions such as (speaker/addressee) evaluation, epistemicity, and evidentiality above TP (Cinque 1999; Speas 2004). This converts the “cognizer” subject (Posio 2011) into a subjective evaluator, which is interpreted in relation to

1 All glosses, translations, and emphatic markings that appear with examples from PRESEEA (2014-) throughout this paper are my own.

2 An anonymous reviewer points out that the question arises of why a special marking of knowing (by means of *saber* ‘know’) should be necessary if unmarked transmitted information is expected to be known by the speaker. Thus, it could be argued that marking by *saber* indicates that the proposition is not (fully) reliable. As will be discussed in section 4.3, there is evidence that evaluative contexts trigger higher overt subject frequencies also with *saber*. However, *creer* is more frequently used in evaluative contexts than *saber*.

a potential set of other perspective holders. The observation that *yo creo (que)* ‘I think (that)’ has degrees of fixation (Posio 2015) will be accounted for by assuming that *creo* can function not only as a main verb moving to the relevant perspectival functional category, but can be directly merged into it. In this configuration, the pronoun *yo* is a perspectival marker, as an overt realization of a “seat of knowledge” (Speas & Tenny 2003) argument.

The structure of this paper is as follows: in section 2, some factors that influence the overt/covert alternation of subject pronouns in Spanish will be discussed. Section 3 outlines the concepts of evidentiality, epistemicity and subjectivity and some problems that arise when studying subject pronouns with cognitive verbs in spoken Spanish. Section 4 presents the corpus study carried out for this paper. The study consists of two parts: (i) an investigation of [verb] (*creer* vs. *saber*), [person] (1SG vs. 2SG) and [polarity] (positive vs. negative) in relation to overt pronoun realization and (ii) a study of the complement type of cognitive verbs. Section 5 presents the syntactic analysis. In this context, the possible means of implementing the different degrees of fixation (Posio 2015) or “pragmaticalization” (Aijmer 1997) of pronoun plus cognitive verb combinations will be discussed, making use of Roberts & Roussou’s (2003) theory. Section 6 discusses potential lines for future research and section 7 offers some concluding remarks.

2 Different factors governing subject pronoun use

In this section, the factors of contrast, topicality, verb type, person and polarity and their relation to the overt/covert alternation of subject pronouns will be discussed.

2.1 Contrast and topic continuity

It has been argued that the overt realization of subjects in Romance-type *pro*-drop languages is triggered by emphasis or contrast, which come in different flavors (cf. Rigau 1989; Luján 1999; Mayol 2010; ex. (1) based on Rigau 1989):

- (1) a. weak contrast:
 Yo iré a Madrid (... los otros, no sé)
 I go.FUT to Madrid the others not know.1SG
 ‘I will go to Madrid (... I don’t know about the others).’
- b. strong contrast:
 YO iré a Madrid (... no Juan).
 I go.FUT to Madrid not Juan
 ‘I will go to Madrid (... not John).’

Mayol (2010) argues for Catalan subject pronouns that they are different types of contrastive topics. One type of overt pronoun expresses an “uncertainty contrast” (*ibid.*: 2506f). It evokes a potential set of topic alternatives, but does not exhaustively resolve them (as in (1a)).

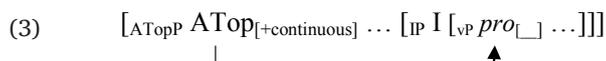
However, it has been argued that overt pronoun realization cannot always be reduced to contrast (Travis & Torres Cacoullós 2012: 723f). In fact, there are several instances of apparently redundant overt pronouns, particularly with first person singular *yo creo*. Consider the following example from PRESEEA (2014–):³

- (2) (PRESEEA-Alcalá, H12_019)
yo no creo ese <alargamiento/> <vacilación/> ese término que le
 I not think.1SG this <lengthening> <hesitation> this term that him
 acabas de dar a la ciudad como ciudad dormitorio ¿eh? / **yo creo**
 finish.2SG of give.INF to the city like dormitory town eh I think
 que es todo lo contrario
 that is.3SG all the contrary
 ‘I don’t think that this... term ‘dormitory town’, that (you) just gave the city, is the right one. I think that it is entirely the opposite.’

³ In the PRESEEA (2014–) transcriptions, <alargamiento> stands for ‘lengthening’, <vacilación> for ‘hesitation’, <simultáneo> indicates that two speakers are talking simultaneously, <cita> ‘citation’ indicates direct style, and <palabra_cortada> indicates that a word was cut off / not completed. Furthermore, a simple slash ‘/’ stands for a short pause and a double-slash ‘//’ for a (long) pause.

Here, *yo* in combination with *creo* cannot easily be argued to be strongly emphatic given its multiple occurrence and the continuity of the speaker’s perspective. However, the pronoun indicates that the addressee (or a potential set of other believers) might have a different opinion with respect to the embedded assertion and, thus, it could be classified as encoding a weak type of contrast.

Another strand of research assumes that (referential or topic) continuity favors ‘smaller’ forms, while non-continuous or shifting contexts favor ‘larger’ ones (see Givón 1983; Levinson 1987). Given that *null* is smaller than *overt*, the former is preferred in continuous and co-referent contexts, while the latter is triggered by shifting contexts. Frascarelli (2007) implements a formal approach to the relevance of topic shift and (dis-)continuity for null subjects in Italian. She assumes that topic types (shift, contrastive, familiar) are encoded by different left-peripheral projections (Frascarelli & Hinterhölzl 2007) in the C-area. In this approach, weak pronouns (destressed overt pronouns and null *pro*) are the result of an *Agree* relation between an Aboutness-Shift topic projection in the left periphery and null *pro* in subject position. Furthermore, the Aboutness-Shift topic can remain phonetically null if it is [+continuous]:



Strong pronouns, on the other hand, indicate a new Aboutness-Shift Topic or reintroduce an insufficiently salient one.

However, the applicability of topic continuity is less straightforward with deictic 1st and 2nd person pronouns. Bentivoglio (1983) and Adli (2019) show that insertion of a 1st or 2nd person pronoun does not necessarily interrupt the formation of 3rd person topic chains. Consider the following example:

- (4) (PRESEEA-Alcalá, M23_010)
 [context: *los soldados* ‘the soldiers’]
 estaban por las calles y estaban todo el día en Alcalá y todo / o sea
 were.3PL in the streets and were.3PL all the day in Alcalá and all that is
 y <vacilación/> y bueno no te digo todo <vacilación/> toda
 and <hesitation> and well not you say.1SG all <hesitation> all
 su vida militar / ahora yo creo que / salen del cuartel y se
 their life military now I think that leave.3PL of-the barracks and REFL
 van a sus casas
 go.3PL to their homes
 ‘(They_i) were in the streets and (they_i) were in Alcalá all day long. That is, and
 ... well (I) don’t say all... their_i military life. Now, I think that (they_i) leave the
 barracks and (they_i) go home.’

As can be seen, *los soldados* ‘the soldiers’, introduced in previous discourse, is continued by a 3PL null subject. Then, the speaker’s perspective is introduced by a 1SG strong pronoun, but it does not break continuity with respect to 3PL ‘the soldiers’, which is continued with a null subject after introduction of the 1SG pronoun. Thus, even though subject pronoun expression influences topic chaining in some contexts, not all instances of overt speaker/addressee pronouns can be explained by the notion of topic continuity.

That 1SG and 2SG pronouns might not necessarily have an effect on referential or topic linking of third person null subjects with verbs taking a CP complement is supported by the observation that the main speech act (or the main assertion) can be located in the subordinate rather than the main clause (see Bianchi & Frascarelli 2010; Krifka 2014; Adli 2019). These so-called “bridge verbs” (see Meinunger 2004 and references) include verbs of communication, verbs of thinking, and verbs of knowing (Class A, B, and E verbs in Hooper & Thompson 1973):

- (5)
$$[_{\text{CP}} \dots (\text{Yo}) \text{digo/creo/sé} [_{\text{CP}} \text{Top que } [_{\text{pro}} \text{ se van a sus casas}]]]$$

 I say/think/know that (they) REFL go.3PL to their home
 ‘I say/think/know that (they) go home.’

If the topic can be situated in the embedded clause with *creer* and *saber*, it cannot be topic-(non-)continuity that sanctions the overt/covert alternation in the main clause in these cases.

The assumption that the topic can be found in the embedded rather than the main clause with *creer* and *saber* is in line with the well-known fact that these verbs have parenthetical uses (see Thompson & Mulac 1991; Aijmer 1997 for English *I think*; Posio 2014 for Spanish), in which their function is evidential or discursive (see Simons 2007). In the following question-answer pair, the embedded rather than the main clause would be the “main point” of the utterance (see Simons 2007 for English examples):

- (6) A: ¿Qué hizo Juan ayer?
 what made.3SG Juan yesterday
 ‘What did John do yesterday?’
 B: Yo creo que *pro* fue al cine.
 I think that went.3SG to-the cinema
 ‘I think that he went to the cinema.’

The main predicate forms part of “non-at-issue” meaning, not answering the current Question Under Discussion (in the sense of C. Roberts 2012). In the terminology of Nuyts (2001), the matrix verb “qualifies” the content of the embedded proposition rather than forming part of it (cf. De Saeger 2008: 67).

In previous corpus studies, one hypothesis is that (inter-)subjectivity, epistemicity and/or evidentiality influence overt pronoun realization, particularly with the verb *creer* in 1SG (cf. Aijón Oliva & Serrano 2010; Hennemann 2012; 2016; Posio 2013; 2014). In the next section, the verb type factor will be discussed.

2.2 Verb type

Several studies of Spanish note that there is variation in overt pronoun frequencies depending on verb type (Enríquez 1984; Morales 1997; Posio 2011; 2013; 2014). These studies agree that verbs like *creer* ‘believe/think’ and *pensar* ‘think’ have the highest overt pronoun frequencies.

Enríquez (1984) postulates four verb classes for her quantitative study of overt pronoun expression in spoken Spanish of Madrid: (i) verbs of mental activity (e.g. *saber* ‘know’), (ii) verbs which express a personal opinion or a judgment (e.g. *creer* ‘think/believe’)⁴ (iii) stative verbs (e.g. *tener* ‘have’), and (iv) verbs of (external or objective) activity (e.g. *hacer* ‘make’) (see Enríquez 1984: 151ff). Within this classification, verbs of personal opinion (type (ii)) have the highest overt pronoun frequency (55%, compared to 33.78% overall overt pronoun frequencies; Enríquez’s 1984: 362). This is partly attributed to their relation to subjectivity and contrast: if overt pronouns imply an individualization of the subject, they would be expected to be especially frequent with those verbs where the personal opinion (in opposition to others) matters (Enríquez 1984: 118).

Posio (2011) also observes differences with respect to verb type and pronoun frequencies. With 1SG, pronoun realization is most frequent with *pensar* ‘think’ (59%) and *creer* (55%), which imply Cognizer subjects (i.e. a “thinker”, “believer”, “knower” or “presumer”; see Van Valin 2001: 31). Posio (2011) accounts for the predominance of overt subjects with these verbs in terms of the notion of “focus of attention”: With verbs like *creer*, the object clause is not prominent, which in turn allows attention to be focused on the subject (see Posio 2011: 786). The high frequency of subject pronoun realization would thus be related to the lower prominence of the object complement with *creer*.

One problem with this reasoning is in distinguishing verbs whose complement is “prominent” from those whose subject is prominent. One case in point is the verb *saber* ‘know’. In several studies (e.g. Enríquez 1984; Posio 2015), it is noted that subject realization frequencies are lower with this type of verb, at least in the peninsular Spanish varieties, than with verbs expressing a personal opinion. However, it is not clear how this is related to a different prominence status of the complement clause, primarily because *saber* also sanctions parenthetical uses.

In the corpus study by Aijón Oliva & Serrano (2010), the subject of 1SG *creo* ‘(I) think’ is less frequently expressed when the verb is used with the epistemic meaning of making a hypothesis

⁴ As Enríquez (1984) notes, some verbs could be classified as belonging to either (i) or (ii), depending on context.

(46.5%/62% overt pronouns in two corpora), compared with when it is used to express a personal opinion (83.7%/81.2%, respectively). Thus, an argumentative use of *creer* facilitates overt pronoun use and *yo* 'I' is a vehicle for "subjectivization" in discourse (Aijón Oliva & Serrano 2010: 11). According to Davidson (1996: 555), the overt pronoun adds "pragmatic weight" to the utterance, making it "more personally relevant". In contrast, *creo* with a null subject implies a more general/objective stance towards the embedded proposition and thus has a merely epistemic value (cf. Aijón Oliva & Serrano 2010: 12). Grajales Alzate (2016: 347), in a study of the corpus PRESEEA-Medellín, also concludes that expression of the 1SG pronoun with *creo* stresses the personal nature of the source of information and, thus, it emphasizes the personal part of the evidential meaning.

De Saeger (2008) draws a distinction between qualificational and representational uses (in the sense of Nuyts 2001). In the former, the matrix verb introduces the opinion of the speaker or expresses a doubt with respect to the embedded proposition (see De Saeger 2008: 64) and, thus, rather than functioning as a main verb, it is an "opinion marker" (*ibid.*: 70). According to the author, an overt pronoun reinforces the subject-perspective in this use, similarly to adverbs such as *personalmente* 'personally' (*ibid.*: 72).

Hennemann (2016) investigates the two constructions *creo* \emptyset and *creo yo* '(I) think' in postposed position. She argues that *creo* fulfils a subjective function, i.e. it appears in contexts where the speaker expresses her or his evidence for "an epistemic evaluation" (*ibid.*: 455) and has mainly a mitigating function (*ibid.*: 461). *Creo yo* appears in contexts of intersubjectivity (e.g. Nuyts 2001; Traugott 2010), i.e. there is an indication of the speaker's awareness of the addressee (see Hennemann 2016: 454f) and "the interlocutor is invited to approve or reject the speaker's opinion" (*ibid.*: 455).

Some issues, however, remain unresolved: (i) Aijón Oliva & Serrano (2010) and Posio (2013; 2014) focus on 1st person (*yo*) *creo* and Hennemann (2016) on *creo* (*yo*). Travis & Torres Cacoullous (2012: 438f) argue that the "*yo* + COGNITIVE VERB" construction belongs to a more schematic "(subject) + COGNITIVE VERB" construction. Even though 1SG *saber* 'know' has lower pronoun frequencies in some studies, it is also a mental/cognitive verb and is highly frequent in spoken language. If we assume a schematic "(subject) + COGNITIVE VERB" construction, the question is why (*yo*) *sé* '(I) know' does not have the same pronoun frequencies as (*yo*) *creo*, at least in the peninsular Spanish varieties investigated here (Madrid and Alcalá).⁵ (ii) Epistemic *creer* has overt pronoun rates of 62% in one corpus from Aijón Oliva & Serrano (2010). Although the frequency is lower than with argumentative *creo*, it is still a high frequency at above 50%; the question, then, is what determines overt pronoun realization in these cases. Lastly, (iii) distinguishing between epistemic and argumentative uses of *creer* is not always a clear-cut issue (as acknowledged by Aijón Oliva & Serrano 2010: 12). It would be interesting to see whether there is any formal variable of perspectival notions like (inter-)subjectivity, epistemicity and evidentiality in the *local* linguistic context of cognitive verbs for the analysis of overt/covert pronouns in a spoken language corpus. This issue is particularly challenging because these perspectival notions rely on speaker-internal information, which cannot always be examined in spoken corpus data.

