



Lozano, Cristóbal & Quesada, Teresa & Charatzidis, Andreas & Papadopoulou, Despina. 2023. What do corpus data reveal about anaphora resolution? Spanish vs. Greek and the Type of Topic Hypothesis. *Glossa: a journal of general linguistics* 8(1). pp. 1–38. DOI: <https://doi.org/10.16995/glossa.9883>



Open Library of Humanities

What do corpus data reveal about anaphora resolution? Spanish vs. Greek and the Type of Topic Hypothesis

Cristóbal Lozano, Universidad de Granada, ES, cristoballozano@ugr.es

Teresa Quesada, Universidad de Granada, ES, teresquesada@ugr.es

Andreas Charatzidis, Aristotle University of Thessaloniki, GR, charatzidisandreas@gmail.com

Despina Papadopoulou, Aristotle University of Thessaloniki, GR, depapa@lit.auth.gr

Anaphora Resolution (AR) is a pervasive phenomenon in natural languages. AR relates to how referring expressions (REs) (e.g., null/overt subject pronouns, and NPs) corefer with their antecedents in discourse. We use corpus methods to simultaneously compare AR in two null-subject languages (Spanish vs. Greek). We analyse a Spanish-native sample (CEDEL2 corpus, N = 341 REs analysed) and an equally-designed Greek-native sample (GLC corpus, N = 400 REs analysed), while keeping constant the text type (Chaplin narrative task), the annotation scheme (tagset), the tagging procedure, and the profile of the natives. Our corpus results reveal similarities in the way Spanish and Greek natives construct their narratives regarding the distribution of the information status of the REs (topic continuity/shift) and the distribution of characters (main/secondary) in discourse. Crucially, our two languages differ in relation to topicality (Greek capitalises on discourse topic whereas Spanish relies more on sentential topic), which leads to a different distribution in the realization of REs in discourse. These similarities and differences are accounted for by a new theoretical proposal, the Type of Topic Hypothesis (TTH), which postulates that there is a tension between discourse-topic vs. sentential-topic oriented languages. The TTH captures the idea that, while narratives are constructed in the same way in both languages, RE realization varies as a result of the discourse-topic orientation of Greek vs. the sentential-topic orientation of Spanish.

Glossa: a journal of general linguistics is a peer-reviewed open access journal published by the Open Library of Humanities. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

OPEN ACCESS



1 Introduction: factors and theories on AR (topicality)

Anaphora Resolution (AR) is the mechanism in natural languages which links referring expressions (REs) such as pronouns and noun phrases (NPs) to their antecedents. In null-subject languages (also known as pro-drop languages) both null and overt pronouns can alternate in subject position. An interesting question is which factors constrain the choice of the RE form. Significant research has focused on the hypothesis that shorter REs refer to more prominent/salient antecedents, whereas longer REs connect with less prominent/salient antecedents (Ariel 1990). Antecedent prominence/saliency is affected by multiple factors: syntactic position (e.g., Wolf et al. 2004), discourse constraints (e.g., Tantos et al. 2014), recency (e.g., Givón 1983) and topichood (e.g., Alonso-Ovalle et al. 2002). In our study, we analyze corpus data from two pro-drop languages, Spanish and Greek, to investigate the effect of topichood and topic transition on AR. Contrary to previous experimental studies on pro-drop languages, we examine the anaphoric behaviour of not only overt and null pronouns, but NPs as well.

We distinguish between sentential topic vs. discourse topic. The sentence topic is the entity around which the meaning of the sentence is constructed. In most utterances a topic is presented and a comment about it follows (Chafe 1976; Gundel 1999).¹ Previous research (e.g., Alonso-Ovalle et al. 2002) shows that sentence-topic antecedents are more prominent and are realized by shorter REs, whereas non-topic antecedents are realized by longer REs. For example, in (1) we would anticipate more attenuated forms (such as the null subject \emptyset of the verb *starts*) to refer to the sentence topic *he* because the non-topic antecedent *a princess* is less salient.

- (1) One day he_i meets a princess_j [...] and \emptyset _i starts to fall in love with her.
[English native: EN_WR_20_3_LH]³

REs are typically used to mark topic transition, i.e., topic *continuity* (maintenance of the sentence topic) and topic *shift* (change of the sentence topic). In pro-drop languages, null pronouns (\emptyset)

¹ In our corpus data analysis (cf. sections 5 and 6), the sentence topic always coincides with the grammatical subject of the sentence. This has been the typical assumption in the literature since topic has been equated with subject, but note that there may be cases where the sentence topic is not a subject. For example, structures like *A Pedro_i lo_i vio Juan_j ayer en el bar* ‘Lit: To Pedro_i, him_i saw Juan_j in the bar’ (*As for Pedro, John saw him in the bar*) are known in null-subject languages as clitic left dislocation (CLLD). It is the dislocated object *A Pedro* that sets the sentence topic (as it appears in first position, which is salient) rather than the postverbal subject *Juan* (for details and a discussion, see de Rocafiguera, 2023).

² The symbol [...] indicates deleted material from the original text so as to shorten the example due to space limitations.

³ The examples from natives in this paper are taken from the following corpora: (i) Spanish natives: CEDEL2 corpus (Corpus Escrito del Español L2) (<http://cedel2.learnercorpora.com>) (Lozano 2022) (cf. section 5.1); (ii) Greek natives: GLC corpus (Greek Learner Corpus) (https://glc.lit.auth.gr/app/GLC_Gateway) (Tantos et al. to appear) (cf. section 5.1); and (iii) English natives: COREFL corpus (Corpus of English as a Foreign Language) (<http://corefl.learnercorpora.com>) (Lozano et al. 2021). We include the corpus filename at the end of each example within square brackets, e.g., [EN_WR_20_3_LH].

typically mark sentential topic continuity, as in (2), whereas topic shift is usually realised via overt pronouns, as in (3). However, unlike what the experimental literature reports (cf. section 2.1), corpus-based studies indicate that NPs are the most frequent and privileged topic-shift marker both in Spanish (e.g., Lozano 2009; Lozano 2016) and Greek (Papadopoulou 2020; Charatzidis et al. in press), as in (4), in detriment of overt pronouns, that are infrequent and, therefore, not a privileged topic-shift marker.

- (2) a. Chaplin_i se acerca al carro y Ø_i deja al bebé_j allí ...
 ‘Chaplin_i approaches the stroller and Ø_i leaves the baby_j there ...’
 [Spanish native: ES_WR_18_14_CRM]
- b. Ο περαστικός_i αφήνει το μωρό_j σε ένα καρότσι που Ø_i βρίσκει τυχαία.
 o perastikós_i αφήνει το moró_j σε ένα karótsi pu Ø_i vrísci tíchéa
 ‘The passer-by_i leaves the baby_j in a stroller that Ø_i finds randomly.’
 [Greek native: GR_WR_CH_007]
- (3) a. Un policía_i ve a Chaplin_j haciendo esto y él_j vuelve ...
 ‘A policeman_i sees Chaplin_j doing that and he_j returns ...’
 [Spanish native: ES_WR_19_14_ABPM]
- b. Ο κύριος_i [...] κυνηγά [τον Τσάρλι]_j αλλά εκείνος_j κρύβεται.
 o círios_i kiniγá ton tsárlj, alá ecínos_j krívete
 ‘The man_i chases Charlie_j but he_j hides.’
 [Greek native: GR_WR_CH_012]
- (4) a. [Charles Chaplin]_i decide entregárselo_j a otro señor_k [...]. El señor_k le_i persigue ...
 ‘[Charles Chaplin]_i decides to give him_j to another man_k. The man_k chases him ...’
 [Spanish native: ES_WR_22_14_CLA]
- b. Στο δρόμο [ο Τσάρλι]_i βλέπει έναν γέρο_j [...]. Ο γέρος_j ψάχνει ...
 sto ðrómo o tsárlj, vlépi énan jéro_j [...]. O jéros_j psáhni ...
 ‘On the street Charlie_i sees an old man_j. The old man_j searches for ...’
 [Greek native: GR_WR_CH_009]

On the other hand, discourse topic results from repeated reference to a specific discourse element, which progressively forms a string of “semantically related topical entities” (Alonso-Ovalle 2006: 21). So, the discourse topic focuses on the “aboutness” of the preceding discourse and, although a discourse may contain several different sentence topics, the discourse topic is always present (Van Dijk 1977). In (5), the sentence-topic sequence is: *Eva / she / she / Ø / the train / she*. Despite the topic shift (*the train*), *Eva* remains the discourse topic of (5).

- (5) **Eva**_i woke at five o’clock that morning. Today **she**_i had to start with her new job in Prague. **She**_i hurriedly took a shower and Ø_i had some breakfast. The train_j would leave at 6:15 and **she**_i did not want to come late the first day.
 [Source: Van Dijk, 1977: 59, our indices]

Distinguishing between sentence vs. discourse topic is crucial for the current study. Our participants wrote a narration of events about a Charles Chaplin's video and the name 'Chaplin' was mentioned in the task title. Thus, we distinguished between main vs. secondary characters. In (6), which is an excerpt from a text similar to the ones we annotated in Spanish and Greek (but this time from an English native speaker for simplicity), there is repeated reference to *Chaplin*. As a result, the discourse topic coincides with the main character, a fact that occurs in all annotated texts. So, the discourse-topic/main-character *Chaplin* would be expected to be more salient than any other character antecedent (e.g., *the police officer, the baby*), so we would anticipate *Chaplin* to be referred to by means of shorter REs.

- (6) **Charles Chaplin_i** is walking when **he_i** finds a baby_j. The baby_j is wrapped up in a blanket on the ground and \emptyset_j is crying. **He_i** picks up the baby_j and as **he_i** tries to set the baby_j back down, a police officer_k approaches. It appears like **he_i** is trying to abandon his own baby_j so **he_i** picks it_j back up and \emptyset_i moves on. **He_i** continues to try to get rid of the baby_j, each time with no luck. Finally, **he_i** finds a note in the baby's blanket saying that he_j's an orphan and \emptyset_j needs to be cared for and loved.
[English native: EN_WR_21_2_SM]

Finally, since the sentence topic is typically associated with the subject position, theories about the effect of syntactic position on antecedent saliency are also relevant to our study. The *Position of Antecedent Strategy* (PAS) (Carminati 2002) claims that in Italian (a pro-drop language like Spanish and Greek) null pronouns tend to refer to preverbal subject antecedents, while overt pronouns present a stronger preference towards non-subject antecedents, as in (7). Moreover, *Centering theory* (Grosz et al. 1995) postulates that reference to the topic/subject leads to topic continuity and, as a result, to a higher degree of discourse coherence. In (8a), the pronoun *she* refers to the subject/sentence-topic *Fiona*, whereas in (8b) *he* refers to the non-subject/non-topic *Craig*. As empirical data have shown, reading times on the pronoun that refers back to the subject of the preceding clause are shorter than when the pronoun refers back to a non-subject antecedent, which supports the claims of *Centering theory* (Gordon et al. 1993) that (8a) could be considered more coherent than (8b).

- (7) Marta_i scriveva frequentemente a Piera_j quando \emptyset_i /**lei_i** era negli Stati Uniti.
'Marta_i wrote frequently to Piera_j when \emptyset_i /**she_i** was in the United States.'
[Source: Carminati 2002: 45]
- (8) a. Fiona_i complimented Craig_j and **she_i** congratulated James.
b. Fiona_i complimented Craig_j and **he_j** congratulated James.
[Source: Wolf et al. 2004: 666, our indices]

The current paper is divided as follows: Section 2 reviews previous experimental and corpus AR studies in native Spanish and Greek. In section 3, we present the motivation and novelty

of the current study. Research questions (section 4) and methodology (section 5) lead to the results (section 6). Section 7 presents a general discussion, the proposal of the TTH and a conclusion.

2 Spanish and Greek studies on AR

This section reviews empirical studies on AR in Spanish and Greek monolingual adults. Most previous AR studies investigated either Spanish monolinguals or Greek monolinguals, while studies comparing both languages simultaneously are scarce. In the following sections, we focus first on experimental studies, which constitute most publications, and then review corpus studies. Finally, we discuss two studies which compared the properties of REs in Greek and Spanish monolinguals simultaneously.

Note that some of the studies reviewed below investigated both native adults as well as L2 learners and/or bilingual speakers, but we restrict our review to native speakers only, who are the target participants of the present study. As for the Spanish studies, we restrict our literature review to peninsular Spanish (which is the variety of our corpus natives), as it has been shown that the use of REs differs in some Spanish varieties (Silva-Corvalán 1994; Toribio 2000) and, thus, results may not be comparable.

2.1 Experimental studies

2.1.1 Spanish experimental studies

Experimental studies on native Spanish have mainly focused on the null/overt pronoun distinction by investigating the PAS, as in (7). While the PAS is confirmed in Italian and in Greek, it is not always confirmed in Spanish. Alonso-Ovalle et al. (2002) investigated the PAS in Spanish monolinguals using several offline experiments in sentences like (9). Null pronouns biased towards subject antecedents, but no clear bias was found for overt pronouns towards non-subject antecedents.

- (9) Juan_i pegó a Pedro_j. Ø_i/Él_j está enfadado.
 ‘John_i hit Pedro_j. Ø_i/He_j is angry.’
 [Source: Alonso-Ovalle et al. 2002: 154, our indices]

Additionally, they tested contexts with one antecedent to investigate the acceptability rate of null vs. overt pronouns in topic-continuity scenarios with only one antecedent, as in (10). They found higher ratings for null than for overt pronouns.

