Chapter 9

Early use of null and overt subjects in L2 Spanish: Evidence from two oral tasks

Laura Domínguez[a] & María J. Arche[b]

[a]University of Southampton [b]University of Greenwich

Recent research has shown that advanced English learners of Spanish can successfully acquire the syntactic, pragmatic and referential properties of null and overt subjects. However, acquiring these structures is problematic at beginner and at intermediate stages of acquisition for these learners. In this study, we investigate the emergence and development of null and overt subjects by 60 English learners of Spanish (20 beginners, 20 intermediate and 20 advanced) in order to understand why these forms are initially difficult to acquire. The oral data for this study were collected using a paired-discussion task and a story retell and are freely available from the SPLLOC project (www.splloc.soton.ac.uk). We argue that the cline of difficulty suggested by Cho & Slabakova (2014), based on whether L1-L2 form-meaning mismatches require reassembly and whether a dedicated morpheme is available, makes appropriate predictions for these structures. We also argue that the type of task used to elicit the oral data and the overall linguistic and narrative abilities of the learners are also likely to influence the rate of use of these forms.

1 Introduction

In Spanish, subjects can be overtly pronounced (1a) or can be null (i.e. not phonetically realised) as in example (1b). Both structures are grammatically correct but they are not felicitous in the same contexts. For instance, the overt pronoun *yo* 'I' in example (1a) usually marks a change in the referent or topic in the discourse or it can signal that it is the speaker (and not someone else) who is going to go to the theatre (i.e. the subject is in contrast with other possible subjects). These

two pragmatic functions are not available when the subject pronoun is null as in (1b).

(1)   a.   Yo voy   a ir   al   teatro.
            I   go.PRES.1SG to go to.the theatre
            'I am going to go to the theatre.'

      b.   Voy   a ir   al   teatro.
            go.PRES.1SG to go to.the theatre
            'I am going to go to the theatre.'

Previous research on the second language acquisition of Spanish subjects has shown that, although English learners find it difficult to use and interpret these forms at the early stages, they eventually acquire them (Pérez-Leroux & Glass 1999, Liceras & Díaz 1999, Lozano 2002, 2006, Hertel 2003, Montrul 2004, Montrul & Rodríguez-Louro 2006, Belletti et al. 2007, Margaza & Bel 2006, Rothman & Iverson 2007, Domínguez 2013, Pladevall Ballester 2013, Clements & Domínguez 2017). The question which remains unresolved is what makes the acquisition of null and overt Spanish subjects a difficult area particularly at the start of the acquisition process.

In this study, we investigate the emergence and development of null and overt subjects in Spanish by three groups of English speakers learning Spanish in an instructed setting in the UK. We focus on the oral production of a group of young beginner (13–14 years old) and intermediate (16–17 years old) learners and compare them with the behaviour of advanced students majoring in Spanish at university level (final-year undergraduate students) and a group of 15 native speakers in Spain of similar ages. Data elicited through oral tasks by young beginner and intermediate groups are scarce in the L2 literature on this topic. The current study aims to fill this gap by analysing oral data elicited by two tasks provided by the Spanish Learner Language Oral Corpora (SPLLOC, www.splloc.soton.ac.uk) (see Mitchell et al. 2008). The datasets have been put together according to principles proposed by Myles (2005) in her pioneering work championing the use of L2 corpora to investigate relevant theoretical questions on L2 development (Myles 2004, 2005, 2007). The results arising from the current study also complement those discussed in Domínguez (2013) on the use of null and overt subjects from the same SPLLOC corpus and students in a semi-spontaneous interview.

In our analysis, we assume that null and overt subjects are constrained by similar discourse-contextual restrictions and thus pose similar processing demands for learners (see Domínguez 2013). Following Slabakova's (2009) and Cho & Slabakova's (2014) proposals on what makes a structure more or less difficult

to acquire we argue that, in contrast to some previous research, null and overt subjects are both potentially difficult to acquire as they do not represent straightforward form-meaning mappings which these learners can transfer from English; a second prediction is that overt subjects are, however, likely to initially pose an additional challenge because they require feature reassembly, whereas null subjects do not.

2 Spanish subjects

2.1 Syntactic properties

Identifying the syntactic principles that regulate the distribution of subjects in languages like Spanish is a very complex issue which has been under debate for decades (see Sheehan 2016 for an overview). For our purposes, it is necessary to understand how exactly the syntax of Spanish subjects differs from English so we can establish the acquisition task required for this structure.

While there is agreement in the field on the status and distribution of English subjects, there is no such consensus for Spanish subjects. Essentially, English subjects are assumed to be generated within the verbal domain (i.e. VP) and then move to the specifier of the Inflectional Phrase (IP) (also known as Tense Phrase (TP)). The specifier of TP position (i.e. [Spec TP]) is considered an A position (i.e. an Argument position). Movement to this site is justified as a way of satisfying the extended projection principle (EPP) which requires some phrase with nominal features to occupy the Tense position. The EPP requirement has been formalised as a feature, (the so-called EPP feature), encoded in Tense since Chomsky 1995. The agreement phi features (number and person) in Tense are considered to be uninterpretable and they get valued by the interpretable phi features of a pronominal (or full DP) subject (Chomsky 1995, 2000, Holmberg 2005, Sheehan 2016, Roberts 2010, a.o.), which is always overt. That is, the [Spec TP] is a position which always needs to be filled, which results in the obligatory presence of preverbal subjects in English. In Spanish, the situation is different. As mentioned above, subjects can be null (i.e. not phonetically realised) and, when overt, can appear either pre- or post-verbally. The discussions about the factors and views on the morpho-syntactic underpinnings of such distribution have mainly focused on the properties of the agreement morphology, the conceivable lack of an EPP requirement in Tense and the possibility that overt subjects do not even actually sit in the same position as they do in English (i.e. [Spec TP]), when they appear preverbally. Below, we summarise the main perspectives about the availability of null subjects (i) and the position of overt subjects (ii).