2.3 Person specification

Several studies have emphasized the opposition between *speech act participant* (SAP) pronouns on the one hand and 3rd person pronouns on the other (see DeLancey 1981: 627ff; see also Harley & Rittner 2002: 486ff for differences between 1st/2nd vs. 3rd person). There is thus an opposition between 1st/2nd person deictic pronouns and potentially (discourse) anaphoric 3rd person (see Posio 2018: 291). In the generative literature, different projections for encoding speaker and hearer vs. topic have been proposed in the left periphery of the clause. According to Sigurðsson (2011), there are projections hosting *logophoric agent* and *logophoric patient*, in addition to a topic category:

5 Overt pronoun frequencies with cognitive verbs are subject to dialectal variation. Posio (2018: 294) discusses differences between 1SG *saber* in Orozco's (2015) study on Colombian Spanish, where this verb form favors pronoun expression, and Posio's (2015) study on Peninsular Spanish, where null subjects are favored.

(7) [CP Force [Top [Δ_A [Δ_P] Fin [TP *pro/pronoun* T ...

Furthermore, person features on the verb must be bound by speaker/addressee coordinates in the CP area of the clause. Frascarelli (2018) suggests that 1st and 2nd person subject pronouns are not sanctioned by a topic projection, but by Logophoric coordinates in the vein of Sigurðsson (2011). However, Frascarelli's approach does not fully clarify what the trigger is for realizing pronouns in first and second person, given that both – overt and null pronouns, being deictic elements – must be linked to Logophoric coordinates in C.

That person is relevant for overt pronoun realization is also demonstrated by several quantitative studies (e.g. Enríquez 1984; Morales 1997; Posio 2011; Erker & Guy 2012; Adli 2019). However, dialectal variation and verb type also play a role. Posio (2011: 795) observes in his study of Peninsular Spanish that 1SG has overall higher overt pronoun frequencies than 2SG, which he attributes to the “egocentric nature of discourse”. However, *creer* ‘think/believe’ is an exception in his study (58% with 2SG vs. 55% with 1SG). Furthermore, in Erker & Guy's (2012: 533) study of a corpus of New York City Spanish, with speakers of Mexican and Dominican origin, (*tú*) *sabes* ‘you know’ had overt pronoun frequencies as high as 92% (*ibid.*: 539). Thus, while person might potentially play a role in overt pronoun realization, it clearly cannot be considered without taking into account both verb type and the features of the particular variety under investigation.

2.4 The role of polarity

Posio (2015) compares the positive and negative forms of *saber* ‘know’ plus a complement clause. He found that overt pronoun realization is lower with negative than with positive forms, at least in 1SG (see Posio 2015: 64). The author argues that frequent verbs such as *creer* and *saber* and their positive and negative forms occur in prefabricated, formulaic sequences, which can have varying degrees of pronoun expression. Aijón Oliva & Serrano (2010: 13) also observe that there seems to be a tendency for negative forms to appear with an overt subject less frequently than positive forms with 1SG *creer* (4 null subjects out of 7 and 10 out of 11 in two corpora; see *ibid.*: 13). Although the number of occurrences of negative *no creo* is generally low, there seems to be a tendency for polarity to also affect overt pronoun realization with this verb.

3 Epistemicity, evidentiality, subjectivity and the study of speaker/addressee pronouns with cognitive verbs

The notion of epistemicity standardly refers to the degree of security/certainty/confidence a speaker has in her or his statements or claims. Evidentiality relates to the source of information or source of evidence (e.g. de Haan, 2001; 2005; Cornillie et al. 2015; Aikhenvald 2018). For many authors, the two notions are closely related, such that evidentiality and epistemicity are seen as subcategories in several works (e.g. Palmer 1986; Boye 2012).

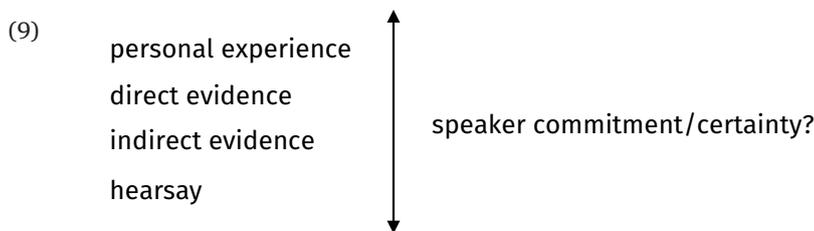
One approach according to which epistemicity and evidentiality belong to one umbrella concept is found in Boye (2012), who uses the term “epistemic support” for epistemic modality and “epistemic justification” for information source (cf. Wiemer 2018). Rooryck (2001: 125) considers evaluatives, subjective epistemic modals and evidentials as interrelated phenomena in that they “all relativize or measure the information status of the sentence”. Lastly, as Wiemer (2018: 86) notes, both concepts relate to “the speaker's cognitive states”, i.e. belief and knowledge.⁶

However, not all languages require a link between epistemicity and evidentiality. It has been argued that the two notions are independent concepts (see Wiemer 2018 and Aikhenvald 2018 for discussion), especially for languages in which evidentials are grammaticalized as discrete morphemes and arranged into specific paradigms. Willett (1988) identifies four different types of source of information (Speas 2004: 257):

⁶ For further discussion of the relation among the notions of epistemicity and evidentiality, see Hennemann (2013).

- (8) personal experience > direct evidence (sensory) > indirect evidence > reported evidence (hearsay)

The categories in (8) reflect different degrees of directness of the evidence that a speaker has for a given assertion. The concept of epistemicity has also been approached in terms of scales (see Givón 1982; Akatsuka 1985; Boye 2012). However, as Speas (2018: 312) observes, although there might be an interdependency between the speaker's degree of certainty/security and the reliability of evidence, these two factors do not form part of the core meaning of an evidential, which is instead dependent on further pragmatic factors. Consequently, while the type of evidence might correlate with different degrees of certainty or reliability in some cases (as informally depicted in (9)), it is possible that the correlation is only a tendency:



Although the differences between evidentiality and epistemicity have been established in the literature, their application to the study of overt/covert pronouns with cognitive verbs in spoken Spanish nonetheless faces some methodological problems. Consider the following example with *creer* 'believe/think':

- (10) (PRESEEA-Madrid, H23_033)
 [Context: talking about whether addressing a person with *tú* might convey excessive familiarity in some situations]
 también **depende** es que <alargamiento/> **no sé** **yo creo que no**
 also depend.3SG is that <lengthening> not know.1SG I think that not
sé que habrá otros factores que <alargamiento/> sean los
 know.1SG that be.FUT other factors that <lengthening> be.SBJV.3PL those
 que te hagan ver <alargamiento/> // con más comodidad
 that you make.SBJV.3PL see.INF <lengthening> with more ease
 o no [...]
 or not
 'It also **depends...**(I) **don't know I think that (I) don't know** that there will be other factors that... will be those that will make you see ... more comfortably or not [...].'

If epistemicity and evidentiality were considered separate factors in the study of overt pronoun realization with *creer*, it would be impossible to determine by means of formal criteria which of the two is decisive in (10). On the one hand, it could be argued that the use of the future form *habrá* indicates a hypothetical situation, which is based on inference (see Squartini 2001 and references for the evidential future). On the other, the degree of speaker commitment cannot be unambiguously determined, although it could be argued that the introduction of *depende* 'it depends' and the repetition of parenthetical *no sé* 'I don't know' indicates that the speaker is not presenting the embedded proposition as fact.

Furthermore, in many cases a fine-grained classification of evidence types as in (8) above cannot be applied to the analysis of cognitive verbs in spoken Spanish. Consider the following example:

- (11) (PRESEEA-Madrid, H21_020)
 [Context: talking about climate changes]:
 que sí sí está cambiando yo creo que sí // la mayoría de la gente
 that yes yes is.3SG changing I think that yes the majority of the people
 piensa que sí /
 think.3SG that yes
 'yes yes, it is changing I think that it does ... most people think that it does'

In the first part of the discourse, the speaker talks about whether he thinks that the climate is changing or not. If we consider this first part only, one could argue that it relies on sensory evidence – climate changes can be perceived by the senses. However, it is impossible to determine whether it is in fact based on personal experience/perception or if it is instead influenced by what others say (scientists, journalists, friends, etc.). In fact, in the second part of the discourse, the speaker explicitly states that the claim of the first utterance is at least partly based on the opinion of others (i.e., hearsay). In many cases, however, this second part is not explicit, rendering the difference between direct and indirect evidence difficult to grasp solely by looking at performance data. Furthermore, the repetition of *sí* could indicate that the speaker’s degree of certainty in asserting that the climate is changing is high. Thus, the information *la mayoría de la gente piensa que sí* ‘most people think that it does’, despite being situated on the level of hearsay, might also be interpreted as supporting the speaker’s opinion, as an anonymous reviewer remarks.

Similarly, with respect to epistemicity/evidentiality and (inter-)subjectivity, it is not always possible to draw a clear distinction when studying overt pronoun realization. For Nuyts (2001), subjectivity has an evidential dimension in that it indicates that the speaker alone draws the conclusions from the evidence s/he has. In (12), although we are dealing with a personal use of *yo creo* ‘I think’, it could also be argued that uncertainty plays a role:

- (12) (PRESEEA-Madrid, H32_043)
- | | | | | | | | | | |
|-----------------------|---------|-----------------|----------------|----------------------|------------|------|----------------|------|---------|
| <simultáneo> | yo | </simultáneo> | creo | que | sí | yo | creo | que | el / |
| <simultaneous> | I | <simulataneous> | think.1SG | that | yes | I | think.1SG | that | the |
| tiempo<alargamiento/> | las | variaciones | climatológicas | afectan | afectan | | | | |
| weather<lengthened> | the | variations | climatological | affect.3PL | affect.3PL | | | | |
| al | al | individuo | creo | mmm | creo | ¿ | eh? | creo | no |
| at-the | at-the | individual | think.1SG | hm | think.1SG | eh | think.1SG | not | can.1SG |
| tampoco | porque | claro | no | tengo<alargamiento/> | base | para | poder | | |
| neither | because | clear | not | have.1SG<lengthened> | base | for | be-able-to.INF | | |
| opinar | ¿ | no? | | | | | | | |
| opinion.INF | no | | | | | | | | |
- ‘I think that it does. I think that the weather... the variations in the climate affect people. (I) think, right? (I) think. (I) cannot really because of course (I) don’t have background so as to voice an opinion, right?’

Posio (2014: 13) argues that *yo creo* with an overt subject pronoun signals a “confident epistemic stance” in some contexts. However, in (12), the speaker relativizes the reliability of his opinion, repeating *creo* ‘I think’ and stating that his opinion might not be sufficiently grounded. Thus, while overt pronoun use might correlate with a higher degree of (inter-)subjectivity, several examples indicate that an adaptation on the scales of epistemicity and/or evidentiality also plays a role.

It is difficult to clearly identify whether the use of (a pronoun plus) *creer* should be interpreted as a hypothesis or personal opinion in many cases. As Aijón Oliva & Serrano (2010: 12) observe, the two concepts form a continuum rather than representing discrete categories (see also De Saeger 2008). One challenge when examining the influence of perspectival factors on overt pronoun realization with cognitive verbs, therefore, is determine the formal criteria by means of which they can be studied. In the corpus study presented in the next section, I decided to focus on factors belonging to the *local* context – the type of information encoded in the embedded clause.

4 The study: Null and overt speaker/addressee subjects with *creer* and *saber*

In this section, the study of pronoun use with *creer* ‘think/believe’ and *saber* ‘know’ will be outlined. After a presentation of the data and a description of the methodology in section 4.1, the results will be presented in sections 4.2 and 4.3.

4.1 Data and methodology

Sentences containing *creer* and *saber* in 1st and 2nd person singular were extracted from the PRESEEA (2014–) corpus: 18 samples from Madrid and 18 from Alcalá (<http://preseea.linguas>).

net). The corpus was suitable because it contains data from (semi-directed) spoken Spanish and, as is well known, the verbs *creer* and *saber* in 1st and 2nd person are particularly frequent in spoken language. In the sociolinguistic interviews, the informants are offered several topics of conversation: greetings, the weather, where the informants live, family and friendship, customs, danger of death, important stories from their lives, and wishes for economic improvement (see Moreno Fernández 2005: 128). What is particularly relevant for the present study is that they contain thematic blocks that stimulate the expression of opinions.

The interviews from the Madrid and Alcalá samples of the PRESEEA (2014-) corpus feature one interviewer and one interviewee. The data from both interlocutors were included in the analysis because many 1SG forms of *creer* and *saber* occurred when the interviewees were speaking, while 2SG with *creer* predominantly occurred when the interviewers were speaking. However, the number of data points is unbalanced between 1SG and 2SG, with the former occurring more frequently (Table 2). Thus, although [person] was integrated into the study of the morpho-syntactic variables, the study of complement type had to be restricted to 1SG.

The data were extracted by means of word form searches, including present, indicative past, imperfective, future and conditional forms. In total, 1076 sentences of the Madrid sample and 836 sentences of the Alcalá sample were extracted, i.e. a total of 1912 sentences containing either 1st or 2nd person singular *creer* or *saber* (excluding repetitions and fixed expressions like *yo qué sé* ‘what do I know’).⁷ However, only verbs with a CP complement (i.e. with the complementizer *que*, *si* or a *wh*-element) were included in the study.⁸ This was necessary for two reasons. First, one main interest was in whether the type of information that a cognitive verb refers to had any effect on overt pronoun realization with the matrix verb. This was least ambiguous if the clausal complement was realized. Second, it was a way to exclude potential semi-fixed parentheticals of the type *no sé* ‘(I) don’t know’ or *¿sabes?* ‘you know?’ which behave very differently with respect to subject pronoun expression. In fact, as Table 1 shows, overt pronoun frequencies differ considerably between [+CP] and [-CP] configurations.

	overt	null	total	%-overt
[+CP]	491	478	969	51%
[-CP]	123	820	943	13%

Table 1 Pronoun realization with 1sg/2sg *creer* and *saber* according to [+CP] vs. [-CP].

Out of these 969 [+CP] sentences, 875 were analyzed, excluding elliptical configurations, cases of repetitions of the embedded event, and modal *saber* + infinitive. Table 2 shows the total number of analyzed sentences.

	total			Madrid			Alcalá		
	1SG	2SG	both	1SG	2SG	both	1SG	2SG	both
<i>creer</i>	457	72	529	288	45	333	169	27	196
<i>saber</i>	264	82	346	136	40	176	128	42	170
both	721	154	875	424	85	509	297	69	366

Table 2 Total of analyzed configurations including *creer/saber* plus CP.

4.1.1 Data classification with respect to morpho-syntactic criteria and phonetic realization

The data were manually classified according to verb (*creer* vs. *saber*), person (1SG vs. 2SG), polarity (positive vs. negative), and null vs. overt realization of the subject pronoun.