- (10) Teresa_i llegó al aeropuerto tarde. Ø_i/Ella_i estaba cansada.
 ‘Teresa_i arrived at the airport late. Ø_i/She_i was tired.’
 [Source: Alonso-Ovalle et al. 2002: 157, our indices]

In a similar vein, Filiaci (2010) and Filiaci et al. (2014) investigated the PAS in Spanish natives using an online experiment with sentences like (11) and found that null pronouns exhibited a clear bias towards subject antecedents, while overt pronouns did not clearly bias towards non-subject antecedents. In particular, null pronouns show a lower reaction time when referring to the subject antecedent (11a) than to the object antecedent (11b), whereas there are no reaction-time differences for overt pronouns. Clements and Domínguez (2017) found parallel results in a Picture Verification Task.

- (11) a. Después de que Bernardo_i criticó a Carlos_j tan injustamente, Ø_i/él_i le pidió disculpas. [SUBJECT BIAS]
 b. Después de que Bernardo_i criticó a Carlos_j tan injustamente, Ø_j/él_j se sintió muy [OBJECT BIAS]
 ‘After that Bernardo has criticised Carlos so unjustly, Ø/he { apologized to him/felt very offended}.’
 [Source: Filiaci et al. 2014: 832]

By contrast, other studies found the opposite pattern (i.e., bias of overt pronouns towards non-subject antecedents), but note that the type of disambiguation differs from previous studies. Chamorro et al. (2015) investigated the PAS in Spanish natives using an Acceptability Judgement Task (AJT) and an Eye-tracking-while-reading experiment in sentences like (12a-b). They found that there was no clear bias of null pronouns towards subject antecedents, but overt pronouns clearly biased towards non-subject antecedents. Using an AJT, Chamorro (2018) confirmed the same pattern for Spanish natives.

- (12) a. La madre_i saludó a las chicas_j cuando Ø_i/ella_i cruzaba una calle con mucho tráfico.
 ‘The mother_i greeted the girls_j when Ø_i/she_i was crossing a busy street.’
 b. Las madres_i saludaron a la chica_j cuando Ø_j/ella_j cruzaba una calle con mucho tráfico.
 ‘The mothers_i greeted the girls_j when Ø_j/she_j was crossing a busy street.’
 [Source: Chamorro 2018: 6–7, our indices]

Other experimental studies in Spanish have investigated AR but focusing on structures other than those examined by Carminati (2002). Lozano (2002) investigated contrastive-focus contexts in Spanish natives using an AJT in sentences like (13) and showed that Spanish natives accepted overt pronouns, but not null pronouns as these were ambiguous. Using a contextualized AJT, Lozano (2018) corroborated these findings by showing higher acceptance of overt than null pronouns in contrastive-focus contexts. Additionally, he tested topic-continuity contexts and found that the acceptance rate was higher for null than for overt pronouns.

- (13) El señor López_i y la señora García_j trabajan en la universidad y en una famosa editorial. No obstante...
- (a) cada estudiante dice que él_i tiene poco dinero.
- (b) cada estudiante dice que Ø_{i/j} tiene poco dinero.
- ‘Mr López_i and Ms García_j work at the university and at a famous publisher. However, each student says that he_i/Ø_{i/j} has little money’.
- [Source: Lozano 2002: 3]

Overall, the experimental findings from native Spanish indicate that null pronouns signal topic-continuity, while overt pronominals signal topic-shift and contrastive focus, even though some studies show a rather flexible behaviour for overt pronouns, which could be related to the variability among the stimuli or methodologies employed (cf. examples (9)-(13) above). In our study, we will further elaborate on these results and, in particular, on such (apparent) flexibility of Spanish overt pronouns by exploring the referential properties of not only overt pronominals but also NPs, which are very frequent in production data and yet have been neglected in experimental research.

2.1.2 Greek experimental studies

The experimental studies with native Greek adults have also mainly explored the referential properties of null and overt pronouns in typical PAS sentences. For example, Papadopoulou et al. (2015) conducted two experiments to investigate the preferences of null and overt pronominals in Greek constructions similar to (7). Results from a sentence-picture matching task showed that, for null pronouns, Greek adults preferred the subject antecedent but also, to some extent, the object. By contrast, the overt pronoun was overwhelmingly linked with the object. Results from a self-paced listening picture verification task corroborated these findings. The listening times revealed a listening advantage for the object over the subject with overt pronouns, but an advantage for the subject over the object with null pronouns. Parallel findings have been reported by Amvrazis (2016) and Kaltsa et al. (2015), who used a similar experimental paradigm. In the same vein, Fleva et al. (2017) employed a self-paced listening task with sentences such as (7) and found a bias for object antecedents with overt pronominals but no bias for either the subject or the object antecedent with null pronouns.⁴ Prentza & Tsimpli (2013) also reported similar biases in an offline written interpretation task.

These experimental findings indicate that, due to its referential flexibility, the null pronoun appears to be the default RE in Greek since it can refer to subject but also to non-subject antecedents,

⁴ Note that the bias of overt pronominals to be linked with object antecedents has also been attested in inter-sentential anaphora (Miltakaki 2007).

whereas overt pronominals exhibit a stronger bias as they refer to non-subject antecedents and are dispreferred with topic referents. Based on this experimental evidence, Tsimpli et al. (2004) and Papadopoulou et al. (2015) argue that overt pronouns are inherently marked for the discourse feature [+topic shift] due to their preference for non-subject antecedents (see the discussion in Papadopoulou 2020: 150–153; and also in Torregrossa et al. 2020). Dimitriadis (1996) made similar remarks based on corpus data (see section 2.2.2). We will come back to this point later in the paper.

Cunnings et al. (2017) investigated overt pronoun resolution in native Greek by means of a visual world paradigm. Their materials resemble those employed by Papadopoulou et al. (2015), but also differ in some respects. In this study, the subordinate clause is preposed and includes the two possible referents of the pronoun, which in turn is located in the main clause, (14). Additionally, the gender of the two referents in the subordinate clause and the overt pronoun in the main clause is manipulated, so that reference to the subject or the object antecedent is unambiguous via gender information or remains ambiguous.

- (14) a. Αφού ο Γιάννης_i μίλησε με την κυρία Ελένη_k / τον κύριο Κώστα_j μπροστά στο ταμείο, **αυτός**_{i/j/*k} πλήρωσε γρήγορα το παγωτό που είχε αγοράσει.
 Afú o Jánis_i mílise me tin ciría Eléni_k / ton círio Kósta_j brostá sto tamío, **aftós**_{i/j/*k} plírose grígora to pagotó pu íze agorási.
 ‘After John_i spoke to Mrs. Helen_k / Mr. Kostas_j by the till, **he**_{i/j/*k} quickly paid for the ice cream that (he) had bought.’
- b. Αφού η κυρία Ελένη_k / ο κύριος Κώστας_j μίλησε με τον Γιάννη μπροστά στο ταμείο, **αυτός**_{i/j/*k} πλήρωσε γρήγορα το παγωτό που είχε αγοράσει.
 Afú i ciría Eléni_k / o círios Kóstas_j mílise me ton Jáni_i brostá sto tamío, **aftós**_{i/j/*k} plírose grígora to pagotó pu íze agorási.
 ‘After Mrs. Helen_k / Mr. Kostas_j spoke to John_i by the till, **he**_{i/j/*k} quickly paid for the ice cream that (he) had bought.’
 [Source: Cunnings et al. 2017: 641]

The Greek natives rapidly used gender information to select the appropriate referent for the overt pronoun in the unambiguous sentences, while no clear preference for either referent was attested in the ambiguous sentences. In the offline comprehension data, however, reference to the object antecedent was preferred over the subject antecedent in both unambiguous and ambiguous sentences, in parallel with the findings of the Greek studies discussed above.

In a more recent study, Fotiadou et al. (2020) explored the reference properties of *aftós* (= he)⁵ in Greek natives by means of a visual world paradigm. The experimental material involved

⁵ Fotiadou et al. (2020) also explored the reference properties of *o ídjos* (= the same.m.sg). However, here we focus on the data from *aftós* (= he), as this pronominal form is investigated in the present study.

inter-sentential anaphora resolution: two referents (*the hunter* and *the fisherman*) in the first sentence (15a) and the pronoun (*aftós* ‘he’) in the second sentence (15b).

- (15) a. Ο κυνηγός_i συναντάει τον ψαρά_j κάθε απόγευμα στο δάσος δίπλα στο ποταμάκι.
Ο ciniγός_i sinadáí ton psará_j κάθε απόγευμα sto dásos dípla sto potamáci.
‘The hunter_i meets the fisherman_j every afternoon in the forest by the river.’
- b. **Αυτός**_{i/j} βρήκε τυχαία εκεί, μετά από πολύ καιρό, τον εργάτη.
Aftós_{i/j} vríce tiçéa ecí, μετά από polí ceró, ton ergáti.
‘**He**_{i/j} found accidentally there, after a long time, the worker.’

The participants did not show any clear referential preferences for *aftós* in accordance with the online results of the study by Cunnings et al. (2017) and in contrast to previous studies (Prentza & Tsimpli 2013; Kaltsa et al. 2015; Papadopoulou et al. 2015; Amvrazis 2016). Fotiadou et al.’s study is the only one that suggests no subject or object biases regarding the overt pronoun *aftós*. We think that this difference probably lies on the fact that Fotiadou et al. investigated inter-sentential, rather than intra-sentential, anaphora by means of short stories, in which additional discourse factors may affected the interpretation of the pronoun.

In an elicited production task, Argyri & Sorace (2007) explored the type of referential expressions used by Greek speakers in topic-continuity contexts. The default response to (16a) would be a null pronoun and not an overt pronoun, because the topic of the two sentences remains the same. Greek native adults overwhelmingly produced null subjects. The same pattern was attested when the participants had to judge the acceptability of (16b): a null pronoun was judged as a significantly more acceptable continuation than an overt pronoun.

- (16) a. Γιατί πήγε η Ελένη_i στο περίπτερο;
jati píje i Eléni_i sto períptero;
‘Why did Eleni_i go to the kiosk?’
- b. Γιατί \emptyset /**αυτή**_i ήθελε να αγοράσει εφημερίδα.
jati \emptyset /**#aftí**_i íthele na agorási efimerída
‘because \emptyset /**she**_i wanted to buy newspaper’
[Source: Argyri & Sorace 2007: 84]

Overall, divergent reference patterns for Greek null and overt pronouns have been attested in previous experimental studies. Overt pronouns show a preference to be used in topic-shift contexts, while null pronouns display rather flexible referential properties. In our study, we will further explore these preferences with corpus data and, importantly, we will widen the range of REs.

2.2 Corpus studies

2.2.1 Spanish corpus studies

Montrul & Rodríguez-Louro (2006) analysed the oral production of Spanish natives and showed that, overall, they produced more null (57.2%) than overt (pronominal and lexical) subjects (42.8%). Such a pattern was also found in the written corpus of Spanish natives in Margaza & Bel (2006). They reported a higher production of null subjects than pronominal subjects, especially with subordinate clauses. These two studies, however, do not provide a clear distinction between the three different REs we consider in the present study (i.e., null pronoun, overt pronouns, and NPs). This distinction is reported in Vande Casteele & Collewaert (2016), who analysed the oral production of Spanish natives. They showed a higher production of null pronouns (42.39%), followed by overt pronouns (30.04%), and noun phrases (20.57%). Crucially, these studies do not examine information status, as we will do in this study.

Some corpus studies in Spanish have added new insights into AR, which were undetected by previous experimental studies. Lozano (2009; 2016) was the first to explore how information status constrains the use of REs by investigating the written production of Spanish natives. Both studies showed that Spanish natives produced mostly null pronouns in topic continuity (Lozano, 2009: 97%; Lozano, 2016: 93.3%), followed by low percentages of overt pronouns (range: 1.8%–2.7%) and NPs (range: 1.2%–4%). By contrast, in topic shift they produced mostly NPs (Lozano, 2009: 87.2%; Lozano, 2016: 70.8%), followed by low production of overt pronouns (range: 12.8%–19.4%) and hardly any null pronouns (range: 0%–2.8%). Importantly, Lozano (2009; 2016) argued that it is crucial to investigate AR in contextualized scenarios to see how other REs like NPs play a role in AR. In fact, Lozano & Quesada (in prep.) investigated PAS in Spanish natives using corpus data and found that NPs biased towards non-subject antecedents more frequently than overt pronouns, which could explain the “apparent” flexibility of overt pronouns in previous PAS experimental studies, a fact to which we return in this study. Regarding topic continuity, Lozano’s findings were corroborated in Martín-Villena & Lozano (2020), who investigated the written production of Spanish natives and found very high production of null pronouns (93.9%), followed by NPs (5.1%) and overt pronouns (1%).

Importantly, other corpus studies investigating the production of REs according to information status found some differences with respect to Lozano’s studies (2009; 2016) in topic shift. Licerias et al. (2010) investigated the written production of Spanish natives and focused on null pronouns in topic shift. They showed similar percentages of null (49%) and overt pronouns (51%) in these contexts. Similarly, Georgopoulos (2017) found that Spanish natives produced null pronouns (33.7%) in topic shift and additionally corroborated previous findings by showing high percentages of null pronouns in topic continuity (86.26%). Likewise, García-Alcaraz & Bel (2019) found null pronouns in topic shift in the written production of Spanish natives. In particular, they analysed null and overt pronouns according to type of construction (intra- vs. inter-sentential)

and found different patterns of production. In intra-sentential scenarios, the natives produced mostly null pronouns in topic continuity (69.01%), but also in topic shift (30.99%), while they produced overt pronouns in topic continuity only (100%). In inter-sentential scenarios, null pronouns were more frequent in topic continuity (64.5%) than in topic shift (35.5%), while the opposite pattern was observed for overt pronouns (topic continuity: 21.05%; topic shift: 78.95%). It is worth highlighting that some of these studies report the production of null pronouns in topic shift but such production could be explained by the implicit causality or the use of directive verbs (García-Alcaraz 2015; Lozano 2016), a factor that is not addressed in these studies. Thus, it is highly likely that null pronouns in topic shift are triggered by the semantics of the verb, though there could be additional reasons, a fact which we address in the present study.

