B O R E A L I S
AN INTERNATIONAL JOURNAL OF HISPANIC LINGUISTICS

SPEAKER SUBJECT EXPRESSION WITH VERBS OF COGNITION – *THINK/BELIEVE* IN ITALIAN AND SPANISH[1]

Peter Herbeck, Andrea Miglietta & Eva-Maria Remberger
University of Vienna

ABSTRACT. In this paper, we study 1st person subject pronoun realization with verbs of cognition in spoken Spanish and Italian corpus data. Previous studies have shown particularly frequent overt subjects with 1st person singular *(yo) creo* '(I) think/believe' in Spanish. However, there have been no in-depth studies of the corresponding Italian structure. We thus examine whether the same patterns can be observed in these two consistent pro-drop Romance languages. The results show that Italian *credo* and *penso* '(I) think' show the opposite pattern to the respective Spanish verb forms. While the Spanish verbs of cognition favor subject expression, Italian predominantly favors subject omission. We link this difference to diverging patterns of pragmaticalization: While in Spanish *creo* with a null and an overt subject pronoun has developed into two semi-fixed expressions with a partial functional specialization, in Italian, pragmaticalization has only targeted *credo* and *penso* with a null subject to express epistemic stance of the speaker. As evidence we focus, among other things, on subject realization with parentheticals. The results show that the pro-drop property can be exploited for pragmaticalization processes in different ways cross-linguistically.

Keywords. first person singular, pragmaticalization, subject realization, verbs of cognition, parentheticals, Spanish, Italian

RESUMEN. En este artículo, estudiamos la expresión del pronombre personal en primera persona con verbos cognitivos en datos de corpus orales del español y del italiano. Estudios previos han demostrado que los sujetos explícitos de primera persona del singular *(yo) creo* son especialmente frecuentes en español. Sin embargo, todavía no se han realizado estudios parecidos sobre la estructura correspondiente en italiano. Por lo tanto, examinamos si se pueden observar los mismos patrones de omisión en estas dos lenguas románicas *pro-drop*. Los resultados muestran que las formas italianas *credo* y *penso* '(yo) creo' siguen el patrón opuesto al de las respectivas formas verbales españolas. Mientras que los verbos cognitivos españoles favorecen la expresión del sujeto, el italiano favorece predominantemente la omisión del sujeto. Relacionamos esta diferencia con patrones divergentes de pragmatización: Mientras que en español *creo* con sujeto nulo y con pronombre explícito se ha convertido en dos expresiones semifijas con una especialización funcional parcial, el proceso de pragmátización en italiano solo ha afectado *credo* y *penso* con sujeto nulo. En nuestro estudio nos centramos, entre otras cosas, en la expresión del sujeto en posición parentética para aportar evidencia adicional para los diferentes patrones de pragmatización con verbos cognitivos en italiano y español. Los resultados muestran que la propiedad de *pro-drop* puede explotarse para los procesos de pragmátización de diferentes maneras en las distintas lenguas.

Palabras clave. primera persona del singular, pragmatización, expresión del sujeto, verbos cognitivos, paréntesis, español, italiano

[1] Acknowledgements: This research was funded by the Austrian Science Fund (FWF) 10.55776/F1003 *Language between Redundancy and Deficiency* (subproject P08). We also would like to thank Sara Suess and Ignasi Planas Villalba for their help with the annotation of the corpus data.
Abbreviations: TAM = Tense, Aspect, Mood; SBJ = subjunctive; IND = indicative.

1. Introduction

The factors governing null vs. overt subjects in null subject languages have been among the most deeply studied topics in linguistic research. In Romance pro-drop languages, the richness of subject-verb agreement morphology has been posited as a possible driving force of null subjects (Rizzi 1982; Speas 1994; Barbosa 1995; among others). Furthermore, various morpho-syntactic, semantic, and discourse-pragmatic factors governing the use of null and overt subject pronouns have been identified in those Romance languages that allow for this alternation. Factors such as *person*/*number*, TAM, *verb type* and *verb semantics*, *emphasis*, *contrast*, *topicality*, *continuity*, and *turn-taking* have been mentioned among the most prominent factors (see Posio 2018 for an overview).

With respect to verb type and semantics, it has frequently been observed in the study of Spanish that verbs of cognition or epistemic verbs,[2] such as *creer* 'to believe/think' and *pensar* 'to think', express the overt 1st person singular subject at an exceptionally high rate (see Enríquez 1984; Morales 1997; Aijón Oliva & Serrano 2010; Posio 2011; Travis & Torres Cacoullos 2012; among many others). In order to account for this tendency, it has been argued that epistemic verbs and the 1st person singular have an inherent subjective tendency and that subjectivity[3] in turn favors 1st person singular overt subjects (Enríquez 1984; Posio 2011). Furthermore, it has been observed that the rate of overt subjects with these verb forms is especially high when they are used in argumentative or opinion contexts (Aijón Oliva & Serrano 2010; Vázquez Rozas & Enríquez Ovando 2020) compared to purely epistemic uses to mitigate speaker commitment (see §2.1 for further discussion). Pursuing this line of reasoning further, Herbeck (2021) argues that high subject expression results from the tendency of 1st person singular epistemic verbs to evoke contrastive contexts between perspective holders. Lastly, it has been postulated that a high rate of overt subjects, especially with Spanish *(yo) creo* '(I) think' is related to a formulaic or routinized use of this sequence (see e.g. Posio 2011, 2013, 2014).

However, while factors such as subjectivity and contrast can account for certain tendencies in the use of subject pronouns with verbs of cognition in Spanish, they cannot account for everything that we observe cross-linguistically. Thus, Posio (2013, 2014) examines 1st person singular overt and null subjects in European Portuguese vs. European Spanish and argues that the former behaves differently in that *acho* 'I think' does not have a high rate of overt subjects. At the same time, the verb of saying *digo* '(I) say' has high frequency of overt subjects in European Portuguese, while the Spanish correlate favors null subjects. Herbeck (2022) analyzes a corpus sample of spoken Valencian Catalan and also observes that 1st person overt subject pronouns are not particularly frequent with verbs of cognition (such as *crec* '(I) think'), but they are with

[2] Of, course, it is debatable whether the group of epistemic verbs and verbs of cognition coincide. Here, we start from the assumption that verbs like THINK and BELIEVE are, in principle, verbs of cognition, including epistemic verbs, and, depending on some of their interpretations (e.g. THINK as a conscious and agentive act of reflection or BELIEVE as a conviction) possibly belong to a sub-group of more specialized verbs (cf. also ex. (1) vs. (2)); another notion used for this group of verbs is "mental state" verbs (see e.g. Nuyts 2001) or verbs of mental activity (cf. e.g. Enríquez 1984).

[3] *Subjectivity* is a widely debated, and also somehow ambiguous, term: in modal semantics, for example, we can distinguish between *objective* and *subjective modal epistemicity* (depending on the modal base and the ordering source; see Kratzer 1981, 1991), but the notion is also used in grammaticalization theory (*subjectivization*, as in e.g. Traugott 1995) with a meaning independent of modality. Furthermore, in the case of 1st person verbal forms of verbs like BELIEVE, *subjectivity* also refers to the *privileged access* that the speaker has to her or his mental state.

verbs of saying in 1st person singular. This raises the question of which factors account for patterns of overt vs. null subjects with high frequency verb forms cross-linguistically. This is particularly interesting within the Romance pro-drop languages Catalan, European Portuguese, Italian and Spanish, given that they are all 'consistent' pro-drop languages (in the sense of Holmberg 2005; Holmberg et al. 2009). According to the undisputed status of Spanish and Italian as a null subject languages, we *a priori* do not expect differences with respect to subject expression frequencies.

One approach is to propose that (null) subject pronoun+verb sequences can yield different patterns of pragmaticalization (in the sense of Aijmer 1997; Diewald 2011; Detges & Waltereit 2016). In pro-drop languages, unlike in English, at least three targets of a pragmaticalization process are potentially possible: pronoun+verb, verb+pronoun or Ø+verb. Different pro-drop languages might thus show different patterns of pragmaticalization of these three configurations to encode functions related to speaker-hearer interactions. One way of testing this, we argue, is to examine subject realization with a lexical verb of cognition in a main clause and compare it to its use as a parenthetical. It is well-known that verbs of cognition in Romance, as in other languages, are often used as parentheticals (see Schneider 2007), i.e. in syntactic positions not directly contributing to the propositional content of the host clause over (the parts of) which they take scope. However, the issue of the use of null and overt subject pronouns in pro-drop languages in this configuration deserves further research.

In this paper, we offer a study of verbs of cognition in spoken Italian and Spanish – in particular, the 1st person singular verb forms of Sp. *creer*/*pensar* and It. *credere*/*pensare*. Alongside the role of *polarity* (affirmative vs. negated), (referential) *continuity*, *clause type* (main, embedded, parenthetical), and *complement type* (e.g., embedded CP-clauses vs. (pro-)nominal complements), we pay particular attention to the notion of *subjectivity* to examine subject expression, drawing on data from two spoken corpora of Italian (KIParla) and Spanish (COSER).

We will show that a comparison of the verb forms Sp. *creo*/*pienso* and It. *credo*/*penso* reveals striking differences: the rate of overt subjects is high with Spanish *creo* and *pienso* (above 59%) while it is low with Italian *credo* and *penso* (below 15%). Furthermore, we demonstrate that parenthetical uses of verbs of cognition – in particular, *creo* – have the highest rate of overt subjects in Spanish (above 70%) while null subjects are the widely preferred option with parentheticals in Italian. This leads us to the conclusion that Ø+*creo*, pronoun+verb and verb+pronoun (i.e. *creo*, *yo creo* and *creo yo*) have all undergone a process of pragmaticalization in Spanish, being used either to mitigate speaker commitment or to shift from an objective to a subjective perspective; in Italian, on the contrary, pronoun+verb shows no signs of fixation. In the latter, it is Ø+*credo* as a parenthetical that is used as a mitigating device to express epistemic stance. We thus conclude that the *pro*-drop property and the co-existence of null and overt subject pronouns can be exploited by different, closely related languages, to encode distinct concepts at the syntax-pragmatics interface, which go beyond the traditionally postulated factors of topic shift and contrast.[4]

Finally, a note on the processes of *grammaticalization* and *pragmaticalization*, notions that will both be used in this paper: we hold that both processes cannot be

[4] In syntax, we would argue that it is not only the relation between the TP- and CP-area that determines subject expression with 1st person singular verbs of cognition, but also the relation between the TP-area and perspectival projections above CP, such as Speas & Tenny's (2003) Speech Act Phrase, Wiltschko's (2021) GroundP, or the layers Miyagawa (2022) and Krifka (2023) proposed for assertive clauses. This relation can be formally expressed by agreement processes between several functional categories. However, a thorough syntactic-pragmatic analysis of this type is beyond the scope of this paper.

clearly distinguished but are intertwined, as are language competence and language use, contrary to their distinction that is commonly made in the literature, as in Detges & Waltereit (2016) based on Lehmann (2002 [1982]); see also Diewald (2011) and Ocampo (2006), who introduces the term of *discoursivization*. Since the "syntactiziation of discourse" (Haegemann & Hill 2013) started, the initially clear distinction between elements becoming more grammatical vs. elements becoming more pragmatic in their use has become blurred. Indeed, with regard to the phenomenon studied in our paper, both types are involved: clear elements of grammaticalization (such as semantic bleaching, morphological fixation, and perhaps eventually also phonological reduction) as well as elements of pragmaticalization (scope extension, greater syntactic mobility/independence, optionality etc.).

This paper is structured as follows: First, we discuss previous studies on overt vs. null subject pronouns, paying particular attention to those studies dealing with 1st person singular verbs of cognition (§2). We then present the data, methodology, and results of our corpus study of Italian and Spanish (§3). In Section 4, we discuss the results in the light of cross-linguistic differences, fixation patterns and functions of (null and overt) subject pronouns in combination with 1st person singular verbs of cognition. In Section 5, we present some ideas for a syntactic analysis, and Section 6 concludes the paper.

2. Background: Subject pronoun realization with verbs of cognition

2.1 Null and overt subject alternations in Spanish

There is an extensive body of research regarding the realization of 1st person singular subjects in spoken (varieties of) Spanish (Enríquez 1984; Davidson 1996; Morales 1997; Posio 2011; Travis & Torres Cacoullos 2012; De Prada Pérez 2015; Vázquez Rozas & Enríquez Ovando 2020; Bessett 2023 and many more; see also Posio 2018 and references therein). In fact, given the functional difference between deictic 1st and 2nd person, on the one hand, and potentially discourse anaphoric 3rd person, on the other, several studies have focused on only one grammatical person with very fruitful results (for Spanish see e.g. Travis & Torres Cacoullos 2012; Lapidus Shin 2014; Vázquez Rozas & Enríquez Ovando 2020). The majority of the investigations have been carried out using spoken corpus data and are variationist (e.g. Erker & Guy 2012; De Prada Pérez 2015) and/or functionalist in nature. The results show a generally high rate of overt subjects with verbs of cognition in Spanish, especially with *creer* 'to think/believe'. In this section, rather than providing a full review of the available studies and approaches, we highlight the reasons that have been proposed for this tendency.[5]

In Enríquez's (1984) seminal study of null vs. overt subject pronouns in spoken Spanish in Madrid, she observes that verbs of 'mental activity' have a high frequency of overt 1st person pronominal subjects (55%), compared to verbs of external activity (25.5%); see Enríquez (1984: 362, Table 14). Among several other observations, Enríquez (1984) also notes that affirmative predicates more frequently have overt subjects in the 1st, but not in 2nd, person (see her Table 17, p. 364), and she argues (1984: 118) that subjectivity and contrast play a crucial role for overt 1st person subjects.

Posio (2011: 795) studies 1st and 2nd person singular verb forms and their subject realization rates in Peninsular Spanish. He observes that 1st person singular has generally higher overt subject pronoun frequencies than 2nd person singular, attributing

[5] We refer the reader to the recent edited volumes by Carvalho et al. (2015) and Posio & Herbeck (2023), and to overview articles such as Posio (2018) for further discussion and references.

this to the "egocentric nature of discourse". Travis & Torres Cacoullos (2012) investigate the frequency of expression of *yo* with verbs of cognition in Colombian Spanish, also observing a high rate of overt subjects (67% with verbs of cognition vs. 44% with other verb types; see Travis & Torres Cacoullos 2012: 726, Table 2). The authors argue against an account that draws a simple correlation between overt subject pronouns and 'contrast' and assume that the sequence *yo* plus cognitive verb belongs to a schematic "(subject) + cognitive verb" construction. This would imply that 1st person subject pronoun + verb of cognition to some degree constitutes a fixed expression or a "formulaic sequence" (see Posio 2011: 785). As Posio (2018: 294) argues, differences in overt pronoun frequencies with different verb types might be due to a boosting effect of either overt or null subjects with high frequency verb forms expressing speaker stance. Thus, the high frequency verb form *creo* favors overt subjects in 1st person singular, boosting the rate of overt pronominal subjects in the category of verbs of cognition and in 1st person singular.[6]

However, alongside the differences between verbs of cognition and other verb types, there are also differences between types of verbs of cognition. Posio (2015) reports a low rate of overt subject pronouns with the sequence *no sé* '(I) don't know' but a high rate with *(yo) creo* '(I) think/believe'. Furthermore, in Herbeck's (2021) study of 1st person singular subjects with *creo* and *sé*, significant differences between these two verb forms with respect to pronoun frequencies were observed even if only non-parenthetical uses were taken into account. This study only examined verbs with a fully realized CP complement and pronoun frequencies still differed significantly (with high overt pronoun rates with *(yo) creo* + CP and low overt pronoun rates with *(yo) (no) sé* + CP).