The question of the classification of null vs. overt subject pronouns required some attention: in the generative literature, a difference is drawn between left-dislocated preverbal overt subjects and non-dislocated ones (e.g. López 2009 and references therein). In some studies, preverbal

⁷ A verb was counted as a repetition if more than one occurrence of it referred to the same event. The same principle applies to repetitions of pronouns.

⁸ Sentences containing *Creo que sí/no* ‘I think [that] yes/no’ were also included. However, for the analysis of complement type (see section 4.1.2), the sentences were only included if the information to which the polarity item refers could be reconstructed from context. Otherwise, it was marked as DOUBT.

overt referential subjects are always analyzed as the topic, the real subject being a null *pro* or agreement morphology on the verb (e.g. Alexiadou & Anagnostopoulou 1998; Ordóñez & Treviño 1999; Frascarelli 2007; Barbosa 2009).

For the quantitative study, every realized pronoun which was related to the main verb, i.e. had the same reference as its external argument, was counted as overt, regardless of distance. A subject pronoun is thus counted as overt whether it is adjacent to the verb or not. This was necessary because of the potential ambiguity of adjacent preverbal subject pronouns with respect to their position.

In section 5.2, some non-quantitative evidence will be provided to show that although 1SG pronouns *can* appear in a high CP position, pronouns adjacent to the verb must also have a derivation in which they appear in the Spec of a perspectival projection within the split-IP. This is in line with arguments in the literature that preverbal overt subjects do not necessarily share the same position as left-dislocated objects (cf. Suñer 2003).

4.1.2 Data classification with respect to complement type

When examining spoken data, it is not always possible to unambiguously determine whether the concept of epistemicity, evidentiality and/or subjectivity has an impact on the use of an overt or a null pronoun with cognitive verbs (see section 3). Therefore, rather than *directly* investigating the relevance of these concepts for the expression or omission of a pronoun, classification criteria were applied to allow the influence of these notions to be examined in an *indirect* way. The type of belief or knowledge that is expressed in the embedded clause of a cognitive verb was therefore examined. In order to classify embedded complements, the main criterion was whether the information encoded in the embedded clause is *potentially* visibly or audibly perceivable or not. While actual direct perception (or direct evidence) can often not be detected unambiguously in spoken data, it is in many cases possible to determine whether the embedded proposition encodes *potentially* directly perceivable information (e.g. a description of an object in the external world) or not (e.g. state-of-mind of somebody else, unreal situations, etc.).

It is important to note that a classification based on this criterion is a simplification in that it diverges from the notion of evidentiality discussed in section 3: an embedded event could be non-visually perceivable in the classification applied here, even though it could potentially rely on direct evidence. Consider the following example:

- (13) (PRESEEA-Alcalá, H13_001)
[Context: talking about changes that have been made in the city]
yo creo que se hacen cosas / que <vacilación/> / que no están bien
I think that SE make.3PL things that <hesitation> that not are well
'I think that there are things that are not well-done.'

In (13), the changes in the city can be directly observed and could thus be based on direct evidence. However, the evaluative predicate *bien* 'well/correct' indicates that we are dealing with a personal evaluation/judgment, which can be inferred from direct evidence but cannot itself be directly, visually perceived. Rather, what can be perceived is the event or state with respect to which the evaluation/opinion has been formed.

On the contrary, the following represents a clear case of an externally perceivable embedded clause:

- (14) (PRESEEA-Madrid, H12_007)
pero sé que el año pasado / creo que hubo una exhibición
but know.1SG that the year past think.1SG that was.3SG an exhibition
'but (I) know that last year ... (I) think that there was an exhibition'

Here, whether there was an exhibition or not can potentially be directly observed, i.e. it can be proven or falsified by means of first-hand, visual information.

In what follows, the categories of embedded clauses of *creer* 'think/believe' and *saber* 'know' will be presented. Several corpus examples were classified as DESCRIPTIONS of local or temporal information (see (15)):

- (15) (PRESEEA-Madrid, M11_004)
 ahí se hace los viernes por la noche creo que son dos
 there SE make.3SG the Fridays at the night think.1SG that are.3PL two
 <vacilación/> do <alargamiento/>s <vacilación/> / dos fines de semana
 <hesitation> two <lengthening> <hesitation> two weekends
 al mes // es por la noche
 per month is.3SG at the night
 ‘It’s done there on Friday evenings (I) think that it is two weekends per month It is
 at night’

With *saber* ‘know’, embedded *wh*-interrogatives with *dónde* ‘where’ and *cuándo* ‘when’ were also included in this category, given that replacing the *wh*-element by a noun phrase would yield a directly perceivable event. The following embedded clauses were also in the category of DESCRIPTIONS: directly perceivable objects, events, states, places, persons, quantities, and past events experienced by the speaker (see Appendix A for examples).

A further type of embedded complement contained an EXISTENTIAL either in the form of the impersonal verb *haber* ‘there is/are’ or with *tener* ‘have’ or *existir* ‘exist’. Here, the embedded verb refers to the existence of an object or state (see (14)).

On the other hand, embedded complements of cognitive verbs were classified according to events or states that are not visually/audibly perceivable. The first category includes complements with an evaluative predicate (such as *importante* ‘important’, *bueno* ‘good’, *malo* ‘bad’, *mejor* ‘better’, *(a)normal* ‘(ab)normal’, *afortunado* ‘lucky/fortunate’, etc.) and was tagged as EVALUATION. We already saw an example with *creer* in (13). Example (16) shows an evaluative complement with *saber*:

- (16) (PRESEEA-Alcalá, H23_007)
 pero <alargamiento/> yo sí sabía que un perro es **muy esclavo** /
 but <lengthening> I yes knew.1SG.IPFV that a dog is very slave
 que te **cambia** <alargamiento/> **mucho la vida**
 that you.DAT change.3SG <lengthening> much the life
 ‘but I did know that a dog is very servile, that it changes your life’

Here, the embedded complement does not encode an objective description of an animal, but a personal evaluation.

Also included in the category of EVALUATION were those complements that contained information about abstract concepts that require a personal definition (such as ‘friendship’, ‘family’, or ‘relations among people’), or when the speaker expressed a recommendation (see Appendix A), often introduced by modal *deber* ‘shall’ or *tener que* ‘have to’.

Further types of embedded propositions encoding non-visual information were MIND_SELF/OTHER (see (17)), expressing the state of mind of the speaker or another individual (or both), and UNREAL or *irrealis* events (often in the conditional mood; see Appendix A):

- (17) (PRESEEA-Madrid, M11_004)
 yo creo que también la gente tampoco está concienciada de que
 I think that also the people also.not is.3SG become.aware-PTCP of that
 realmente puede llegar a pasar algo
 really can arrive at happen.INF something
 ‘I think that the people are also not aware that something could really happen’

Apart from encoding non-visual information, several of these examples imply an evaluation by the speaker, even though it is not marked by an evaluative predicate.

In a few cases, there was a combination of the categories of EVALUATION and UNREAL, particularly when an embedded event in the conditional mood also contained an evaluative predicate, or of the categories of EVALUATION and MIND_SELF/OTHER:

- (18) (PRESEEA-Madrid, M11_004)
 yo creo que es igual / que la gente no <alargamiento/> <vacilación/> no
 I think that is same that the people not <lengthening> <hesitation> not
 se conciencia
 REFL become-aware
 ‘I think it’s the same ... that the people do not become aware’

If these cases clearly expressed an evaluation by the speaker (as in (18)), they were annotated as EVALUATION, otherwise they were marked as doubt cases.

The final class was composed of NONFINITE complements of negative *saber*, which are introduced by *si* ‘if/whether’ or a *wh*-element:

- (19) No sé qué decirte.
 not know.1SG what say.INF-you
 ‘(I) don’t know what to tell you.’

Even though these control complements have an inherent future or *irrealis* interpretation (see Landau’s 2000 Partial Control), they are subject to configurational restrictions which finite *irrealis* complements are not: they only appear with negative *saber*. It therefore seems legitimate to consider these separately.⁹ Table 3 summarizes the classification criteria of embedded complements.

category	criteria	embedded information = sensorily perceivable?
EVALUATION	evaluative predicate, abstract concept requiring a personal definition, recommendation, evaluation of veracity	no
MIND_SELF/OTHER	emotions, thoughts, state of mind of an individual	no
UNREAL	conditional mood, <i>irrealis</i> , not fulfilled events	no
DESCRIPTIONS	temporal and local indications, descriptions of external events (present/past), objects, quantities, manners, persons/names, places, states	yes
EXISTENTIALS	<i>haber, tener, existir</i> plus object/state	yes
NON-FINITE (negative <i>saber</i> only)	introduced by <i>si</i> or a <i>wh</i> -element	unclear

Table 3 Classification of embedded complements of 1SG *creer* and *saber*.

It should be noted that there were several cases that could not be assigned an unambiguous category of annotation. For example, it was sometimes not possible to establish whether the complement in question was an EVALUATION or a DESCRIPTION:

- (20) (PRESEEA-Madrid, H12_007)
 yo creo que antes estábamos mucho mejor comunicados // que ahora ¿no?
 I think that before were.1PL much better connected than now no
 ‘I think that before (we) were better connected than now, right?’

Although it could be argued that including *mejor* ‘better’ in the embedded clause implies a subjective evaluation on the part of the speaker, it could also be argued that the embedded clause contains a description in terms of ‘more public transport’, which can be objectively falsified. These ambiguous cases were excluded as DOUBT cases (see also Appendix A).

The classification of complements of cognitive verbs outlined in this section is designed to investigate whether the type of embedded belief or (lack of) knowledge has any effect on overt pronoun realization with a matrix cognitive verb. However, it must be emphasized that this classification cannot examine the relevance of the concepts of epistemicity, evidentiality and/or subjectivity for overt pronoun realization in a *direct* way. It instead attempts to achieve this indirectly by looking at the type of belief that is expressed in the embedded clause. Furthermore, it reduces clause-external information to a minimum, focusing on the material contained in the

⁹ Nonfinite complements of *creer* are rare – there were only two examples of *Creo recordar* ‘As far as I remember’ which were classified as MIND_SELF.

embedded clause. The results should thus be considered with these reservations in mind; the advantage of this approach, however, was that the classification could be applied to a larger data set and could to a large extent reduce, though not fully eradicate, potential ambiguities.

4.1.3 Statistical analysis

The study consists of two parts: the first investigates whether there is an association between overt pronoun realization and the morpho-syntactic variables of [person] and [polarity] with *creer* and *saber*, as has been postulated in previous studies. All variables are nominal/binary: overt vs. covert, *creer* vs. *saber*, 1SG vs. 2SG, negative vs. positive. Pearson’s chi-squared test was applied to the respective (sub-)contingency tables in R (R Core Team 2018) to extract the *p*-values.¹⁰ For the effect strength, Cramer’s V was extracted with the *VCD* package (Meyer at al. 2020). For those categories with a low number of data points, Fisher’s Exact Test was used.¹¹ Given the application of multiple testing (11 comparisons), Bonferroni-Holm correction was applied for *p*-value adjustment.

The second part of the study consists of a detail examination of the type of complement of a 1SG cognitive verb (section 4.1.2). Fisher’s Exact Test was used to check whether there is an association between overt pronoun frequencies in the matrix clause and the type of complement. In order to explore the data more in detail, multiple comparisons were made (with the *RVAIDEMEMOIRE* package; Hervé 2021) between the different categories of complement type. Also in this case, *p*-values were adjusted by means of Bonferroni-Holm correction. The data frames used for the statistical analysis can be consulted at <https://doi.org/10.5281/zenodo.5035307>.

4.2 Results of [verb], [person], and [polarity]

Table 4 presents the results of the variables [verb], [person] and [polarity] with 1SG/2SG *creer* and *saber* [+CP].

verb	polarity	person	overt	null	total	%
<i>creer</i>	pos	1SG	336	100	436	77%
		2SG	38	32	70	54%
		TOT	374	132	506	74%
	neg	1SG	9	12	21	43%
		2SG	0	2	2	0%
		TOT	9	14	23	39%
	TOT	1SG	345	112	457	75%
		2SG	38	34	72	53%
		TOT	383	146	529	72%
<i>saber</i>	pos	1SG	20	23	43	47%
		2SG	10	62	72	14%
		TOT	30	85	115	26%
	neg	1SG	43	178	221	19%
		2SG	0	10	10	0%
		TOT	43	188	231	19%
	TOT	1SG	63	201	264	24%
		2SG	10	72	82	12%
		TOT	73	273	346	21%

Table 4 Overt pronoun realization with 1SG/2SG *creer* and *saber* [+CP].

¹⁰ The significance level $p_{critical}$, which functions as “the threshold value for rejecting or sticking to H_0 [the Null Hypothesis; added by PH]” (Gries 2013: 27) is set to 0.05.

¹¹ Thanks to an anonymous reviewer for this suggestion.

Figures 1–3 depict the effect of verb, person and polarity in the full sample.

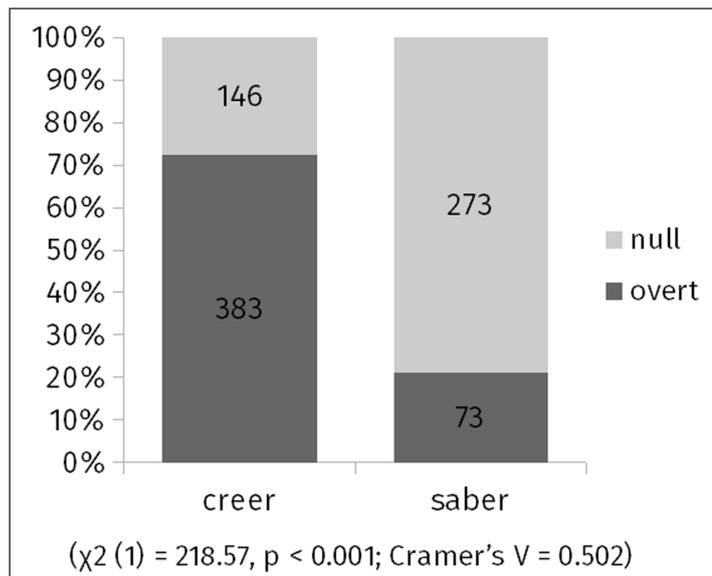


Figure 1 Null and overt Speech Act Participant pronouns with *creer* and *saber*.

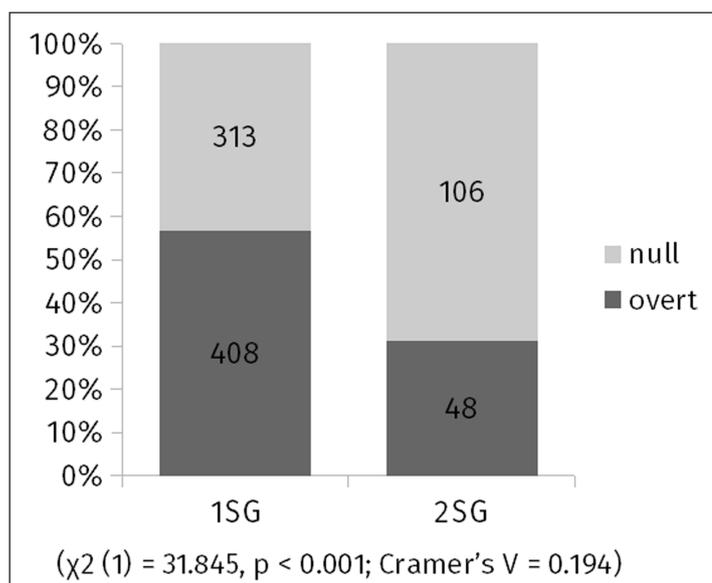


Figure 2 Null and overt Speech Act Participant pronouns according to person.