2.2.2 Greek corpus studies

Contrary to Spanish, there are just a few Greek corpus studies that explore AR. Dimitriadis (1996) was the first to investigate the reference patterns of null and overt (*aftós* ‘he, this one’ and *ecínos* ‘that one’)⁶ Greek pronouns in a written corpus. The data showed that the null pronouns preferably referred to a topic antecedent (78%), whereas the overt pronouns were overwhelmingly linked to non-topic antecedents (87%). Based on these results, Dimitriadis proposed the *Overt Pronoun Rule*, according to which overt Greek pronouns should not be construed with the topic of the previous utterance. Notice that a corresponding rule for null pronouns has not been proposed due to their relatively less firm referential properties.

In a more recent corpus study, Charatzidis et al. (in press) explored the reference properties of three types of REs, null and overt pronouns as well as NPs, in relation to the syntactic position and the sentential and discourse topichood of the antecedent, among other factors. The corpus consisted of descriptive, narrative and argumentative written texts produced by native adult speakers of Greek. The predominant RE used in this corpus was null pronouns (68%), followed by NPs (26%), while overt pronouns represented the least preferred option (4%). This observation comes into sharp contrast with most experimental studies, which explored only null vs. overt pronouns, even though NPs appear to be a much more frequent RE than overt pronouns, at least in written corpora. Additionally, the results showed that null pronouns referred to subject, sentential-topic and discourse-topic antecedents, as well as to referents located one and two clauses away from the RE. By contrast, the NPs and the overt pronominals exhibited the opposite reference pattern since they were preferably used to refer to non-subject

⁶ As an anonymous reviewer points out, the two pronominal forms have distinct features. *Aftós* is a personal and a demonstrative pronoun and indicates an entity that is close to the speaker. *Ecínos* is a demonstrative pronoun used for an entity that is not as close to the speaker. Indeed, these two forms have been shown to show different behaviours since *ecínos* may refer to a referent previously mentioned in the discourse but not within the same sentence (Dimitriadis, 1996).

and non-topic antecedents. A difference between the NPs and the overt pronominals, however, emerged regarding the distance between the antecedent and its RE: NPs were used for referents located even four clauses away, while overt pronouns were restricted to referents being up to two clauses away, patterning more with null pronominal forms in that respect. Even though the referential properties of null and overt referential expressions appear to be divergent, the corpus data indicate that the Greek null pronominal subjects may also refer to non-subject and non-topic antecedents. Interestingly, in topic-shift contexts the null pronoun is still involved in 39% of these cases, while NPs and overt pronouns in such cases amount to 48% and 13% respectively (see Papadopoulou 2020 for more details). This finding further points to the flexibility of Greek null pronouns, as has been shown by some experimental studies, an issue that will be further explored and discussed in the current study.

2.3 Studies comparing AR in Greek vs. Spanish

To the best of our knowledge, there are only two studies that have compared native speakers of Greek and Spanish simultaneously.

First, Margaza & Gavarró (2020) investigated natives' interpretations of null and overt pronouns using an offline multiple-choice task that included topic-continuity and topic-shift contexts, (17a-b) (among other conditions that are not of interest for the present study). Results showed that both groups preferred null pronouns in topic continuity (Greek: 89.33%; Spanish: 100%), while in topic shift null pronouns were mostly preferred in Spanish (78.67%) but not in Greek (56%). The data indicate that, even though both Spanish and Greek are null-subject languages, they manifest differences in the referential properties of the REs. Overt subject pronouns appear to be more prone to such differentiation, since Greek overt subject pronouns were more frequently used in contexts such as (17a) and (17b).

- (17) a. Primero, Rosa_i prepara la comida y luego Ø_i/ella_i hace los deberes del colegio.
 'First, Rosa_i prepares the meal and then Ø_i/she_i does her homework.'
- b. Ángela_i quiere publicar un libro y los editores_j explican que Ø_i/ella_i precisa completar un manuscrito de su obra.
 'Ángela_i wants to publish a book and the editors tell (her) that Ø_i/she_i has to complete a manuscript of her work.'
- [Source: Margaza & Gavarró 2020: 9–10, our indices]

Second, Giannakou & Sitaridou (2020) also investigated Greek and Spanish natives using an oral production task and an offline interpretation listening task. Importantly, their Spanish natives speak the Chilean variety. Firstly, in the production task the authors found that both groups of natives produced significantly more null pronouns (Greek: 61.71%; Spanish: 56.52%) than lexical subjects, also known as NPs in the literature (Greek: 32.99%; Spanish: 32.58%),

while overt pronouns were scarce (Greek: 0.61%; Spanish: 2.7%), as has been pointed out in several corpus studies (cf. our review in section 2.2). Additionally, the production data on topic continuity showed that both Spanish and Greek natives produced null pronouns (Greek: 95.79%; Spanish: 93.67%), while the use of NPs (Greek: 4.04%; Spanish: 5.91%) was minimal and the production of overt pronominals was close to zero (Greek: 0.18%; Spanish: 0.42%). In topic shift, both groups produced more NPs (Greek: 74.71%; Spanish: 66.52%) than null pronouns (Greek: 23.35%; Spanish: 25.32%) and overt pronouns (Greek: 1.95%; Spanish: 8.15%). However, there were more NPs in Greek than in Spanish and more overt pronouns in Spanish than in Greek.

In Giannakou & Sitaridou's offline interpretation listening task, they explored the effects of PAS (see example (7)) and definiteness (the object of the main clause was either definite or indefinite) in Spanish and Greek. They found no clear bias of null pronouns towards subject antecedents neither in Greek (definite object: 46.25%; indefinite object: 50%) nor in Spanish (definite object: 47.5%; indefinite object: 46.25%). Importantly, they found a clear bias of overt pronouns towards non-subject antecedents in Greek (definite object: 82.5%; indefinite object: 77.5%), but not in Spanish (definite object: 48.75%; indefinite object: 58.75%). Giannakou & Sitaridou (2020) claimed that null pronouns exhibit similar referential properties in both languages, as they can be used in topic-continuity and topic-shift contexts, which contrasts with the results from the production task discussed in the previous paragraph. Furthermore, Greek overt pronouns were found to be related to topic shift both in production and interpretation tasks, whereas Spanish overt pronouns showed variability, particularly in the interpretation task. This finding also highlights differences in the distribution of REs in two null subject languages. Giannakou & Sitaridou (2020) attribute these referential differences in the features of overt pronouns. Greek overt subject pronouns also have a deictic interpretation and, thus, are excluded in topic-continuity contexts, whereas Spanish overt subject pronouns are weaker and can therefore refer to topics.

The aim of the present study is to further explore the referential patterns in Spanish and Greek by comparing two written corpora in the two languages. We restrict our analyses to 3rd person singular [+human] REs, and we keep all other variables, apart from the language, constant across the two corpora (see Sections 3 and 5), in order to isolate and highlight the referential similarities and differences between Spanish and Greek. Additionally, we investigate peninsular Spanish natives (and not Chilean Spanish), which could be also a potential source of differences between some of the previous studies and ours.

3 Motivation and novelty of our study

We incorporated the following innovations in our study to explore and systematically compare AR in Spanish vs. Greek:

- (i) Everything except for the participant's L1 is kept constant (task and text type, methodology, target structures, tagset, analysis, participant profile) so that results can be attributed exclusively to the typological differences between the two languages.
- (ii) We examine whether the story is narrated similarly in the two languages to determine whether potential differences stem from differences in narrative patterns.
- (iii) We conduct fine-grained analyses to thoroughly compare our datasets. The information-status factor is presented in two formats, for reasons that will be clear later:
 - a. (analysis #1): RE form by information status
 - b. (analysis #2): information status by RE form
- (iv) We discuss the effect of both sentence and discourse topichood, a previously neglected but crucial distinction.
- (v) Unlike previous research, we focus on the effect of character on its own but also on character by information status.

4 Research questions

Previous corpus-based studies on AR in Spanish (inter alia: Vande Castele & Collewaert 2016; García-Alcaraz & Bel 2019; Martín-Villena & Lozano 2020) and Greek (Dimitriadis 1996; Papadopoulou 2020; Charatzidis et al. in press) and the production study comparing Spanish and Greek (Giannakou & Sitaridou 2020) have not fully examined the way discourse is constructed (i.e., the distribution of topic-continuity vs. topic-shift for all REs as well as the distribution of main vs. secondary characters for all REs). This fact leads to our first research question:

RQ1: Given the same narrative task, how do Spanish vs. Greek natives construct their discourse regarding the distribution of:

- a. the information status (topic continuity/topic shift/focus new introduction) of all REs?
- b. the information status of each RE form (overt/null/NP)? (Analysis #1: RE x info status)
- c. the characterhood (main/secondary characters) of all RE forms?

The overall distribution and proportion of REs in previous corpus-based studies shows that Spanish and Greek natives produce mostly null pronouns, followed by NPs and overt pronouns (cf. section 2.2). Crucially, the same pattern of production (despite differences between groups) emerged in the only study simultaneously comparing Spanish and Greek, Giannakou & Sitaridou (2020), who investigated AR in Greek and Chilean Spanish, not Peninsular Spanish. Thus, we formulated our second research question:

RQ2: Given the same narrative task, do Peninsular-Spanish and Greek natives produce the same proportion of REs (overt/null pronominal subjects and repeated NPs) in narratives?

Previous experimental studies have shown that, in both Greek and Spanish, the null pronoun biases towards subject/sentence-topic antecedents but the two languages differ as to the preference of the overt pronoun towards subject/sentence-topic or non-subject/non-sentence-topic antecedents. These differences could be accounted for if we also consider lexical subjects (NPs) as an additional RE form, a fact that has not been considered in the experimental literature. On the other hand, even though some previous corpus-based studies have included NPs in their analysis, there is no corpus-based study that simultaneously compares Greek and Peninsular Spanish. Thus, our study addresses this research question:

RQ3: Given the same narrative task, do Spanish and Greek natives produce similar proportions of REs in Topic Continuity vs. Topic Shift contexts? (Analysis #2: Info status x RE)

Finally, in previous corpus-based studies in Spanish and Greek, antecedent characterhood is not considered to be a factor affecting saliency. Given that the main character frequently happens to be the discourse topic, it is worth examining how this factor affects RE choice:

RQ4a: Given the same narrative task, does characterhood affect the choice of REs in Spanish vs. Greek?

RQ4b: Additionally, is there an interplay between the RE to refer to the character and its information status?

5 Methodology

5.1 Corpus sample

We analysed a sample of Spanish and Greek adult natives.⁷ The Spanish sample was taken from *Corpus del Español como L2* (CEDEL2) (<http://cedel2.learnercorpora.com>), which is a large, freely available Spanish corpus that follows strict design criteria (see Lozano 2022 for an overview). The Greek sample was taken from Greek Learner Corpus (GLC) (https://glc.lit.auth.gr/app/GLC_Gateway), which is also a thoroughly designed large and freely online available corpus (see Tantos et al. to appear for more information). The native speakers of both languages met a series of criteria so as to ensure maximum comparability across both corpus samples: age range (18–30); proficiency in L2 (either no knowledge of an L2 or very low proficiency level in an L2); and language variety (Peninsular Spanish; Standard Greek). Following these requirements,

⁷ For ethical issues, see section *Ethics and consent* below.

the Spanish texts were downloaded from CEDEL2 and the Greek texts from GLC. As shown in **Table 1**, a total of 29 participants (15 Greek and 14 Spanish) were included in this study and a total of 741 REs (400 in Greek and 341 in Spanish) were tagged for different factors (cf. tagset in **Figure 2**).

Language	N participants	Tagged REs	Mean age	Age range
Greek	15 (F: 11)	N = 400	19,2	18–23
Spanish	14 (F: 13)	N = 341	20,7	18–25

Table 1: Participants.

5.2 Task

Participants watched a Charles Chaplin’s silent video⁸ and had to retell it in a written format. This task was deemed suitable to elicit AR contexts since: (i) it promoted the introduction, maintenance, and reintroduction of characters to investigate topic-continuity vs. topic-shift contexts, and (ii) it included several [+human] characters promoting the production of 3rd person REs. As for the characters, there is one main character (i.e., Chaplin) and four secondary characters (i.e., the baby, the woman, the old man, and the policeman).

5.3 Tagset

Tagging and analyses were performed with version 3.3 of UAM Corpus Tool (O’Donnell 2008) (<http://www.corpustool.com>) via two tagsets. The first tagset classified each text according to the natives’ L1 (Spanish|Greek) (**Figure 1**). The second tagset (**Figure 2**) was used to annotate all REs and included the relevant linguistic features to address our research questions. Some of these features were previously annotated and investigated in corpus-based studies (Lozano 2009; Lozano 2016; Martín-Villena & Lozano 2020; Quesada & Lozano 2020; Charatzidis et al. in press).

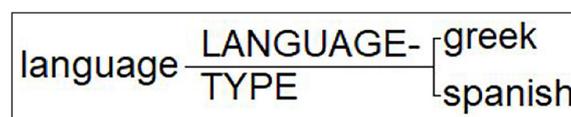


Figure 1: Tagset to classify participants.

⁸ Charles Chaplin’s 4-minute excerpt from *The Kid*: <https://www.youtube.com/watch?v=4QkTNJFhu-g>.