A further complicating factor is that the same verb form can show different context-dependent meanings (see De Saeger 2008; Aijón Oliva & Serrano 2010; Vázquez Rozas & Enríquez Ovando 2020; Herbeck 2021 for discussion): De Saeger (2008: 65) distinguishes between (speaker-oriented) "modal uses" and (subject-oriented) "opinion markers" and Aijón Oliva & Serrano (2010) between epistemic and argumentative uses of 1st person singular *creo* (see Vázquez Rozas & Enríquez Ovando 2020; see also Schneider 2007 for discussion in the context of parenthetical verbs).

The following two examples demonstrate the two different uses of *creo* in the data from our corpus study:

(1) Epistemic use:
Y entonces, en nuestros tiempos se pagaba tre-, trece, trece, trece monedas. Creo que es trece monedas. (COSER-2005_01; Guipúzcoa)
'And before, in our times, we paid thirteen coins. I think that it's thirteen coins.'

(2) Doxastic (opinion) or argumentative use:
Y yo creo que una iglesia debería sé al revés, ¿no? A mi punto de vista, eh. (COSER-2106_01; Huelva)
'And I think that a church should be the reverse, right? From my point of view.'

Given that the classification of these context-dependent notions in corpus data is challenging, Vázquez Rozas & Enríquez Ovando (2020) and Herbeck (2021) furthermore propose systematic criteria for distinguishing between the two uses of *(yo)*

[6] As an anonymous reviewer correctly points out, frequency cannot be the only factor. As argued throughout the paper, the interaction of verb semantics, discourse function and frequency is crucial in understanding this phenomenon.

creo in spoken data (see §3). Vázquez Rozas & Enríquez Ovando (2020: 205f), building on Schneider (2007) among others, point out that uses of *creo* with an epistemic modality meaning might have a higher degree of grammaticalization (related to the factor of desemanticization) and that parenthetical uses of *creo* more easily allow the epistemic meaning even though, as the authors acknowledge (2020: 226), the doxastic meaning is not excluded in parenthetical position (see also Aijón Oliva & Serrano 2010). It is interesting to note in this context that the two different meanings of *creo* seem to correlate with different patterns of subject expression (Aijón Oliva & Serrano 2010; Vázquez Rozas & Enríquez Ovando 2020): the epistemic meaning does not trigger overt subjects as frequently as the opinion (i.e. doxastic) meaning. Herbeck (2021) relates this observation to an interplay between epistemicity, subjectivity and contrast: the expression of an opinion and of beliefs with a high degree of subjectivity favor contrasting perspectives, which has an impact on subject realization in 1st person singular.

Cross-linguistic differences pose a further challenge (see Posio 2018 for discussion and references). In Peninsular Spanish, *yo* is frequently expressed with *creer* 'to think/believe' but not with *saber* 'to know', as we have discussed above. However, Orozco (2015) observes a high rate of overt subjects with 1st person singular verb forms of *saber* 'know' in Colombian Spanish. Posio (2013, 2014) compares European Spanish with European Portuguese, and finds opposing trends in the rates of overt 1st person subjects with the verbs THINK (i.e. Sp. *creer* and Pt. *achar*) and SAY (i.e. Sp. *decir* and Pt. *dizer*): In Spanish, the rate of 1st person overt subject pronouns is high with *creo* but in European Portuguese, the rate of overt subject pronouns with *acho* is average. On the contrary, 1st person *yo* is not frequent with *digo* in Spanish, but the European Portuguese equivalent triggers a high rate of overt subject pronouns. Furthermore, Herbeck (2022) examines spoken Valencian Catalan data and observes in a similar vein that *dic* '(I) say' frequently cooccurs with an overt subject pronoun, but *crec* '(I) think' has average pronoun realizations. There is a corresponding difference in frequency of these verb forms in the Valencian Catalan corpus, with high frequency of use for *dic* but low frequency of use for *crec*. Erker & Guy (2012) observe that high frequency verb forms in Spanish varieties might have subject realization frequencies and patterns that differ from other verb forms (e.g. *tú sabes* is frequent in the Spanish of New York[7]).

This indicates that Ø+verb and pronoun+verb can have different degrees of fixation and might have undergone different patterns of grammaticalization or pragmaticalization. A similar reasoning applies to the sequence verb+pronoun: *Creo yo* 'I think' with a postverbal pronoun has been related to intersubjectivity in the literature (see Hennemann 2016), i.e. it is used, for example, to invite the interlocutor to express his/her perspective (Hennemann 2016: 464). Given this functional specialization, Hennemann (2016) argues that the Ø+*creo* and *creo*+pronoun should be considered different "constructions" in the sense of Construction Grammar.

While English only has the option pronoun+verb as the possible target of a grammaticalization process (see Thompson & Mulac 1991 for *I think* and *you know* as epistemic phrases; see also Aijmer 1997 for pragmaticalization in this context), *pro*-drop languages potentially allow different options – Ø+verb, pronoun+verb or verb+pronoun – as targets of pragmaticalization. While sequences such as *yo creo*, *no sé*, *sabes*, etc. have been investigated in depth, either from the perspective of discourse markers (see e.g. Enghels 2018 for *sabes*), from the perspective of subject realization

[7] As an anonymous reviewer notes, this might originate from a "diatopically conditioned calque of 'you know'."

(Posio 2015), or from the perspective of parenthetical uses (Schneider 2007), the issue of cross-linguistic variation with these elements requires further investigation.

2.2. Null and expressed subject alternations in Italian[8]

To the best of our knowledge, the literature on Italian null and overt subject pronoun alternation does not specifically focus on verbs of cognition, probably because the increase in overt subject pronouns with this type of verb is not a statistically significant phenomenon. Nevertheless, data concerning unconstrained sets of predicates suggest that although both Spanish and Italian are consistent *pro*-drop languages, null subjects are more frequent in Italian than in Spanish. Contemori & Di Domenico (2021) compare the production of Italian speakers and Mexican Spanish speakers in a picture description task, concluding that "Italian speakers produced significantly more null pronouns than Spanish speakers in all conditions, whereas Spanish speakers produced a higher number of explicit pronouns and full NPs than Italian-speaking participants" (Contemori & Di Domenico 2021: 1017). In Schmitz et al.'s (2016) study of production data, Italian monolingual speakers use the null subject in 65.6% of occurrences, while Spanish monolingual speakers do so in 62.5%. The gap increases with the 1st person subject pronoun, where the null subject rate is 87.4% in Italian and 77.9% in Spanish.

In the Italian data, the rate of null subject production is generally never lower than 58%. In Lorusso's et al. (2005) analysis of a corpus of spontaneous productions of adult Italian native speakers, null subjects are attested in 74% of cases, increasing to 80% with transitive verbs. Di Domenico et al. (2020) show that, in a video description task, adult native speakers of Italian use null subjects in 66.46% of occurrences. Similarly, in their investigation of spontaneous productions of child-directed speech, Schmitz & Müller (2008) find a rate of null subject of 67%.

The ratio of null subjects is even higher with deictic 1st and 2nd person. In the analysis of spontaneous interaction of Italian monolingual adult speakers carried out by Schmitz et al. (2011), the 3rd person subject is null in 58% of occurrences, 2nd person is null in 80%, and 1st person is null in 82% of occurrences. In Scherger & Schmitz's (2019) analysis of semi-structured interviews with adult Italian native speakers, the rate of null subjects is 69.7% with the 3rd person, and 85.8% with the 1st person.

2.3 First person and verbs of cognition: General properties

In declaratives,[9] 1st person has "primary authority" over the validity of the proposition expressed (Frana 2025, who uses this term also for expressions of epistemic modality) and has a particular "epistemic privilege"[10] especially with verbs of personal taste (cf. Lasersohn 2005, Stephenson 2007a), evidentials (which are always based on the speaker's evidence), expressives (which are directly linked to the speaker's stance or attitude), as well as, of course, verbs of cognition.

[8] In this section, the quantitative data all refer to null subjects rather than realized subjects. This is because, while the category of null subjects remains consistent throughout the quoted literature, the category of realized subjects varies from one paper to another, encompassing different syntactic categories (e.g., pronouns, full DPs, etc.). Therefore, presenting the data on null subjects rather than those on overt subjects allows us to report the percentages as given in the literature without adding any further layer of complexity.

[9] Note that 2nd person usually takes over the authority in interrogatives, a phenomenon called "interrogative flip" by Frana (2025); see e.g. the corresponding expression in Italian *sai?*, which is also used as a discourse marker (see e.g. Biasio & Del Fante 2024).

[10] An anonymous reviewer would prefer "condition of (primary) epistemic accessibility" over "epistemic privilege". We do not have a preference for one or the other term.

Epistemic expressions are interpreted according to the perspective of a cogent entity, i.e. they are anchored to a mental state of information. While this entity is not always unambiguous in the case of e.g. epistemic modality (Stephenson 2007a), with inflecting epistemic verbs in the 1st and 2nd person the epistemic perspective is clearly grammatically encoded.

Epistemic modality introduced by propositional attitude verbs like *believe* or *think* in the 1st person is naturally bound to subjective (and not objective or circumstantial) epistemic modality (for the distinction see Lyons 1977 e.g. for epistemic *must*). Furthermore, verbs like *think* and *believe* are different from epistemic verbs proper like *know*, since, as said above, they may contain a doxastic modal base. The former can hence be used as hedges for the expression of uncertainty, which would be impossible for a purely knowledge-based epistemic modal base (compare *Silvio is in Italy, I think.* vs. *?Silvio is in Italy*, *I know*), cf. (3) with a hedging expression in the adversative clause:

(3) I believe the Giants will win, but I'm not sure they will.
(Koev et al. 2021: 487)

Koev et al. (2021) argue that *believe* (along with *think*) is a verb that "conveys high certainty but qualifies this certainty as subjective or lacking evidence."[11] For them, subjective certainty is high, in the sense that the speaker is positively committed to the complement of *believe*, but it is not publicly undisputable and can thus be hedged, leading to the expression of modal uncertainty. However, as Koev et al. (2021) have shown in an experiment, hedging is most natural only in the 1st person singular present tense in main clauses. In the present tense, hedging is less natural with the 3rd person than with the 1st person, but in the past, hedging is more natural with the 3rd person than with the 1st person but is still less natural than in the present tense (for both persons). The authors argue that 3rd person in this context implies "a secondary speech context", since "people have not direct access to other people's mental states, so one's words are often reported as one's beliefs" (Koev et al. 2021: 493). With the verb in the 1st person, the speaker obviously does not need to rely on a secondary speech context for her or his own belief. The most natural context for hedging is 1st person singular present tense (in main clauses), when the subjective epistemic modality is attached to the utterance context and thus naturally connects to current beliefs. This, of course, makes verbs like *think* and *believe* also susceptible to grammaticalization processes in the 1st person singular, in declaratives, and in the present tense.

As Hooper (1975) observed, *think* and *believe* belong to the group of non-factive weak assertive predicates. Assertive predicates, i.e. verbs whose complements are assertions, have several common properties: in English, assertive predicates allow complement preposing and thus can easily be used parenthetically (Hooper 1975: 94–

[11] Cf. Nuyts (2001: 390–391) partially quoted also by Koev et al. (2021): "The mental state predicates systematically express subjectivity. Thus, they typically and predominantly occur in contexts in which the speaker voices personal opinions, very often about topics in the realm of strictly individual experiences or concerns, or also in contexts involving antagonism between the views of speaker and hearer. […] Because the mental state predicates are inherently subjective, they are frequently used as mitigating or hedging devices […]. In such uses, it is usually quite obvious that speakers are absolutely certain about or convinced of what they are saying, but by using the mental state predicate they suggest that they are voicing a tentative and personal opinion which may be wrong, thus 'officially' leaving room for another opinion or for a reaction on the part of the hearer."
As an anonymous reviewer notes this can also be interpreted as a lack of, or reduction in, personal commitment on the part of the speaker.

95); this transforms the assertive complement into the main assertion, whereas the parenthetical is not an assertion of its own, i.e. it is not part of the at-issue meaning. According to Hooper (1975), the non-parenthetical use of assertive verbs would still contain two assertions. Bary & Maier (2021), discussing reporting verbs (i.e. strong assertives) in their non-parenthetical use, show that either the main clause with the assertive verb or – more commonly[12] – the embedded clause of the assertive verb[13] is at-issue, whereas in the parenthetical use only the formerly embedded clause can be at-issue, cf. e.g. the Challengeability test (see 4 and 5). Note that an epistemic verb in the 1st person cannot be challenged even in the non-parenthetical position, although it may represent an assertion on its own.

(4) A: I think that Silvio is in Italy.
B1: #No, that's not true, you don't think that.
B2: No that's not true, I saw him yesterday here in Vienna.
(5) A: She thinks that Silvio is in Italy.
B1: No, that's not true, she doesn't think that.
B2: No that's not true, I saw him yesterday here in Vienna.

Thus, 1st person singular forms of assertive verbs in declaratives are predisposed to develop two properties: First, they can easily be interpreted as not at-issue; and second, they can hence be used parenthetically. Weak assertives such as *think* and *believe* are furthermore inclined to develop into markers of uncertainty, especially in the 1st person singular present tense in declaratives. This is also valid for Sp. *(yo) creo*/It. *credo,* in the 1st person. However, as will be shown in our study, there are some interesting differences between Spanish and Italian as far as subject realization is concerned.

3. The study: Speaker subject realization with verbs of cognition in Italian and Spanish

Given the differences observed between Italian and Spanish, and the lack of previous specific studies on overt vs. null subject pronouns with 1st person singular verbs of cognition in the former, we carried out a study of the specific verb forms Sp. *creo* / It. *credo* '(I) believe/think' and Sp. *pienso* / It. *penso* '(I) think' in spoken corpora of both languages.

3.1. Corpora

Our study of 1st person singular subject realization with Sp. *creo*/*pienso* and It. *credo*/*penso* examines data from two speech corpora. For Spanish, we extracted all occurrences of the word forms *creo* and *pienso* from the COSER corpus (Fernández Ordóñez 2005-). All examples were checked manually, and non-relevant cases (e.g. nominal *pienso* 'fodder') excluded. COSER is a speech corpus of speakers that were of interest in traditional dialectology studies,[14] i.e. older speakers (with an average age of

[12] Koev et al. (2021) show this for English in a second experiment.
[13] According to Hooper (1975: 100–101), assertive verb forms can still receive the parenthetical interpretation (i.e. as not at-issue and not containing an assertion) even if they are not syntactically in a parenthetical position but are e.g. sentence initial.
[14] In this paper, we do not investigate dialectal variation, as that would require a larger database. Some previous studies (e.g. Enrique Ovando & Vázquez Rozas 2020) note that there might be dialectal variation between European and Latin American Spanish varieties in the rate of null subject use with *creo*. The COSER corpus only contains data from European varieties, so we necessarily focus on differences between Spanish and Italian and not on dialectal variation.