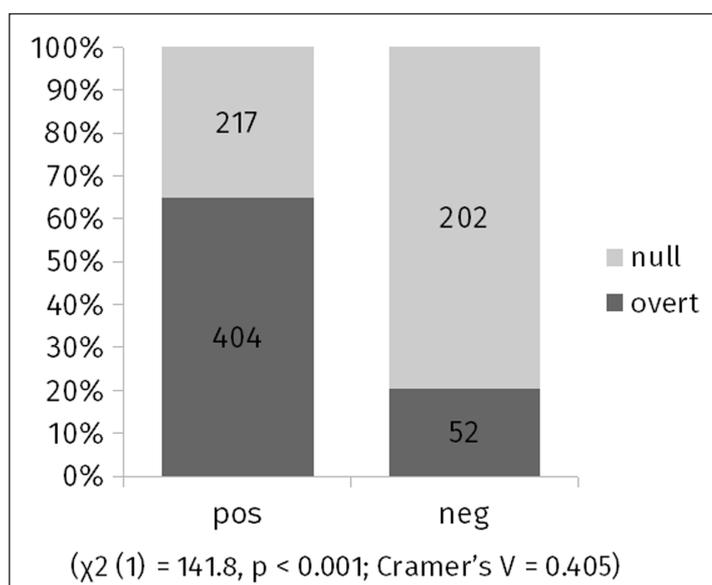


Figure 3 Null and overt Speech Act Participant pronouns according to polarity.

As indicated under each figure, the values obtained through statistical analysis show that there is a significant association between subject pronoun expression and (i) verb (*creer/saber*), (ii) person (1SG/2SG) and (iii) polarity (neg/pos), respectively. Overt pronoun realization is higher with *creer* than with *saber* (72% vs. 21%; $p < 0.001$), higher with 1SG than with 2SG (57% vs. 31%; $p < 0.001$) and higher with positive than with negative verb forms (65% vs. 20%; $p < 0.001$). The association between subject pronoun expression and [verb] as well as the association between subject pronoun expression and [polarity] has a moderate effect size (Cramer's $V = 0.502$ and 0.405 , respectively), but the association between subject pronoun expression and [person] has only a small effect size (Cramer's $V = 0.194$).¹²

Considering the association between [person] and subject pronoun expression, results are only significant with affirmative *creer* and *saber*; the effect not being significant with the negative verb forms. **Figures 4** and **5** depict pronoun frequencies with positive verb forms.

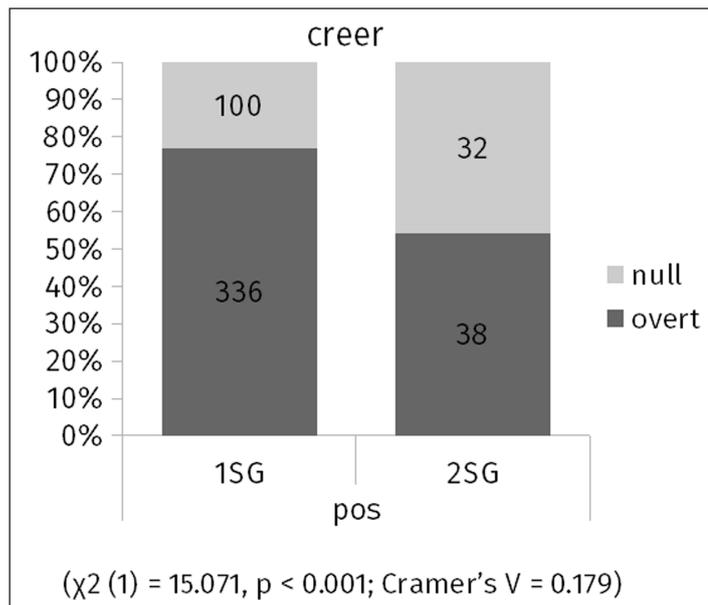


Figure 4 [person] with positive *creer* [+CP].

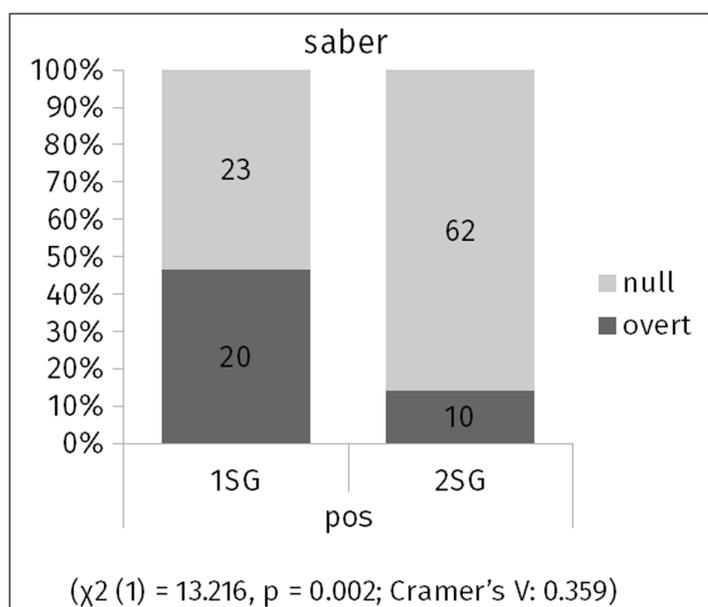


Figure 5 [person] with positive *saber* [+CP].

As **Figure 5** shows, positive 1SG (*yo sé* '(I) know' [+CP]) has overt subject pronoun frequencies of 47%, compared with 14% overt pronouns with 2SG (*tú sabes* '(you) know'). The difference is significant ($\chi^2 (1) = 13.216, p = 0.002; \text{Cramer's } V: 0.359$). With respect to positive *creer* [+CP] (see **Figure 4**), 1SG (*yo creo* '(I) think') also has significantly higher overt pronoun rates

¹² See Levshina (2015: 209) for guidance on interpreting Cramer's V .

than 2SG (*tú crees* ‘(you) think’ (77% vs. 54%; $\chi^2(1) = 15.071, p < 0.001$), but the effect strength is lower (Cramer’s $V = 0.179$).

With respect to the negative verb forms, there are only a small number of data points for 2SG (see [Table 4](#)). Thus, there are only two occurrences of 2SG *no crees* ‘(you) don’t think’ [+CP] in the sample (both appearing with a null pronoun) and the comparison with 1SG *no creo* (9/21 = 43% overt pronouns) is not significant ($p = 0.86$). With negative *saber* ‘know’ [+CP], subject expression has a tendency towards being higher in 1SG (43/221 = 19%) than in 2SG (0/10 = 0%) but the effect is not significant ($p = 0.86$).

Turning to the association between subject pronoun expression and polarity, Aijón Oliva & Serrano (2010) report a tendency for overt pronouns to be more frequently used with positive 1SG *creo* than with negative *no creo* (see section 2.4). Although there are only a small number of data points for negative *no creo que* ‘(I) don’t think that’ in the sample investigated here, the results show a tendency in the same direction ([Figure 6](#)).

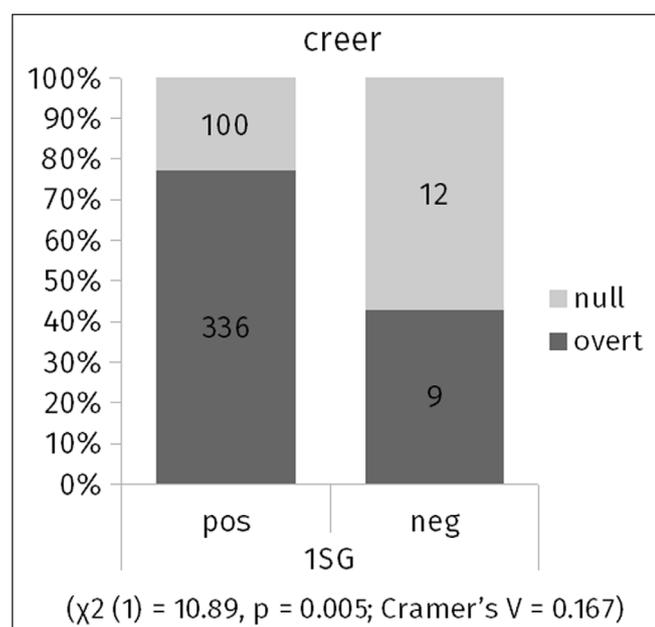


Figure 6 Positive and negative (yo) *creo* [+CP].

Positive 1SG (*yo creo que* ‘(I) think that’ occurred with an overt pronoun more often (336/436 = 77%) than negative 1SG (*yo no creo que* ‘(I) don’t think that’ (9/21 = 43%). The association is significant ($p = 0.005$), even though the effect strength is small (Cramer’s $V = 0.167$).

Posio (2015) observes higher overt pronoun frequencies in positive 1SG (*yo sé* ‘(I) know’ than in negative (*yo no sé* ‘(I) don’t know’. Again, the results in [Figure 7](#) show the same tendency: Positive forms of 1SG *saber* ‘know’ [+CP] have significantly higher overt pronoun rates than negative forms (47% vs. 19%; $p = 0.002$). The association between overt pronoun realization and polarity does not prove significant in either 2SG *creer* ‘think’ or 2SG *saber* ‘know’ ($p = 0.86$).

4.2.1 Discussion

The results with respect to pronoun frequencies, verb, and person confirm the findings of previous studies on the peninsular Spanish varieties (e.g. Enríquez 1984, Davidson 1996, Posio 2013, 2015). They indicate that overt pronoun realization is especially favored with ‘believer/thinkers’ and less so with ‘knowers’. However, a further factor is polarity, in that in the 1SG, positive verb forms trigger higher overt pronoun rates than negative forms. This confirms findings of Aijón & Oliva Serrano (2010) and Posio (2015). In fact, 1SG positive (*yo sé* ‘(I) know’ [+CP] has an overt pronoun rate of 47% (see [Figure 5](#)). It should be noted that *saber* is generally more frequently used with negation than *creer* (see also Posio 2013: 280). As shown in [Table 4](#), only 23/529 sentences with *creer* appeared with negation, compared to 231/346 negative forms of *saber*. Thus, the general tendency to use *creer* in affirmative contexts could also have an effect on overt pronoun realization, as well as the semantic difference between the two verbs. With respect to the effect of polarity with 2SG forms, we cannot draw firm conclusions, given the low number of negative 2SG in the data.

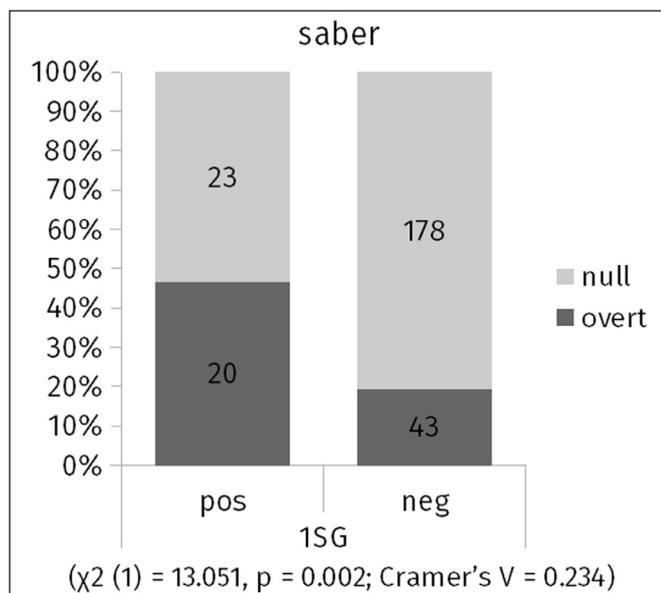


Figure 7 Positive and negative (yo) sé [+CP].

Posio (2013; 2015) argues that negative *no sé* ‘(I) don’t know’ with a null subject appears in formulaic sequences. On a more general level, the relevance of polarity with 1SG could also be explained if perspectival notions are taken into account: overt pronoun realization is more frequent in the person specification reflecting the speaker’s perspective. Furthermore, positive verb forms imply a higher degree of speaker commitment or “speaker involvement” (De Saeger 2008), than negative ones. As Aijón Oliva & Serrano (2010: 9) point out, sometimes subject pronoun expression equips the embedded event with a higher “assertivity” or “pragmatic force”. Negative verb forms often relativize the importance or truth value of the embedded proposition (see *ibid.*: 13). Thus, the assertivity of the embedded complement of positive vs. negative matrix cognitive verbs might play a role (see also section 6.2).

4.3 Results of the analysis of the complement of (yo) *creo* and (yo) *sé*

Table 5 presents the main numbers of the annotation of complement type with 1SG (yo) *creo* and (yo) *sé*.

complement type	(yo) <i>creo</i>	(yo) <i>sé</i>
EVALUATION	113 (33%)	23 (10%)
MIND_SELF/OTHER	70 (20%)	44 (20%)
UNREAL	45 (13%)	36 (16%)
EXISTENTIAL	22 (6%)	7 (3%)
DESCRIPTION	94 (27%)	94 (42%)
WH/SI-INFINITIVE	---	18 (8%)
TOTAL (analyzed)	344	222
doubt (excluded)	113	42
TOTAL	457	264

Table 5 Complement types with (yo) *creo* and (yo) *sé*.

Independently of overt pronoun realization, 1SG *creer* more frequently co-occurs with EVALUATIONS (113/344 = 33%) than 1SG *saber* (23/222 = 10%). Complements expressing DESCRIPTIONS, in contrast, appeared more frequently with 1SG *saber* (94/222 = 42%) than with 1SG *creer* (94/344 = 27%).

Figure 8 shows overt pronoun frequencies depending on complement type with 1SG *creer*.