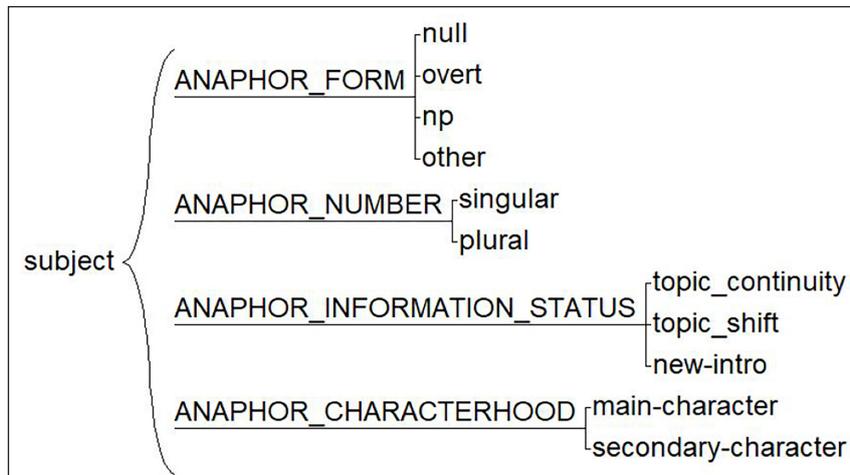


Figure 2: Linguistically-informed tagset.

For each 3rd person animate REs in subject position, we assigned the following tags. We firstly tagged the form of the RE (shown in bold in the examples below), which included NPs (e.g., *un policía* ‘a policeman’), null pronouns (\emptyset), and overt pronouns (e.g., *él* or *αυτός* ‘he’), as shown in (18). Note that the tag NP can refer to either a proper name (e.g., *Chaplin*) or a common noun (e.g., *a policeman*). We found extremely low frequencies of other RE forms (quantifiers, demonstratives) (Greek: 3/400 [0.75%]; Spanish: 8/341 [2.35%]), so we excluded them from our analyses.

- (18) a. **Chaplin_i** lo_j coge y...**un policía_k** ve a Chaplin_i haciendo esto y **él_i** vuelve a cogerlo_j y \emptyset_i se lo_j deja en brazos a ... [Spanish native: ES_WR_19_14_ABPM]
 ‘**Chaplin_i** takes him_j and...**A policeman_k** sees Chaplin_i doing this and **he_i** takes him_j again and \emptyset_i leaves him_j in the arms of a ...’
- b. **Αυτός_i** βάζει το μωρό στο καρότσι [...], \emptyset_i βλέπει τη γυναίκα και της \emptyset_i εξηγεί ότι της έπεσε το **μωρό_j**. [Greek native: GR_WR_CH_007]
aftós_i vázi to moró sto karótsi \emptyset_i vlépi ti jinéka ce tis \emptyset_i eksiyí óti tis épese **to moró_j**
 ‘**He_i** puts the baby in the stroller [...], \emptyset_i sees the woman and \emptyset_i explains to her that **the baby_j** fell.’

Next, we tagged the anaphor number including both singular and plural REs, as in (19). Importantly, the frequencies of plural REs were very low, so we also excluded them from the analyses.

- (19) a. **La señora_i** que conduce el carrito decide golpearlo_j con un paraguas porque \emptyset_i cree que **él_j** ha sido el que ha abandonado al niño_k. \emptyset_{ij} Comienzan a discutir ... [Spanish native: ES_WR_22_14_CLA]
 ‘The lady_i who is driving the pram decides to hit him_j with an umbrella because (she_i) thinks that he_k has left the baby_k. (They_{ij}) start arguing...’

- b. **Ο Τσάρλι**_i περπατάει στον δρόμο και [...] από τα παράθυρα πετούν **οι ιδιοκτήτες**_j διάφορα πράγματα ... [Greek native: GR_WR_CH_012]
ο Tsárli_i perpatáí ston δρόμο ce [...] από ta paráthira petún **i iðioktíte**_j διάφορα πράγματα
 ‘**Charlie**_i walks on the street and [...] the **homeowners**_j throw various items from the windows ...’

We also tagged the information status of the RE, including topic-continuity contexts as in (2a-b) above, topic shift contexts as in (3a-b) and (4a-b) above. Note that topical transitions (topic continuity/shift) were always annotated at the clausal level. So, the REs in both (2a) and (2b) are annotated as topic-continuity contexts: the null pronoun refers back to the subject of the preceding clause, irrespective of whether we have a coordinate clause (2a) or a subordinate clause (2b). Therefore, the subject of the preceding clause marks the topic. This topic has been traditionally referred to as ‘sentence topic’ in the literature, so we stick to this term.

Additionally, we included focus-new-introduction contexts as in (20a-b) in subject position, which represent cases of the first introduction of a character into the story.⁹

- (20) a. Esta secuencia pertenece a la película “The Kid”. En ella **un vagabundo**_i solitario (Charles Chaplin) se encuentra tirado en la calle a un bebé_j. [Spanish native: ES_WR_20_14_LVB]
 ‘This sequence is from the film “The Kid”. In it, a lonely **vagabond**_i (Charles Chaplin) finds a child_j lying in the street.’
- b. Στην ιστορία παρουσιάζεται **ο Τσάρλι Τσάπλιν**_i όπου βρίσκει ένα μωρό_j δίπλα από τους κάδους απορριμμάτων. [Greek native: GR_WR_CH_014]
 stin istoría parusiázete **ο Tsárli Tsáplin**_i ópu vrísci éna moró_j dípla apó tus káðus aporimáton
 ‘In the story, **Charlie Chaplin**_i is presented, who finds a baby_j next to the rubbish bins.’

Finally, we tagged the type of character the RE referred to: main character (i.e., *Charles Chaplin*), as in (18), or secondary characters (i.e., *the man*, *the baby*, *the woman*, or *the policeman*) as shown in bold in (21).

⁹ Note that in Focus-new-introduction, the NPs tend to be indefinite (e.g., *un vagabundo* ‘a vagabond’) as they are introduced in the discourse for the first time and they represent discourse-new information. By contrast, in topic continuity/shift, NPs tend to be either repeated NPs (e.g., *el vagabundo* ‘the vagabond’), which are definite NPs by virtue of being discourse-old information, or repeated proper Ns (e.g., *Chaplin*).

- (21) a. ...**este hombre**_i decide dejar al **bebé**_j en el mismo carrito donde lo dejó **Chaplin**_k antes. **La mujer**_i que está [...] **El policía**_m aparece de nuevo y Chaplin_k recoge...
[Spanish native: ES_WR_19_14_JMR]
'...**this man**_i decides to leave **the baby**_j in the same pram where Chaplin_k left him_j before. **The woman**_i who is [...] **The policeman**_m appears again and Chaplin_k picks up...'
- b. Ενώ λοιπόν [**Ο Τσάρλι**_i] βρίσκεται σε αμηχανία, ξαφνικά περνά μπροστά του **μία γυναίκα**_j με ένα καρότσι [και] \emptyset _i υποθέτει ότι **το βρέφος**_k είναι δικό της ... [Greek native: GR_WR_CH_005]
enó lipón [**o Tsárlí**_i] vríscete se amixanía ksafniká perná brostá tu mía **jinéka**_j me éna karótsi [ce] \emptyset _i ipothéti óti to vréfos_k íne dikó tis
'While [**Charlie**_i] is feeling awkward, suddenly a **woman**_j passes in front of him and \emptyset _i supposes that **the infant**_k is hers ...'

Once we agreed on the features, we manually annotated the texts. Importantly, the original tagset included additional tags (e.g., sentence type, number of activated antecedents, or syntactic structure), but we leave these additional features for future research. For each corpus, the annotations were checked by the four researchers. Additionally, difficult cases were discussed and agreed on. This process ensured 100% inter-annotator reliability.

After tagging all the texts, we carried out the statistical analyses with the same software (UAM Corpus Tool), which provided descriptive (raw frequencies and percentages) and inferential (χ^2) statistics based on the tag frequencies. The software reports medium significance (95%) and high significance (98%), which corresponds to $p < 0.05$ and $p < 0.02$ respectively.

6 Results and discussion

6.1 RQ1: Properties of the narratives

The rationale behind RQ1 is that, if Spanish and Greek natives construct their narrative discourse similarly, any anaphoric differences in the rest of RQs cannot be attributed to different narrative structures/styles/strategies/patterns between the two languages.

The discourse distribution of the information status (topic continuity/shift) of all RE forms (RQ1a) reveals that Greek and Spanish natives behave similarly by producing around two thirds of topic continuity and one third of topic shift in their narratives (**Figure 3**): topic continuity (Greek: 66.7% vs. Spanish: 61.39%, $\chi^2 = 2.100$, n.s.), topic shift (Greek: 33.33% vs. Spanish 38.61%, $\chi^2 = 2.100$, n.s.). This entails that (i) topic continuity predominates in the narratives of Greek and Spanish speakers, since discourse coherence is better achieved via topic continuation than shift, as explained in the introduction; (ii) both Greek and Spanish natives build their narratives in a similar way as far as the information status of the REs is concerned, a fact that

has not been reported in the literature. These similarities in narrative construction in Greek and Spanish ensure that future differences between the two languages (as we will see below) are not due to different narrative styles, but rather to differences having to do with the referential properties of the REs.

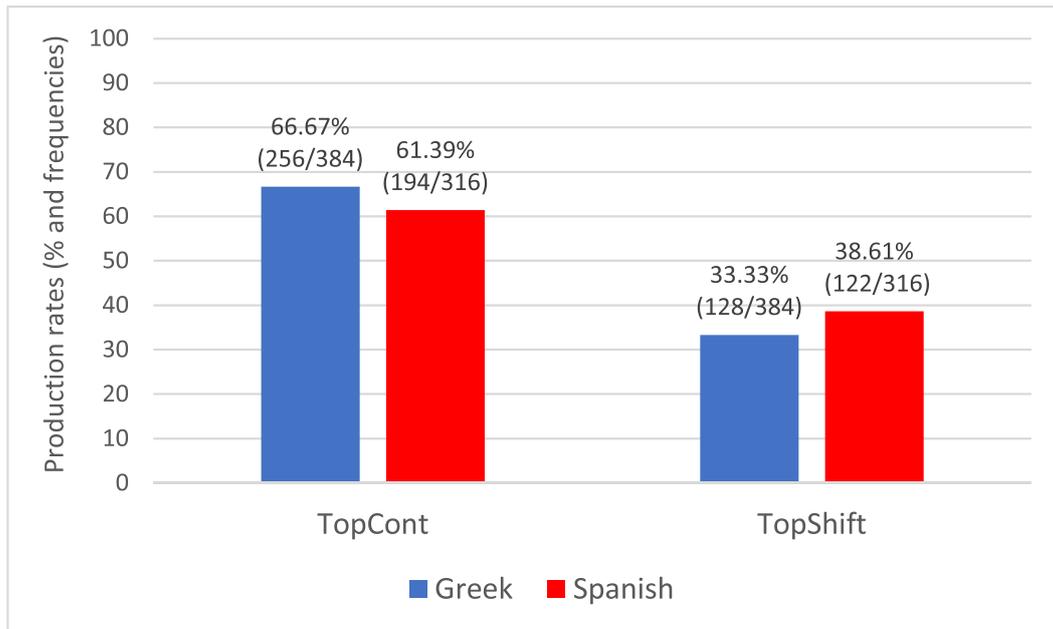


Figure 3: Discourse distribution of the information status (topic continuity/shift) of all RE forms.

Next, we analyse how each RE form is used across information-status scenarios (RQ1b) (analysis #1: form x info status, **Figure 4**). Null pronouns are predominantly used to mark topic continuity (around 80% of the time in both groups) and occasionally to mark topic shift (around 20%, though *prima facie* these rates are lower than reported in the literature for Greek, where null pronouns can mark both topic continuity and topic shift cf. discussion in section 6.3). By contrast, overt pronouns predominantly mark topic shift (around 85%), which is in line with the experimental literature of both languages. Importantly, repeated NPs are mainly used to mark topic shift, a fact that has not been fully explored in experimental studies but has been extensively reported in corpus-based studies in Spanish and Greek (cf. section 2). NPs also mark focus new introduction (i.e., when introducing new characters in the discourse for the first time, as expected) and, occasionally, they mark topic continuity (these typically refer to *Chaplin* and we termed this as the ‘Chaplin’ main-character effect (MCE), cf. results in **Figure 9**, section 6.4). Importantly, there are no differences between Greek vs. Spanish natives for any of the RE forms: null pronouns ($\chi^2 = 0.214$, n.s. for both topic continuity and shift); overt pronouns ($\chi^2 = 0.024$,

n.s. for both topic continuity and shift); NPs ($\chi^2 = 3.062$, n.s. for topic continuity; $\chi^2 = 1.003$, n.s. for topic shift; $\chi^2 = 0.099$, n.s. for focus new introduction), which suggests that Greek and Spanish natives construct narratives similarly in this respect. This type of analysis (analysis #1: RE form x info status) is somewhat misleading since overt pronouns would appear to specialise in Topic Shift and not so much on Topic Continuity, as the experimental literature reports, but note that the raw-frequency production of overt pronouns is very low (N = 15 overt pronouns out of 397 RE forms in Greek; N = 13 overt pronouns out of 333 RE forms in Spanish). Thus, a different data analysis (analysis #2: Info Status x RE form) will provide a more informative picture (cf. RQ3, section 6.3 below).