74 years) and from rural regions. As of 2025, the available data covers more than 351 hours of recordings from a total of 5,519,285 tokens (https://corpusrural.es/descripcion.php; 14.07.2025). The corpus contains semi-spontaneous interviews, and the topics are centered around the habits and customs of rural life, but several conversations also touch on other topics, such as education, wishes, personal experiences, family, opinions, etc.

For Italian, we analyzed the data from the KIParla corpus (Mauri et al. 2019). KIParla is an open-access corpus of spoken Italian that includes more than 150 hours of conversation and a total of 1,991,903 words. It comprises a variety of data conditioned by socio-linguistic diversity: the speakers are of different ages, education levels, and different geographical origins within Italy. The types of productions vary, and include semi-structured interviews, spontaneous conversations, university lessons, exams, etc.

The fact that the two corpora contain different types of speakers and different types of conversational topics obviously limits the comparability of the data. However, we will show that the differences observed between Italian and Spanish with respect to the frequency of 1st person overt subject pronouns are sufficient to draw several conclusions on a comparative basis. We will discuss further evidence that it is not the age of the informants that is the crucial factor by comparing the results from COSER (which contains older speakers) with previous studies on Spanish, which consider a variety of data types (see §4).

In total, 1,490 sentences containing Sp. *(yo) creo* and 90 sentences with Sp. *(yo) pienso* from the COSER corpus were analyzed. For Italian, 419 sentences containing *(io) credo* and 669 sentences with *(io) penso* from KIParla were analyzed. The difference in overall frequency of the verb forms might be relevant for the patterns of subject pronoun realization and parenthetical uses that we will discuss later.

3.2 Data annotation

For the study of subject expression with 1st person singular verbs of cognition, we annotated the following categories:

(6) Annotated categories:[15]

[15] We also wanted to annotate the factor of *referential continuity* in the data set, given that it has been repeatedly observed in previous studies that continuous contexts favor subject omission while non-continuous or shifting reference favors subject realization (see Givón 1982; Cameron 1992; Frascarelli 2007, among many others). However, most previous studies that explore (topic) continuity focus on 3rd person subjects. Furthermore, the complexity of annotating factors relating to information structure in (semi-spontaneous) spoken data with various speakers made it impossible to codify the whole data set in a reliable fashion. Additionally, prosodic criteria, which are crucial, could not be taken into account. We thus leave aside a full investigation of this variable for future research. For Spanish, we carried out a preliminary analysis of a random sample of 400 sentences with regard to *continuity*, using simplified coding: Reference was annotated as continuous when the previous sentence contained reference to the speaker and non-continuous if it lacked mention of the 1st person singular. This initial analysis suggests that continuity does not affect subject realization: In the random sample of 400 sentences, continuous contexts had an overt pronoun rate of 66%, compared to 69% in non-continuous contexts. In Italian, we annotated 412 sentences of *(io) credo* with the same simplified coding and the result is that both continuous and non-continuous contexts had an overt pronoun rate of 9% (which is the average in the Italian data of *credo*; see above), showing that continuity does not affect 1st person singular pronoun realization. With It. *penso*, the simplified annotation of 593 sentences yielded no significant difference either: 11% overt *io* in continuous contexts and 16% in non-continuous contexts (compared to the average of 15% overt *io* with this verb from). From this preliminary simplified analysis, then, *continuity* does not seem to affect the expression of 1st person singular subject pronouns with verbs of cognition in our data.

(i) null vs. overt subject (and pre- vs. postverbal position in the case of overt subjects),
(ii) polarity (affirmative vs. negated),
(iii) complement type (embedded clause with or without complementizer, pronominal, NP, clitic, no complement, etc.),
(iv) co-occurring adverbs or discourse markers,
(v) clause type in which the verb of cognition appears in (main clause, coordinated clause, subordinated clause, parenthetical),
(vi) complement semantics (only in those cases in which the CP complement was fully realized).

While factors (i)–(iv) are straightforward, categories (v)–(vi) require further explanation: Regarding (v) clause type, we classified parentheticals separately, adapting the definition of parenthetical verb provided by Schneider (2007: 77). According to Schneider, two of the hallmarks of parenthetical occurrences of a verb, his "reduced parenthetical clauses", are the absence of an overt syntactic link with the host clause and the structural self-sufficiency of the host structure. Building on these two notions, we counted as parentheticals all those occurrences in which the epistemic verb is found in medial or final position and, more broadly, when it takes scope over a constituent whose syntactic category is incompatible with the verb's selectional requirements, e.g., an AP.[16]

With respect to the category of *complement semantics* and function (vi), we use the classification in Herbeck (2021), which is based on the following distinction: whether the type of information encoded in the complement clause of THINK is concrete, falsifiable or abstract, subjectively assessed information. The idea behind this classification is that with an internal, subjective perspective, subject pronoun realization rates are expected to be higher in 1st person singular, given that the speaker perspective is stressed. In order to be able to classify these notions in spoken language data, the following categories of complements (A)-(E) were annotated (Herbeck 2021). Each category is accompanied by examples from the Italian and Spanish corpus data.

(7) Complement semantics:

A) DESCRIPTIONS (of things, persons, states, events): Objectively falsifiable, externally perceivable events, states, objects, beings, etc.

Spanish:
a. Y el kilo de pan aquí, creo que cuesta ciento noventa pesetas. (COSER-1232_01; Cantábria)
'And a kilo of bread, I think that it costs ninety pesetas.'
b. […] ayer llovió creo que un poco, y decían que iba a llover hoy también (COSER-0310_01; Alicante)

However, as mentioned above, an in-depth study of this factor with verbs of cognition remains to be carried out.

[16] Some authors (e.g. Hooper 1975) would also distinguish between main clause and parenthetical use in verb-initial position (based, among others, on the criterion of at-issueness, cf. fn. 13). Dehé & Wichmann (2010) make a distinction (for English) between *main clause*, *comment clause* and *integrated parenthetical* on prosodic grounds. Hedberg & Elouazizi (2015) use semantic relationship (scope), syntactic position and prosodic phrasing to distinguish between *main clause use* and use as a *parenthetical verb phrase* (including for English). The notion of *parenthetical verb* goes back to Urmson (1952), while the notion of *comment clause* goes back to the English grammar of Quirk et al. (1972).

'yesterday, it rained, I think [that], a little and they said that it was going to rain today too.'

Italian:

c. E credo che un'iniziativa del genere l' abbiano fatta anche quest' anno. (KIParla_TOD2012)
'And I believe they have done a similar event this year.'

d. Credo che facesse di cognome proprio Verdi (KIParla_BOA3021)
'I believe his/her surname was actually Verdi.'

B) EXISTENTIALS: (with *haber, esistere* or other existential verbs)

e. Spanish:
Antes, casi todo eran mineros, ahora ya mineros aquí creo que no haiga, no sé si habrá uno solo. (COSER-0509_01; Asturias)
'Before, almost everyone was a miner, but now I don't think there are any miners here anymore. I don't know if there's a single one left.'

f. Italian:
E ho fatto scienze del turismo che è una era un corso di laurea che credo che esista ancora adesso (KIParla_PTD006)
'And I studied tourism sciences, which is a/was a degree course that I believe still exists today'

C) EVALUATIONS: Subjective evaluations or assessments of states, events, persons, etc., e.g. by means of an evaluative predicate or opinion markers

g. Spanish:
No van las parejas así mejor que si van así uno y así otro como si estuvieran reñíos? [HS:E1 Claro…] Como si estuvieran reñíos. Pues yo creo que van mejor así, ¿no? (COSER-1401_01; Ciudad Real)
'Don't couples get along better when they go like this, one like this and the other like that, as if they were arguing? [Of course …] As if they were arguing. Well, I think they are better this way, right?'

h. Italian:
il vecchio nietzche diceva che la vita senza la musica sarebbe un errore credo che sia una delle cose più belle che abbia scritto (KIParla_TOD1002)
'Old Nietzsche said that life without music would be a mistake. I think that's one of the most beautiful things he ever wrote.'

D) UNREAL events, states, etc. (in the conditional, future, …)

i. Spanish:
Yo creo que no me casaría, [RISAS] creo que no, no me casaría, creo que no, no sé. (COSER-2001_02; Guipúzcoa)
'I think that I wouldn't get married, I think no, I wouldn't get married, I think not, I don't know'

j. Italian:
Non credo che sarete in tanti al primo appello (KIParla_BOA1001)
'I don't think there will be many of you at the first roll call.'

E) MIND (of the speaker or other referents): Statements with respect to thoughts, feelings, etc. of the speaker or other referents

k. Spanish:
Ah, no, yo tampoco, yo creo que tampoco lo sabría. (COSER-4904_01; Mallorca)
'Ah no, me neither, I think that I wouldn't know it either'

l. Italian:
Porti michilino parleremo di lui credo che oramai sia pronto (KIParla_BOD1001)
'Bring Michilino! We'll talk about him. I think he's ready now.'

Apart from these categories, we annotated some cases as combinations of DESCRIPTION and EVALUATION. This mostly affected comparisons that are objectively falsifiable, but that also in several cases imply (subjective) assessments on the part of the speaker:

(8) COMPARISONS/ASSESSMENTS

a. Spanish:
Que daba muchísima independencia tener la bicicleta, ¿no? Cambiaba, yo creo que cambiaba la vida. (COSER-0204_01; Albacete)
'Having a bike gave you a lot of independence, didn't it? It changed, I think that it changed life'

b. Italian:
Credo che in confronto a molte altre città ci sia molto verde (KIP_PTA001)
'I think that compared to many other cities, there is a lot of green space.'

Although (8)a and (8)b describe an event that can be perceived by means of external evidence, we are dealing with a (subjective) assessment of the differences between the past and the present or between various places.

While the morpho-syntactic variables of *null* vs. *overt*, *polarity*, *clause type*, and *complement type* were annotated for the whole dataset, only those sentences with THINK that have a fully realized CP complement could be annotated for the variable of *complement semantics*. Furthermore, several doubtful cases had to be excluded. The complements were annotated separately by two annotators with respect to their semantics. For Spanish, there were in total 590 complement clauses of *creo* for which the two annotators were in agreement and the embedded CP was fully realized (i.e. not truncated). These form the basis of the study of subject expression in relation to complement semantics in Spanish.[17] Furthermore, there were some specific cases which we treated separately: 168 cases contained only the polarity item *sí* or *no* and 49 cases

[17] The other cases were excluded from the study for one of the following reasons: the annotators did not agree; the sentences were truncated or repeated; or the classification was doubtful.

contained a specific, repetitive configuration at the end of the interviews, in which the interviewers expressed that they thought that the interview was over. Given that the last configuration is very specific to the type of interview situation, we consider it separately. With *pienso*, there are far fewer data points (only 41 full CP clauses in total) so we can only examine them qualitatively (see the discussion around example (18)).

For Italian, 122 occurrences of full CP complements selected by *credo,* and 264 occurrences of full CP complements selected by *penso* were annotated semantically, with agreement between the two annotators.

3.3. Results

In this section, we present the results of the corpus study on a descriptive quantitative basis. We look first at the Spanish data from the COSER corpus, focusing on *(yo) creo*. We then present the study of *(io) credo* and *(io) penso* in the Italian corpus KIParla.

3.3.1. Spanish – *(yo) creo* and *(yo) pienso* in the COSER corpus

As stated above, *creo* has a much higher frequency in the spoken data than *pienso*, so our discussion will mainly concentrate on the former, with some observations about the latter towards the end of the section. In the COSER corpus, the rate of 1st person overt subject pronouns is generally high with present indicative *creo* with a frequency of 62% of overt *yo* in total. With respect to **polarity**,[18] the observation of previous studies that polarity affects subject realization (e.g., Enríquez 1984; Posio 2015; Herbeck 2021) also holds for the present data:

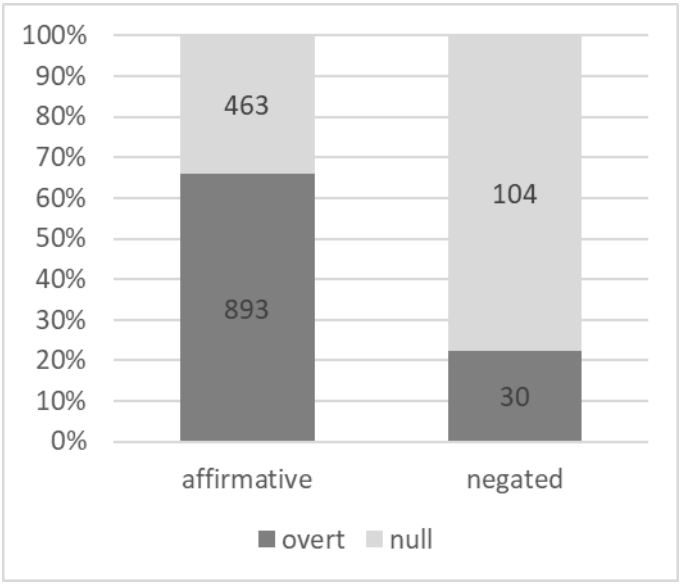

Figure 1: Overt and null subjects with *creo* according to **polarity**

As can be seen, *creo* appears with overt *yo* in 66% of affirmative sentences and in only 22% of negated sentences.

[18] Note that we only consider the matrix position of negation for the annotation, i.e. *Creo que no lo harán* 'I think that they will not do it' with negation inside the embedded clause would count as 'affirmative' in our classification. However, we are aware of the fact that *believe*-verbs are typically neg-raising, which means that a negated matrix verb could also encode negation of the embedded clause (i.e. *no creo que lo harán* could be interpreted as both 'I do NOT believe that they will do it' and 'I believe that they will NOT do it').