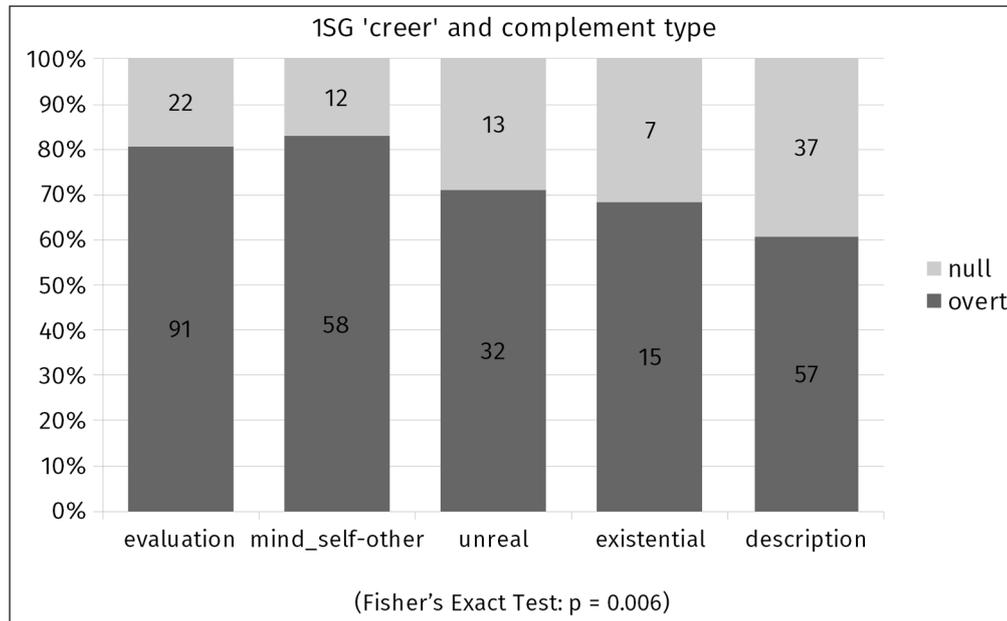


Figure 8 Subject expression with 1SG *creer* according to complement type.

The overall association between complement type and 1SG pronoun expression with *creer* is significant ($p = 0.006$). Subject pronoun expression is highest with 1SG *creer* if it takes complements tagged as EVALUATION (91/113 = 81%) and MIND_SELF/OTHER (58/70 = 83%) and lowest with complements expressing DESCRIPTIONS (57/94 = 61%). An anonymous reviewer asks Fisher's Exact Test to be applied to comparisons of subject expression with different complement types. If pairwise comparisons are made (Fisher's Exact Test with Bonferroni-Holm correction), the only two comparisons that result significant are those between EVALUATION and DESCRIPTION ($p = 0.02$) and between MIND_SELF/OTHER and DESCRIPTION ($p = 0.028$). In fact, these are important comparisons if the status of the embedded complement as [sensorily perceivable] vs. [non-sensorily perceivable] (see Table 3) plays a role for 1SG subject expression with matrix cognitive verbs. The other comparisons are not significant (see Appendix B).

Figure 9 shows overt pronoun frequencies with 1SG *saber* according to complement type.

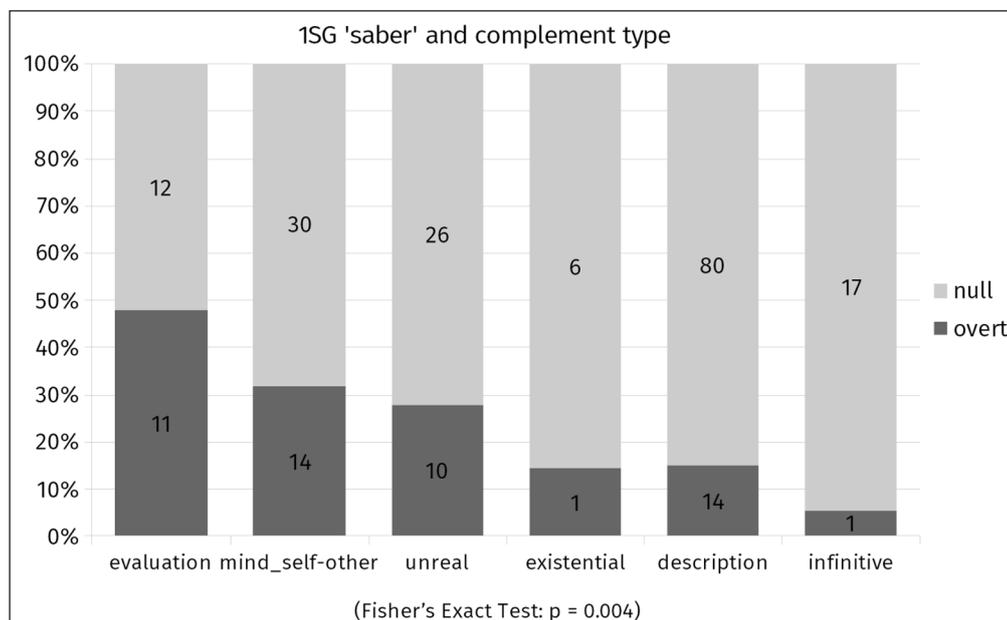


Figure 9 Subject expression with 1SG *saber* according to complement type.

The overall association between overt 1SG subject pronoun expression with *saber* and complement type is significant ($p = 0.004$). It is interesting to note that the highest overt pronoun rates ($11/23 = 48\%$) are also found with *saber* if its complement is tagged as EVALUATION. However, the total number of cases is substantially lower than with *creer*. Complements classified as MIND_SELF/OTHER and UNREAL have overt pronoun frequencies of 32% ($14/44$) and 28% ($10/36$), respectively. Just as with *(yo) creo que*, 1SG *saber* has lower overt pronoun frequencies if its complement is tagged as DESCRIPTION ($14/94 = 15\%$). The lowest overt pronoun frequencies were found with 1SG *saber* selecting an infinitive introduced by a complementizer ($1/18 = 6\%$).

If pairwise comparisons are made between different complement types, the comparison between EVALUATION and DESCRIPTION is significant ($p = 0.021$). This indicates that, also with *saber* ‘know’, evaluative complements favor 1SG pronoun expression. The other comparisons are not significant (see Appendix B), which indicates that they have to be tested against a larger data set in future studies.

4.3.1 Discussion

The analysis of complement type indicates that 1SG pronoun expression is frequent with *creer* and *saber* when the cognizer is an evaluator. The high rates of 1SG pronoun expression with evaluative complements supports the findings of Aijón Oliva & Serrano (2010), Posio (2014) and Hennemann (2016) that *yo creo* ‘I think’ is favored in contexts of personal opinion and (inter-) subjectivity. Furthermore, it indicates that the role of evaluator might be relevant for 1SG *saber* as well, even though fewer data points are available. With 1SG *creer*, also the comparison between MIND_SELF/OTHER and DESCRIPTION results significant. The evidence stemming from this comparison is less clearly related to argumentative uses of *yo creo*. However, it can still be argued that subjectivity and epistemicity/evidentiality are a factor: we are dealing with less concrete, more abstract information, which cannot be falsified by means of visual evidence. This is in contrast to the category of DESCRIPTION which, apart from encoding externally perceivable information, expresses beliefs with respect to information that is less subjectively debatable. However, as I have underlined throughout the paper, the data can only give an indirect indication of the relevance of these concepts for overt pronoun use.

Figure 9 shows that overt pronoun realization frequencies are lowest with 1SG *saber* when it selects infinitives introduced by a complementizer, although the comparisons with other complement types are not significant. If this tendency is confirmed by future studies, there could be two potential reasons for low overt pronoun frequencies. First, negative forms of cognitive verbs have low overt pronoun frequencies and *si + inf* or *wh + inf* appear only in the complement of negative *saber*. Second, if overt pronoun realization is related to assertivity (Aijón Oliva & Serrano 2010), one reason for low overt pronoun frequencies could be that infinitives are not asserted (see Heycock 2006: 189, citing Hooper & Thompson 1973).

With *(yo) creo*, overt pronoun expression still has a percentage of 61% in the DESCRIPTION category. This figure should be considered in light of the following: one specific context within the category of DESCRIPTION that favors overt pronoun expression is when the speaker talks about climate changes (see example (11)). In fact, if the context of climate change is separated from other subtypes of DESCRIPTION (object, local, temporal, etc.), the overt pronoun rates are as shown in Table 6:¹³

	overt	null	total	%-overt
DESCRIPTION (without meteorological)	31	36	67	46%
DESCRIPTION (with meteorological)	26	1	27	96%

Table 6 *(yo) creo* and “descriptions” of climate changes.

When the climate change context is removed from the category of DESCRIPTION, the latter has 1SG overt pronoun frequencies below 50%, while contexts of ‘beliefs’ with respect to climate change have a rate of 96%. One explanation is that these complements often involve

13 No cases of *(yo) sé* ‘(I) know’ were identified in this context.

comparatives, e.g. comparing two temporal spaces (past vs. present) or quantities. Thus, we are dealing with an evaluation of temporal spaces, effects, and causes.

One type of sentence occurred multiple times at the end of an interview in the corpus, which could not be assigned a clear category of complement type:

- (21) (PRESEEA-Madrid, M12_010)
 bueno pues yo creo que ya hemos terminado
 well well I think.1SG that already have.1PL finished
 ‘Well ... I think that (we) have already finished.’

This type of sentence often occurred with an overt pronoun; one explanation for this could be that the speaker is not merely reporting that the interview has finished, but is implicitly awaiting the addressee’s approval. This becomes explicit in the following example:

- (22) (PRESEEA-Alcalá, H32_033)
 claro / pues yo creo que ya hemos terminado ¿qué dices tú?
 sure well I think.1SG that already have.1PL finished what say.2SG you
 ‘Sure... well I think that (we) have already finished. What do you say?’

Overt pronoun use in (21)/(22) might be influenced by the intersubjective relation between speaker and addressee (as argued by Hennemann 2016 for *creo yo*).

Thus, the results shown in [Figures 8](#) and [9](#) should be considered in the context of the annotation of complement type as outlined in section 4.1.2, which reduces sentence-external information to a minimum. If further contextual factors are considered in detail, it seems to be the case that the factor of (inter-)subjectivity might also play a role in some cases beyond the category EVALUATION. In the next section, we will have a look at how the relevance of perspectival factors for overt subject expression can be encoded in syntax.

5 Towards an analysis of overt/covert alternations with cognitive verbs

In this section, an analysis of the interpretative and syntactic properties of overt 1SG pronouns with cognitive verbs and their relation to contrast will be outlined. On the interpretative side, it will be argued that evaluative contexts favor contrastive interpretations, evoking perspectival alternatives (section 5.1). This is in line with assumptions in the literature that verbs of personal opinion have high frequencies of overt subject pronouns because they favor contrastive contexts (see Enríquez 1984; Fernández Soriano 1999). In syntax, it will be argued that overt *yo* ‘I’ is generated in the specifier of a perspectival projection, encoding speaker evaluation, epistemicity and evidentiality (Cinque 1999; Speas 2004) in the extended IP (section 5.2). Furthermore, the status of *yo creo (que)* ‘I think (that)’ as an “epistemic phrase” in the sense of Thompson & Mulac (1991) and Posio (2015) will be analyzed as the outcome of a pragmatization process, in which *yo + creo* directly merge in the perspectival functional projection (section 5.3).

5.1 Epistemicity, evidentiality, subjectivity and the extended IP

Cinque (1999) and Speas (2004) argue that pragmatically relevant features, such as speech act, evaluative, evidential and epistemic mood are encoded as projections in a designated order:

- (23) [_{SpeechActP} Speech Act Mood [_{EvalP} Evaluative Mood [_{EvidP} Evidential Mood
 [_{EpistP} Epistemic Mood]]]]

In sections 4.2. and 4.3, it was shown that overt pronoun expression is frequent when the subject of 1SG *creer* is not merely a cognizer but a subjective evaluator. Moreover, there is non-quantitative evidence that the use of *yo creo* is not only preferred when expressing speaker confidence, but also when subjectivity is a strategy for presenting the embedded proposition as information whose truth value depends on the speaker’s perspective. The close relation between speaker perspective and epistemic/evidential values for overt pronoun realization could be encoded in what Davis et al. (2007), building on Lewis (1976), call the “quality threshold”:

- (24) Lewis (1976: 297):
 “The truthful speaker wants not to assert falsehoods, wherefore he is willing to assert only what he takes to be very probably true. He deems it permissible to assert that A only if $P(A)$ is sufficiently close to 1, where P is the probability function that represents his system of degrees of belief at the time. Assertability goes by subjective probability.”
- (25) Davis et al. (2007: 78):
 Every context c has a quality threshold $C_T \in [0,1]$.
 An agent A can felicitously assert p in context c only if $C_{A,c}(p) \geq C_T$.

If assertability is related to the subjective truth probability, the “quality threshold” of an expressed belief should be dependent on the speaker’s relation to the embedded information. Note that evaluative contexts often imply a contrast with respect to a set of potential other evaluators (which can include the addressee in intersubjective settings):

- (26) (PRESEEA-Alcalá, H23_007)
 yo es que / creo que es una cosa que no tiene tantas variaciones
 I is that think.1SG that is a thing that not have.3SG so-many variations
como a veces se dice que hay ¿no?
 as sometimes SE say.3SG that there-are no
 ‘Me, the thing is that I think that it is something that doesn’t have so much
variation as (they) sometimes say that there is, right?’

Here, the perspective of the speaker is explicitly contrasted with respect to an undefined set of individuals and, thus, an alternative set is evoked (see Rooth 1992; Krifka 2007 for *focus alternatives* and Büring (2003) for *topic alternatives*).

Mayol (2010) argues that Catalan overt subject pronouns encode different types of contrast, one being a “weak contrast”. The author draws on Büring’s (2003) analysis of Contrastive Topic (CT). However, unlike Büring (2003), who argues that CTs introduce alternative sets of questions (i.e. complex sets of alternatives), Mayol (2010) follows Hara & van Rooij (2007) in assuming that CTs create a simple set of topic alternative propositions, combined with a CT-implicature which indicates that one of the topic alternatives is not known to be true by the speaker (cf. Mayol 2010: 2505). This implicature is called an “uncertainty contrast” (*ibid.*: 2506).

Several examples with *yo creo que* ‘I think that’ closely match the interpretation of “weak contrast”. Consider (13), repeated here for convenience:

- (27) [talking about changes in the city]:
 Yo creo que se hacen cosas que no están bien.
 ‘I think that there are things that are not well-done.’

Realization of the 1SG pronoun in this context of evaluation implies an uncertainty contrast. It indicates that the opinion of the speaker may or may not be shared by others, i.e. it evokes a set of alternative perspectives, including other perspective holders:

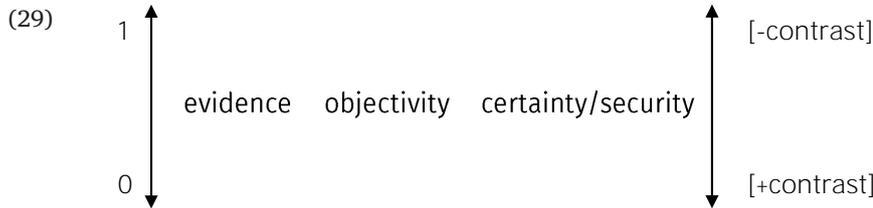
- (28) {The speaker thinks that there are things that are not well-done, The addressee thinks that there are things that are not well-done, ... thinks that there are things that are not well-done, ...}

Realization of the 1SG pronoun contrasts the speaker’s perspective towards the embedded proposition with a potential set of other ‘evaluators’, which might agree with the speaker or not, i.e. the evoked alternatives in (28) might be true or false. This type of contrast is made explicit in (26). It also becomes apparent in examples in which intersubjectivity is crucial (such as (21)), where realizing the 1SG pronoun evokes the addressee as an alternative perspective holder and the question of whether the two perspectives agree remains open.