2.1.1 The status of the null subjects

Some authors (Barbosa 1995, 2009, Alexiadou & Anagnostopoulou 1998) argue that the rich agreement morphology suffices to satisfy the EPP in T(ense); under this account, a null subject (i.e. pro) is not needed and lexical subjects do not have to be in Spec TP. Others propose that there is a null pro arguably occupying [Spec TP]. In either case, there are a few new features (or characteristics of features) that an English learner will have to acquire: regulating the movement of the verb to T (instead of the one that results in T lowering to V as it is assumed in English); the potential syntactic consequences of a rich morphology, namely, that they satisfy EPP in T and no movement is syntactically necessary, and maybe the realisation of a null pro.

Example (2) shows how complex the relation between agreement morphology and overt/null subjects can be.

(2) Los lingüistas disfrut-an   /disfrut-amos /disfrut-áis   con una
     the linguists enjoy-PRES.3PL /enjoy-PRES.1PL /enjoy-PRES.2PL with a
     coma.
     comma
     'The linguists/us linguists/you linguists enjoy a comma.'

In this example, there is a third person plural DP which can co-appear not only with a third person plural agreement form (as usual) but with a first or with second person plural form as well, contrary to expectation (Torrego & Laka 2015, Villa-García 2018). The lack of agreement between the phi features of the overt DP (*los lingüistas*) and the morphology shown on the verb in the latter two cases suggests that the DP cannot be the element satisfying all the relevant features in T. That is, the DP cannot check the verbal morphological phi features. One way to account for this is to assume that a null pro (with a set of phi features different from those shown by the DP but agreeing with those shown on the verb) occupies the [Spec TP] position and values the phi features in the verbal agreement. For our purposes, the co-occurrence of a pro with an overt lexical DP suggests that achieving a full command of subject distribution in Spanish entails more than the mastery of pro as a null subject. It cannot be reduced to a dichotomy "overt DP vs pro" since both may occur at the same time.

2.1.2 The position of overt subjects

Spanish overt subjects can appear pre- or post-verbally. In order to account for the postverbal position, most authors nowadays assume that subjects are spelled

out in their original position. For example, in cases such as (3a), with un unerga-tive verb, the DP subject is argued to occupy the [Spec vP] position; in contrast, in (3b), with an un-accusative verb, the DP is argued to remain in the position which is known as the "sister" to the verb.

(3)   a. Ha llora-do Marta.
              has cry-PTCP Marta
              'Marta has cried.'

        b. Ha llega-do   Marta.
              has arrive-PTCP Marta
              'Marta has arrived.'

For the preverbal position, there does not seem to be full consensus regard-ing the exact position the overt DP occupies. While English preverbal subjects are deemed to consistently occupy [Spec TP] (an A-position), some authors have argued that overt preverbal subjects in Spanish occupy a discourse sensitive po-sition, which would be an A-bar (i.e. non-argument) position. In support of this hypothesis, Alexiadou & Anagnostopoulou (1998) mention that only in null sub-ject languages adverbs can occur between the verb and the subject as shown in the Spanish/French contrast below. The contrast in grammaticality can be accounted for by assuming that subjects in Spanish occupy an A-bar position, higher than [Spec TP].

(4)   a. Spanish
              Juan ya   quier-e   ir-se.
              Juan already wants-PRES.3SG go=REFL
              'Juan wants to leave already.'

        b. French
              Jean {*déjà} veu-t   {déjà} s'en   aller.
              Jean already want-PRES.3SG already REFL.CL go
              'Jean wants to leave already.'

However, it has also been pointed out that it may not be the case that all sub-jects occupy such an A-bar position. For instance, SVO structures in out of the blue contexts or with wide scope in response to a question such as "What hap-pened?" (see 5), suggest an analysis of overt subjects roughly equivalent to the English position in [Spec TP].

(5)   Marta ha   comprado un libro.
       Marta has bought   a   book
       'Marta has bought a book.'

Goodall (2001) discusses evidence suggesting that not all preverbal subjects in Spanish are actually left dislocated. Examples in (6) from Goodall (2001) show a contrast between an undisputable left dislocated phrase (6a) and a subject (6b). Clauses with fronted topics are islands for extraction and result in ungrammaticality, whereas clauses with preverbal subjects are not. If the subject Juan in (6b) was to be considered to occupy a left dislocated position (instead of [Spec TP]), the contrast would remain unexplained.

(6)   a.   * A quién cre-es   que el premio se   lo   dieron?
            to whom think-PRES.2SG that the prize   3.DAT 3SG.ACC gave
            'Who do you think that the prize they gave it to?'

      b.   A quién cre-es   que Juan le   dio el premio?
            to whom think-PRES.2SG that Juan 3SG.DAT gave the prize
            'Who do you think Juan gave the prize to?'

Villa-García (2018) also concludes that Spanish preverbal subjects may be in TP or above. One of the pieces of evidence he shows is based on bare NPs. These seem disallowed in positions that can be argued to be [Spec TP], as shown in the contrast between (7a) and (7b), but are grammatical in unequivocally topic positions, as shown in (7c). This points to the conclusion that the overt subject in (7b) is in [Spec TP].