Turning to 1st person subject pronoun expression with respect to the **complement type** selected by *creo*, the highest frequency of overt pronouns is found when *creo* selects an embedded CP complement (71% overt *yo*),[19] and the lowest when it selects clitic complements (i.e. *lo*; only 8% overt *yo*):[20]

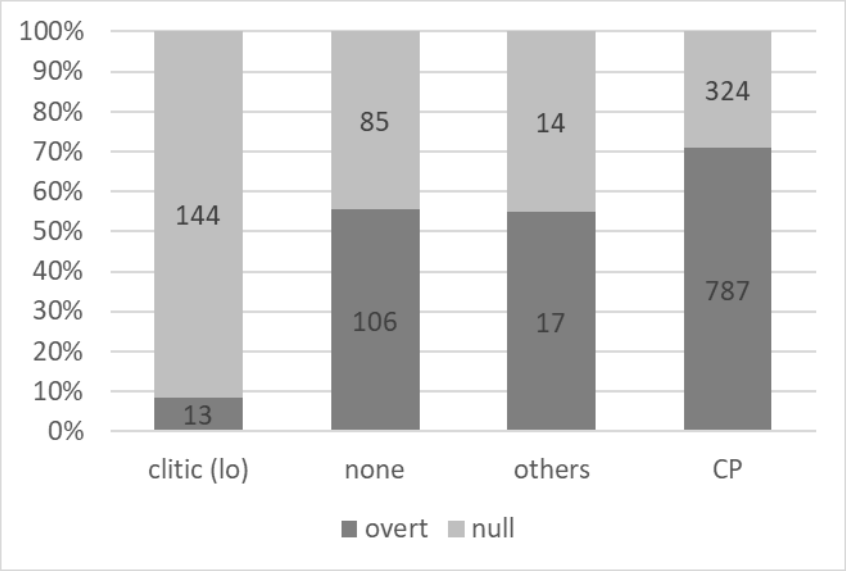

Figure 2: Overt and null subjects with *creo* according to the **function of the complement clause**

We take the low frequency of overt *yo* with *creo* when it has a clitic complement to be related to a formulaic or idiomatic use of the sequence *ya lo creo*. This sequence can be used to express *verum focus*, similarly to other expressions like *sí que* in Spanish (Escandell-Vidal & Leonetti 2009):[21]

(9) a. Creo que me escucha.
'I think that s/he listens to me.'
b. Ya lo creo que me escucha.
'S/he really listens to me.'
(Escandell-Vidal & Leonetti 2009 [translations added])

While the first sentence is neutral, the latter with clitic doubling of the complement clause indicates that the embedded complement conveys given information as part of the background (Leonetti 2008) and an emphatic affirmation on the speaker's side. Escandell-Vidal & Leonetti (2009) thus analyze the sequence as a fixed expression – as an emphatic affirmative marker. In this use, it seems to be the case that a fixation

[19] Note that complementizer deletion is not frequent in spoken Modern Spanish (see e.g. Pountain 2015), in contrast to Italian, which licenses it productively (see Giorgi & Pianesi 2005 for Italian). In fact, in our data, complement clauses in Spanish are almost always introduced by *que*, with only a few exceptions (11 in total). Furthermore, as is well-known, complement clauses with affirmative matrix *creo* usually trigger indicative mood while negative *no creo* tends to select subjunctive mood, with some exceptions (see NGLE 25.3h and 25.8d-e, 25.8k). In Italian, on the contrary, positive *credo* can select either indicative or subjunctive, depending on a variety of semantic, stylistic and sociolinguistic factors (see Mari & Portner 2021; Cerruti & Ballarè 2023).

[20] The category 'others' includes demonstratives, PPs and NPs as complements.

[21] Note that *ya lo creo* also has an entry in the RAE dictionary, which defines it as a colloquial expression used to affirm or agree emphatically.

process affects Ø+*creo*. Also without the complementizer, a degree of fixation of *creo* can be observed: *ya lo creo*, apart from expressing agreement on the speaker's side, indicates willingness to support the expressed reality and thus adds a subjective value (cf. Delbecque 2006: 54; for the use of *ya* with epistemic verbs, see also Urdiales 1973).

The following example from our corpus study supports these assumptions:

(10) I1 : Sí, claro que se hiela.
E2 : ¿Hace aquí frío en invierno?
I1 : Claro que ha-, ya lo creo que hace. Sí, sí, sí que hace frío, sí.
'I1: Yes, of course it freezes. E: Is it cold here in winter? I1: Of course… Indeed it's cold. Yes yes, of course it's cold, yes.'
(COSER-0310_01; Alicante (Salinas))

In (10), the speaker expresses agreement with the content expressed by E2. At the same time, emphasis is put on the affirmative value and the truth of the embedded proposition, which is taken as given information. In (10), *claro que* 'sure', *ya lo creo que* 'indeed' and *sí que* 'of course' all have a similar function of emphasizing the truth of the embedded proposition and agreement with previously introduced, given information (see e.g. Kocher 2017). The following example demonstrates the same function without duplication of the *que*-complement, but its content is retrieved from the immediately preceding context:

(11) I1: Aquí se bebía, sí. [RISAS] Sí, ya lo creo.
'I1: Here they used to drink, yes [LAUGHTER] Yes, indeed.'
(COSER-3106_01; Murcia)

In these cases, it is the speaker's agreement with the embedded proposition of *creo*, but not the speaker himself, that is in the foreground. Given that the clause including the speaker subject is backgrounded and the focus is on the *verum* part, it follows that *creo* has pragmaticalized into a fixed expression together with a null subject (and not with overt *yo*) in these cases. Evidence for this reasoning comes from the observation that in 100% of the 124 cases in which the adverb *ya* 'already', the clitic *lo* 'it' and *creo* are combined, the subject is null.[22]

In the other cases in which *creo* takes an accusative clitic complement, the number of overt 1st person subjects is average (20/33; 61%) compared to the whole sample, but the overall numbers are low. In these cases, *creo* often appears with the reflexive pronoun (i.e. pronominal *creerse*) and has the meaning of 'believe' (i.e. 'accepting the truth of a proposition' or 'having faith in something')[23] and not the meaning of 'think' or 'holding an opinion'.

(12) I: Cualquiera de aquí que se le encontró, dice: "[NP] tiene novio". Pues rápidamente: "No, no me lo creo." (COSER-1823_01; Granada)
'Anyone here who has met her says, '[NP] has a boyfriend' Then quickly: 'No, I don't believe it.'

[22] The cases of *ya lo creo* have been included in the overall frequency of 1st person subject realization. The reason is that the focus of this paper is precisely this type of (semi-)fixed uses of verbs of cognition and their (null) subjects.

[23] Notably, according to Anvari et al. (2019), when *creerse* is used not in the 1st person in main clause contexts, it seems to convey that the content of the embedded CP is false.

However, there are configurations in which the frequency of subject expression with *creo* is higher, as is the case when the verb of cognition takes an embedded CP complement (71% of overt *yo*). As has been discussed in Aijón Oliva & Serrano (2010), Vázquez Rozas & Enríquez Ovando (2020) and Herbeck (2021), this seems to be partly related to a high degree of subjectivity and a use of *creo* as a verb for expressing personal opinions, rather than a verb expressing epistemic uncertainty, which mitigates commitment to the truth of the embedded proposition.

As described in Section 3.1, we annotated the CP complement of *creo* with respect to the type of information that is encoded within it. Figure 3 presents the 1st person singular subject realization frequencies in the 590 analyzable cases with regard to **complement semantics** (plus 168 cases of complements that only contain a polarity item):

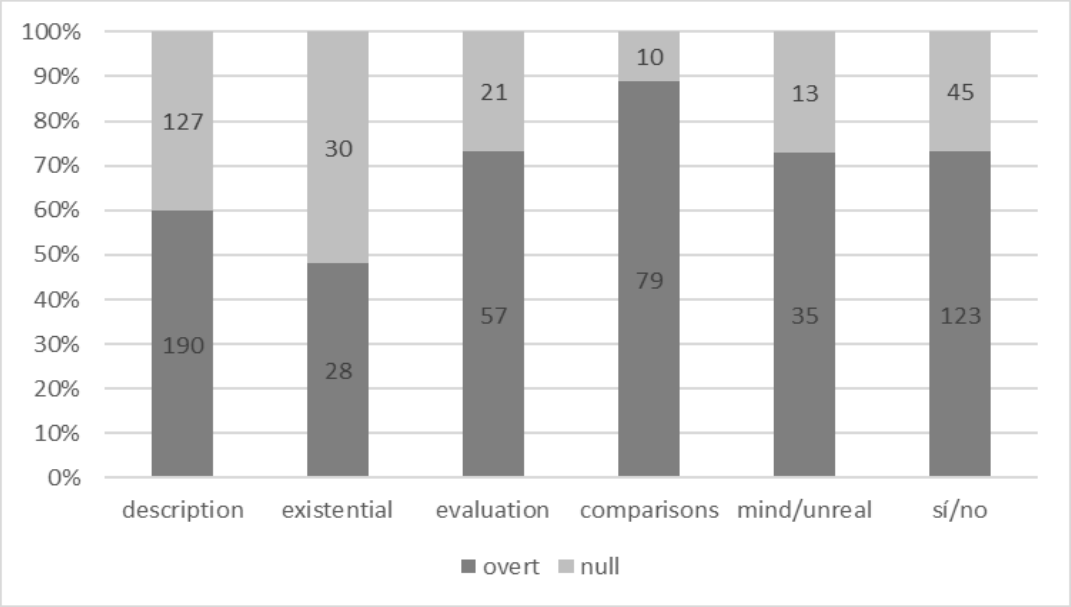

Figure 3: Frequency of overt *yo* with *creo* according to the **function of the complement clause**

From a descriptive standpoint, it can be observed that *yo* is most frequently realized with *creo* when the complement expresses a *comparison* (89% overt *yo*) or a subjective *evaluation* (73%). Furthermore, abstract concepts, i.e. *state-of-mind* or *unreal*, and complements containing a polarity item, also have a high rate of overt subject pronouns (73% of overt *yo*). With respect to the last of these, complements containing *sí* or *no* often express agreement or disagreement with the interlocutor, meaning that the notion of *opinion* is also relevant here. The lowest frequency of overt *yo* is observed with *existentials* (= 48%) and *descriptions* (60% overt *yo*). Abstract information and subjective evaluations thus tend to correlate with a high rate of overt subject pronouns, to a greater extent than concrete, objectively falsifiable information (*descriptions* and *existentials*). This confirms observations from previous studies, such as those by DeSaeger (2008), Aijón Oliva & Serrano (2010), Vázquez Rozas & Enríquez Ovando (2020), and Herbeck (2021).

Note that one particular exceptional configuration at the end of the interviews, used by the interviewer to terminate the conversation, has an unusually high rate of overt subject pronouns – out of 49 cases, 46 appear with overt *yo*, i.e. a frequency of 94% (see also Herbeck 2021 for a similar observation in the PRESEEA data):

(13) E1: Bueno, pues nada, yo creo que ya hemos preguntao todo. (COSER-1417_01, Ciudad Real)
'Well, that's it. I think that we have already asked everything.'

It is important to note that in these cases, the speaker is not merely expressing uncertainty with respect to the truth of the embedded proposition. The speaker (here, the interviewer) is implicitly also asking for the approval of the addressee (the informant) or of another referent in this communicative situation (e.g. another interviewer). Thus, we have an interactive and intersubjective use of *yo creo* which in our data has a very clear tendency towards overt expression of *yo*.[24] In fact, in the following example, this dimension of intersubjectivity is made explicit by the interaction of the two interviewers and the informant:

(14) E1: [HCruz: Yo creo que sí, ¿no? [HS:E3 [Asent].] [HS:I1 [L-Otra: Què més?]] Yo creo que, yo creo que estamos, [HS:E3 Sí.] empezamos a y media aproximadamente]. [HCruz: [HS:E3 [Asent].]] […] (COSER-4904_01, Mallorca)
'Interviewer 1: I think yes, right? [Interviewer 3 nods]
Informant: What else?
Interviewer 1: I think that, I think that it's enough. [Interviewer 3: Yes]
Interviewer 1: We started at approximately half past. [Interviewer 3 nods]'

Turning to the **type of clause** in which *(yo) creo (que)* appears, some interesting data arise with regard to subject expression in parenthetical uses:[25]

clause type	null	overt	total	%-overt
main clause	470	762	1232	62%
subordinate	22	24	46	52%
coordinated	34	38	72	53%
parenthetical	41	98	139	71%
total	567	922	1489	62%

Table 1: First person subject realization with *creo* according to clause type

In the sample, parenthetical uses of *creo* have the highest frequency of overt subject pronouns. Interestingly, these parenthetical uses allow all three options: a null subject, a preverbal overt pronoun or a postverbal pronoun (see also Schneider 2007: 121).

As is well-known, **parentheticals** can appear in clause-medial and clause-final position. In both positions, Spanish allows versions with a null subject or with an overt 1st person singular pronoun:

(15) *(yo) creo* in clause final position
a. Null pronoun:
La, la que es fina como el dedo la llamamos la salchicha…, eh.
Después hay la, la normal y después la, una de gorda también, creo
(COSER-0805-01; Barcelona)

[24] See Hennemann (2016) for an analysis of *creo yo* along the lines of intersubjectivity. In our data, however, intersubjectivity also seems to play a role with preverbal *yo*.
[25] We excluded one case of *creo* in an interrogative clause.

'The one that's as thin as a finger, we call it the *salchicha* [sausage]... eh. Then there's the normal one, and then there's a thick one too, I think.'

b. Overt *yo* (preverbal):
Si se casan cien, se, se apartan cincuenta fácil, [HS:E1 [Asent].] fácil, yo creo. (COSER-0103_01; Ávala)
'If a hundred get married, fifty will easily separate, [Interviewer nods] easily, I think.'

c. Overt *yo* (postverbal)
de momento le ponemos fuego, en la mayoría le pondrán, creo yo. (COSER-3924_02; Soria)
'For now, we're setting it on fire, and most people will do the same, I think.'

(16) *(yo) creo* in clause-medial position:

a. Null pronoun:
Ya le digo, pa allí pa abajo hay una mujer que las hace creo mu bien.
'I tell you, down there there is a woman that makes them, I think, very well.' (COSER-4617_01; Zamora)

b. Overt *yo* (preverbal):
E2: Vamos a ir a comer, yo creo ya. (COSER-0107_01; Ávala)
'E2: We will go and have lunch I think already.'

c. Overt *yo* (postverbal):
Ah, esa costumbre habrá existido siempre, creo yo, de regalarle el novio a la novia y la novia al novio. (COSER-0934_01; Burgos)
'Ah, this custom will have always existed, I think, [that] of giving the groom to the bride and the bride to the groom.'

Thus, *(yo) creo* as a parenthetical has a certain flexibility with respect to the realization and position of the subject. Use as a parenthetical is one sign of a grammaticalization process (Posio 2011; Vázquez Rozas & Enríquez Ovando 2020: 222) and, as such, it is *a priori* surprising that (i) a high degree of variation is observed with respect to subject realization and position and (ii) the preferred option in terms of frequency is the option with an overt subject pronoun in our sample.

Vázquez Rozas & Enríquez Ovando (2020: 230) also report a considerable frequency of expressed subjects with parenthetical *(yo) creo* and, interestingly, subject realization in this context is higher when *creo* has a clearly doxastic meaning, i.e. when it expresses an opinion (66.7% and 76.5% *yo* respectively in their Mexican and Galician Spanish corpora) than with the epistemic meaning (34.9% in the Mexican corpus and 20% in their Galician Spanish corpus). It might thus be that it is not only *creo* that has a certain degree of grammaticalization (as an epistemic marker), but that *yo creo* has also undergone a fixation process, arguably pragmaticalization, towards an opinion marker.

Another issue that is worth noting is that parenthetical uses of *(yo) creo* do not only exist without complementizer *que*; the complementizer can also be maintained (see also Schneider 2007):

(17) pero se ha recorrido yo creo que el mundo, casi. (COSER-4503_01 ; Vizcaya)
'But he has travelled I think [that] almost the entire world.'