Evoking alternative perspective holders means that the truth probability of the embedded proposition is consequently adapted to the speaker as an evaluator. The reason for the high frequency of this strategy with the verb *creer* ‘think/believe’ is that it leaves the (subjective)

truth probability of the complement clause largely undefined. This paves the way for further strategies of lowering or elevating the “quality threshold” according to context. Contrast assignment to the subject and correlated 1SG overt pronoun expression can thus be considered one such strategy.

In our data, evaluative contexts favor weak contrast assignment. With the verb *creer* ‘think’, also the category MIND_SELF/OTHER favors 1SG subject expression (see [Figure 8](#)), which indicates that abstract information which is not based on direct perception favors contrast assignment to the subject. Thus, the probability of contrast assignment and the evoking of alternative perspective holders seems to increase in contexts of low objectivity, evidence, and/or certainty:

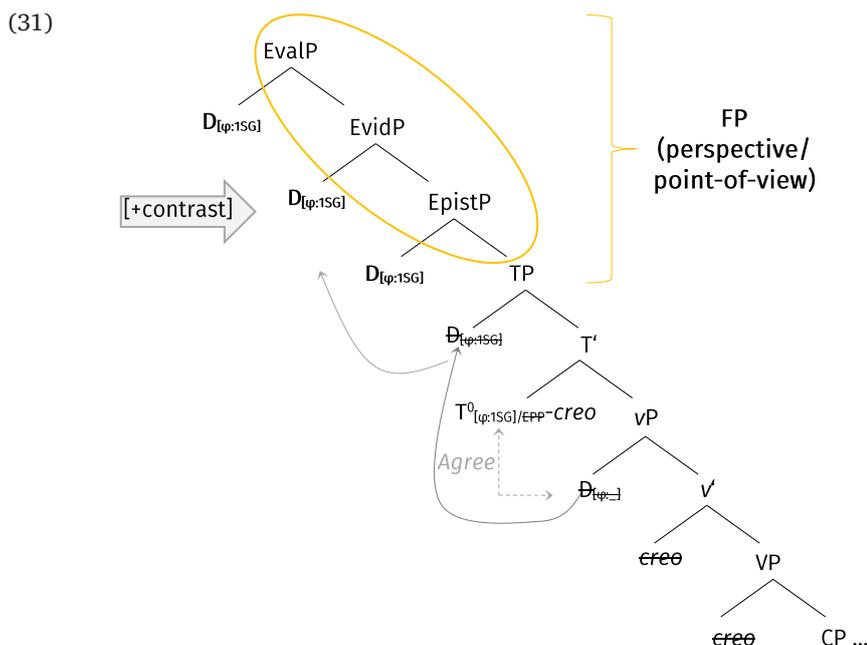


With respect to assignment of the feature [contrast] in syntax, this means that the following preference relations obtain:

- (30)
- a. Epist = [-certainty] → [+contrast]
 - b. Evid = [-direct evidence] → [+contrast]
 - c. Eval = [-objective] → [+contrast]

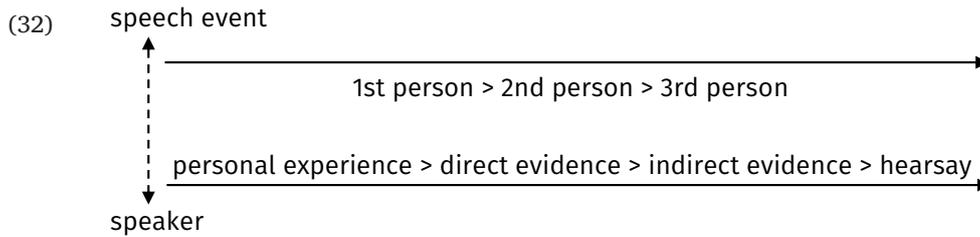
This scale might have a correlate in Cinque’s (1999: 128ff) marked and unmarked values of the Epist and Evid projections, where [direct evidence] is the unmarked value for Evid and [commitment] the unmarked value for Epist (see also Roberts & Roussou 2003 for discussion).

There is evidence that the interaction of the notions of epistemicity, evidentiality and subjectivity has an impact on contrast assignment to the subject position and, consequently, on the use of (*yo*) *creo* and (*yo*) *sé*. Contrast can thus yield an overt, strong pronoun in a functional category encoding perspective or “point-of-view” (e.g. Uriagereka’s 1995 FP; Speas & Tenny’s 2003 Sentience Phrase), in which epistemicity, evidentiality and subjectivity interact to yield contrast assignment:



FP as conceived of here is the interaction between perspectival notions relating to epistemicity, evidentiality and subjectivity, as an extension of the morpho-syntactic IP containing [person] (see section 5.2). A relation between person and perspectival notions such as evidentiality has in fact been postulated by de Haan (2005), who argues that evidentiality is a deictic category,

and Rooryck (2019), who argues that evidentiality encodes *distal* relations.¹⁴ In (32), I depict these views of the deictic nature of evidentiality:



If perspectival notions are projected in syntax as functional categories (Cinque 1999, Speas & Tenny 2003; Speas 2004), it is reasonable to assume that [+contrast] can be assigned to their specifier (Spec,F), in addition to being assigned to Spec,TopP in the high C-domain. The next section lays out the evidence for a low, IP-related position for preverbal *yo* with *creo*.

5.2 Deriving preverbal *yo* with cognitive verbs in syntax

The approach adopted so far assumes that preverbal *yo* ‘I’ is in a position below CP. This is problematic in light of the general assumption that notions such as [contrast] and [focus] form part of the (high) C-domain, as an anonymous reviewer notes.

For Spanish, evidence that preverbal subjects do not share all their properties with left dislocated, topicalized objects in the C-domain has been widely discussed (see Suñer 2003; López 2009). One argument against a uniform dislocation analysis is that SVO, in contrast to OVS, is the unmarked word order and can be the answer to the question ‘What happened?’ (see Zagona 2002).

In the case of *yo creo que*, some evidence that *yo* can appear in a high CP-related or a low IP-related position is functional in nature. In the few cases, in which *yo* appears in a clearly dislocated position, separated from the verb by a non-clitic constituent (8 cases of Subj-XP-V, where XP = phrasal), the overt pronoun seems to have a shifting function and could in fact be argued to be situated in the high left periphery:

- (33) (PRESEEA-Madrid, H32_047)
- E: ¿hay mucho problema en el barrio? [...] concretamente en el
 there-is much problem in the neighborhood concretely in the
 barrio de Salamanca
 neighborhood of Salamanca
- I: no **yo en el barrio de Salamanca** aah no creo
 no I in the neighborhood of Salamanca aah not think.1SG
 que <alargamiento/> no creo si hacemos un término comparativo
 that <lengthening> not think.1SG if make.1PL a term comparative
 con otros barrios
 with other neighborhoods
- ‘E: Are there a lot of problems in the neighborhood? To be precise, in the neighborhood of Salamanca?’ ‘I: No. No. **Me, in the Salamanca neighborhood**, I don’t think that ... I don’t think so, if we compare it to other neighborhoods.’

Here, *yo* shifts from impersonal *hay* ‘there is’ to the speaker’s perspective. The preverbal PP *en el barrio de Salamanca* ‘in the Salamanca neighborhood’ is continuous, referring to information already contained in E’s question. This could be explained if shifting topics are situated above familiar topics in the syntactic tree (see Frascarelli & Hinterhölzl 2007).

In 4 out of 8 examples with unambiguously peripheral *yo*, the discourse marker *es que* ‘[the thing] is that’ appears between the pronoun and the verb (see (26)). It has been argued that *es que* introduces rhematic or focal material (see Fernández Leborans 1992: 239). Sentences with peripheral *yo* followed by *es que* therefore represent configurations in which *yo* is shifted out of the focus projection. However, *yo creo que* can also appear inside the clause introduced by *es que*:

¹⁴ Givón (1982: 44) states that his scale of “subjective certainty” is built upon several hierarchies, including the “personal/deictic hierarchy”, the “sensory/source hierarchy”, and the “proximity hierarchy”.

- (34) (PRESEEA-Alcalá, M31_054)
es que yo creo que ahora la gente son más decididas porque [...]
 is that I think that now the people are more determined because
 ‘[The thing] is that I think that the people are more determined now because [...]’

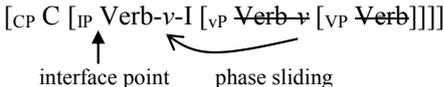
If *es que* introduces focal/rhematic material, preverbal *yo* cannot be a left-dislocated topic in this case.¹⁵ This can be explained if *yo* is a perspectival marker that appears in a position below TopP, i.e. in the extended IP.

Furthermore, in the word order *XP-yo-creo*, the fronted *XP* can be a contrastive topic:

- (35) (PRESEEA-Madrid, M22_030)
 y <alargamiento/> observo que por la calle / mmm sí tienen los dos
 and <lengthening> observe.1SG that on the street mmm yes have.3PL the two
 cubos el amarillo y el normal ¿no? /pero el amarillo yo creo que no se
 bins the yellow and the normal no but the yellow I think that not SE
 saca tanto como <alargamiento/> como el normal /
 take-out so-much like <lengthening> like the normal
 ‘And ... (I) observe that (they) do have both garbage cans on the street – the yellow
 [one] and the normal [one], right? But **the yellow [one]**, I think that (they) don’t
 take it out as often as the normal [one]’

In (35), left-peripheral *el amarillo* ‘the yellow [one]’ is a contrastive topic and *yo creo que* appears below it. If the ContrP is not recursive (Frascarelli & Hinterhölzl 2007: 97), *yo* cannot be a dislocated contrastive topic in this case. As argued here, it is a perspectival marker within the extended IP.

There is therefore evidence that the preverbal 1SG subject pronoun in *yo creo que* is not uniformly a dislocated, C-related topic, but can appear in a low position despite being licensed by (weak) [contrast].¹⁶ In fact, it has been argued for Spanish that Spec,TP/IP can be a pragmatic interface point (see Zubizarreta 1998 for the assumption that [topic]/[focus] phrases can move to Spec,TP). According to Gallego’s (2010) *phase sliding*, V-to-T movement has the effect of extending the vP phase level to TP/IP in Spanish. If pragmatically relevant features are assigned at the phase level (López 2009), it is expected that [contrast] can be assigned to the Spec of (the extended) IP:

- (36) [CP C [IP Verb-v-I [VP Verb-v [VP Verb]]]]


Technically, movement of the subject pronoun would not be triggered by [Case] + EPP, but by weak [contrast] associated with an Edge Feature (Chomsky 2008).

If Spanish allows the assignment of pragmatically relevant features at the IP as well as the CP level, it straightforwardly follows that preverbal *yo* can be licensed in Spec,IP/FP or in Spec,CP, correlating with different discourse functions (see López 2009 for the assumption that subjects can be in Spec,CP or Spec,TP). Generating *yo* with *creo* in Spec,IP/FP corresponds to a weak contrast, motivated by notions relating to epistemicity/evidentiality/subjectivity.

¹⁵ A reviewer points out that the subject pronoun can also be left null in (34) and focused constituents cannot be silent. However, while null pronouns are incompatible with narrow focus, the following example from Brucart (1987: 216) indicates that null and strong pronouns can be part of sentences with wide focus, in which the whole sequence represents new information:

- (i) (Brucart 1987: 216 [translations added])
 A: ¿A qué se debe tanto revuelo? (‘What’s all the fuss about?’)
 B: {Ella / pro} le acaba de pedir el divorcio.
 she him finish.3SG of ask.INF the divorce
 ‘(She) has just asked him for a divorce.’

This supports the view that subject pronouns are not *obligatorily* left-dislocated topics.

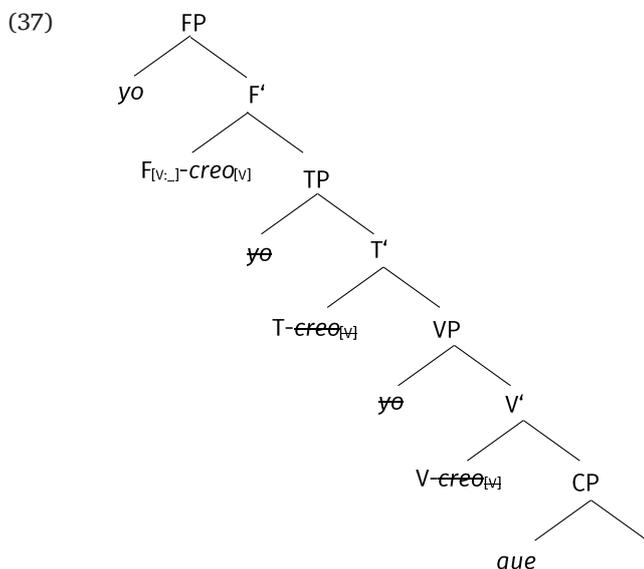
¹⁶ The data of this section only applies to 1SG/2SG pronouns with cognitive verbs. A study of the position of 3rd person referential and non-referential subjects is beyond the scope of this paper.

The analysis outlined up to this point has assumed that subject + cognitive verb combinations are fully transparent. However, it has been argued by Thompson & Mulac (1991) for English that expressions like *I think* do not always function as a main verb plus clausal complement, but as an “epistemic phrase”, i.e. they behave similarly to adverbs (see Aijmer 1997 for “pragmaticalization” and uses as “speech act adverbs”). Posio (2015) also argues that frequent verb forms occur in “prefabricated, formulaic sequences” in Spanish and that sequences like *no sé* ‘(I) don’t know’, *yo creo* ‘I think’, etc. can be seen as “epistemic/evidential markers” (*ibid.*: 74). One piece of evidence for a certain degree of fixation of *yo creo que* comes from its low internal flexibility: in addition to the fact that material only rarely intervenes between preverbal *yo* and the cognitive verb, postverbal *yo* is also almost absent in the data. Only 1 instance of 1SG *creer que* (in imperfect aspect) had the pronoun in postverbal position. Furthermore, *(yo) creo (que)* ‘(I) think (that)’ does not always function as a lexical verb with a prototypical external argument. Apart from the literal meaning of ‘believe (in God)’ and ‘think’ as a mental process, it can have more functional (epistemic/evaluative) meanings (see Aijón Oliva & Serrano 2010) and its subject pronoun is a perspectival marker rather than an agentive subject.

In what follows, an analysis of degrees of fixation and functional uses of *(yo) creo (que)* will be outlined, based on Roberts & Roussou’s (2003) theory and Cruschina’s (2015) analysis of grammaticalization of Italian *dice che* and Sicilian *dicica* (see Cruschina & Remberger 2008 for SAY+Comp in Romance). According to Aijmer (1997: 2), there is a distinction between grammaticalization and pragmaticalization in that elements such as *you know*, *you see*, etc. “involve the speaker’s attitude to the hearer”. According to Diewald (2011: 384), pragmaticalization is a subordinate process of grammaticalization, i.e. it can be considered as “grammaticalization of discourse functions”. Both processes are closely related to frequency and “routinization” (see Detges & Waltereit 2016). As we have discussed, subject + COGNITIVE VERB sequences are frequent in spoken discourse and their use depends on the speaker’s perspective.