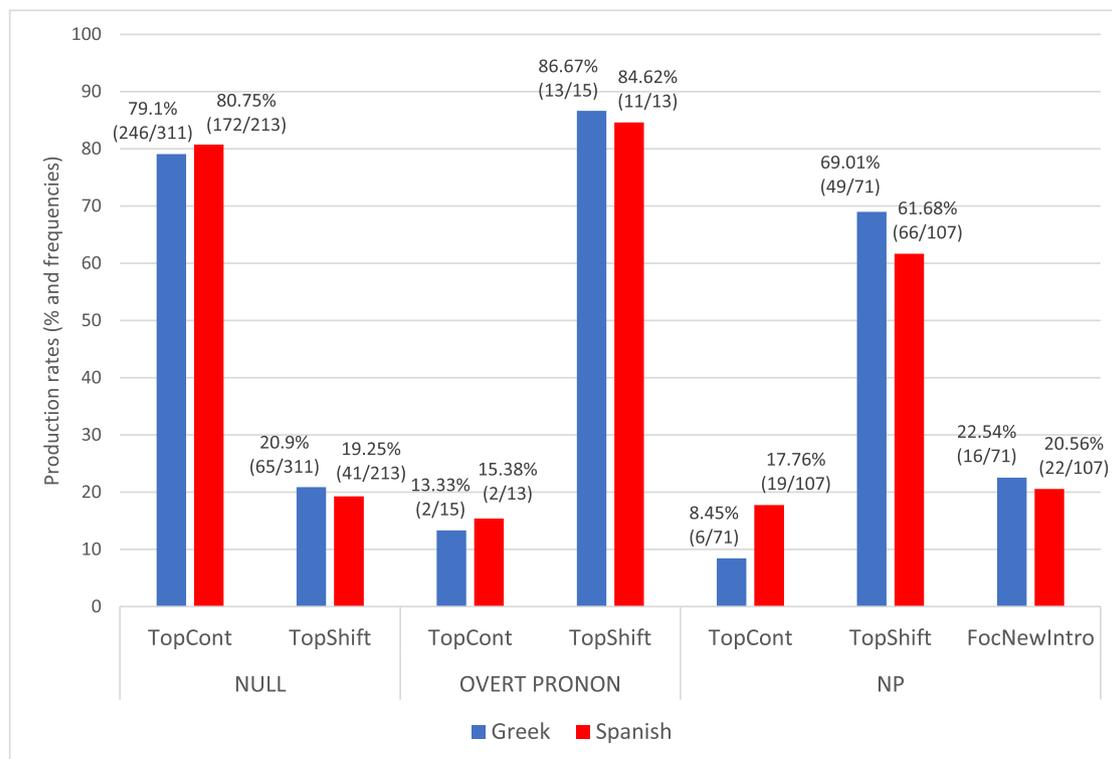


Figure 4: The distribution of RE forms according to topic continuity/shift.

Finally, regarding RQ1c (**Figure 5**), both Greek and Spanish natives' REs refer more to the main character (Chaplin) (Greek: 73%; Spanish: 69.21%) than to the secondary characters of the story (Greek: 27%, Spanish: 30.79%). Chaplin is obviously not only the main character but also the discourse topic, as discussed in section 1, since the video is about Chaplin, and this is reflected in the narratives. There are no significant differences between Greek and Spanish natives ($\chi^2 = 1.292$, n.s.), which indicates that the story is being narrated (in terms of characters) in the same way in both languages, a fact that has not been reported in previous studies.

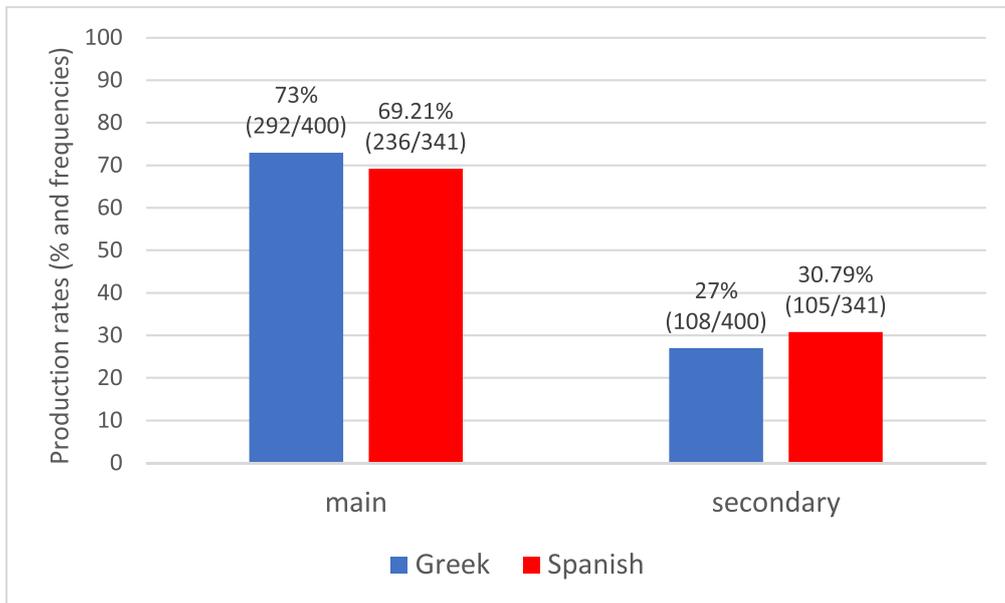


Figure 5: Discourse distribution of the characterhood (main/secondary characters) of all RE forms.

To summarise, findings for RQ1 show that Greek and Spanish natives construct their narratives in a similar way, though the next sections report on differences in their use of REs.

6.2 RQ2: Overall RE forms

As for RQ2, **Figure 6** shows the overall production of RE forms (irrespective of the information status of the RE):

- (i) *Null pronouns*: Greek natives significantly produce more null pronouns than Spanish natives (Greek: 78.34%; Spanish: 63.96%; $\chi^2 = 18.471$, $p < 0.02$, highly sig.). This is in line with previous studies (cf. section 2) since in PAS scenarios, Greek null pronouns typically refer to a subject antecedent but also to a non-subject antecedent, whereas in Spanish null pronouns typically refer to a subject antecedent but may less frequently refer to a non-subject antecedent Spanish. Null pronouns thus show more “flexibility” in Greek than in Spanish in their choice of antecedent (as we will discuss below), hence their higher production in corpus data.
- (ii) *Overt pronouns*: Their production is very low and non-significantly different in both languages (Greek: 3.78%, Spanish: 3.9%; $\chi^2 = 0.008$, n.s.). However, NPs play a more important role in AR, as we discuss next.
- (iii) *NPs*: Their frequency (and their corresponding percentages) is higher than overt pronouns. Additionally, Spanish natives significantly produce more NPs than Greek

natives (Spanish: 32.13%; Greek: 17.88%; $\chi^2 = 19.939$, $p < 0.02$, highly sig.), probably as a result of Greek null pronouns being more flexible than in Spanish. While some of these NPs represent introduction of new characters (i.e., focus new intro), they mainly mark topic shift, as reported above. Therefore, as we will discuss in detail in the next section, NPs are a privileged RE form (in comparison to overt pronouns) to mark topic shift.

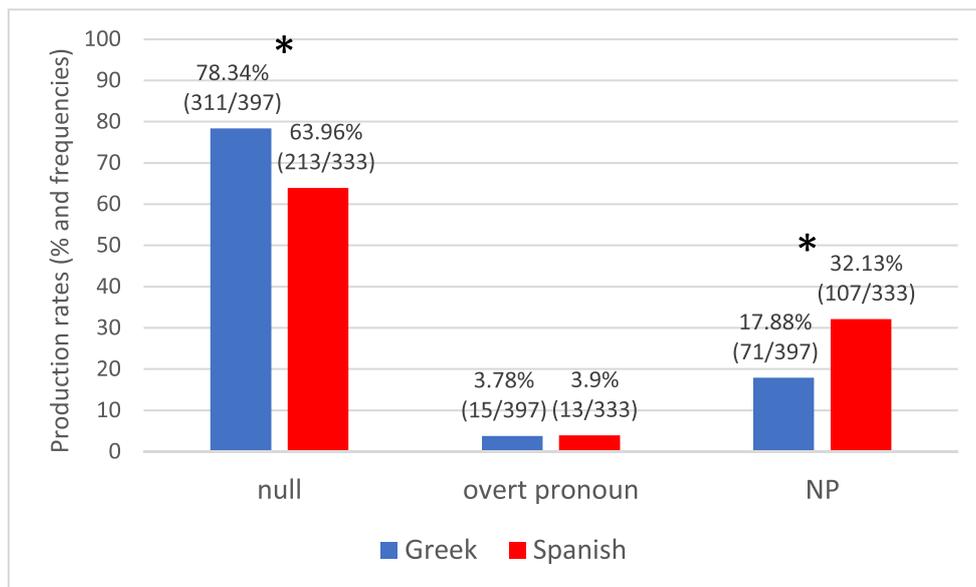


Figure 6: Overall production of RE forms.

In short, null pronouns are the privileged RE form, largely followed by NPs and, to a lesser extent, overt pronouns. NPs have not been the focus of experimental studies but have been examined in corpus studies in native Spanish (e.g., Margaza & Bel 2006; Lozano 2009; Lozano 2016; Vande Castele & Collewaert 2016; Martín-Villena & Lozano 2020) and Greek (Papadopoulou 2020; Charatzidis et al. in press). Importantly, as just reported above, Greek natives significantly differ from Spanish natives since the former produce more null pronouns but less NPs in their narratives. This key difference is partially in line with Giannakou & Sitaridou (2020), who found differences for null pronouns (as we do) and overt pronouns (i.e., Spanish produced more overt pronouns than Greek), but not for NPs. We will explore the reasons behind such crucial difference in the following sections.

6.3 RQ3: The effect of info status (topic continuity/shift)

The analysis for RQ1b was *form x info status* (analysis #1). We focus next on *info status x form* (analysis #2) to answer RQ3, which provides more fine-grained insights.

In topic-continuity scenarios (**Figure 7**), both groups produce almost exclusively null pronouns to mark sentential topic continuation (Greek: 96.85%; Spanish: 89.12%), as expected, and some NPs (Greek: 2.36%; Spanish: 9.84%), with a negligible production of overt pronouns (only 2 tokens in each group). Topic continuity is exclusively marked via null pronouns in native Greek, but via null pronouns and some NPs in Spanish. Our findings are in line with previous experimental studies that showed a clear bias of null pronouns towards subject-antecedents (Spanish: Alonso-Ovalle et al. 2002; Filiaci et al. 2014; Clements & Domínguez 2017; Lozano 2018) (Greek: Papadopoulou 2020; but see Fleva et al. 2017 for different results). This means that null pronouns are clear markers of topic continuity, which was also found in Spanish and Greek corpus studies (Spanish: Lozano 2009; Lozano 2016; Georgopoulos 2017; Martín-Villena & Lozano 2020) (Greek: Charatzidis et al. in press). However, in our data Greek natives produce significantly more null pronouns than Spanish natives ($\chi^2 = 10.805$, $p < 0.02$, highly sig.), but Spanish natives produce significantly more NPs than Greek natives ($\chi^2 = 11.628$, $p < 0.02$, highly sig.), a finding to which we will return in the following section on the effect of character. Contrary to our findings, previous studies comparing these two languages did not report differences in topic continuity (Giannakou & Sitaridou 2020; Margaza & Gavarró 2020).

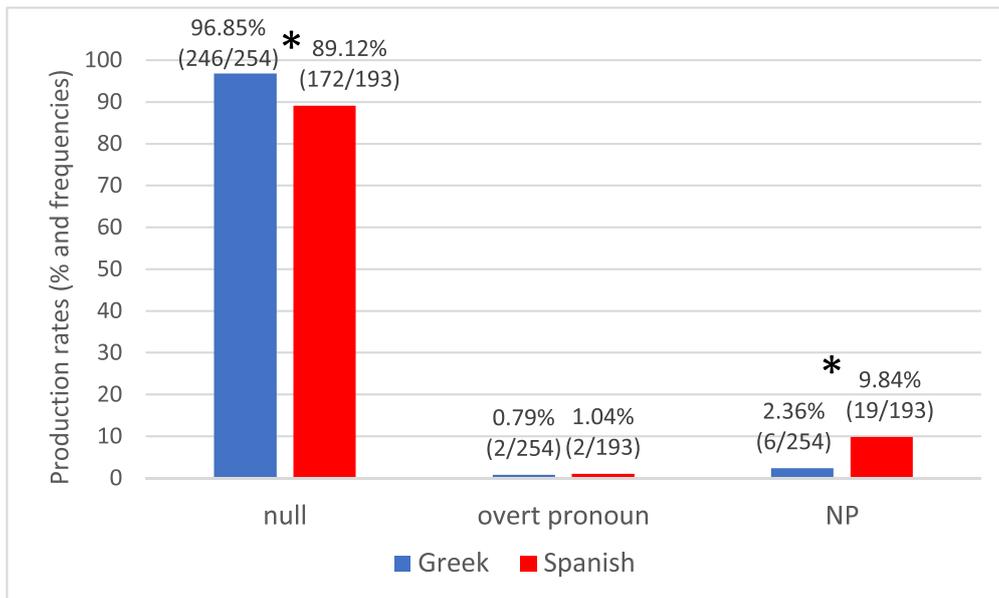


Figure 7: Production of RE forms in Topic Continuity.

In topic-shift scenarios (**Figure 8**), Spanish natives produce more NPs (55.93%) than null pronouns (34.75%), whereas the opposite holds true for Greek natives (null 51.18%, NPs 38.58%). This mirror-image pattern is significantly different for both null pronouns (Greek 51.18% vs. Spanish 34.75%; $\chi^2 = 6.731$, $p < 0.02$, highly sig.) and NPs (Spanish 55.93% vs.

Greek 38.58%; $\chi^2 = 7.392$, $p < 0.02$, highly sig.) but not for overt pronouns (Greek 10.24%, Spanish 9.32%), whose production rates are low in each language. Crucially, these differences diverge from those previously reported. In particular, Giannakou & Sitaridou (2020) found that in topic shift (i) both Spanish and Greek natives produced mostly NPs, though the Greek rates (74.71%) were significantly higher than the Spanish rates (66.52%), which is the opposite from what is observed in our data (38.58% Greek vs 55.93% Spanish); (ii) Spanish natives (8.15%) produced significantly more overt pronouns than Greek natives (1.95%), whereas our data reveal similar (and low) percentages in both languages (9.32% Spanish, 10.24% Greek); (iii) no differences were observed between Greek (23.35%) and Spanish (25.32%) in the production of null pronouns, but we found more null pronouns in Greek (51.18%) than in Spanish (34.75%), a fact that will be accounted for in the general discussion.