(7)   a.   * Niños juga-ban   en la playa.
            Kids   play-IPFV.3PL on the beach
            (intended) 'Kids were playing on the beach.'

      b.   Los niños juga-ban   en la playa.
            The kids   play-IPFV.3PL on the beach
            'The kids were playing on the beach.'

      c.   Niños, no creo que jueg-uen   muchos en la playa.
            Kids   not think that play-PRES.3PL.SUBJ many   on the beach
            'As for kids, I do not think many play on the beach.'

This particular issue goes beyond the scope of this paper, but based on these examples and the comprehensive overviews of subject positions in Sheehan (2016) and Villa-García (2018), the evidence for the type of position that overt preverbal subjects occupy in Spanish is mixed and different constructions seem to favour different analyses. It may be the case that not all apparently preverbal subjects are located in the same syntactic position. The important point is that the distribution of overt subjects poses a rather complicated task for an L2 learner of

Spanish whose native language is English. The array of structures available from the input is not uniform and it seems to entail the acquisition of new syntactic features that regulate a complex picture concerning the distribution of subjects. We can conclude that a one-to-one mapping between English and Spanish cannot be established for overt subjects and that this may be a difficulty for learners.

2.2 Pragmatic and referential properties of null and overt subjects

The distribution of overt and null subject pronouns is dependent on discourse-contextual factors, mainly to help maintain continuity in the discourse. Generally, overt subjects are preferred in contexts signalling a change of referent, contrast (i.e., contrastive focus) or emphasis (see 8b), whereas null subjects are preferred if the subject can be properly identified in the discourse (see 8a) (Luján 1985, 1986, Fernández Soriano 1989, Alonso-Ovalle & D'Introno 2001).

(8)   a.   Ayer jugué al tenis con mi hermano. *Pro* Se enfadó cuando *pro* perdió.
          'Yesterday I played tennis with my brother. (He) got upset when he lost.'
      b.   Ayer jugué al tenis con Juan y Marta. *Ella* es muy buena pero *él* tiene que practicar más.
          'Yesterday I played tennis with Juan and Marta. She is very good but he has to practice more.'

Sorace (2000) has proposed that the pragmatic distinction between null and overt subjects can be captured by the [+/− topic shift] feature. In her analysis, overt subjects introduce a new referent in what she refers to a [+ topic shift] context, and thus carry a [+ topic shift] feature. This is, however, not the full picture as native speakers of Spanish have been found to use null subjects to introduce new referents in [+ topic shift] contexts in informal conversations quite often (Silva-Corvalán 2001, Blackwell 2003, Lubbers Quesada & Blackwell 2009, Liceras et al. 2010, Domínguez 2013, Clements & Domínguez 2017). Lubbers Quesada & Blackwell (2009) discuss the complexity surrounding the pragmatic and referential properties of null and overt subjects and conclude that both forms can be used in the same contexts. For instance, in the following example from the SPLLOC project a native speaker of Spanish (NS6) in a conversation with one of the researchers chooses to use a null *tú* 'you' as a generic or impersonal referent. The null pronoun is used even though this could be considered a [+ topic shift] context:[1]

---

[1] The symbols [/] and [//] are used in the transcriptions to signal interruptions in the oral speech.

(9) NS6: Yo cuando llegué aquí *pro* estaba un poco así solo y tal y entonces con los españoles cuando *pro* los ves *pro* te [//] te *pro* cierras más y te [//] *pro* se te queda como grupo de amigos. Básicamente *pro* salimos por ahí también a tomar algo, cenar, *pro* hacemos excursiones para ver el país.

'When I got here (I) *was* a bit alone and then with the Spaniards when (you) *see* them (you) *focus* on them and (you) *are left* with a group of friends. Basically, (we) also *go out* to eat something, have dinner (we) *go on* trips to explore the country.' (example from Domínguez 2013)

A quantitative analysis of the uses of null and overt subjects reported in Domínguez (2013) reveals that 14.3% of the null subjects produced by the native Spanish speakers are indeed used in what Sorace would consider to be [+ topic shift] contexts. This corroborates the argument that these forms can be used in both types of pragmatic contexts. Lubbers Quesada & Blackwell (2009) also suggest that null subjects can be used as epistemic parentheticals, expressions which do not bring the referent into focus. The example below shows null subjects used as epistemic parentheticals by a native Spanish speaker (NS5) as reported by Domínguez (2013):

(10) NS5: sí *pro* estamos aquí en verano allí *pro no sé pro* tiene que ser al [//] justo al contrario o no?

'Yes, (we) are here in the summer. Over there (I) *don't know* (it) has to be just the opposite isn't it?'

Lubbers Quesada & Blackwell (2009) also argue that overt pronouns are often used even though they are not introducing a new referent. This applies in particular to *yo* "I", which the authors argue can be used with speech act verbs of claiming, belief, opinion, emotion, or knowledge "to add pragmatic weight to an utterance, to take a firmer stance, to express a greater stake in, or emotional commitment to your assertion or to express that your utterance is highly relevant" (Lubbers Quesada & Blackwell 2009: 122). This non-referential use of *yo* was also found in the native data of the SPLLOC corpus Domínguez (2013), shown in the following example:

(11) NS6: bueno *yo creo* que todos los que estudiamos Historia eh la salida de profesor es una [//] es una opción.

'Well, *I think* that for all of us who study History–eh–becoming a teacher is an option.'