Let us briefly turn to ***(yo) pienso*** '(I) think'. We cannot offer a detailed analysis of this verb form because it is overall much less frequent than *(yo) creo* – only 90 cases in total (compared to the 1,490 cases of *(yo) creo*) in the sample). In this small sample, *(yo) pienso* has an overt subject pronoun rate of 56%, slightly lower than for *(yo) creo* (62%). Furthermore, the rate of overt subject pronouns seems to be highest with *que*-complements (61%) and when there is no complement (75%), but there are very few data points. Given the low frequency of *pienso* in the sample, we also only find a small number of cases of parenthetical uses (12, of which 9 have an overt subject pronoun). However, in the case of parenthetical *pienso*, all overt subjects appear in postverbal position in our data, i.e. *pienso yo*:

(18) Claro, tampoco la vida en los pueblos es como la capital, pienso yo, porque claro, aquí tampoco había mucha cosa de… (COSER-3203_01; Navarra)
'Of course, life in the villages isn't like life in the capital, I think, because, of course, there wasn't much of…'.

This combination of *pienso* plus a postverbal 1st person singular subject pronoun seems to have a similar function as other verbs of cognition or saying plus *yo* (such as *creo yo* or *digo yo* '(I) say'). Hennemann (2016) argues for an intersubjective function. Although this use can be observed in our data, it seems that there is a high degree of polyfunctionality with the verb form *pienso* as well as *creo + yo*: in (18), the opinion of the addressee is not necessarily directly involved. Rather, the speaker's perspective is emphasized, and the question of whether agreement is obtained with other referents is left open, i.e. its meaning could be paraphrased with 'at least that's what I think (I don't know about others – they might be of a different opinion)'. The construction can thus also evoke a type of contrast similar to Mayol's (2010) "uncertainty contrast", in which the subjective perspective is stressed; other alternative perspectives are evoked, but they are not exhaustively negated.

The lower frequency of *pienso*, the less clear functional characteristics of its use and low frequency of parentheticals (with the exception of *pienso yo*) indicates a lower degree of pragmaticalization of this verb form compared to *creo*.

3.3.2. Italian – *(io) credo* and *(io) penso* in the KIParla corpus

From the KIParla corpus, a total of 422 occurrences of the 1st person singular *credo* were annotated. In total, 91% of these appear with a null subject, and 9% with an overt pronominal subject. This rate of overt subject realization is not only significantly lower than the Spanish counterpart but is also lower than has been reported in other literature on subject realization in Italian (see §2.2).

The rate of overt subject pronouns does not seem to covary with polarity; it appears in 9% of affirmative sentences and 11% of negated sentences.

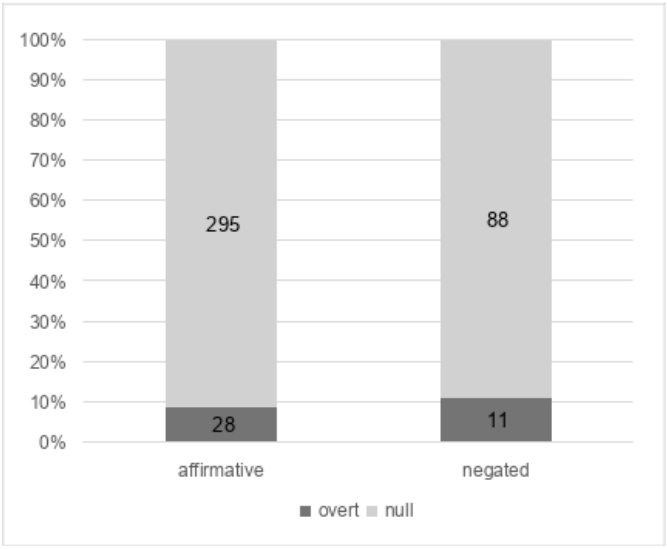

Figure 4: Overt and null subjects with *credo* according to the **polarity**

With regard to **clause type**, the rate of overt subject pronoun realization is close to the average value in both main (11%) and coordinate (9%) sentences. Pronoun realization is highest in subordinate clauses (17%), and lowest in parenthetical positions (2%).

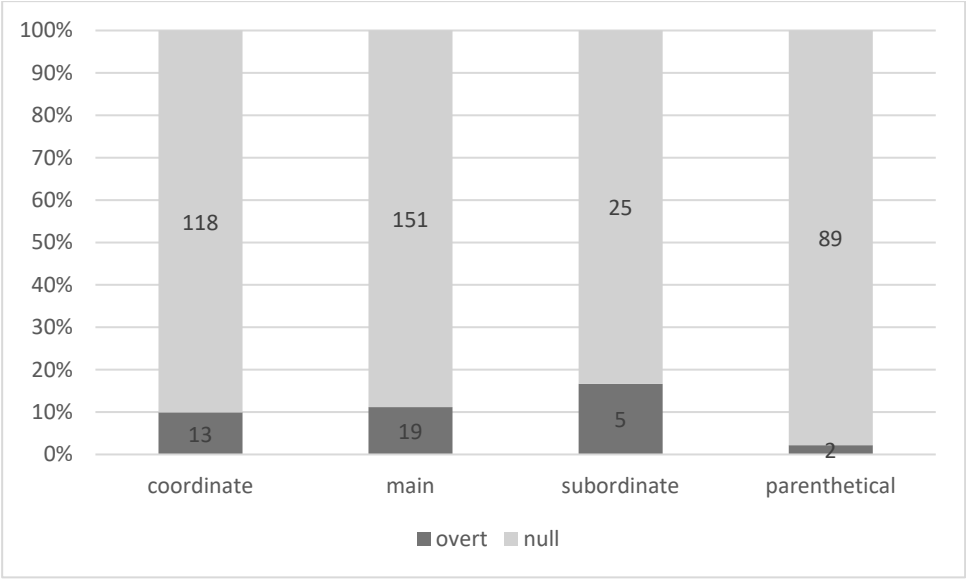

Figure 5: Overt and null subjects with *credo* according to **clause type**

It is interesting to note that, in **parentheticals**, subject expression frequencies in the Italian data show the opposite trend to the Spanish data, confirming a tendency observed previously by Schneider (2007). While parentheticals more frequently appear with overt *yo* in Spanish, the Italian data show that parentheticals almost exclusively appear with a null subject, the only exception being two cases in final position and with a postverbal subject:

(19) a. Quando tocca a lui Coltrane lo fa suonare sugli accordi originali del pezzo questo è il motivo credo io. (KIParla_BOD1010)
'When it's his turn, Coltrane plays it on the original chords of the piece. This is why, I believe.'

b. Perché adesso non c'è più questa cosa an~ sotto quell'aspetto dalle case popolari qui attualmente credo io. (KIParla_ PTB018)
'Because now there is no longer this thing under that aspect from the public housing here, I believe.'

With respect to occurrences of *credo* selecting a fully realized CP complement, the role of complement semantics in determining subject realization cannot be fully assessed because of the generally low number of instances of overt *io* in the data. At first glance, some functions might play a role in in determining the frequency of overt subject pronouns, as is the case for the category of 'unreal' complements (4/13 = 31% overt *io*), in which overt *io* appears more frequently than in 'descriptions' (7/60 = 12%). However, this tendency needs to be checked against a database that includes more cases of overt *io* with *credo*.

Turning to *penso*, the number of total analyzed occurrences was 608; in contrast to the patterns observed in Spanish, this form is overall more frequent than *credo* in the KIParla corpus. Of these 608 occurrences, 86% appear with a null subject, and 14% with an overt subject pronoun, showing a slightly higher rate of overt subject pronouns than *credo*.

The overt pronoun appears in 15% of affirmative sentences, and in 6% of occurrences with negative **polarity**. This data differs from *credo,* where the overt pronoun appears more frequently in negated contexts.

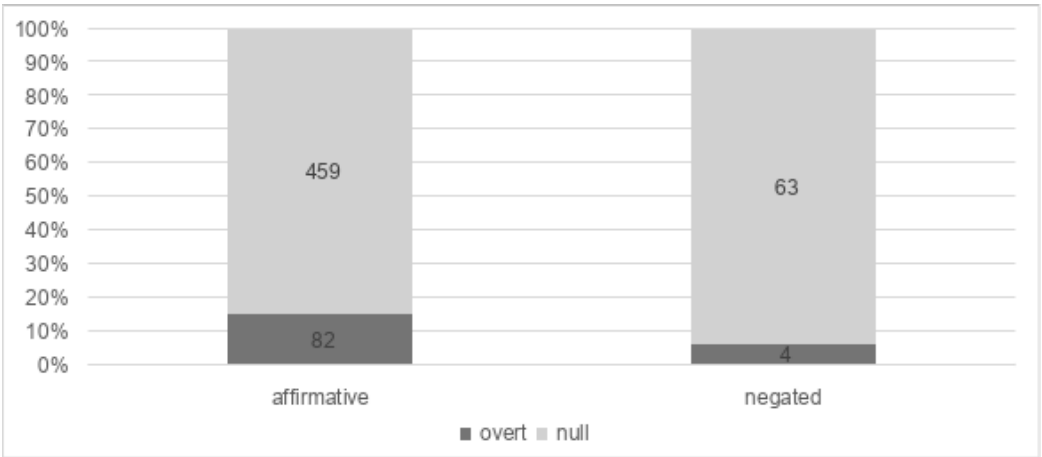

Figure 6: Overt and null subjects with *penso* according to **polarity**

Turning to **clause type**, the overt pronoun appears more frequently in main clauses (24%), and at a lower rate in coordinate (11%) and subordinate clauses (12%). What remains consistent between the two Italian epistemic verbs is the drop in pronoun realization in parenthetical positions, i.e. *credo* (2%), *penso* (0%). Thus, with both verbs, parentheticals almost exclusively appear with a null subject, contrary to Spanish.

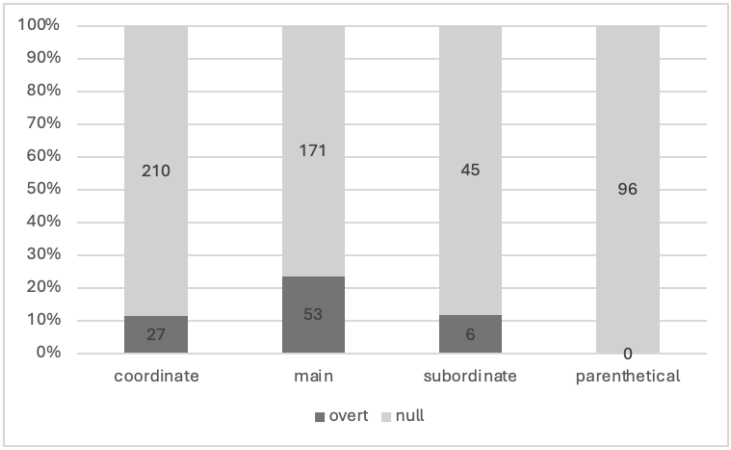

Figure 7: Overt and null subjects with *penso* according to the **clause type**

Interestingly, the rate of overt subject pronouns does covary with the **complement type**. Full CP complements have one of the lowest subject realization rates in the sample (17% with complements introduced by *che* and only 4% overt *io* when *penso* takes a complement with *che*-deletion). The frequency of overt subject pronouns increases when the complement is an infinitive (25%) (introduced by *a* or *di*), a clitic (29%), or a complement expressing direct thought, i.e., quotation (29%), and is at its highest with PP complements (33%), see Table 2.

complement type	null	overt	total	%-overt
che	220	44	264	17%
che deletion	77	3	80	4%
clitic	20	8	28	29%
a/di+inf	24	8	32	25%
di_si/no	30	6	36	17%
*a/di*_DP	29	14	43	33%
none	117	1	118	1%
direct thought	5	2	7	29%
total	522	86	608	14%

Table 2: Overt and null subjects with *penso* according to **complement type**

For full CP complements, in line with what we have already seen with *credo*, the lowest rate of overt subject pronouns is observed when the complement semantics belong to the *description* category (10%), and the highest with the *unreal* (16%), and *description+evaluation* categories (36%), although there are few data points for the latter (see Figure 8). In keeping with what has been seen in the Spanish data, there seems to be a tendency for subject realization to occur with abstract information and subjective evaluations. However, given the lower overall frequency of overt 1st person subject pronouns, this tendency needs to be interpreted with caution.

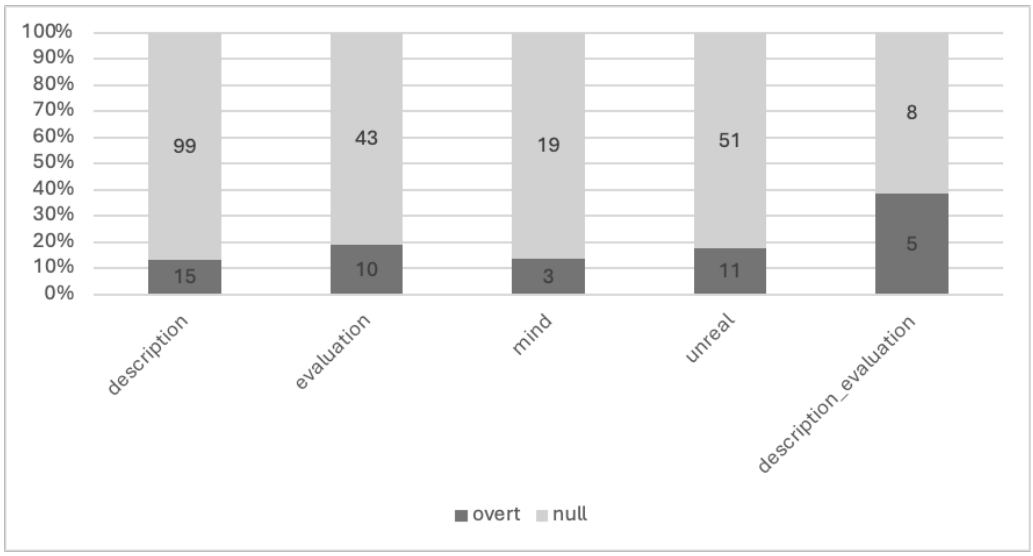

Figure 8: Overt and null subjects with *penso* according to the **complement semantics**

In the next section, we provide a comparative discussion of the results and make some tentative suggestions regarding the reason for the differences between Italian and Spanish with respect to 1st person subject expression with verbs of cognition.

4. Discussion

The greatest difference between Italian and Spanish subject expression with 1st person singular verbs of cognition can be found in the general frequency of overt subject pronouns: in Spanish, overt *yo* is particularly frequent with *creo* (= 62%), but is infrequent with Italian *credo* (= 9%) and is thus even lower than general overt subject frequencies reported in previous studies (see §2.2). This difference is surprising because there is no clear morpho-syntactic reason for these tendencies, e.g. with respect to subject verb agreement morphology. There are only a small number of data points for *pensar* in Spanish, but overt *yo* also tends to be frequent with this verb form (56% overt subjects with 90 verb forms), contrary to what we see with Italian *pensare* (14% overt *io* with 610 verb forms). What, then, is the origin of the differences in frequency between the Italian and Spanish data?