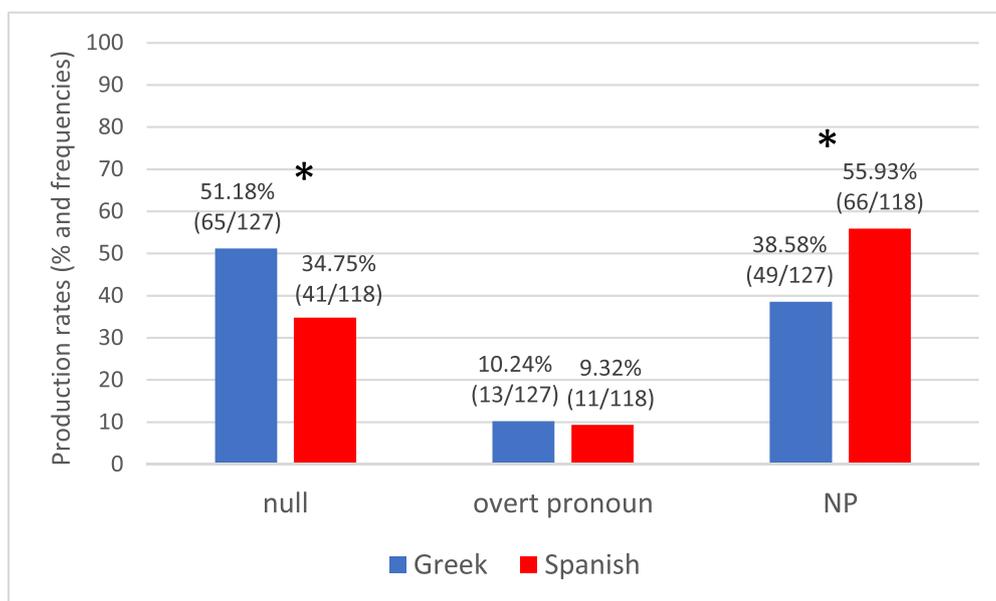


Figure 8: Production of RE forms in Topic Shift.

Interestingly, previous experimental studies in Spanish and some Greek have reported no clear bias of overt pronouns towards non-subject antecedents (Spanish: Alonso-Ovalle et al. 2002; Filiaci et al. 2014; Clements & Domínguez 2017) (Greek: Cunnings et al. 2017; Fotiadou et al. 2020, but see section 2.1.2 for different evidence). We argue that this is a reflection of the low frequency and unprivileged status of overt pronouns to mark topic shift in both languages (below 10%). By contrast, our high frequency of NPs in topic shift in Spanish (and to a less extent in Greek) is in line with previous corpus studies (Spanish: Lozano 2009; Lozano 2016) (Greek: Charatzidis et al. in press), which reveals the privileged status of NPs (instead of overt pronouns) in topic-shift. We will get back to this issue in the general discussion.

In short, the results from topic-continuity and topic-shift scenarios reveal that in Greek null pronouns are more “flexible” in their antecedent selection than in Spanish since they can select a topic (subject) antecedent, as expected, but also a non-topic (non-subject) antecedent more often than in Spanish. The low production of overt pronouns confirms again that, contrary to the experimental literature, overt pronouns are not the privileged RE to mark topic shift, which is rather marked via an NP in Spanish (as previously reported in corpus-based studies), but via an NP and also a null pronoun in Greek, which is a finding not reported in the previous literature. This mirror-image pattern regarding null vs. NP in topic-shift scenarios may be caused by a character effect, a factor to which we turn next.

6.4 RQ4: The effect of characterhood and info status

Regarding RQ4a (Figure 9), both Greek and Spanish predominantly use null subjects to refer to the main character of the story (*Chaplin*), some NPs (i.e., repeated proper name *Chaplin*), as well as overt pronouns in rare cases. Null pronoun rates are significantly higher in Greek (87.59%) than in Spanish (73.08%) ($\chi^2 = 17.545$, $p < 0.02$, highly sig.), the opposite being true for NPs (Spanish 23.93% vs. Greek 10.34%; $\chi^2 = 17.329$, $p < 0.02$, highly sig.). Although both Greek and Spanish native speakers predominantly produce null pronouns to refer to the main character, it appears that reference to the main character by means of a repeated proper noun (Chaplin) is also an option for Spanish natives. We refer to this tendency as MCE, which is stronger in Spanish than in Greek (a fact to which we will return in this section when we discuss main- vs. secondary-character effects).

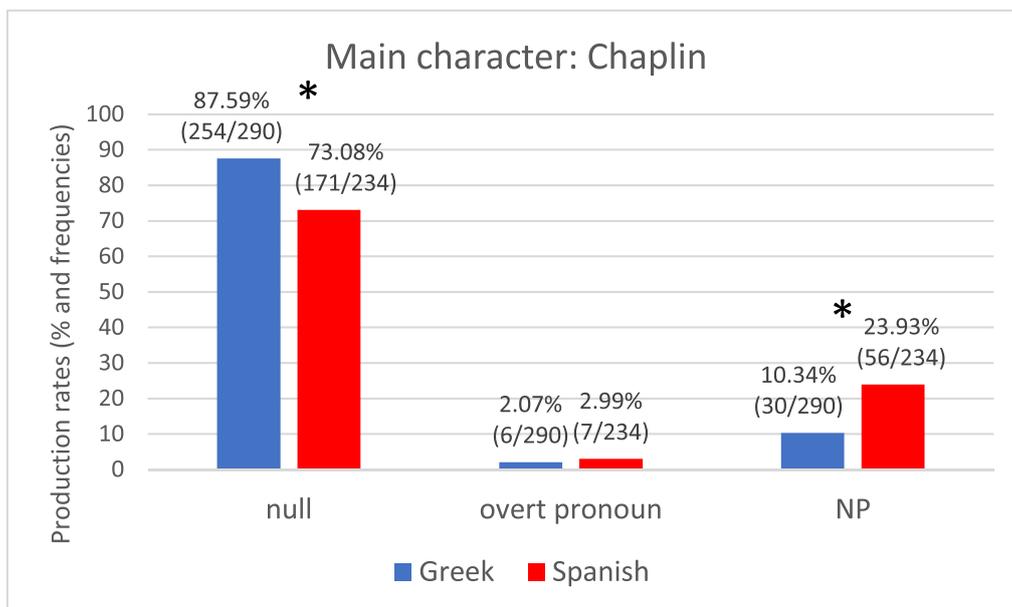


Figure 9: Production of RE for the main character.

The division of labour when referring to secondary characters is less clear-cut, since both null pronouns and NPs are used (**Figure 10**), but a mirror-image pattern emerges again: null pronouns (53.27%) are higher than NPs (38.32%) in Greek, but the inverse pattern holds in Spanish (NPs 51.52%, null pronouns 42.42%). Overt pronoun production is very low again. Once again, no significant differences between languages are found ($p > 0.10$ in all cases).

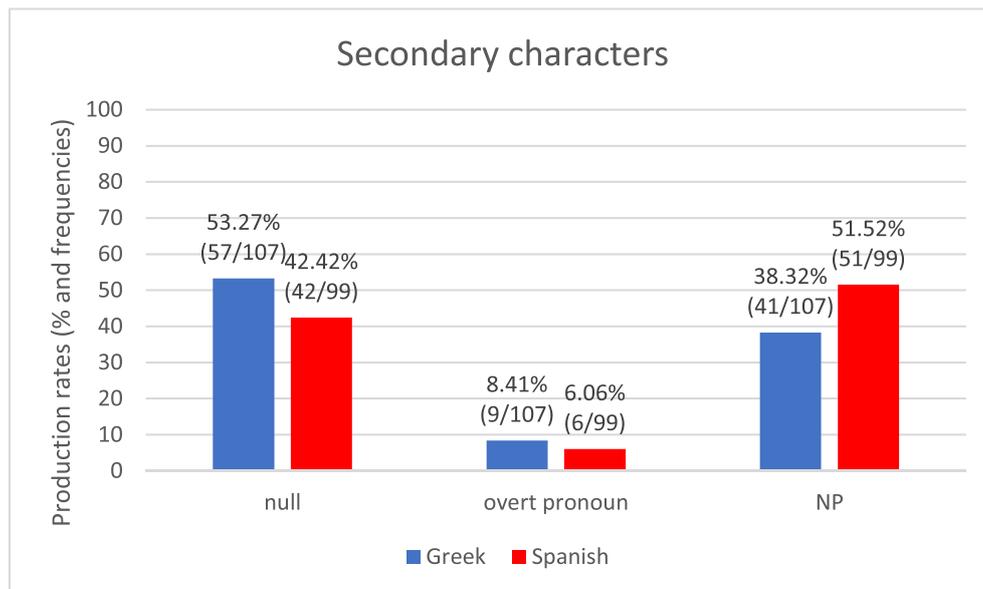


Figure 10: Production of RE for the secondary characters.

In both languages, secondary characters show a higher % of NPs than the main character (cf. **Figure 9** and **Figure 10**). This may be triggered by focus new introduction (FNI) since there are more FNIs for secondary characters than for the main character Chaplin, i.e., every new secondary character needs to be introduced in the story via an NP, but the main character is introduced only once at the beginning of the narrative. We conducted an additional analysis and found that this tendency is not caused by FNI since, when FNI is excluded from the analysis, the rates are similar to the results just reported in the previous paragraph: null (Greek 58.16%; Spanish 47.73%), overt (Greek 9.18%, Spanish 6.82%), NP (Greek 31.63%, Spanish: 42.05%), with no significant differences. So, it appears that, after all, there is a specific trend regarding reference to secondary characters in the narratives of Greek and Spanish natives.

We analyse next the characterhood effect (main/secondary) in each information-status scenario: topic continuity vs. topic shift (RQ4b). In topic continuity (**Figure 11**), analyses reveal that:

- (i) Null pronouns clearly encode topic continuity in both languages, independently of whether they refer to Chaplin (**Figure 11**, first half), or to secondary characters (**Figure 11**,

second half), or even to all characters together (**Figure 7**). Differences between languages are significant only for the main character ($\chi^2 = 13.776$, $p < 0.02$, highly sig.).

- (ii) The production of overt pronouns is extremely low, which again confirms that overt pronouns are not a popular RE form to resolve anaphora.
- (iii) Spanish natives significantly produce more repeated Ns to refer to the main character in topic continuity than Greek natives do (cf. first half of **Figure 11**: $\chi^2 = 14.621$, $p < 0.02$, highly sig.). This was also reported in **Figure 9**. This effect is diluted when it comes to secondary characters (**Figure 11**, second half), since NP rates are similar (and not significantly different) in Greek and Spanish. So, the proportion of repeated Ns reflect a MCE in Spanish, but not in Greek.
- (iv) The raw frequencies for secondary characters are lower than for the main character, which is expected in topic-continuity scenarios since the narrative is about Chaplin (**Figure 11**).

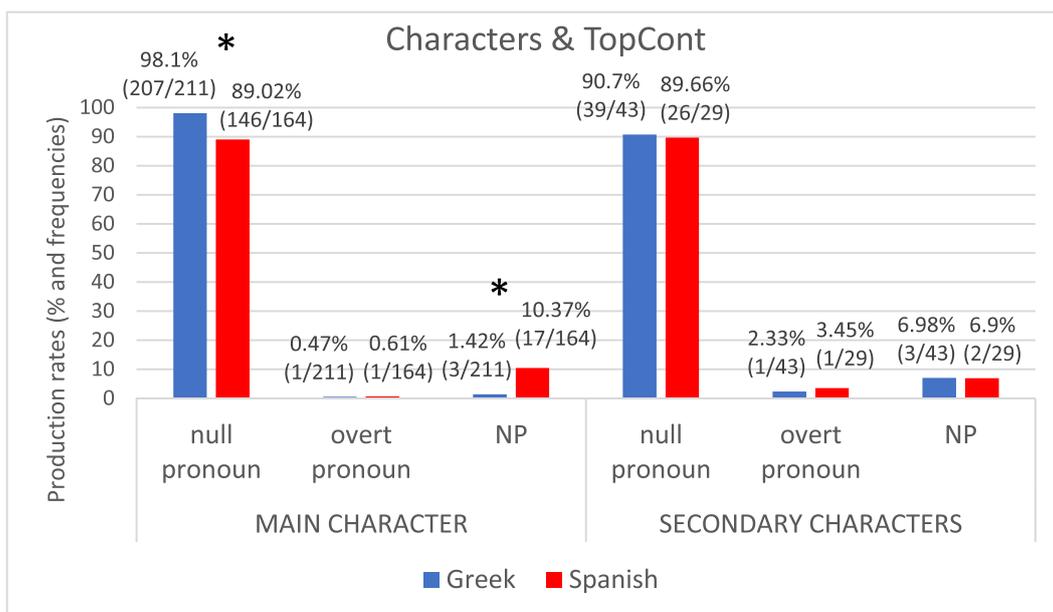


Figure 11: Production of RE for main vs. secondary characters in Topic Continuity.

In topic-shift scenarios (**Figure 12**), between-language differences are significant again for the main character (null pronouns and NPs), but not for secondary characters:

- (i) When referring to the main character (Chaplin), we observe a clear difference between languages. When shifting the topic, a null pronoun is used in Greek (64.38%) significantly more than in Spanish (40.32%) ($\chi^2 = 7.798$, $p < 0.02$, highly sig.), but a repeated N is preferred in Spanish (50%) significantly more than in Greek (28.77%) ($\chi^2 = 6.382$, $p <$

0.02, highly sig.). In short, when shifting the topic to Chaplin, Greek natives produce a null pronoun but Spanish natives produce a repeated N.