In this example, the overt pronoun *yo* is optional. It does not mark a change of referent and thus does not carry a [+ topic shift] feature. Thus, both null and overt subjects can be used to introduce a new referent [+ topic shift] (see details Domínguez 2013 and Clements & Domínguez 2017). As shown in Table 1, null and overt subjects can be used in both [+/− topic shift] contexts as well as in in non-referential settings as epistemic parentheticals and to add pragmatic weight.

Table 1: Summary of pragmatic and referential properties of null and overt subjects from (Domínguez 2013)

	[+ topic shift]	[− topic shift]	Non-referential
Null subjects	Yes	Yes	Epistemic parenthetical (e.g. *No sé, digo*)
Overt subjects	Yes	Yes	Pragmatic weight (e.g. *Yo creo*)

In summary, null subjects are subject to similar contextual and pragmatic restrictions as overt subjects and can be used in [+topic shift] contexts, too. Since both null and overt subjects can be used in an array of pragmatic contexts, it is difficult to distinguish between null and overt subjects based on whether they carry a pragmatic feature or not. Consequently, it is also difficult to predict whether learners may find one form more problematic than another based on the pragmatic status of each of the forms (see Clements & Domínguez 2017).

3 Previous research on the acquisition of Spanish null and overt subjects

It is well documented that even though the acquisition of Spanish subject expression is somewhat problematic for some learners, advanced English speakers are able to behave target-like in an array of tests and tasks (Liceras 1988, Phinney 1987, Pérez-Leroux & Glass 1999, Liceras & Díaz 1999, Lozano 2002, 2006, Hertel 2003, Montrul 2004, Montrul & Rodríguez-Louro 2006, Belletti et al. 2007, Rothman & Iverson 2007, Domínguez 2013, Pladevall Ballester 2013, Clements & Domínguez 2017). Most of these studies have elicited and analysed comprehension or judgement data. Whether the same results would be obtained from oral data elicited through different task types remains an open question which this study directly addresses.

The first studies investigating the acquisition of null and overt subjects in Spanish were interested in testing whether English speakers could successfully reset the value of the null subject parameter (NSP) (Chomsky 1981, Jaeggli 1982, Jaeggli & Safir 1989, Rizzi 1982, 1986) to the correct setting (Spanish instantiates the + option whereas English instantiates the – option). These early studies focused on the acquisition of the null pronoun pro as this is the form which is not available in English (see review in Domínguez 2013). Al-Kasey & Pérez-Leroux (1998) found that English speakers may initially transfer the value of the setting from English to Spanish. According to this evidence, the resetting of the [+] value of the NSP may not be as straightforward as initially argued by authors such as Phinney (1987). Although acquisition of these properties is achievable, it is not without problem, particularly early on in the process. For instance, some of those early studies revealed a tendency to overuse both overt and null subjects (Almoguera & Lagunas 1993, Díaz & Liceras 1990). Liceras & Díaz (1999) show how the Japanese and Chinese (i.e. [+ topic languages]) learners of Spanish in their study overuse null-subject pronouns, and Almoguera & Lagunas (1993) also report variation in the correct and incorrect use of pro by seven participants. For some of these speakers, the problem was an overproduction of null subjects showing that null subjects can be difficult to acquire as well. Bini (1993) examined the first stages in the acquisition of null and overt subjects in L2 Italian by a group of beginners and a group of low-intermediate Spanish speakers (both Italian and Spanish allow null subjects). Learners initially overuse pronouns during the first six weeks of instruction. Problems shown by an overproduction of null subjects in L2 grammars have in fact been extensively reported in the literature (see Díaz & Liceras 1990, Liceras et al. 1999, Pérez-Leroux & Glass 1999, LaFond et al. 2001, Montrul & Rodríguez-Louro 2006, Rothman & Iverson 2007, Lubbers Quesada & Blackwell 2009). Cases of underproduction of null subjects (Lozano 2009) as well as individual variation in their use amongst the least proficient learners (Liceras & Díaz 1999, Rothman & Iverson 2007) have been reported, as well. These studies show early problems with the acquisition of null subjects in L2 Spanish.

Pérez-Leroux & Glass (1999) and Liceras & Díaz (1999) correctly pointed out that an examination of the acquisition of pragmatic constraints is necessary in order to understand the acquisition of null/overt subjects, as first argued by White (1989) and Liceras (1988, 1989). Liceras & Díaz (1999) argued that even though the use of null subjects may be in place from early on, their status in interlanguage grammars may not be the same as in native grammars, in particular with regard to the mechanisms that learners employ to identify them (as well as overt pronouns) in discourse (see also Lozano 2002, 2006, Hertel 2003, Montrul 2004, Montrul & Rodríguez-Louro 2006, Pladevall Ballester 2013). An important body

of literature on this topic has shown that overt subject pronouns especially are more difficult to acquire than null pronouns (see Sorace 2004, 2011) as their realisation depends on features that belong to the syntax/pragmatics interface (based on Sorace's [+ topic shift] feature).

Accordingly, Sorace & Filiaci (2006) argue that when acquiring overt pronouns learners access inadequate processing resources or "shallow" parsing strategies, which indicates a processing problem that linger even at advanced stages of acquisition (see the "interface hypothesis", Sorace 2011). Crucially, null subjects are spared from these problems as they are purely syntactic phenomena according to these authors (see Belletti et al. 2007). An early study which casts some doubt on this claim was presented by Montrul & Rodríguez-Louro (2006). These authors examined whether constraints at the syntax–pragmatics interface are intrinsically more difficult for learners for the acquisition of subjects in L2 Spanish. A crucial point of departure from previous research is that these authors assume that a pragmatic deficit can affect the use of null subjects as well. Their findings show an incremental learning of the appropriate discourse properties of both overt and null subjects which is not expected by the interface hypothesis.