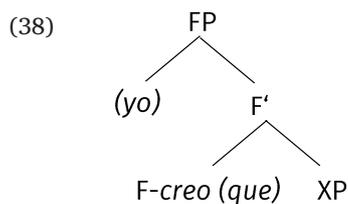
In the generative architecture adopted here, discourse-sensitive features are encoded in FP or CP. The process of pragmaticalization should thus affect these functional projections. We have shown that fixation of *yo + creo (+ que)* is predominantly associated with its evaluative use, which is closely related to contrastive interpretations. It will be argued that the close connection between contrast, evaluative uses, and speaker perspective has the consequence that the subject pronoun and cognitive verb are not always generated within *vP*, but that they can be directly merged in the functional category F encoding perspective.

Roberts & Roussou (2003) assume in the context of modals that (lexical) verbs come to be more functional by means of movement of lexical V to functional T with subsequent loss of the movement option. Thus, modals can be generated by direct merge into a functional category (see Roberts & Roussou 2003: chapter 2). For *creo* ‘(I) think’, let us assume that, in its more functional, epistemic/evaluative meaning, this verb is moved into the category $F_{[epist/eval]}$ triggered by a V-feature (in the sense of Chomsky 1995):



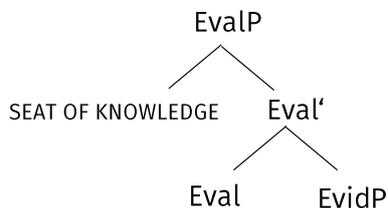
As Roberts & Roussou (2003: 47) note, a system along the lines of Cinque (1999) allows a single lexical item to receive different interpretations according to its syntactic position. In the case of *(yo) creo*, the epistemic or evaluative meaning can thus be derived via movement to the relevant functional head F marked for [epist]/[evid] or [eval].

Roberts & Roussou (2003: 198) argue that certain instances of grammaticalization involve loss of movement and simplification of morphological features: prior to reanalysis, a lexical element bears a feature combination X + Y and after reanalysis, the element itself becomes Y. In (37), *creo* bears [V] by means of its lexical specification and $F_{[epist/eval]}$ by means of movement. We have argued that the subject pronoun in *(yo) creo* does not function as a typical argument within the VP but, rather, it is a perspectival marker. This correlates with a more functional use of *creo* as a verb expressing the personal opinion of the speaker. If this use as an “evaluative phrase” is the outcome of a pragmatization process, it can be implemented by means of loss of movement to the functional category F. This way, *creo (que)* does not function as V bearing $F_{[epist/eval]}$ but it constitutes an $F_{[epist/eval]}$ element itself, encoding speaker perspective:¹⁷



In contrast to (37), where a theta-role is assigned to the subject pronoun within *vP*, *creo* in (38) is a functional element, lacking an external as well as internal argument. However, if *creo*, in its use as an evaluative/epistemic phrase, does not have argument structure, its appearance with an overt pronoun is unexpected. Following Speas & Tenny (2003: 332), the functional category encoding epistemic, evidential and evaluative mood projects a “speaker” or “seat of knowledge” argument in its specifier:

(39) (Speas & Tenny 2003: 332)



If [contrast] associates with the “seat of knowledge”=“speaker”, the result is an overt 1SG pronoun.

There is furthermore evidence from spoken Spanish that *que* ‘that’ does not always function as a subordinating conjunction; *yo creo que* is instead used parenthetically:¹⁸

(40) (PRESEEA-Madrid, H23_033)
 ahí hay una <vacilación/> un defecto yo creo que de previsión /
 there there-is a <hesitation> a flaw I think.1SG that of foresight
 ‘there is a flaw, I think [that], of foresight’

Here, the whole phrase *yo creo que* appears clause-internally, and thus has a certain configurational freedom, similarly to adverbs.

For the grammaticalization of Spanish *dizque* ‘say + comp’, Demonte & Fernández Soriano (2013) propose that *que* ‘that’ as a subordinating conjunction has the features SUB + OP (building on Roussou 2000), i.e. *que* functions as an operator and a subordinating conjunction. With grammaticalized *dizque*, the complementizer loses its SUB feature and *que* moves to the verb contained in Evid. In fact, if *creo* can be directly merged as an F category, the CP projected

¹⁷ For Italian *dice che*, Cruschina (2015) argues that it can directly be generated in a SpeechActPhrase.

¹⁸ Thanks to an anonymous reviewer.

by *que* would appear in the non-prototypical configuration of not being the complement of V. This might consequently lead to *que* losing its SUB feature and incorporating into the F-head.

Admittedly, although there is evidence in our data for a certain degree of fixation of *yo creo que*, more research is needed to determine the exact productivity of each of the two proposed configurations (37) and (38). In fact, an investigation of the degree of pragmaticalization of (yo) + *creo* + COMP would ideally integrate a diachronic perspective, which I leave for future research. However, the use of *yo creo que* ‘I think that’ and *no sé* ‘(I) don’t know’ as *perspectival phrases* is compatible with the approach outlined in this paper: in addition to a derivation in which the pronoun and verb originate in the VP and move to FP, they can directly merge in the functional projection encoding perspective.

6 Some issues for future research

In this section, two issues will be discussed that remain open for future research.

6.1 2SG pronouns

Pronoun realization with 2SG is less frequent than with 1SG in the positive forms of *creer* and *saber* in the study presented here. However, pronouns are realized with positive 2SG *creer* with a high frequency of 54% (Table 4). Because 2SG *creer* mostly appeared when the interviewers were speaking, there are only a small number of data points, which only allow a suggestion of a possible reason for null and overt pronoun use. 2SG *creer* frequently appears in interrogatives. It might be the case that overt pronoun realization is a strategy on the speaker’s side to indicate acceptance of a subjective perspective towards the embedded information. Hence, the use of an overt 2SG pronoun might be an explicit request for an answer to the current Question Under Discussion (in the sense of Roberts 2012), regardless of its evidential basis. Thus, the intersubjective relation between speaker and addressee could be of crucial importance (as Hennemann 2016 argues for *creo yo*). The following example supports the idea that a reasoning along these lines might be interesting to pursue:

- (41) (PRESEEA-Madrid, H13_013)
 y tú crees que eso eeh eso que se cuenta de del / del cambio
 and you think.2SG that this eh this that SE tell of of-the of-the change
 climático ¿tú todo eso cómo lo ves?
 climatic you all that how it see.2SG
 ‘And you think that what they say about climate change, you, all that, what do you think about it?’

Here, the speaker explicitly asks the addressee for her/his personal view compared with other perspectives. Thus, similar mechanisms could be at stake in yielding 1SG and 2SG overt pronouns with *creer* ‘believe/think’.

6.2 The cognitive verb class and ‘bridge verbs’

The cognitive verbs investigated here form part of the class of “bridge verbs”, whose complement displays main clause characteristics (see Emonds 2004; Meinunger 2004; Heycock 2006; Bianchi & Frascarelli 2010). Meinunger (2004) argues that verbs of thinking are among the contexts that license embedded clauses with root-like properties, such as embedded V2 in German and “root transformations” (Hopper & Thompson 1973) in English. Furthermore, Hooper & Thompson (1973) argue that “root transformations” in embedded contexts are possible if the embedded clause introduces an assertion (cf. Meinunger 2004: 217).

Meinunger (2004) observes that certain contexts, including negation, block embedded V2 in German. He furthermore draws a parallel with the trigger of embedded indicatives (with positive *creer*) and subjunctives (with negative *no creer*) in Romance languages:

- (42) a. *German* (Meinunger 2004: 314, citing Reis 1977)
 Ich glaube er hat recht.
 I believe.1SG he has right
 ‘I think, he’s right.’

- b. *German* (Meinunger 2004: 317)
 Ich glaube nicht, *er hat recht / *k^hdass er recht hat.
 I believe.1SG not, *he has right / that he right has
 ‘I don’t believe he’s right.’

- (43) a. (Yo) creo que tiene razón.
 I think.1SG that has.IND right
 b. (Yo) no creo que tenga razón.
 I not think.1SG that has.SBJV right
 ‘I (don’t) think that he’s right.’

We have seen that overt 1SG pronoun frequencies are lower with negative *no creo* than with positive *creo*. If overt pronoun realization is related to assertivity (Aijón Oliva & Serrano 2010), the lower overt pronoun frequencies with negative forms of *creer* could be related to the non-assertive character of the embedded clause, in parallel with the (im-)possibility of “root transformations” and embedded V2 with bridge verbs in other languages. If this line of reasoning is pursued further, the expectation would be that those matrix contexts that have high 1SG pronoun frequencies in Spanish correlate to an extent with those contexts that license embedded clauses with main clause characteristics in other languages. This comparative approach is worthy of further investigation.

7 Conclusions

This paper has investigated speaker/addressee pronouns with cognitive verbs in the Madrid and Alcalá samples of PRESEEA (2014-). First, the influence of the factors of [verb] (*creer* vs. *saber*), [person] (1SG vs. 2SG) and [polarity] (positive vs. negative), which have been investigated in previous studies, have been examined. The results show that *creer* [+CP], in comparison to *saber* [+CP], favors subject expression and that overt pronouns are more frequent with 1SG than with 2SG in the samples examined. With respect to polarity, the tendency towards higher overt pronoun rates with positive than with negative forms of 1SG *creer* and *saber* indicates that the assertivity status of the embedded complement plays a role. The fact that *creer* is more frequently used than *saber* in affirmative contexts might thus play a role in overt pronoun realization, in addition to the lexical semantics of the matrix verb.

The study of complement type has shown that the type of embedded belief plays a role in the realization of 1SG pronouns with matrix cognitive verbs. Concrete, visually falsifiable information (DESCRIPTIONS in the classification proposed here) triggers lower overt 1SG pronoun frequencies with matrix cognitive verbs than beliefs which are expressed with respect to EVALUATIONS and non-visual, abstract information. Thus, (inter-)subjectivity is an important factor (Enríquez 1984; Aijón Oliva & Serrano 2010; Posio 2013; 2014; Hennemann 2016) in that *yo creo que* often encodes a subjective *evaluator* rather than a mere cognizer. I have argued, building on Mayol (2010), that overt 1SG pronoun insertion with *creer* is motivated by “weak contrast”. This leads to the interpretation of the subjective evaluator in relation to a set of potential alternative perspective holders. Subjectivity through overt 1SG pronoun realization might thus also relate to an epistemic/evidential scale: interpreting the subject pronoun in relation to an alternative perspective holder adapts the truth probability (or “quality threshold”; Davis et al. 2007) of the embedded proposition to the speaker’s perspective.

In syntax, overt subjects of cognitive verbs, although highly frequent, are not a special case, but fit into the general mechanism of contrast assignment. It has been proposed that weak contrast is assigned to the Spec of a functional category encoding epistemicity, evidentiality and subjectivity (Cinque 1999; Speas 2004) within the extended IP. This position might be available for preverbal strong pronouns in Spanish because V-to-I movement converts the IP-domain into a phasal pragmatic interface point (in the vein of Gallego’s 2010 *phase sliding*).

The observation that sequences like *yo creo (que)* and *no sé* show degrees of fixation (Posio 2015) can be encoded in this system by assuming that not only does the verb move to the relevant perspectival functional category, but the cognitive verb (together with the complementizer) can be directly merged in it. In this configuration, *creo* is a spell out of the head of the functional

category encoding perspective and the overt 1SG pronoun *yo* is a spell out of a “seat of knowledge” (Speas & Tenny 2003) argument. It might thus be due to the close relation between *contrast* and *perspective* that degrees of “pragmaticalization” and fixation can be observed with *yo creo (que)* in 1SG.

Additional files

The additional files for this article can be found as follows:

- **Appendix A.** Sample annotation of complement type with the cognitive verbs *creer* ‘think/believe’ and *saber* ‘know’ in 1SG. DOI: <https://doi.org/10.16995/glossa.5873.s1>
- **Appendix B.** Table of pairwise comparisons using Fisher’s Exact Test with *creer* and *saber*. DOI: <https://doi.org/10.16995/glossa.5873.s2>
- Dataframes used for the quantitative analysis of subject expression according to (i) verb, person and polarity and (ii) complement type: <http://doi.org/10.5281/zenodo.5035307>

Abbreviations

CP = complementizer phrase, CT = contrastive topic, SAP = speech act participant, SE = impersonal clitic *se*.

Standard abbreviations have been taken from the Leipzig Glossing Rules.

Acknowledgements

I would like to thank three anonymous reviewers for their critical and helpful comments, which have led to substantial improvement of this paper. Parts of this research were presented at *Going Romance 2018* (University of Utrecht), the *II Workshop on Spoken Corpus Linguistics* (Universitat d’Alacant), *LKR* (Universität zu Köln) and *CoRoLi* (Humboldt-Universität zu Berlin). Thanks to the participants for helpful comments, especially Anne Wolfsgruber. Parts of this research were carried out at the University of Cologne. I would like to thank Amalia Canes Nápoles, Alessia Cassarà, and Eric Engel for many interesting and helpful discussions and Maximilian Hörl for his help with the program R. All potential errors are fully my own. Parts of the research for this paper have been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB “Prominence in Language” in the project C01 “Prominence and information structure” at the University of Cologne.

Competing interests

The author has no competing interests to declare.