- (ii) When referring to secondary characters, both groups produce mostly a repeated NP to shift the topic (Greek: 51.85%; Spanish: 62.5%) and some null pronouns as well (Greek: 33.33%; Spanish: 28.57%) and a few overt pronouns (Greek: 14.81%, Spanish: 8.93%). No significant differences were found between the two languages.

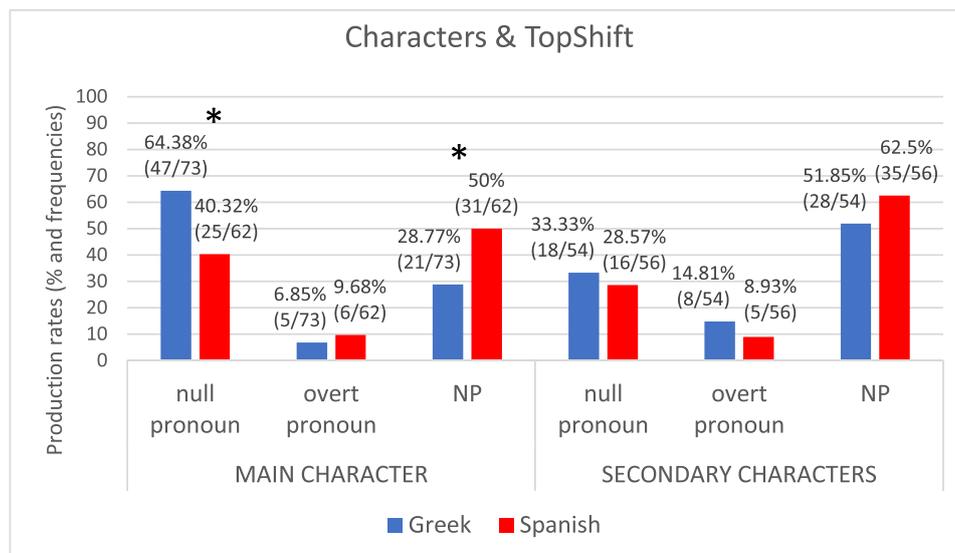


Figure 12: Production of RE for main vs. secondary characters in Topic Shift.

Our analysis of topic continuity/shift in relation to main/secondary characters reveals (i) no differences between Greek and Spanish natives when it comes to secondary characters (a null pronoun encodes topic continuity but an NP marks topic shift), but (ii) clear differences regarding the main character (*Chaplin*): null pronouns are used to continue the topic in both languages (with Spanish natives also using some NPs), but a shift in topic to Chaplin is marked via a null pronoun in Greek but via a repeated N in Spanish. Therefore, secondary characters are more “aseptic” than the main character. Crucially, these findings are novel in our understanding of Spanish vs. Greek anaphora, since character effects have not been previously reported. These differences between Spanish and Greek are further discussed next.

6.5 Summary of results

Our results can be summarised as follows:

[RQ1] *Narration and characterhood*: Spanish and Greek natives construct narrations similarly in terms of (i) the information status of all RE forms (topic continuity is more frequent

than topic shift), (ii) the division of labour for each RE (in terms of their topic continuity/shift, focus new intro), and (iii) characterhood (the main character being more frequent than secondary characters). This is a robust finding since the same task was administered to both groups. This similarity may be a reflection of universal cognitive strategies speakers use to construct and cohere narrations around main/secondary characters, in line with Centering Theory.

[RQ2] *Overall distribution of RE forms*: Overt pronominal subjects are infrequent in both languages. Null pronouns are more frequent in Greek than Spanish while NPs are more frequent in Spanish than Greek.

[RQ3] *Information status*: In both languages, overt pronouns are not a privileged form to mark topic shift. Null pronouns are more “flexible” in Greek than in Spanish, as they specialise in topic continuity but, crucially, are also the prevalent form in topic shift. In Spanish, it is NPs that are more “flexible” than in Greek since they specialize in topic shift, though they can occasionally signal topic continuity. These effects, however, are modulated by characterhood and topichood, as we will see next.

[RQ4] *Characterhood and topichood*: Importantly, one of the key differences between Greek and Spanish lies in the way *discourse* topic (as opposed to *sentential* topic) is marked. In Greek null pronouns signal *discourse* topic (Chaplin) both in topic continuity and even in topic shift, which explains the significant difference between Spanish and Greek for null pronouns when referring to Chaplin, but not when referring to secondary characters.

In the next section, we postulate the Type of Topic Hypothesis (TTH), in an attempt to theoretically capture the similarities and differences across the two languages.

7 General discussion and concluding remarks

Our contrastive corpus-based study reveals both similarities and differences between Spanish and Greek that have not been previously accounted for in a systematic and unified way. For example, the deictic behaviour of Greek overt subject pronouns as compared to the Spanish overt pronouns (see Giannakou & Sitaridou 2020) cannot explain the dataset presented here, because the frequencies of the overt subject pronouns are extremely low and their referential patterns are parallel in both languages. To account for our Greek and Spanish corpus data, we put forward the TTH.

In particular, we found similarities in narrative construction but differences in AR realization in discourse despite Spanish and Greek being null-subject languages. Such differences are not necessarily a reflection of micro-parametric syntactic variation, as assumed in some recent

studies (Giannakou & Sitaridou 2020), because otherwise the variation should be observed across the board (i.e., in both main and secondary characters), which runs against the findings in our study. Rather, Spanish/Greek differences are more discursive than syntactic in nature and emerge only when there is reference to the main character (discourse topic): Greek natives produce null pronouns to refer to *Chaplin* irrespective of its information status (topic continuity/shift), whereas Spanish natives produce a null pronoun in topic continuity but an NP in topic shift, irrespective of whether they are talking about the *discourse* topic (main character *Chaplin*) or secondary characters.

Languages have been typologically classified into topic-prominent languages vs. subject-prominent languages (cf. Li & Thompson 1976 for the original formulation and Paul and Whitman 2017 for an updated overview and discussion). For example, Chinese and Japanese have been argued to represent topic-prominent languages whereas English and other Indo-European languages (e.g., Spanish and Greek) are typically classified as subject-prominent languages. In subject-prominent languages like Spanish and Greek, our corpus results crucially reveal that *discourse* topic constrains AR in Greek more than in Spanish, but *sentential* topic constrains AR in Spanish more than in Greek. Thus, we propose the Type of Topic Hypothesis (TTH), which postulates that in subject-oriented languages like Spanish and Greek there is a tension between discourse-topic vs. sentential-topic oriented languages. The TTH (**Figure 13**) captures the following ideas:

Narratives: Narratives are constructed in the same way in both languages in terms of the discourse distribution of the information status (topic continuity/shift) of all RE forms; the information status (topic continuity/shift) of each RE form; and the characterhood (main/secondary characters) of all RE forms. By contrast, the realization of REs varies as a result of the discourse-topic orientation of Greek vs. the sentential-topic orientation of Spanish. We develop this idea below.

Secondary characters: When it comes to referring to secondary characters, and thus non-discourse topics, both languages show the same AR behaviour: a null pronoun marks topic continuity (i.e., sentential topic is marked via a null pronoun in both languages), whereas an NP (and not an overt pronoun) marks a topic shift.

Main character (*Chaplin*): Several observations are in order:

(i) **Null pronouns:** In Greek, null pronouns refer to the main character (*Chaplin*) significantly more frequently than in Spanish, in both topic continuity (98.1% Greek vs. 89.02% Spanish) and particularly in topic shift (64.38% Greek vs. 40.32% Spanish), as we saw earlier. So, null pronouns appear to have a stronger discourse-topic orientation in Greek than in Spanish.

Additionally, the behaviour of null pronouns to mark topic continuity is different in the two languages. Spanish uses null pronouns to mark topic continuity irrespective of character status (main 89.02% or secondary 89.66%; $\chi^2 = 0.010$, $p > 0.05$, non sig.), whereas Greek uses null pronouns in topic continuity to refer to the main character (98.1%) significantly more often than to the secondary character (90.7%) ($\chi^2 = 6.424$, $p < 0.02$, highly sig.). Since the main character coincides with the discourse topic (*Chaplin*), Greek null pronouns appear to be the default RE not only for sentential topics but also for discourse topics. Thus, discourse topicality seems to play a bigger role in Greek than in Spanish, so that is why Greek could be considered more of a discourse-topic language more than Spanish.

Regarding the behaviour of null pronouns to mark topic shift, we also find differences between the two languages. There seems to be a slight tendency for Spanish null pronouns to refer to the main character (40.32%) more often than to the secondary character (28.57%) in topic shift, but crucially this 11.75% difference is not significant ($\chi^2 = 1.792$, $p > 0.05$, non sig.). By contrast, Greek null pronouns are used to refer to the main character (64.38%) more frequently than to the secondary character (33.33%), the difference (31.05%) being highly significant ($\chi^2 = 11.977$, $p < 0.02$). Thus, null pronouns in topic shift are more heavily constrained by discourse referential properties in Greek than in Spanish.

Thus, in both topic continuity and topic shift scenarios, null pronouns are constrained by discourse topicality more in Greek than in Spanish, but by sentential topicality more in Spanish than in Greek. However, note that arguing that sentential topicality plays a stronger role in the choice of null pronouns in Spanish than in Greek, does not imply that discourse topicality does not play a role at all in Spanish, as the percentages above show. Instead, what we have is rather a gradience, as the dotted arrow in **Figure 13** illustrates. So, our data suggest that reference in human languages is not categorized as either fully discourse-oriented or fully sentential-oriented, but rather languages are closer to (or farther away from) one edge of the continuum than the other.

(ii) NPs: As we saw in **Figure 12**, there is a significant difference between Greek (28.77%) and Spanish (50%) in topic shift when referring to the main character (*Chaplin*) only, but not to secondary characters (51.85% and 62.5% respectively).

Additionally, the behaviour of NPs to mark topic shift is differently modulated by character-hood in the two languages. In Spanish, characterhood does not modulate the use of NPs since these are not significantly different for the main character (50%) vs. secondary characters (62.5%) ($\chi^2 = 1.865$, $p > 0.05$). By contrast, characterhood significantly modulates NPs in Greek: main character (28.77%) vs secondary characters (51.85%) ($\chi^2 = 6.980$, $p < 0.02$). Thus, as just seen with null pronouns, discourse topicality appears to be a more significant factor in Greek than in Spanish in relation to NPs. Once again, Greek behaves more like a discourse-topic oriented language than Spanish.

	Discourse-Top orientation (Greek)	Sentential-Top orientation (Spanish)
Main char. (Chaplin)	\emptyset [\pm TopCont]	\emptyset [+TopCont] NP [-TopCont]
Secondary chars.	\emptyset [+TopCont] NP [-TopCont]	
Narrative construction	<i>Info status of all REs:</i> Top Cont > Top Shift <i>Info status of each RE:</i> \emptyset (Top Cont > Top Shift); overt & NP (Top Shift > Top Cont) <i>Characterhood:</i> Main char > Secondary chars	

Notes:

1. [+TopCont] = Topic Continuity.
2. [-TopCont] = Topic Shift.
3. The '>' symbol implies 'more frequent than'.
4. The dark shaded area represents divergence across languages, whereas the light shaded areas represent convergence.

Figure 13: TTH (Discourse-topic vs Sentential-topic orientation).

The TTH is an empirically falsifiable hypothesis that can be put to the test in languages other than Spanish and Greek in future research. Importantly, researchers should ensure that, in doing so, all factors must be kept constant (the task and text type, the annotation scheme, the tagging procedure, and the profile of the natives), while maintaining as variable the natives' mother tongue. We leave further corroboration of the TTH for future research.

Data availability

The corpora used as the main source of data in this study are freely available/downloadable online in accordance to the Open Science and FAIR (Findable, Accessible, Interoperable, Reusable) principles:

CEDEL2 corpus: <http://cedel2.learnercorpora.com>

GLC corpus: https://glc.lit.auth.gr/app/GLC_Gateway

Ethics and consent

Corpus data collection was approved by their respective ethics committees: CEDEL2 corpus (Universidad de Granada, *Comité de ética en investigación humana* ('Human Research Ethics Committee'), authorization no. 1165/CEIH/2020 and 1794/CEIH/2020), GLC corpus (Aristotle University of Thessaloniki, *Επιτροπή Ηθικής και Δεοντολογίας της Έρευνας, ΕΗΔΕ* (Research Ethics Committee) authorization no. 62503/2022). The corpus informants read an informed consent form and voluntarily agreed to participate anonymously in the study.

Funding information

This research was funded by the following agencies/institutions:

CEDEL2 corpus (ANACOREX R&D project no. PID2020-113818GB-I00 and ANACOR R&D project no. FFI2016-75106-P granted to the first author) funded by MINECO (Ministerio de Economía y competitividad), AEI (*Agencia Estatal de Investigación*), [10.13039/501100011033](https://doi.org/10.13039/501100011033), Spain and ERDF (European Regional Development Fund, A way of making Europe).

GLC corpus (*LAL2A* project no. 3161) funded by the Hellenic Foundation for Research & Innovation, Greece, granted to Alexandros Tantos, the fourth author being a member of the research project.

Competing interests

The authors have no competing interests to declare.

References

- Alonso-Ovalle, Luis. 2006. *Disjunction in alternative semantics*. Amherst: University of Massachusetts. (Doctoral dissertation).
- Alonso-Ovalle, Luis & Fernández-Solera, Susana & Frazier, Lyn & Clifton, Charles. 2002. Null vs. overt pronouns and the Topic-Focus articulation in Spanish. *Italian Journal of Linguistics* 14(2). 151–170.