More recently, Domínguez (2013) and Clements & Domínguez (2017) also assume that both null and overt pronouns are subject to similar pragmatic restrictions and that both forms can bear a [+ topic shift] feature as explained in §2. These studies also cast some doubt on the predictions of the interface hypothesis for the acquisition of Spanish pronominal subjects. Domínguez (2013) reports on the oral production of null and overt subjects from the same SPLLOC dataset as in the current study and from the same learners. The data were elicited by means of an interview. Learners show some problems with null subjects that mostly disappear at advanced levels although some learners overproduce and some learners underproduce both forms when compared to native controls. Individual differences were found in the data from the beginner and intermediate groups. Clements & Domínguez (2017) report on data obtained by a group of 20 advanced English learners of Spanish who completed a picture verification task and a context-matching preference task. The results show that these learners allow null subjects in certain [+ topic shift] contexts and that they show less felicitous judgements affecting the use of both overt and null pronouns in some contexts. These authors speculate about the possibility that performance-related problems affect the use of null and overt subjects in context. It is possible that pro may be used as a default form by learners in these cases, a phenomenon also attested in the data of monolingual Spanish children (Grinstead 1998, Villa-García 2013). Furthermore, Pladevall Ballester (2013) also reports that English (instructed) advanced learners of Spanish have problems with both null and

overt subjects. She proposes that processing difficulties and the lack of positive evidence available in the input may be the explanation for these findings.

In conclusion, problems with both null and overt subjects have been observed in the data reported for Spanish learners, particularly during the early stages of acquisition, using mostly judgment data as evidence. Investigating what happens at early stages of acquisition and focusing on oral data can be useful to advance our understanding of the nature of this problem.

4 The role of input and the cline of difficulty in L2 acquisition

When acquiring the properties of null and overt subjects in Spanish, English speakers need to determine whether a similar form exists in their native grammar (for both null and overt subjects) and whether that specific form has the same properties and distribution. This form-meaning mapping can be explained as a form of feature-reassembly as proposed by Lardiere (2005, 2008, 2009) and Hwang & Lardiere (2013). Lardiere assumes that L2 speakers initially transfer their full native grammar and that L2 acquisition involves the mapping of features into the correct functional categories and lexical items. In some cases, and for some properties, this mapping can be done in a straightforward manner but in other cases (as an effect of transfer) a process of feature reassembly is needed. This process entails the effective reconfiguration of L1 syntactic features which do not have the exact same morpholexical expression in the L2. In the case of overt subjects, English learners of Spanish need to figure out that there are key differences in the syntactic properties of overt subjects in these two languages.

Slabakova (2009) acknowledges the role that feature reassembly plays in L2 acquisition but proposes a cline of difficulty of properties dependant on whether the target properties are encoded by a morpheme (these will be easier to acquire) or whether they are fixed by discourse context (more difficult to acquire). Following Ramchand & Svenonius (2008), she assumes a universal syntax/semantics system that feeds the conceptual-intentional interpretational mechanisms. According to these authors, variation exists regarding whether the features are present in the syntax/semantics or whether they are contextually filled. This is the kind of crosslinguistic variation that is relevant for establishing correct form-meaning mappings during second language acquisition. In cases where a certain feature is not morphologically visible, its meaning can or needs to be recovered by the discourse context. The thrust of Slabakova's proposal (see Cho

& Slabakova 2014 as well) is that whether features (which exist in both the native and target grammars) are overtly or covertly expressed has to be taken into account alongside feature reassembly. Thus, the two dimensions which are relevant in predicting whether a certain structure will be easy or difficult to acquire are the need for feature reassembly and whether the form is overtly expressed by a dedicated form or not (i.e. the meaning can be assigned by the context). The easiest scenario is one in which there is a one-to-one relationship between certain dedicated functional morphology and its grammatical meaning. This could be the case of overt subjects in Spanish and Catalan which are overtly realised by a dedicated form in both languages, have similar syntactic and distribution properties and do not require reassembly. At the other end, the feature (F) associated with specificity and shown in definite articles are covertly expressed by discourse means in languages like English and Russian. This would be a hard property to acquire by speakers of these languages according to the cline of difficulty of Cho & Slabakova (2014) as shown in Figure 1.
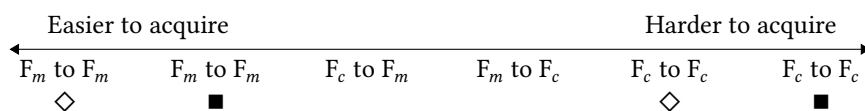
Easier to acquire					Harder to acquire
F_m to F_m	F_m to F_m	F_c to F_m	F_m to F_c	F_c to F_c	F_c to F_c
◇	■			◇	■

Figure 1: Cline of difficulty of acquisition of features by Cho & Slabakova (2014) adapted from Slabakova (2009). F_m = $F_{morpheme}$, F_c = $F_{context}$. ◇: no re-assembly required; ■: re-assembly required