Parenthetical uses of verbs of cognition in the two languages offer important data in this respect. In Spanish, *creo* as a parenthetical has an overt subject pronoun rate of 71% (see Table 1), even higher than in main clause position (62%). We have further seen that overt *yo* can appear with the parenthetical in clause-medial and clause-final position and that the pronoun itself can appear in pre- and postverbal position (see (15)b,c and (16)b,c). In Italian, on the contrary, parenthetical use of *credo* almost always appears with a null subject – there are only two cases with overt *io*, both of which have the pronoun in postverbal position.[26] It therefore seems to be the case that in Spanish the verb of cognition has undergone a process of pragmaticalization and fixation together with the overt subject pronoun, but this has not happened to the same extent in Italian. The use of *yo creo (que)* as a semi-fixed expression consequently boosts overt subject pronoun frequencies in Spanish while the use of Ø+*credo* boosts the number of null subjects in Italian.

[26] This does not mean that preverbal *io* in parenthetical clauses is ruled out – however, we have found no positive evidence in the corpus data, unlike in Spanish, where preverbal *yo* in parenthetical clauses is actually the most frequent pattern.

Spanish *(yo) creo* appears to have developed into two (semi-fixed) markers with partly specialized functions[27] – an argumentative vs. epistemic or doxastic vs. mitigating device (see Aijón Oliva & Serrano 2010; Vázquez Rozas & Enrique Ovando 2020 for discussion). In Spanish, *yo creo* has thus undergone a process of pragmaticalization together with the overt pronoun, with the speaker subject expressing his/her opinion, even though this process is arguably not yet complete. The corresponding combination Ø+*creo* has undergone a parallel process, with the resulting expression encoding epistemic stance, through which the speaker mitigates his/her commitment towards the truth of the proposition (see hedging and the explanation given by Nuyts 2001 in our fn. 11). Evidence in favor of this view comes from the relevance of what we have annotated as 'function of the complement clause': the rate of 1st person singular overt subject pronouns with *creo* is higher in Spanish when the verb takes a complement that contains an evaluative element, compared to complements in which the information can be objectively falsified. In Italian, on the other hand, we propose that *io credo* has not reached the same degree of fixation and specialization for discourse functions as its Spanish counterpart: only *credo* with a null subject has pragmaticalized into a marker of epistemic stance.

A further difference between Italian and Spanish is that the parenthetical use of verbs of cognition is connected to complementizer deletion in the former, but in the latter the complementizer can be maintained (see (17)), even though its absence is the more frequent option. The possibility of maintaining the complementizer in Spanish shows that cross-linguistically it cannot be assumed that complementizer deletion is a necessary intermediate step in the pragmaticalization cline of a verb into an epistemic marker (see e.g. Aijmer 1997: 7 for English).

Although we cannot offer a full explanation for the differences between Italian and Spanish verbs of cognition in the expression of null and overt subject pronouns in relation to (semi-)pragmaticalized expressions, we would like to discuss some possible lines of reasoning: one possible explanation could lie in the semantic and functional differences between the respective verbs in the two languages. In our data, *creo* is in general far more frequent in the Spanish spoken data than *pienso* (1,490 analyzable tokens of *creo* vs. 90 tokens of *pienso*). Thus, as expected, the highly frequent verb form shows more signs of fixation and reduction towards an opinion or epistemic marker than the low frequency verb form. Given the role that frequency plays in the routinization of forms, which in turn is crucial for the process of pragmaticalization (see e.g. Detges & Waltereit 2016), these data fit with the special role of *(yo) creo* in Spanish. In Italian, on the other hand, neither *credo* nor *penso* show a clear advantage in terms of frequency, with 420 occurrences of the former and 610 of the latter.[28]

[27] The term 'semi-fixed' refers to the assumption that the pragmaticalization process is arguably not complete, as stated below. We do find cases of *creo* with a null subject to express opinions, but the use of overt *yo* in this context is more frequent, even in parenthetical clauses.

[28] Note that the frequency of our verbs of cognition depends on the type of corpus used. Without going into statistical details, in the general corpora and databases available for Spanish and Italian we find the following approximate numbers:

Spanish: The word frequencies in the *Corpus del Español: NOW* (search for frequency of the verb forms "creo_v" and "pienso_v" <14.09.2025>) are:

creo 1,208,990 *pienso* 155,812

The *Corpus del Español del Siglo XXI (CORPES)* of the RAE has the following absolute numbers (for Spain only):

creo 50,875 *pienso* 844

The SUBTLEX-ES database shows a similar relative frequency count:

creo: 71,358 *pienso:* 6,254

Italian: The BADIP, which also contains the LIP, is currently not available.

Note also that both Spanish *(yo) creo* and Italian *credo*, usually translated into English as either 'I think' or 'I believe', express subjective descriptions, evaluations or judgments, whereas Sp. *pienso*/It. *penso* 'I think' can have an additional semantic component of mental calculation or activity (like English *think,* cf. It. *Penso che mi comprerò una macchina* 'I'm thinking of buying myself a car.'). Be that as it may, it might be the case that there are cross-linguistic differences with regard to the modal strength of these different verbs of cognition, i.e. not only between Sp. *pienso*/It. *penso* and Sp. *(yo) creo*/It. *credo*, which is probably also mirrored in their cross-linguistically different frequencies, but also between the cognates Sp. *(yo) creo*/It. *credo* themselves.[29]

Another explanation might also be found in the morpho-phonological properties of the 1st person singular full pronoun (Sp. *yo* / It. *io*), which appears frequently in Spanish but almost never in Italian with parenthetical uses of verbs of cognition. For example, the Spanish pronoun is the short monosyllabic [ʝo] or [j̯o] consisting of a consonant and a full vowel, but the Italian strong pronoun is bi-syllabic and contains a hiatus that, furthermore, has stress on the first – lower – vowel ['i:.o]. It is well known that even strong (non-clitic) pronouns, being functional elements, can be prosodically reduced. Therefore, it can be assumed that vowel reduction, as it occurs in reduced parentheticals[30] in English like *y'know* (Schiffrin 1987) or French *t'sais* (cf. Schneider 2007) is more easily applicable to Spanish *yo* than to Italian *io*. Moreover, prosodically integrated (reduced) parentheticals probably do not contain a phase edge, which means that there is no mechanism that would rescue the underlying (floating) vowels (for Spanish) by additional syllable space for the combination pronoun+V.

This is an idea discussed within the CVCV framework by Newell & Scheer (2025): Functional elements contain floating segments, which can be linked in certain environments, e.g. when stressed. In its underlying representation Spanish *yo* is probably just a monosyllabic CV with the consonant linked to C and the vowel floating, and thus the vowel can be reduced when unstressed (as it is in parentheticals) and in preverbal position (in postverbal position the vowel would be finally linked to V and would thus be realized). Italian *io* is already underlyingly bi-syllabic, i.e. CVCV [i.o] and is thus already a minimal word, which cannot reduce. In Italian, then, stress adds a further CV to its right, which actually lengthens the lower first vowel with the result CVCVCV [i:.o]. The first vowel, which can never be reduced to a glide (*[jo] vs. [iǫ]) is also underlyingly linked to a V-position (and thus does not float). However, a more thorough prosodic and phonological analysis of the sequences *yo creo* and *io credo* would be necessary to fully assess this line of reasoning.[31]

In the SUBTLEX-IT database, *credo* is more frequent than *penso* (frequency count *credo*: 93,087, *penso*: 62,897), but the overall frequency of forms of the lemma *pensare* is more frequent than that of the lemma *credere*. The absolute frequency in the *Repubblica*-Corpus shows a similar relationship:

pensare	201,875	*penso*	26,740
credere	122,609	*credo*	55,622

[29] Stephenson (2007b: 59ff) e.g. notes that in English there are semantic differences between the cognitive verbs *think*, *believe* and *find*, and that the latter two have more "fine-grained lexical semantics." *Find* seems to require a personal experience of the attitude holder (the subject), whereas *believe* does presuppose that there is no such direct experience. *Think* is neutral in this regard.

[30] Heine & Kaltenböck (2021) analyze comment clauses that are prosodically separated from the main clause and argue that the prosodic "comma intonation" is what defines them as belonging to the level of thetical grammar (Kaltenböck et al. 2011); our parentheticals, as noted above, are integrated parentheticals (see again Dehé & Wichmann 2010 for a prosodic distinction between main clauses, comment clauses and integrated parentheticals in English; also fn. 16).

[31] See also Portuguese *eu* /eu/, Romanian *eu* /jeu/ which both satisfy word minimality, a fact that would also be compatible with the data presented in Section 2. These tentative insights stem from personal

Yet another possibility would be to consider the syntactic nature of (pronominal) subjects in the two languages. Although both are consistent *pro*-drop languages, which allow 'free' subject verb inversion, it has been repeatedly noted that there are syntactic differences: Spanish allows more freedom with respect to postposed subjects (e.g., the VSO order) than Italian (see Ordóñez 2000; Belletti 2004). Furthermore, the A- vs. A'-status and the exact position of preverbal subjects have been debated in both languages (see Ordóñez & Treviño 1999; Cardinaletti 2004, among many others). It might be that fine-grained differences in the subject position between the two languages also have consequences for the process of pragmaticalization that targets the combination of subject+verb.

Before concluding the discussion, one potential confounding factor must be addressed: it could be argued that we are not dealing with cross-linguistic differences but, rather, differences that are due to the type of data. Although both corpora contain spoken data, the type of speakers is different, COSER being a corpus whose methodology is based on more traditional dialectology, containing recordings of older speakers from rural regions, while KIParla comprises data from speakers of different ages, education levels, and different origins within Italy. While the differences in data types might indeed have an influence, some facts indicate that these differences are not what determines the patterns in the use of 1st person singular verbs of cognition and their subjects in the two languages: While there are, to our knowledge, no previous specific studies on the topic in Italian, there are various preexisting studies on (different dialects of) Spanish (see §2). The studies on Spanish contain a variety of datasets and speakers and the results in terms of frequency are similar: They report a particularly high frequency of overt *yo* with *creo*. In Italian, the data show that *pro*-drop applies surprisingly consistently despite its regional (micro-)variation. In the study by Herbeck (2021), the database is the PRESEEA corpus which contains speakers of different ages and with different education levels. The results from that database are confirmed by the present study, despite the difference in speaker types. This indicates that the results with respect to the high frequency of *yo* with verbs of cognition in Spanish are not due to the variable of age. However, future comparative studies with datasets that are fully balanced with regard to sociolinguistic variables will hopefully shed further light on this issue.

5. First steps towards a syntactic analysis: Subject expression in formulaic sequences

As stated previously, further research, particularly on the prosodic and morpho-phonological properties of our formulaic expressions in Spanish and Italian, is required in order to identify the reasons for the differences in expression of the subject. Many factors are certainly at play, as we have outlined in Section 4. Our proposal for a possible analysis is therefore necessarily only a first step towards the interpretation of a development that is influenced by further language-specific factors (like morpho-phonology, frequency, semantic idiosyncrasy etc.).

We will restrict our tentative analysis to Sp. *creer/* It. *credere* in the 1st person singular. In Italian and Spanish, there are two major syntactic differences in this type of verb. First, in Italian, but not in Spanish, the complement clause of *credere* is usually in the subjunctive, at least with respect of the written norm that also influences the spoken language, although in spoken language the complement is often in the indicative

discussion with Heather Newell and need to be further elaborated on the basis of our data in a future paper.

(Wandruszka 1991; Squartini 2010; Mari & Portner 2021). In Spanish, the complement of *creer* is in the subjunctive under negation, i.e. the subjunctive is triggered by an operator and is not lexically chosen by the verb.[32] Second, Italian allows complementizer deletion with *credere* (as does English, but under different conditions, including the realization of the subjunctive; see Giorgi & Pianesi 2005), whereas Spanish does not, which might explain examples like (17) above where the complementizer is maintained even in parenthetical position. Interestingly, it is only with 1st person Italian *credo* (and perhaps also imperfect *credevo* 'I thought') that complementizer deletion is allowed for all speakers in syntactic contexts where it would not be allowed in the other persons (e.g. **Crede*/*Credo Paolo abbia*$_{SBJ}$ *telefonato* '*He thinks/I think Paolo called', with an overt preverbal subject in the embedded clause; see Giorgi & Pianesi 2005: 113).

Both Spanish and Italian allow left dislocation, e.g. of subjects or other arguments, which might already be the first step towards the positional flexibility of our formulaic expression. The grammaticalization process that has been proposed for English by Thompson & Mulac (1991), among others, can easily be applied to the Romance languages under discussion here.[33]

In what follows, we use Italian as a model as this allows us to take complementizer deletion and the mood of the complement clause into account. A syntactic pathway toward the grammaticalization of *credo* in Italian could look as shown in (20):

(20) a. ***Credo*** ***che*** *Gianni lavori (lavora) all'ospedale di Mestre.*
think.1sg **that** G. works.SBJ (works.IND) at-the hospital of M.
b. *Gianni* ***credo che*** *lavori (lavora) all'ospedale di Mestre.*
(subject topicalization)
c. ***Credo*** *lavori all'ospedale di Mestre.*
(bridge verb/C-deletion (Italian))
d. *Gianni* ***credo*** *lavori (lavora) all'ospedale di Mestre.*
(subject topicalization with bridge verb/C-deletion (Italian))
e. *Gianni,* ***credo,*** *lavora all'ospedale di Mestre.*
(parenthetical position, no subjunctive)
f. *Gianni lavora all'ospedale di Mestre*, ***credo***.
(parenthetical position, no subjunctive)

[32] For mood choice under *believe*-verbs, Mari & Portner (2021) claim that in spoken Italian the choice of the subjunctive is semantically driven (see especially their example (6) from Italian): When the epistemic statement in the complement is linked to the mental state of the subject (attitude holder) the indicative can be chosen and a personal opinion is expressed (which cannot be challenged as such); if it is linked to the discourse context (common ground) the subjunctive must be chosen and the content of the complement is proposed to be added to the common ground (truth-conditions). This correlation seems to be parallel to the distinction between subjective and objective epistemic modality mentioned in Section 2.3 and differs from the distinction proposed in (1) and (2). Note that the objective reading should therefore be excluded for Spanish *creer*, since it does not take the subjunctive in a non-negated root context.

[33] Note that this – theoretical – model was developed independently from Thompson & Mulac (1991). The latter posit two factors conditioning a similar (but not identical) development in English: First, the epistemic interpretation of the main verb, and second, the topicality (or rather, at-issueness) of the complement clause. However, Italian has information structural syntactic devices different from English and can furthermore use the subjunctive for the expression of syntactic dependencies. In response to an anonymous reviewer concerning the acceptability of (20)i: We have, indeed, all kinds of examples for different positional and scope properties of prosodically integrated *credo* in Italian, even in this position; an instance of the use of *credo* between a preposition and its complement is, for example, *una [pizza] bianca con credo panna* 'a white [pizza] with I think cream' (KIParla TOI012).

g. *Gianni lavora **credo** all'ospedale di Mestre*.
(parenthetical position, no subjunctive)
h. *Gianni lavora all'ospedale **credo** di Mestre*.
(parenthetical position, no subjunctive)
i. *Gianni lavora all'ospedale di **credo** Mestre*.
(parenthetical position, no subjunctive => adverbial modifier?)