- Amvrazis, Nikolaos. 2016. Anaphora resolution in Serbian learners of Greek. Germany. (Presented at the International Symposium on Bilingual Processing in Adults and Children 16, Germany.)
- Argyri, Efrosyni & Sorace, Antonella. 2007. Crosslinguistic influence and language dominance in older bilingual children. *Bilingualism: Language and Cognition* 10(1). 79–99. DOI: <https://doi.org/10.1017/S1366728906002835>
- Ariel, Mira. 1990. *Accessing Noun-Phrase Antecedents*. London and New York: Routledge.
- Carminati, Maria Nella. 2002. *The processing of Italian subject pronouns*. Amherst: University of Massachusetts. (Doctoral dissertation).
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, Charles N. (ed.), *Subject and Topic*, 25–55. New York: Academic Press.
- Chamorro, Gloria. 2018. Offline interpretation of subject pronouns by native speakers of Spanish. *Glossa: a journal of general linguistics* 3(1). DOI: <https://doi.org/10.5334/gjgl.256>
- Chamorro, Gloria & Sorace, Antonella & Sturt, Patrick. 2015. What is the source of L1 attrition? The effect of recent L1 re-exposure on Spanish speakers under L1 attrition. *Bilingualism: Language and Cognition* 19(3). 520–532. DOI: <https://doi.org/10.1017/S1366728915000152>
- Charatzidis, Andreas & Georgopoulos, Athanasios & Papadopoulou, Despina & Tantos, Alexandros. in press. Anaphora resolution in L1 Greek: A corpus-based study. In Fotiadou, Georgia & Tsimpli, Ianthi Maria (eds.), *Individual Differences in Anaphora Resolution: Language and Cognitive Effects*. John Benjamins.
- Clements, Maria & Domínguez, Laura. 2017. Reexamining the acquisition of null subject pronouns in a second language: Focus on referential and pragmatic constraints. *Linguistic Approaches to Bilingualism* 7(1). 33–62. DOI: <https://doi.org/10.1075/lab.14012.cle>
- Cummings, Ian & Fotiadou, Georgia & Tsimpli, Ianthi Maria. 2017. Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the Visual World Paradigm. *Studies in Second Language Acquisition* 39(4). 621–652. DOI: <https://doi.org/10.1017/S0272263116000292>
- de Rocafiguera, N. (2023). Information Structure and Language Dominance in Subject Pronoun Resolution in Catalan-Spanish Bilingualism [PhD dissertation, Universitat Pompeu Fabra]. <http://hdl.handle.net/10803/687754>
- Dimitriadis, Alexis. 1996. When pro-Drop Languages Don't: On Overt Pronominal Subjects in Greek. In Dobrin, Lise M. & Singer, Kora & McNair, Lisa (eds.), *Proceedings of CLS 32*, vol. 2, 33–47. Chicago: Chicago Linguistic Society.
- Filiaci, Francesca. 2010. Null and overt subject biases in Spanish and Italian: A cross-linguistic comparison. *Selected proceedings of the 12th Hispanic linguistics symposium*, 171–182.
- Filiaci, Francesca & Sorace, Antonella & Carreiras, Manuel. 2014. Anaphoric biases of null and overt subjects in Italian and Spanish: a cross-linguistic comparison. *Language, Cognition and Neuroscience* 29(7). 825–843. DOI: <https://doi.org/10.1080/01690965.2013.801502>
- Fleva, Eleni & Fotiadou, Georgia & Katsiperi, Eleni & Peristeri, Eleni & Mastropavlou, Maria & Tsimpli, Ianthi Maria. 2017. Language experience and memory effects in anaphora resolution in Greek. In Escobar, Linda & Torrens, Vicenç & Parodi, Teresa (eds.), *Language Processing and Disorders*, 75–91. Newcastle: Cambridge Scholars Publishing.

- Fotiadou, Georgia & Muñoz, Ana I. Pérez & Tsimpli, Ianthi Maria. 2020. Anaphora resolution and word-order across adulthood: Ageing effects on online listening comprehension. *Glossa: a journal of general linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.997>
- García-Alcaraz, Estela. 2015. *Comprensión y producción de los pronombres nulos y explícitos de tercera persona en posición de sujeto en la adquisición temprana del español L2*. Barcelona: Universitat Pompeu Fabra. (Doctoral dissertation).
- García-Alcaraz, Estela & Bel, Aurora. 2019. Does empirical data from bilingual and native Spanish corpora meet linguistic theory? The role of discourse context in variation of subject expression. *Applied Linguistics Review* 10(4). 491–515. DOI: <https://doi.org/10.1515/applirev-2017-0101>
- Georgopoulos, Athanasios. 2017. *Anaphora in the interlanguage of English and Greek learners of L2 Spanish: A study based on the CEDEL2 corpus*. Granada: Universidad de Granada. (Doctoral dissertation).
- Giannakou, Aretousa & Sitaridou, Ioanna. 2020. Microparametric variation in the syntax of Spanish and Greek pronominal subjects. *Glossa: a journal of general linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.960>
- Givón, Thomas. 1983. *Topic Continuity in Discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/tsl.3>
- Gordon, Peter C. & Grosz, Barbara J. & Gilliom, Laura A. 1993. Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science* 17(3). 311–347. DOI: https://doi.org/10.1207/s15516709cog1703_1
- Grosz, Barbara J. & Joshi, Aravind K. & Weinstein, Scott. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*. Cambridge, MA: MIT Press 21(2). 203–225. DOI: <https://doi.org/10.21236/ADA324949>
- Gundel, Jeanette K. 1999. On different kinds of focus. In Bosch, Peter & Sandt, Rob van der (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*, 293–503. Cambridge University Press.
- Kaltsa, Maria & Tsimpli, Ianthi Maria & Rothman, Jason. 2015. Exploring the source of differences and similarities in L1 attrition and heritage speaker competence: Evidence from pronominal resolution. *Lingua* 164. 266–288. DOI: <https://doi.org/10.1016/j.lingua.2015.06.002>
- Liceras, Juana M. & de la Fuente, Anahí Alba & Martínez Sanz, Cristina. 2010. The distribution of null subjects in non-native grammars: Syntactic markedness and interface vulnerability. In Iverson, Michael & Ivanov, Ivan & Judy, Tiffany & Rothman, Jason & Slabakova, Roumyana & Tryzna, Marta (eds.), *Proceedings of the 2009 Mind/Context Divide Workshop*, 84–95. Somerville MA: Cascadilla Press.
- Lozano, Cristóbal. 2002. The interpretation of overt and null pronouns in non-native Spanish. *Durham Working Papers in Linguistics* 8. 53–66.
- Lozano, Cristóbal. 2009. Selective deficits at the syntax-discourse interface: evidence from the CEDEL2 corpus. In Snape, Neal & Leung, Yan-kit Ingrid & Sharwood Smith, Michael (eds.), *Representational Deficits in SLA: Studies in honor of Roger Hawkins*, 127–166. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/iald.47.09loz>

- Lozano, Cristóbal. 2016. Pragmatic principles in anaphora resolution at the syntax-discourse interface: Advanced English learners of Spanish in the CEDEL2 corpus. In Alonso-Ramos, Margarita (ed.), *Spanish Learner Corpus Research: State of the Art and Perspectives*, 236–265. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/scl.78.09loz>
- Lozano, Cristóbal. 2018. The Development of Anaphora Resolution at the Syntax-Discourse Interface: Pronominal Subjects in Greek Learners of Spanish. *Journal of Psycholinguistic Research* 47(2). 411–430. DOI: <https://doi.org/10.1007/s10936-017-9541-8>
- Lozano, Cristóbal. 2022. CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research* 38(4). 965–983. DOI: <https://doi.org/10.1177/02676583211050522>
- Lozano, Cristóbal & Díaz-Negrillo, Ana & Callies, Marcus. 2021. Designing and compiling a learner corpus of written and spoken narratives: The corpus of English as a Foreign Language (COREFL). In Bongartz, Christiane & Torregrossa, Jacopo (eds.), *What's in a Narrative? Variation in Story-Telling at the Interface between Language and Literacy*, 21–46. Bern: Peter Lang.
- Lozano, Cristóbal & Quesada, Teresa. in prep. Anaphora resolution in native and L2 Spanish: What corpus data reveal about the Position of Antecedent Strategy (PAS).
- Margaza, Panagiota & Bel, Aurora. 2006. Null Subjects at the Syntax-Pragmatics Interface: Evidence from Spanish Interlanguage of Greek Speakers. In O'Brien, Mary Grantham & Shea, Christine & Archibald, John (eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, 88–97. Somerville, MA: Cascadilla Press.
- Margaza, Panagiota & Gavarró, Anna. 2020. Null/overt subject alternations in L2 Spanish and L2 Greek. *Glossa: a journal of general linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.969>
- Martín-Villena, Fernando & Lozano, Cristóbal. 2020. Anaphora resolution in topic continuity: Evidence from L1 English–L2 Spanish data in the CEDEL2 corpus. In Ryan, Jonathon & Crosthwaite, Peter (eds.), *Referring in a Second Language. Studies on Reference to Person in a Multilingual World*, 119–141. New York: Routledge. DOI: <https://doi.org/10.4324/9780429263972-7>
- Miltsakaki, Eleni. 2007. A rethink of the relationship between salience and anaphora resolution. In Branco, António & McEnery, Tony & Mitkov, Ruslan & Silva, Fátima (eds.), *Proceedings of the 6th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2007)*. Porto: Centro de Linguística de Universidade do Porto.
- Montrul, Silvina & Rodríguez Louro, Celeste. 2006. Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In Torrens, Vincent & Escobar, Linda (eds.), *Language Acquisition and Language Disorders*, vol. 41, 401–418. Amsterdam: John Benjamins Publishing Company. DOI: <https://doi.org/10.1075/lald.41.19mon>
- O'Donnell, Michael James. 2008. The UAM CorpusTool: Software for corpus annotation and exploration. *Proceedings of the XXVI Congreso de AESLA*. 1433–1447.
- Papadopoulou, Despina. 2020. Reference patterns in language comprehension & production: the “Greek case”. In Bongartz, Christiane & Torregrossa, Jacopo (eds.), *What's in a narrative? Variation in Storytelling at the Interface Between Language and Literacy*, 147–182. Berlin: Peter Lang.

- Papadopoulou, Despina & Peristeri, Eleni & Plemenou, Evagelia & Marinis, Theodoros & Tsimpli, Ianthi. 2015. Pronoun ambiguity resolution in Greek: Evidence from monolingual adults and children. *Lingua* 155. 98–120. DOI: <https://doi.org/10.1016/j.lingua.2014.09.006>
- Prentza, Alexandra & Tsimpli, Ianthi Maria. 2013. Resolution of Pronominal Ambiguity in Greek: Syntax and Pragmatics. In Stavrou-Sifakim Melita & Michlopoulou, Stamatia (eds.) *Studies in Greek Linguistics 33, Proceedings of the 33rd Annual Meeting of the Department of Linguistics, School of Philology, Faculty of Philosophy, A.U.Th*, 197–208. Thessaloniki: Institute of Modern Greek Studies.
- Quesada, Teresa & Lozano, Cristóbal. 2020. Which factors determine the choice of referential expressions in L2 English discourse?: new evidence from the COREFL corpus. *Studies in Second Language Acquisition* 42(5). 959–986. DOI: <https://doi.org/10.1017/S0272263120000224>
- Silva-Corvalán, Carmen. 1994. *Language Contact and Change: Spanish in Los Angeles*. Clarendon Press.
- Tantos, Alexandros & Amvrazis, Nikolaos & Drakonaki, Elena. to appear. Greek Learner Corpus II: Design and development of an online corpus for L2 Greek. *Journal of Applied Linguistics*.
- Tantos, Alexandros & Papadopoulou, Despina & Charatzidis, Andreas. 2014. Discourse and grammatical effects on pronoun resolution in Greek. In Kotzoglou, Georgios & Nikolou, Kalomoiria & Karatzola, Eleni & Frantzi, Katerina & Galantomos, Ioannis & Georgalidou, Marianthi & Kourti-Kazoullis, Vasileia & Papadopoulou, Ch. & Vlachou, Eleni (eds.), *Selected Papers from the 11th International Conference on Greek Linguistics*, 1655–1667. Rhodes: Laboratory of Linguistics of the SE Mediterranean.
- Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 110(5). 315–341. DOI: [https://doi.org/10.1016/S0024-3841\(99\)00044-3](https://doi.org/10.1016/S0024-3841(99)00044-3)
- Torregrossa, Jacopo & Andreou, Maria & Bongartz, Christiane M. 2020. Variation in the use and interpretation of null subjects: A view from Greek and Italian. *Glossa: a journal of general linguistics* 5(1). DOI: <https://doi.org/10.5334/gjgl.1011>
- Tsimpli, Ianthi Maria & Sorace, Antonella & Heycock, Caroline & Filiaci, Francesca. 2004. First Language Attrition and Syntactic Subjects: A Study of Greek and Italian Near-Native Speakers of English. *International Journal of Bilingualism* 8(3). 257–277. DOI: <https://doi.org/10.1177/13670069040080030601>
- Van Dijk, Teun Adrianus van. 1977. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse*. Longman.
- Vande Castele, Ann & Collewaert, Kim. 2016. The use of referring expressions in Spanish language learners' oral productions: an exploratory study. In Almeida, Francisco Alonso & Ortega Barrera, Ivalla & Quintana Toledo, Elena & Sanchez Cuervo, Margarita (eds.), *Input a Word, Analyze the World: Selected Approaches to Corpus Linguistics*, 415–424. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Wolf, Florian & Gibson, Edward & Desmet, Timothy. 2004. Discourse coherence and pronoun resolution. *Language and Cognitive Processes* 19(6). 665–675. DOI: <https://doi.org/10.1080/01690960444000034>