Cho & Slabakova explain that other variables such as the availability of consistent or inconsistent input can make acquisition of new L1-L2 mappings harder. Slabakova (2013) has also argued that problems with certain structures can be linked to the fact that the input provides evidence for alternate structures with similar frequency, and that this can lead to divergence in L2 grammars. Following Papp (2000), Domínguez & Arche (2008, 2014) also argue that problems acquiring new mappings can persist at advanced levels of acquisition if L2 input is non-robust, parametrically ambiguous or simply not transparent or systematic enough. These authors explain that the type of input available for each structure has to be taken into account as well as learners' sensitivity to the frequency and consistency in which a certain structure appears. In the case of acquiring null and overt subjects in Spanish, these are forms that are abundant in the input but less experienced learners may not have had access to all of the scenarios in which a null and an overt subject pronoun can be used in Spanish. It is also possible that the type of evidence needed may not be obvious in the input. Since

the input has evidence of both null and overt pronouns being grammatical in the same position, it is possible to assume that figuring out in which exact context each of these two forms can be used will take some time. According to these observations, it is very likely that learning when to use null and overt subjects in an L2 when these forms are not available in a speaker's native language will be a gradual process which takes time and requires sufficient exposure to the right evidence in the input.

If we take into account the role of feature reassembly in modulating L1-L2 mappings, the role of the input and whether the L1 and L2 express the same structure with a dedicated morphological expression or not, we can predict that both null and overt subjects in Spanish would be somewhat problematic for English speakers but, nevertheless, would not constitute a particularly hard property to be acquired. A second prediction is that overt subjects may take longer to be used properly since they require reassembly (overt subjects also exist in English but with different syntactic characteristics). Null subjects do not require reassembly since there is no form in English which overtly expresses the syntactic features associated with pro. Crucially, we predict problems at the early stages of acquisition where reassembly is starting to take place and when learners have not had abundant exposure to input.

5 The current study

In the current study, we examine the emergence and development of null and overt subjects in the oral data of three groups of L2 Spanish speakers (60 in total) taking into account that both forms can be used in contexts where there is a switch in reference if this is salient enough. The data are part of the Spanish Learner Language Oral Corpora (SPLLOC) project (www.splloc.soton.ac.uk) and are freely available to the research community to investigate the acquisition of Spanish morphosyntactic properties by three groups of English learners in the UK. The whole database contains a total of 333,491 words (269,262 from learners and 64,229 from native speakers) and a total of 561 digital audio files (461 from learners and 100 from native speakers). Details on the rationale and principles for the design of the corpus can be found in Mitchell et al. (2008). The recordings were transcribed using CHAT conventions and analyses were carried out using the CLAN software suite (MacWhinney 1991, 2000). The analysis below was based on those transcripts that had been POS-tagged by means of the Spanish MOR and POST programs. MOR adds a %mor tier to provide a complete part-of-speech tagging for every word in the transcript so that researchers can carry out morohosyntactic analyses on the data. In the current study, we analysed the data elicited by a story retell and a paired-discussion task.

5.1 Participants

The participants were 60 native speakers of English learning Spanish in a school or university in the UK. In order to track the first uses of the target forms, we analysed the data from a group of 20 beginners (13–14 years of age) which at the time of testing were in Year 9 of the UK school system (third year into their secondary school education) and had received around 180 hours of instruction. 20 intermediate students were in Year 13 (the last year of school before university) at the time that the data were collected (ages 17–18). The SPLLOC website shows the accumulative hours of instruction as around 750 for this group. Finally, a group of 20 final-year undergraduate students majoring in Spanish are part of the advanced group (ages 21–22) which had around 895 hours of instruction. The three groups are meant to represent three key stages in the acquisition of a second language in an instructed setting.

The control group was formed by 15 native speakers from Spain of similar ages as the three learner groups. These participants were mainly in Madrid and Alicante when the data were collected although a small number were in Southampton (UK) as they had just arrived in the UK to participate in a period of study abroad.

Only participants who had started learning Spanish in Year 7 (around 11 years of age) and who had declared Spanish as their main foreign language were included in the study. Even though all of the native speakers had had some exposure to English through schooling, none of them considered themselves to be bilingual Spanish-English speakers.

5.2 Tasks

5.2.1 The story-retell task

The story-retell was based on a series of pictures depicting a story in which a family (mum, grandma and three children) go on holiday in Scotland. The story is named the "Loch Ness" story because the characters think that they can see the Loch Ness monster only to find out that grandma had painted some car tyres to make them look like the monster. The last picture depicts the family going into the house and the real monster swimming in the lake. The story had been used successfully in the French Learner Language Oral Corpora (FLLOC www.flloc.soton.ac.uk), a sister site to SPLLOC with the same design principles. Overall, there were 12 colour pictures which clearly depicted the story that participants had to tell. These pictures were chosen with the younger participants in mind and were meant to show a story simple enough that this group of learners could describe. To aid the Year 9 and Year 13 participants (Y9 and Y13 from here

on), a member of the research team read a script of the story in Spanish and had access to a list of main vocabulary words if they needed them. The script was read whilst the learners looked at the pictures to ensure that they had understood the story and had something to say. The task was piloted with native controls and learners of the same proficiencies to ensure that all of the participants would be able to complete the task as planned. Figure 2 shows the first and last pictures of the story that the participants saw.

Figure 2: Two pictures used in the "Loch Ness" task. Illustrations by Alex Brychta for *A Monster Mistake* by Roderick Hunt (Oxford Reading Tree 2003) used by permission of Oxford University Press.