We will concentrate here on the semi-fixed structures and consider only the integrated parentheticals in formulaic expressions. Although we have not yet discussed this in detail, *credo* (and also Sp. *(yo) creo*), at least in the final stages of the development, are reduced parentheticals that are prosodically integrated in the phrase. They also contribute to the meaning of the clause, since they function as epistemic markers, i.e. the host clause of the integrated parentheticals is not a canonical assertion but is epistemically marked by subjective certainty (cf. §2.3). Furthermore, the scope of the epistemic marker can vary, and can refer either to the whole host clause (in intermediate or final position) or to parts (constituents) of it: In (20)g it scopes over the PP *all'Ospedale di Mestre,* in (20)h it scopes over the PP *di Mestre,* and in (20)i it scopes over the toponym *Mestre.* It also might scope over adverbials within the clause, which means that these are certainly not completely "outside the clause" (as Kaltenböck et al. 2011 would say for thetical grammar; see the title of Kaltenböck et al. 2016) but quite integrated, both semantically and prosodically.[34] Integration also occurs at the level of information structure, since, as Steinbach (2007: 79) states (for German), "the integrated parentheticals form a single focus-background structure with the host."

The main question is the status of the formulaic expression in relation to the host clause. Various syntactic solutions have been proposed for these epistemic markers, but they all create connections between the marker and the left periphery. Badan (2021) argues that they are phrases in the specifier of a functional projection (IP/CP) agreeing with a speech act head encoding the Addressee feature, an approach that is consistent with Cinque's (1999) adverbial hierarchy. Giorgi & Pianesi (2005) interpret *credo* as a functional head, whereas for Newmeyer (2014) the corresponding English form *I think* is lexical (probably similar to adverbs).

Other proposals for similar elements e.g. in German have claimed that the argument structure of the verb of cognition is also still maintained for the integrated parenthetical and that there is a free variable for the required argument in the complement (see Steinbach 2007 for German verb initial parentheticals like *glaub ich* 'believe I'), which is linked to the host clause by non-canonical licensing (cf. Steinbach 2007: 79–80).

Hedberg & Elouazizi (2015) analyze the parenthetical use of the corresponding English forms of our formulaic expressions in sentence medial position as CP adjuncts, which can either right- or left-adjoin (more frequently the latter), depending on their scope properties. In any case, they must c-command the material in their scope and this material is associated with focus. In sentence final position the parenthetical is ambiguous since it usually has focus over the whole host clause, but it could also have focus on the final constituent. While Steinbach (2007) assumes a free variable for the missing argument, Hedberg & Elouazizi (2015: 128) propose a structure for the CP in

[34] Of course, a separate study is required for the prosodic properties, which could not only help to identify the categorial status of our markers at phase edges or otherwise, but could also shed light on the question whether the overt Spanish subject pronoun *yo* and the overt Italian subject pronoun *io* are different in prosodic structure and thus also in their ability to undergo vowel reduction, as discussed in Section 4.

which the complement of the verb of cognition is a trace bound by an operator sitting in [Spec, CP].

Whether we interpret our formulaic expressions as constituents in specifier positions of a dedicated epistemic head (à la Cinque 1999) or as scope-taking adjuncts, as proposed by Hedberg & Elouazizi (2015), or as – functional or lexical – heads is a question parallel to the discussion of adverbs and adverbials in general. Whatever the correct explanation may be, our epistemic formulaic expressions are distinct from adverbs like Sp. *probablemente*/It. *probabilmente* 'probably' or Sp. *siguramente*/It. *sicuramente* 'surely', which could also be expressions of objective epistemic modality: In our formulaic expressions it is the presence of a subject – be it overt or null – encoding the speaker that indisputably anchors the epistemic judgment to the mental state of the speaker.

In the discussion in Section 2.3, we have already seen that the 1st person singular subject in combination with a present tense verbal form of an assertive verb can become an epistemic marker. The differences in subject realization between Italian and Spanish, however, still need further investigation, as they are not explained either by our first tentative analysis nor by the literature. The more frequent realization of these overt pronominal subjects in Spanish vs. their very rare realization in Italian seems to be dependent on a set of grammatical, lexical and language-dependent properties which we could only touch upon in this paper.

6. Conclusions and issues for future research

In this paper, we have investigated subject realization in one specific context in two Romance *pro*-drop languages: the use of null and overt 1st person singular subject pronouns with the verbs of cognition *creer*/*credere* and *pensar*/*pensare*. We started from the observations in previous literature that Spanish verbs of cognition, especially *(yo) creo*, have a high rate of overt subject pronoun realization in comparison to other verbs and that cross-linguistic differences might exist, even between closely related consistent *pro*-drop languages. We therefore examined verbs of cognition and their 1st person singular subjects in spoken Spanish, comparing the observed patterns to those of spoken Italian. The data from the COSER corpus confirm the tendencies in previous studies on spoken Spanish, in that high subject expression frequencies can be observed with *(yo) creo*. We have furthermore delved into the specific context of parentheticals and have seen that these are also frequently used with overt *yo*. The high use of overt pronouns seems to correlate with a higher degree of subjectivity, which has led us to the conclusion that Ø+*creo* and *yo*+*creo* have partly developed into markers of specialized functions – argumentative/opinion vs. mitigating/epistemic (see Schneider 2007; De Saeger 2008; Aijón Oliva & Serrano 2010; Vázquez & Rozas & Enrique Ovando 2020; Herbeck 2021). This means that the *pro*-drop property yields two potential and separate clines of pragmaticalization processes, targeting verb forms with or without an overt subject pronoun. In the Italian KIParla corpus, on the other hand, the results are *a priori* surprising because they show the exact opposite pattern to Spanish verbs of cognition: the frequency of overt *io* is especially low with *credo* (9%) and *penso* (14%). We have tentatively suggested that the differences observed between Italian and Spanish subject pronoun realization are related to the existence of semi-fixed patterns of pronoun+*think* in the Spanish data and its absence in Italian, where Ø+*think* is the only parenthetical epistemic marker. However, we have also noted that future studies will need to draw on datasets that are more balanced with respect to sociolinguistic factors in order to find out to what extent these factors should also be taken into account.

Furthermore, a thorough morpho-phonological investigation of our data is needed in order to establish their prosodic and hence also syntactic integration into the main clause. Since the process of grammaticalization is ongoing and may also be more or less advanced depending on the language and the specific verb of cognition involved, there might be also different grades of integration (see also the distinction made for English, cf. fn. 30). Prosodic properties and clausal integration might also be a decisive factor for overt 1st person singular subject pronouns, since they could also be connected to the status of the formulaic expression as an interactional comment or an epistemic mitigator. We leave this research for future studies.[35]

We can thus conclude that null and overt subject pronouns, apart from fulfilling the standardly observed functions of contrast, (topic) shift, and emphasis, among others, can have crucial functions in the marking and management of speaker perspective. Together with particular verb forms, they can have an epistemic, mitigating, or subjectivizing function. However, this is a particular property of the 1st person singular subject, which is semantically different from 2nd or 3rd person, since it refers to the speaker who has direct access to his or her mental state.

Peter Herbeck
Universität Wien
Institut für Romanistik
Spitalgasse 2
Campus (Hof 8)
1090 Vienna
Austria
peter.herbeck@univie.ac.at

Andrea Miglietta Universität Wien
Institut für Romanistik
Spitalgasse 2
Campus (Hof 8)
1090 Vienna
Austria
andrea.miglietta@univie.ac.at

Eva-Maria Remberger Universität Wien
Institut für Romanistik
Spitalgasse 2
Campus (Hof 8)
1090 Vienna
Austria
eva-maria.remberger@univie.ac.at

References

Aijmer, K. (1997). *I think*: An English modal particle, in T. Swan & O. J. Westvik (eds.), *Modality in Germanic languages*: *Historical and comparative*

[35] Note that in the relevant chapter of the *Grande Grammatica Italiana di Consultazione* comment clauses and integrated (verbal) parentheticals are not distinguished (cf. Borgato & Salvi 2001).

perspectives, Berlin: De Gruyter, pp. 1-47. https://doi.org/10.1515/9783110889932.1

Aijón Oliva, M. Á. & M. J. Serrano (2010). El hablante en su discurso: Expresión y omisión del sujeto de *creo*. *Oralia* 13, pp. 7-38. https://doi.org/10.25115/oralia.v13i.8100

Anvari, A., M. Maldonado & A. Sora Ruiz (2019). The puzzle of reflexive belief constructions in Spanish, in: M. T. Espinal et al. (eds.), *Proceedings of Sinn und Bedeutung 23,* vol. 1, Bellaterra, Universitat Autònoma de Barcelona, pp. 57-74.

Badan, L. (2021). Verb-based discourse markers in Italian *Guarda, vedi, guarda te, vedi te,* in D. Van Olmen & J. Šinkūnienė (eds.), *Pragmatic markers and peripheries*. Amsterdam/Philadelphia, John Benjamins, pp. 144-170. https://doi.org/10.1075/pbns.325.05bad

Barbosa, P. (1995). *Null subjects*. Doctoral dissertation, Massachusetts Institute of Technology.

Bary, C. & E. Maier (2021), The landscape of speech reporting. *Semantics & Pragmatics* 14(8), pp. 1-54. https://doi.org/10.3765/sp.14.8

Belletti, A. (2004). Aspects of the low IP area, in L. Rizzi (ed.), *The structure of CP and IP*. Oxford, NY, Oxford University Press, pp. 16-51. https://doi.org/10.1093/oso/9780195159486.003.0002

Bessett, R. M. (2023). A cross dialectal comparison of first person singular subject pronoun expression in Southern Arizona and Southeast Texas, in P. Posio & P. Herbeck (eds.), *Referring to discourse participants in Ibero-Romance languages.* Berlin, Language Science Press, pp. 25-39. https://doi.org/10.15460/repohh/sub.2024100009

Biasio, M. & Del Fante, D. (2024). Once known, always known. Turn-final *sai* in North-East regional Italian. *Folia Linguistica* 59(1), pp. 207-253. https://doi.org/10.1515/flin-2024-2001

Borgato, G. & G. Salvi (2001). Le frasi parentetiche, in L. Renzi, G. Salvi, & A. Cardinaletti (eds.), *Grande grammatica italiana di consultazione. Vol 3. Tipi di frase, deissi, formazione delle parole* 2nd ed., Bologna, Il Mulino, pp. 165-174.

Cameron, R. (1992). *Pronominal and Null Subject Variation in Spanish: Constraints, Dialects, and Functional Compensation*. Doctoral dissertation, University of Pennsylvania at Philadelphia.

Cardinaletti, A. (2004), Toward a Cartography of Subject Positions, in L. Rizzi (ed.), *The structure of CP and IP*. Oxford, NY, Oxford University Press, pp. 115-165. https://doi.org/10.1093/oso/9780195159486.003.0005

Carvalho, A. M., R. Orozco & N. Lapidus Shin (eds.) (2015). *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, Georgetown University Press.

Cerruti M. & Ballarè S. (2023). Sociolinguistic variation, or lack thereof, in the use of the Italian subjunctive. *Sociolinguistica* 37(1), pp. 75-93. https://doi.org/10.1515/soci-2022-0023

Cinque, G. 1999. *Adverbs and functional projections: A cross-linguistic perspective*, New York, Oxford University Press. https://doi.org/10.1093/oso/9780195115260.001.0001

Contemori, C. & Di Domenico, E. (2021). Microvariation in the division of labor between null- and overt-subject pronouns: The case of Italian and Spanish. *Applied Psycholinguistics 42*, 1028. https://doi.org/10.1017/S0142716421000199

Davidson, B. (1996). 'Pragmatic weight' and Spanish subject pronouns: The pragmatic and discourse uses of 'tú' and 'yo' in spoken Madrid Spanish. *Journal of Pragmatics* 26, pp. 543-565 https://doi.org/10.1016/0378-2166(95)00063-1

Dehé, N. & A. Wichmann (2010), The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary**.** *Functions of Language* 17(1), pp. 1-28. https://doi.org/10.1075/fol.17.1.01deh

Delbecque, N. (2006). *Ya*: aclaración cognitiva de su uso y función. *Revista española de lingüística (RSEL)*, pp. 43-71.

De Prada Pérez, A. (2015). First person singular subject pronoun expression in Spanish in contact with Catalan, in A. M. Carvalho, R. Orozco & N. Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish. A Cross-Dialectal Perspective*. Washington D.C, Georgetown University Press, pp. 121-142.

De Saeger, B. (2008). Speaker involvement through cognition verbs in Spanish, in P. De Brabanter & P. Dendale (eds.), *Commitment – Belgian Journal of Linguistics* 22, pp. 63-81. https://doi.org/10.1075/bjl.22.04sae

Detges, U. & R. Waltereit (2016). Grammaticalization and pragmaticalization, in S. Fischer & C. Gabriel (eds.), *Manual of grammatical interfaces in Romance*. Berlin: De Gruyter, pp. 635-658. https://doi.org/10.1515/9783110311860-024

Diewald, G. (2011). Grammaticalization and pragmaticalization, in B. Heine & H. Narrog (eds.), *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press, 450–461. https://doi.org/10.1093/oxfordhb/9780199586783.013.0036

Di Domenico, E., Baroncini, I. & Capotorti, A. (2020). Null and overt subject pronouns in topic continuity and topic shift: An investigation of the narrative productions of Italian Natives, Greek Natives and near-native second language speakers of Italian with Greek as a first language. *Glossa (London) 5*, 35. https://doi.org/10.5334/gjgl.1009

Diewald, G. (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49(2), pp. 365-390. https://doi.org/10.1515/ling.2011.011

Enghels, R. (2018). On the grammaticalization of the deverbal epistemic pragmatic marker *sabes*: a study in recent language change. *Revue Roumaine de Linguistique – Romanian Review of Linguistics* 63(4), pp. 341-360.

Enríquez, Emilia V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid, Instituto Miguel de Cervantes.

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

Escandell-Vidal, M. V. & M. Leonetti (2009). La expresión del *verum focus* en español. *Español actual* 92, pp. 11-46.

Frana, I. (2025). Bias and evidential meaning in questions. The view from Italian. Talk at the *Jezik & Linguistics Colloquia series* at the University of Nova Gorica, 6.6.2025.

Frascarelli, M. (2007). Subjects, topics and the interpretation of referential *pro*: An interface approach to the linking of (null) pronouns. *Natural Language and Linguistic Theory* 25, pp. 691-734. https://doi.org/10.1007/s11049-007-9025-x

Giorgi, A. & F. Pianesi. (2005). *Credo* (I Believe): Epistemicity and the syntactic representation of the speaker. *Working Papers in Linguistics,* Venice, University of Venice.