Author affiliation

Peter Herbeck  orcid.org/0000-0003-4763-3959
University of Vienna, AT

References

- Adli, Aria. 2019. Topic chains in dialogue. *Journal of Pragmatics* 154 (Special Issue on *Prominence in discourse*, Klaus von Heusinger & Petra Schumacher (eds.)), 39–62. DOI: <https://doi.org/10.1016/j.pragma.2019.07.022>
- Aijmer, Karin. 1997. *I think*: An English modal particle. In Toril Swan & Olaf J. Westvik (eds.), *Modality in Germanic languages: Historical and comparative perspectives*, 1–47. Berlin: de Gruyter. DOI: <https://doi.org/10.1515/9783110889932.1>
- Aijón Oliva, Miguel Ángel & María José Serrano. 2010. El hablante en su discurso: Expresión y omisión del sujeto de *creo*. *Oralia* 13. 7–38.
- Aikhenvald, Alexandra Y. 2018. Evidentiality: The framework. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, 1–43. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780198759515.013.1>

- Akatsuka, Noriko. 1985. Conditionals and the Epistemic Scale. *Language* 61(3). 625–639. DOI: <https://doi.org/10.2307/414388>
- Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parameterizing AGR: Word order, V-movement, and EPP-checking. *Natural Language and Linguistic Theory* 16. 491–539. DOI: <https://doi.org/10.1023/A:1006090432389>
- Barbosa, Pilar. 2009. Two kinds of subject *pro*. *Studia Linguistica* 63(1). 2–58. DOI: <https://doi.org/10.1111/j.1467-9582.2008.01153.x>
- Bentivoglio, Paola. 1983. Topic continuity in discourse: A study of spoken Latin-American Spanish. In Talmy Givón (ed.), *Topic continuity in discourse*, 255–312. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.3.06ben>
- Bianchi, Valentina & Mara Frascarelli. 2010. Is Topic a root phenomenon? *Iberia: An International Journal of Theoretical Linguistics* 2(1). 43–88.
- Boye, Kasper. 2012. *Epistemic meaning: A cross-linguistic and functional-cognitive study*. Berlin: Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110219036>
- Brucart, José M. 1987. *La elisión sintáctica en español*. Bellaterra: Publicacions de la Universitat Autònoma de Barcelona.
- Büring, Daniel. 2003. On D-trees, beans, and B-accent. *Linguistics and Philosophy* 26. 511–545. DOI: <https://doi.org/10.1023/A:1025887707652>
- Chomsky, Noam. 1981. *Lectures on Government and Binding – the Pisa lectures*. Dordrecht: Foris Publications.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & María Luisa Zubizarreta (eds.), *Foundational issues in linguistic theory*, 133–166. Cambridge: MIT Press.
- Cinque, Guglielmo. 1999. *Adverbs and functional heads: a cross-linguistic perspective*. Oxford: Oxford University Press.
- Cornillie, Bert, Juana I. Marín Arrese & Björn Wiemer. 2015. Evidentiality and the semantics-pragmatics interface: An introduction. *Belgian Journal of Linguistics* 29(1). 1–17. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/bjll.29.001int>
- Cruschina, Silvio. 2015. The expression of evidentiality and epistemicity: Cases of grammaticalization in Italian and Sicilian. *Probus* 27. 1–31. DOI: <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.
- Davidson, Brad. 1996. ‘Pragmatic weight’ and Spanish subject pronouns: The pragmatic and discourse uses of ‘tú’ and ‘yo’ in spoken Madrid Spanish. *Journal of Pragmatics* 26. 543–565. DOI: [https://doi.org/10.1016/0378-2166\(95\)00063-1](https://doi.org/10.1016/0378-2166(95)00063-1)
- Davis, Christopher, Christopher Potts & Margaret Speas. 2007. The pragmatic values of evidential sentences. In: T. Friedman & M. Gibson (eds.), *Proceedings of SALT XVII*, 71–88. Ithaca, NY: Cornell University. DOI: <https://doi.org/10.3765/salt.v17i0.2966>
- De Haan, Ferdinand. 2001. The relation between modality and evidentiality. In Reimar Müller & Marga Reis (eds.), *Modalverben und Modalität* (Linguistische Berichte, Sonderhefte 9), 201–216. Hamburg: Helmut Buske Verlag.
- De Haan, Ferdinand. 2005. Encoding speaker perspective: Evidentials. In Zygmunt Frajzyngier, Adam Hodges & David S. Rood (eds.), *Linguistic diversity and language theories*, 379–397. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.72.18haa>
- DeLancey, Scott C. 1981. An interpretation of split ergativity and related patterns. *Language* 57(3). 626–657. DOI: <https://doi.org/10.2307/414343>
- 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. 211–234.
- De Saeger, Bram. 2008. Speaker involvement through cognition verbs in Spanish. In Philippe De Brabanter & Patrick Dendale (eds.), *Commitment – Belgian Journal of Linguistics* 22. 63–81. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/bjll.22.04sae>
- Detges, Ulrich & Richard Waltereit. 2016. Grammaticalization and pragmaticalization. In Susann Fischer & Christoph Gabriel (eds.), *Manual of grammatical interfaces in Romance*, 635–658. Berlin: de Gruyter. DOI: <https://doi.org/10.1515/9783110311860-024>
- Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2). 365–390. DOI: <https://doi.org/10.1515/ling.2011.011>
- Emonds, Joseph. 2004. Unspecified categories as the key to root constructions. In David Adger, Cécile De Cat & George Tsoulas (eds.), *Peripheries*, 75–120. Dordrecht: Springer. DOI: https://doi.org/10.1007/1-4020-1910-6_4
- Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Instituto Miguel de Cervantes.

- Erker, Daniel & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3). 526–557. DOI: <https://doi.org/10.1353/lan.2012.0050>
- Fernández Leborans, María Jesús. 1992. La oración del tipo: “es que...”. *Verba* 19. 223–239.
- Fernández Soriano, Olga. 1999. El pronombre personal. Formas y distribuciones. Pronombres átonos y tónicos. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española* 1. 1209–1273. Madrid: Espasa Calpe.
- Frascarelli, Mara. 2007. Subjects, topics and the interpretation of referential *pro*: An interface approach to the linking of (null) pronouns. *Natural Language and Linguistic Theory* 25. 691–734. DOI: <https://doi.org/10.1007/s11049-007-9025-x>
- Frascarelli, Mara. 2018. The interpretation of *pro* in consistent and partial null subject languages: A comparative interface analysis. In Frederica Cognola & Jan Casalicchio (eds.), *Null subjects in generative grammar: A synchronic and diachronic perspective*, 211–239. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oso/9780198815853.003.0009>
- 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. DOI: <https://doi.org/10.1075/la.100.07fra>
- Gallego, Ángel. 2010. *Phase theory*. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/la.152>
- Givón, Talmy. 1982. Evidentiality and epistemic space. *Studies in Language* 6(1). 23–49. DOI: <https://doi.org/10.1075/sl.6.1.03giv>
- Givón, Talmy. 1983. Topic continuity in discourse: An introduction. In Talmy Givón (ed.), *Topic continuity in discourse*, 1–42. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.3.01giv>
- Grajales Alzate, Róbinson. 2016. Los verbos de actitud proposicional como estrategias evidenciales en el español de Medellín. *Lingüística y Literatura* 69. 339–361. DOI: <https://doi.org/10.17533/udea.lyl.n69a15>
- Gries, Stefan Th. 2013. *Statistics for linguistics with R*. Berlin: de Gruyter. DOI: <https://doi.org/10.1515/9783110307474>
- Hara, Yurie & Robert van Rooij. 2007. Contrastive topics revisited – A simpler set of topic alternatives. Talk given at NELS 38, University of Ottawa.
- Harley, Heidi & Elizabeth Rittner. 2002. Person and number in pronouns: A feature-geometric analysis. *Language* 78(3). 482–526. DOI: <https://doi.org/10.1353/lan.2002.0158>
- Hennemann, Anja. 2012. The epistemic and evidential use of Spanish modal adverbs and verbs of cognitive attitude. *Folia Linguistica* 46(1). 133–170. DOI: <https://doi.org/10.1515/flin.2012.5>
- Hennemann, Anja. 2013. *A context-sensitive and functional approach to evidentiality in Spanish or why evidentiality needs a superordinate category*. Frankfurt a.M.: Peter Lang. DOI: <https://doi.org/10.3726/978-3-653-02066-3>
- Hennemann, Anja. 2016. A cognitive-constructionist approach to Spanish *creo* \emptyset and *creo yo* ‘[I] think’. *Folia Linguistica* 50(2). 449–474. DOI: <https://doi.org/10.1515/flin-2016-0017>
- Hervé, Maxime. 2021. RVAideMemoire: Testing and plotting procedures for biostatistics. R package version 0.9-79.
- Heycock, Caroline. 2006. Embedded root phenomena. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, 174–209. Oxford: Blackwell Publishing. DOI: <https://doi.org/10.1002/9780470996591.ch23>
- Hooper, Joan B. & Sandra A. Thompson. 1973. On the applicability of root transformations. *Linguistic Inquiry* 4(4). 465–497.
- Krifka, Manfred. 2007. Basic notions of information structure. In Caroline Féry, Gisbert Fanselow & Manfred Krifka (eds.), *Interdisciplinary studies on information structure* 6. 13–55. Potsdam: Universitätsverlag Potsdam.
- Krifka, Manfred. 2014. Embedding illocutionary acts. In Thomas Roeper & Margaret Speas (eds.), *Recursion: Complexity in cognition*, 59–87. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-05086-7_4
- Landau, Idan. 2000. *Elements of control*. Dordrecht: Kluwer. DOI: <https://doi.org/10.1007/978-94-011-3943-4>
- Levinson, Stephen C. 1987. Pragmatics and the grammar of anaphora. *Journal of Linguistics* 23. 379–434. DOI: <https://doi.org/10.1017/S0022226700011324>
- Levshina, Natalia. 2015. *How to do linguistics with R*. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/z.195>
- Lewis, David. 1976. Probabilities of conditionals and conditional probabilities. *The Philosophical Review* 85(3). 297–315. DOI: <https://doi.org/10.2307/2184045>
- López, Luis. 2009. *A derivational syntax for information structure*. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199557400.001.0001>
- Luján, Marta. 1999. Expresión y omisión del pronombre personal. In Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española*, 1275–1315. Madrid: Espasa.

- Mayol, Laia. 2010. Contrastive pronouns in null-subject Romance languages. *Lingua* 120(10). 2497–2514. DOI: <https://doi.org/10.1016/j.lingua.2010.04.009>
- Meinunger, André. 2004. Verb position, verbal mood, and the anchoring of (potential) sentences. In Horst Lohnstein & Susanne Trissler (eds.), *The syntax and semantics of the left periphery*, 313–342. Berlin: de Gruyter. DOI: <https://doi.org/10.1515/9783110912111.313>
- Meyer, David, Achim Zeileis & Kurt Hornik. 2020. vcd: Visualizing categorical data. R package version 1.4-8.
- Morales, Amparo. 1997. La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico. *Hispania* 80(1). 153–165. DOI: <https://doi.org/10.2307/345995>
- Moreno-Fernández, Francisco. 2005. Corpus para el estudio del español en su variación geográfica y social. El corpus “PRESEEA”. *Oralia* 8. 123–140.
- Nuyts, Jan. 2001. Subjectivity as an evidential dimension in epistemic modal expression. *Journal of Pragmatics* 33(3). 383–400. DOI: [https://doi.org/10.1016/S0378-2166\(00\)00009-6](https://doi.org/10.1016/S0378-2166(00)00009-6)
- Ordóñez, Francisco & Esthela Treviño. 1999. Left dislocated subjects and the pro-drop parameter: A case study of Spanish. *Lingua* 107. 39–68. DOI: [https://doi.org/10.1016/S0024-3841\(98\)00020-5](https://doi.org/10.1016/S0024-3841(98)00020-5)
- Orozco, Rafael. 2015. Pronominal variation in Colombian Costeño Spanish. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish. A cross-dialectal perspective*, 17–58. Washington: Georgetown University Press.
- Palmer, Frank R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- Posio, Pekka. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3). 777–798. DOI: <https://doi.org/10.1016/j.pragma.2010.10.012>
- Posio, Pekka. 2013. The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1). 253–291. DOI: <https://doi.org/10.1515/flin.2013.010>
- Posio, Pekka. 2014. Subject expression in grammaticalizing constructions: The case of *creo* and *acho* ‘I think’ in Spanish and Portuguese. *Journal of Pragmatics* 63. 5–18. DOI: <https://doi.org/10.1016/j.pragma.2013.07.001>
- Posio, Pekka. 2015. Subject pronoun usage in formulaic sequences: Evidence from Peninsular Spanish. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 59–78. Washington: Georgetown University Press.
- Posio, Pekka. 2018. Properties of pronominal subjects. In Kimberly L. Geeslin (ed.), *The Cambridge handbook of Spanish linguistics*, 286–306. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/9781316779194.014>
- 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: April and June 2019].
- R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [<https://www.R-project.org/>].
- Reis, Marga. 1977. *Präsupposition und Syntax*. Tübingen: Niemeyer. DOI: <https://doi.org/10.1515/9783111344843>
- Rigau, Gemma. 1989. Connexity established by emphatic pronouns. In Maria-Elisabeth Conte, János Sándor Petöfi & Emel Sözer (eds.), *Text and discourse connectedness*, 191–205. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/slcs.16.17rig>
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro*. *Linguistic Inquiry* 17(3). 501–557.
- Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics* 5. 1–69. DOI: <https://doi.org/10.3765/sp.5.6>
- Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511486326>
- Rooryck, Johan. 2001. Evidentiality, Part I. *Glott International* 5(4). 125–133.
- Rooryck, Johan. 2019. ‘Recycling’ evidentiality: A research program. In Metin Bağrıaçık, Anne Breitbarth & Karen De Clercq (eds.), *Mapping linguistic data: Essays in honour of Liliane Haegeman*, 242–261. Ghent. [<http://hdl.handle.net/1854/LU-8625919>].
- Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. DOI: <https://doi.org/10.1007/BF02342617>
- Roussou, Anna. 2000. On the left periphery: Modal particles and complementisers. *Journal of Greek Linguistics* 1(1). 65–94. DOI: <https://doi.org/10.1075/jgl.1.05rou>
- Sigurðsson, Halldór Á. 2011. Conditions on argument drop. *Linguistic Inquiry* 42(2). 267–304. DOI: https://doi.org/10.1162/LING_a_00042
- Simons, Mandy. 2007. Observations on embedding verbs, evidentiality, and presupposition. *Lingua* 117(6). 1034–1056. DOI: <https://doi.org/10.1016/j.lingua.2006.05.006>
- Speas, Margaret. 2004. Evidentiality, logophoricity and the syntactic representation of pragmatic features. *Lingua* 114(3). 255–276. DOI: [https://doi.org/10.1016/S0024-3841\(03\)00030-5](https://doi.org/10.1016/S0024-3841(03)00030-5)

- Speas, Margaret. 2018. Evidentiality and formal semantic theories. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, 286–315. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780198759515.013.15>
- 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. DOI: <https://doi.org/10.1075/la.57.15spe>
- Squartini, Mario. 2001. The internal structure of evidentiality in Romance. *Studies in Language* 25(2). 297–334. DOI: <https://doi.org/10.1075/sl.25.2.05squ>
- 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. DOI: <https://doi.org/10.1075/cilt.238.25sun>
- Thompson, Sandra A. & Anthony Mulac. 1991. A quantitative perspective on the grammaticalization of epistemic parentheticals in English. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to grammaticalization* 2. 313–339. Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/tsl.19.2.16tho>
- Traugott, Elizabeth C. 2010. (Inter)subjectivity and (inter)subjectification: A reassessment. In Kristin Davidsen, Lieven Vandelanotte & Hubert Cuyckens (eds.), *Subjectification, intersubjectification and grammaticalization*, 29–74. Berlin: de Gruyter. DOI: <https://doi.org/10.1515/9783110226102.1.29>
- 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. DOI: <https://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.
- Van Valin, Robert D. 2001. *An introduction to syntax*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139164320>
- Wiemer, Björn. 2018. Evidentials and epistemic modality. In Alexandra Y. Aikhenvald (ed.), *The Oxford handbook of evidentiality*, 85–108. Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/oxfordhb/9780198759515.013.4>
- Willett, Thomas. 1988. A cross-linguistic survey of the grammaticization of evidentiality. *Studies in Language* 12(1). 51–97. DOI: <https://doi.org/10.1075/sl.12.1.04wil>
- Zagona, Karen. 2002. *The syntax of Spanish*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511613234>
- Zubizarreta, María Luisa. 1998. *Prosody, focus, and word order*. Cambridge: MIT Press.

TO CITE THIS ARTICLE:

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): 122. 1–34. DOI: <https://doi.org/10.16995/glossa.5873>

Submitted: 16 December 2019

Accepted: 05 July 2021

Published: 11 October 2021

COPYRIGHT:

© 2021 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/>.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by Open Library of Humanities.