5.2.2 The paired-discussion task

This task was modelled after a similar task used by Dippold (2006). Each participant was presented with a topic which was chosen by the research team for their likelihood to generate discussion (e.g. What can be done to help the environment? How can we help eradicate street violence? etc). Each topic was followed by four propositions of actions that could help solve each of the problems which each participant was asked to rank in order of preference. Participants were also asked to suggest one more solution or proposal to address the issue being discussed. In this task, each participant was paired with another participant from the same proficiency group. Each participant had to defend their ranking of propositions, and both had to work together to agree on a ranking. Only the intermediate group was provided with the translations of key vocabulary items to aid their discussions. This task was designed to offer a high probability of oral productions between the pairs. Due to the demands of this kind of task, in which learners are required to construct and defend an argument in real time, the beginner group was not asked to participate.

6 Results

In this section, we report the results of the production of null and overt subjects (both pronouns and full DPs) by the four groups of participants. The average and median for each form was calculated for both tasks and for each individual task. We first report the combined results for both tasks together. Overall, we see that the number of null and overt subjects increases with proficiency and that the advanced undergraduate group (UG) perform like the controls (N). Table 2 shows the means of use of null subjects for all the participant groups. The Y9 participants use very few null subjects (mean 2.0) when compared with the controls (mean 18.0).[2]

Table 2: Means of use of null subjects (both tasks)

Group	Mean	Conf. level	Trad. lower	Trad. upper
N	18.0	0.95	13.100	23.00
UG	18.2	0.95	13.700	22.60
Y13	8.1	0.95	6.210	9.99
Y9	2.0	0.95	0.814	3.19

Table 3: Means of use of overt subjects (both tasks)

Group	Mean	Conf. level	Trad. lower	Trad. upper
N	19.10	0.95	16.5	21.8
UG	17.20	0.95	14.8	19.6
Y13	9.48	0.95	7.7	11.2
Y9	8.90	0.95	7.1	10.7

Table 3 shows the means of use of overt subjects. Again, we see a difference in use between the native controls and the advanced group on the one hand and the beginner and intermediate groups on the other. The Y9 participants clearly use more overt subjects than null subjects.

The use of null and overt subjects for all the groups is shown in Figures 3 and 4. These figures clearly show that the use of these two forms increases with proficiency and that the advanced speakers show similar rates of use as the controls.

---

[2]The tables show the means, confidence levels and confidence intervals which are indicated by Trad.lower and Trad.upper.

The beginner and intermediate groups show lower use of both forms, particularly for null subjects.

For the null subject results, an independent in-between groups ANOVA yielded a statistically significant effect ($F(3, 123) = 15.82, p < 0.001$). Tukey multiple comparisons of means at 95% confidence level revealed that the only two comparisons which were not statistically significant were between Natives (N) and the advanced group (UG) and between Y13 and Y9 learners (see Table 4). This confirms that the advanced speakers' performance was indistinguishable from that of the controls and that any problems that learners experience using null subjects early on can persist after years of instruction but can be ultimately overcome.

Table 4: Results of the Tukey multiple comparisons (null subjects)

Group	diff	lwr	upr	p adj
UG-N	0.1504762	−6.673095	6.974048	0.9999316
Y13-N	−9.9400000	−16.826930	−3.053070	0.0014794
Y9-N	−16.0400000	−24.143829	−7.936171	0.0000058
Y13-UG	−10.0904762	−16.058373	−4.122580	0.0001332
Y9-UG	−16.1904762	−23.529279	−8.851674	0.0000004
Y9-Y13	−6.1000000	−13.497750	1.297750	0.1440460

Table 5: Results of the Tukey multiple comparisons (overt subjects)

Group	diff	lwr	upr	p adj
UG-N	−1.929524	−6.060342	2.201294	0.6175541
Y13-N	−9.645000	−13.814174	−5.475826	0.0000001
Y9-N	−10.220000	−15.125853	−5.314147	0.0000017
Y13-UG	−7.715476	−11.328290	−4.102662	0.0000009
Y9-UG	−8.290476	−12.733202	−3.847751	0.0000206
Y9-Y13	−0.575000	−5.053411	3.903411	0.9870526

For overt subjects, the independent in-between groups ANOVA also yielded a statistically significant effect ($F(3, 123) = 20.45, p < 0.001$). Tukey multiple comparisons of means at 95% confidence level also revealed the UG-Native comparison and the Y9–Y13 comparison to not be statistically significant as shown in Table 5.

Null subjects by group
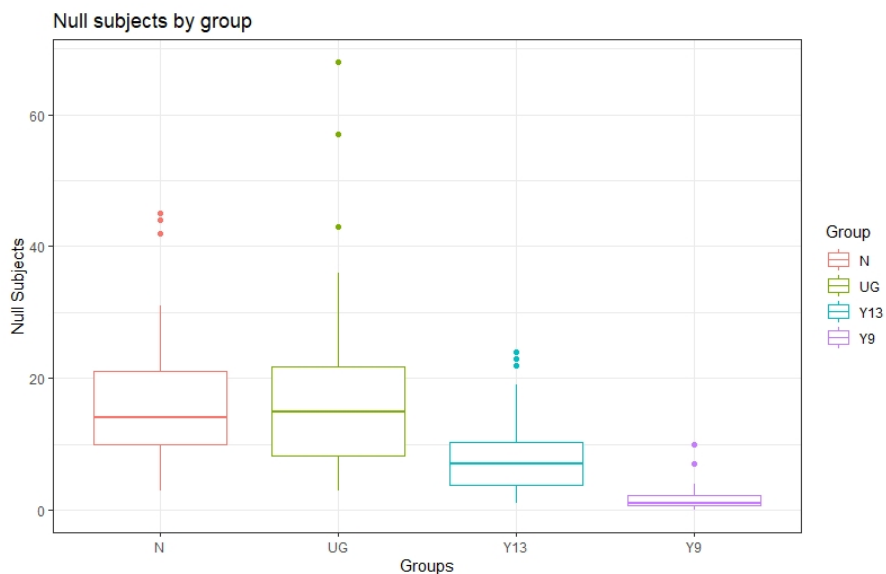
Figure 3: Null subjects for each participant group (both tasks)