Givón, T. (1982). Evidentiality and epistemic space. *Studies in Language* 6(1), pp. 23-49. https://doi.org/10.1075/sl.6.1.03giv

Haegeman, L. & Hill, V. (2013). 'The Syntacticization of Discourse', in R. Folli, C. Sevdali & R. Truswell (eds.), *Syntax and its Limits*, Oxford, Oxford University Press, 370–390. https://doi.org/10.1093/acprof:oso/9780199683239.003.0018

Heine, B. & G. Kaltenböck (2021). From clause to discourse marker: on the development of comment clauses. *Language Sciences* 87, 101400. https://doi.org/10.1016/j.langsci.2021.101400

Hennemann, A. (2016). A cognitive-constructionist approach to Spanish *creo Ø* and *creo yo* '[I] think'. *Folia Linguistica* 50(2), pp. 449-474. https://doi.org/10.1515/flin-2016-0017

Hedberg, N. & N. Elouazizi (2015). Epistemic parenthetical verb phrases: C-command, semantic scope and prosodic phrasing, in S. Schneider, J. Glikman & M. Avanzi (eds.), *Parenthetical verbs*, Berlin, De Gruyter, pp. 121-136.

Herbeck, P. (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. https://doi.org/10.16995/glossa.5873

Herbeck, P. (2022). Variable First Person Singular Subject Expression in Spoken Valencian Catalan. *Catalan Journal of Linguistics* 21, pp. 129-157. https://doi.org/10.5565/rev/catjl.380

Holmberg, A. (2005). Is There a Little pro? Evidence from Finnish. *Linguistic Inquiry* 36(4), pp. 533-564. https://doi.org/10.1162/002438905774464322

Holmberg, A., A. Nayudu & M. Sheehan (2009). Three partial null-subject languages: a comparison of Brazilian Portuguese, Finnish and Marathi. *Studia Linguistica* 63(1), pp. 59-97. https://doi.org/10.1111/j.1467-9582.2008.01154.x

Hooper, J. B. 1975. On assertive predicates, in J. P. Kimball (ed.), *Syntax and Semantics*, Vol. 4, New York, Academic Press, pp. 91-124. https://doi.org/10.1163/9789004368828_005

Kaltenböck, G., B. Heine & T. Kuteva (2011). On thetical grammar. *Studies in Language* 35(4), pp. 848-893. https://doi.org/10.1075/sl.35.4.03kal

Kaltenböck, G., E. M. Keizer, A. Lohmann (2016). *Outside the clause: form and function of extra-clausal constituents*, Amsterdam/Philadelphia, John Benjamins. https://doi.org/10.1075/slcs.178

Kocher, A. (2017). From Verum to Epistemic Modality and Evidentiality: On the Emergence of the Spanish Adv+C Construction. *Journal of historical linguistics* 7 (1/2), pp. 77-110. https://doi.org/10.1075/jhl.7.1-2.04koc

Koev, T., C. Bill & M. Mohammadi (2021). *Believe* is strong: Experimental experience from hedging, in P. G. Grosz, L. Martí, H. PearsonY. Sudo & S. Zobel (eds.), *Proceedings of Sinn und Bedeutung* 25, University College of London and Queen Mary University of London, pp. 497-514.

Kratzer, A. (1981). The Notional Category of Modality, in H.-J. Eikmeyer & H. Rieser (eds.), *Words, Worlds, and Contexts: New Approaches in Word Semantics*, Berlin, De Gruyter, pp. 38-74. https://doi.org/10.1515/9783110842524-004

Kratzer, A. (1991). Modality, in A. v. Stechow & D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch zeitgenössischer Forschung*, Berlin, De Gruyter, pp. 639-650. https://doi.org/10.1515/9783110126969.7.639

Krifka, M. (2023). Layers of assertive clauses: propositions, judgements, commitments, acts, in J. M. Hartmann & A. Wöllstein (eds.), *Propositionale Argumente im Sprachvergleich / Propositional Arguments in Cross-Linguistic Research: Theorie und Empirie / Theoretical and Empirical Issues,* Tübingen, Narr, pp. 116-183.

Lasersohn, P. (2005). Context Dependence, Disagreement, and Predicates of Personal Taste. *Linguistics and Philosophy* 28(6), pp. 643-686. https://doi.org/10.1007/s10988-005-0596-x

Lehmann, C. (2002 [1982]). *Thoughts on grammaticalization,* 2nd revised edition, Erfurt, Universität Erfurt [Arbeitspapiere für Sprachwissenschaft der Universität Erfurt, vol. 9].

Lapidus Shin, N. (2014). Grammatical complexification in Spanish in New York: 3SG pronoun expression and verbal ambiguity. *Language Variation and Change* 26, pp. 303-330. https://doi.org/10.1017/S095439451400012X

Leonetti, M. (2008). Specificity in Clitic Doubling and in Differential Object Marking. *Probus* 20, pp. 35-69. https://doi.org/10.1515/PROBUS.2008.002

Lorusso, P., Caprin, C., & Guasti, M. T. (2005). *Overt Subject Distribution in Early Italian Children*. 29th Annual Boston University Conference on Language Development-BUCLD 29, Online Proceedings Supplement, Boston UK.

Lyons, J. (1977). Semantics, vol. 2. Cambridge, Cambridge University Press.

Mari, A. & Portner, P. (2021), Mood variation with belief predicates: Modal comparison and the raisability of questions. *Glossa: a journal of general linguistics* 6(1), 132, pp. 1-51. https://doi.org/10.16995/glossa.5726

Mayol, L. (2010). Contrastive pronouns in null-subject Romance languages. *Lingua* 120(10), pp. 2497-2514. https://doi.org/10.1016/j.lingua.2010.04.009

Miyagawa, S. (2022). *Syntax in the tree tops*. Cambridge Massachusetts: MIT Press. https://doi.org/10.7551/mitpress/14421.001.0001

Morales, A. (1997). La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico. *Hispania* 80(1), pp. 153-165. https://doi.org/10.2307/345995
DOI: http://doi.org/10.2307/345995

Newell, H. & T. Scheer (2025). Syntactic and phonological underspecification in function words: The case of English pronoun reduction. Ms.

Newmeyer, F. (2014), Parentheticals and the grammar of complementation, in S. Schneider, J. Glikman & M. Avanzi (eds.), *Parenthetical verbs*, Berlin, De Gruyter, pp. 13-38. https://doi.org/10.1515/9783110376142-003

NGLE = Real Academia Española (2009). *Nueva gramática de la lengua Española*. Madrid, Espasa. <https://www.rae.es/gramática/>

Nuyts, J. (2001). Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics* 33, pp. 383-400. https://doi.org/10.1016/S0378-2166(00)00009-6

Ocampo, F. A. (2006). Movement towards discourse is not grammaticalization: The evolution of *claro* from adjective to discourse particle in spoken Spanish, in N. Sagarra & A. J. Toribio (eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium: Cascadilla Press Proceedings Project.* Online: http://www.lingref.com/cpp/hls/9/paper1388.pdf [accessed 11/9/2025)

Ordóñez, F. (2000). *The clausal structure of Spanish: a comparative study*. New York, Garland.

Ordóñez, F. & E. Treviño (1999). Left dislocated subjects and the *pro*-drop parameter: a case study of Spanish. *Lingua* 107, pp. 39-68. https://doi.org/10.1016/S0024-3841(98)00020-5

Orozco, R. (2015). Pronominal variation in Colombian Costeño Spanish, in A. Carvalho, R. Orozco & N. Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish. A Cross-Dialectal Perspective*, Washington D.C., Georgetown University Press, pp. 17-58.

Posio, P. (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), pp. 777-798. https://doi.org/10.1016/j.pragma.2010.10.012

Posio, P. (2013). The expression of first-person-singular subjects in spoken Peninsular Spanish and European Portuguese: Semantic roles and formulaic sequences. *Folia Linguistica* 47(1), pp. 253-291. https://doi.org/10.1515/flin.2013.010

Posio, P. (2014). Subject expression in grammaticalizing constructions: The case of *creo* and *acho* 'I think' in Spanish and Portuguese. *Journal of Pragmatics* 63, pp. 5-18. https://doi.org/10.1016/j.pragma.2013.07.001

Posio, P. (2015). Subject pronoun usage in formulaic sequences: Evidence from Peninsular Spanish, in A. M. Carvalho, R. Orozco & N. Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington, Georgetown University Press, pp. 59-78.

Posio, P. (2018). Properties of pronominal subjects, in K. L. Geeslin (ed.), *The Cambridge Handbook of Spanish linguistics*. Cambridge: Cambridge University Press, pp. 286-306. https://doi.org/10.1017/9781316779194.014

Posio, P. & P. Herbeck (eds.) (2023). *Referring to discourse participants in Ibero-Romance languages.* Berlin, Language Science Press. https://doi.org/10.15460/repohh/sub.2024200013

Pountain, C. J. (2015). *Que*-deletion: the rise and fall of a syntactic fashion, in F. Dubert García, G. Rei-Doval & X. Sousa (eds.), *En memoria de tanto miragre*. *Estudos dedicados ó profesor David Mackenzie*, Santiago de Compostela, Universidade de Santiago de Compostela, Servizo de Publicacións e Intercambio Científico, pp. 143-159.

Quirk, R., S. Greenbaum, G. Leach & I. Svartvik (1972). *A Grammar of Contemporary English.* London, Longman.

Rizzi, L. (1982). *Issues in Italian syntax*. Dordrecht: Foris Publications. https://doi.org/10.1515/9783110883718

Scherger, A.-L., & Schmitz, K. (2019). The Role of Age in the Domain of Subject Expression in Young Italian-German Bilinguals. *Heritage Language Journal 16*. https://doi.org/10.46538/hlj.16.1.5

Schiffrin, D. (1987). Information and participation: *y'know* and *I mean*, in: D. Schiffrin, *Discourse Markers*, Cambridge, Cambridge University Press, pp. 267-311. https://doi.org/10.1017/CBO9780511611841.010

Schmitz, K. & Müller, N. (2008). Strong and clitic pronouns in monolingual and bilingual acquisition of French and Italian. *Bilingualism: Language and Cognition* 11 (1), pp. 19-41. https://doi.org/10.1017/S1366728907003197

Schmitz, K., M. Patuto & N. Müller (2011). The null-subject parameter at the interface between syntax and pragmatics: Evidence from bilingual German–Italian, German–French and Italian–French children. *First Language 32*, 238. https://doi.org/10.1177/0142723711403880

Schmitz, K., L. Di Venanzio & A.-L. Scherger (2016). Null and overt subjects in Italian and Spanish heritage speakers in Germany. *Lingua 180*, 123. https://doi.org/10.1016/j.lingua.2016.04.004

Schneider, S. (2007). *Reduced Parenthetical Clauses as Mitigators: A corpus study of spoken French, Italian and Spanish*. Amsterdam, John Benjamins. https://doi.org/10.1075/scl.27

Speas, M. (1994). Null arguments in a Theory of Economy of Projection. *University of Massachusetts Occasional Papers in Linguistics* 17, pp. 179-208.

Speas, P. & C. Tenny (2003). Configurational properties of point of view roles, in A. M. Di Sciullo (ed.), *Asymmetry in grammar* 1, Amsterdam: John Benjamins, pp. 315-344. https://doi.org/10.1075/la.57.15spe

Squartini, M. (2010). Mood in Italian, in B. Rothstein & R. Thieroff (eds.), *Mood in the languages of Europe*, Amsterdam/Philadelphia, John Benjamins, pp. 237-250. https://doi.org/10.1075/slcs.120.13squ

Steinbach, Markus (2007). Integrated parentheticals and assertional complements, in N. Dehé & Y. Kavalova (eds.), *Parentheticals*, Amsterdam/Philadelphia, John Benjamins, pp. 53-87. https://doi.org/10.1075/la.106.06ste

Stephenson, T. (2007a). Judge Dependence, Epistemic Modals, and Predicates of Personal Taste. *Linguistics and Philosophy* 30(4), pp. 487-535. https://doi.org/10.1007/s10988-008-9023-4

Stephenson, T. (2007b). *Towards and theory of subjective meaning* (Doctoral dissertation). Massachusetts Institute of Technology. Available at http://semanticsarchive.net/Archive/2QxMjk0O

Thompson, S. A. & A. Mulac (1991). A quantitative perspective on the grammaticalization of epistemic parentheticals in English, in E. C. Traugott & B. Heine (eds.), *Approaches to grammaticalization* 2, Amsterdam, John Benjamins, pp. 313-339. https://doi.org/10.1075/tsl.19.2.16thoo

Traugott, E. (1995). Subjectification in grammaticalization", in D. Stein & S. Wright (eds.), *Subjectivity and Subjectivisation.* Cambridge: Cambridge University Press, 37-54. https://doi.org/10.1017/CBO9780511554469.003

Travis, C. E. & R. Torres Cacoullos (2012). What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4), pp. 711-748. https://doi.org/10.1515/cog-2012-0022

Urdiales, J. M. (1973). Valores de *ya*. *Archivum* 23, 149-199.

Urmson, J. O. (1952), Parenthetical verbs. *Mind* 61, 280–496. https://doi.org/10.1093/mind/LXI.244.480

Vázquez Rozas, V. & A. Enríquez Ovando (2020). *(yo) creo* en el español de la Ciudad de México y de Galicia: Diferencias de gramaticalización, in R. Maldonado & J. de la Mora (eds.), *Evidencialidad. Determinaciones léxicas y construccionales*, Ciudad de México, Universidad Nacional Autònoma de México, pp. 199-239. https://doi.org/10.19130/iifl.evidDlyc.2021.1.462718

Wandruszka, U. (1991). Frasi subordinate al congiuntivo, in L. Renzi & G. Salvi (eds.), *Grande grammatica italiana di consultazione, vol. 2: I sintagmi verbale, aggettivale, avverbiale. La subordinazione*, Bologna, il Mulino, pp. 415‑481.

Wiltschko, M. (2021). *The Grammar of Interactional Langua*ge. Cambridge, Cambridge University Press. https://doi.org/10.1017/9781108693707

Corpora

CORPES = Real Academia Española: Corpus del Español del Siglo XXI. <https://www.rae.es/banco-de-datos/corpes-xxi>

Corpus del Español NOW = Davies, M. (2012-2019). Corpus del Español NOW (2012–2019). <https://www.corpusdelespanol.org/now/>

COSER = Fernández Ordóñez, I, (dir.) (2005-). *Corpus Oral y Sonoro del Español Rural*. <www.corpusrural.es> [June & July 2025]

KIParla = Mauri, C., S. Ballarè, E. Goria, M. Cerruti & F. Suriano (2019). KIParla corpus: a new resource for spoken Italian, in R. Bernardi, R. Navigli & G. Semeraro (eds.), *Proceedings of the 6th Italian Conference on Computational Linguistics CLiC-it.* <www.kiparla.it>

LIP = De Mauro, T., F. Mancini, M., Vedovelli & M. Voghera (1993). *Lessico di frequenza dell'italiano parlato.* Milano, Etaslibri.

PRESEEA. 2014-. *Corpus del proyecto para el estudio sociolingüístico del español de España y de América*. Alcalá de Henares: Universidad de Alcalá. [http://preseea.uah.es].

SUBTLEX-IT = Crepaldi, D., Amenta, S., Mandera, P., Keuleers, E., & Brysbaert, M. (2015). Subtitle-based word frequency estimates for Italian. <https://osf.io/zg7sc/files/osfstorage>.

SUBTLEX-ES = Cuetos, F., M. Glez-Nosti, A. Barbon & M. Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicologica* 32, pp. 133-143. < https://osf.io/xp6sz/files/osfstorage>