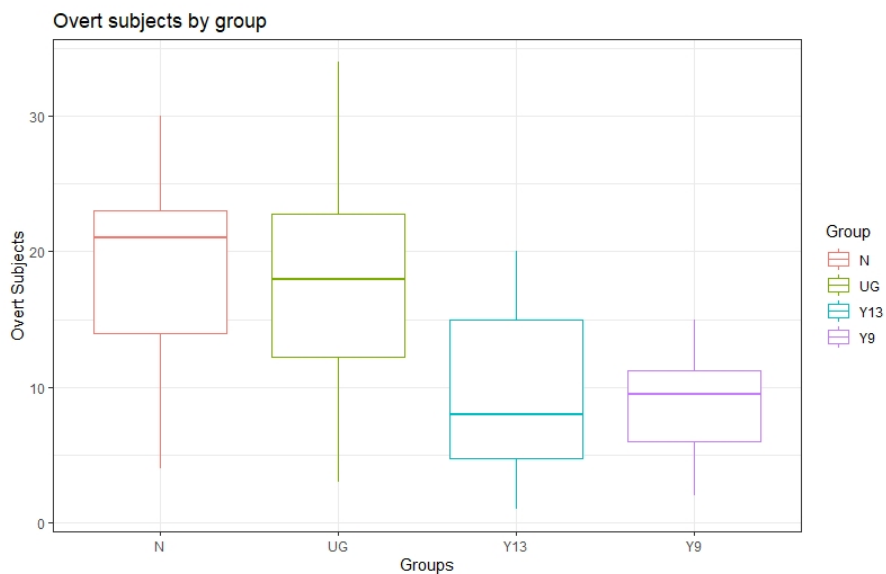**Overt subjects by group**

Figure 4: Overt subjects for each participant group (both tasks)

These results suggest that both null and overt subjects are equally problematic for the beginner and intermediate groups and that targetlike use is ultimately achievable.

The next two tables show the results for each of the two tasks separately (see Table 6 for null subjects and Table 7 for overt subjects).

Table 6: Mean use of Null Subjects

Group	Loch Ness	Paired-discussion
N	11.7	27.6
UG	9.9	25.7
Y13	5.5	10.7
Y9	2.0	-

Table 7: Mean use of Overt Subjects

Group	Loch Ness	Paired-discussion
N	19.6	18.40
UG	20.1	14.50
Y13	13.8	5.15
Y9	8.9	-

Overall, we see that the paired-discussion task elicited more null subjects than the Loch Ness task for all groups which indicates that the type of oral task used to investigate this property can have an effect on the results obtained.

6.1 Loch Ness task

Figures 5 and 6 show the use of null and overt subjects for all the participant groups in this task. The Y9 learners show low production of both target forms, particularly of null subjects. The rates of use of both forms for the advanced group is similar to that found for the controls.

In this task, five Y9 speakers did not produce any null subjects and seven only produced one null subject. In contrast, this is the group in which we find the highest rate of use of null subjects by one single participant (10 instances which is 77% of all of the preverbal subjects they used). This shows that there is variability of use of null subjects at this early stage.
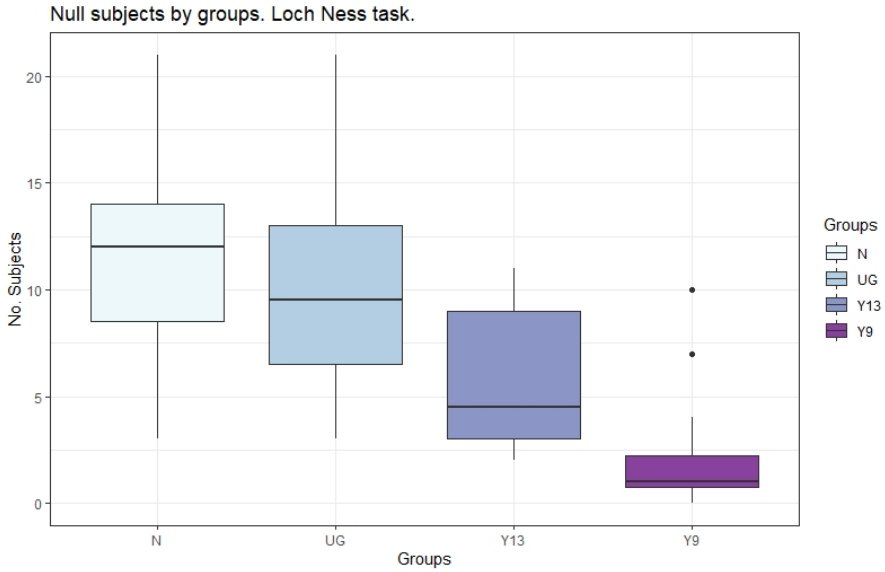
Null subjects by groups. Loch Ness task.

Figure 5: Null subjects for each participant group (Loch Ness task)

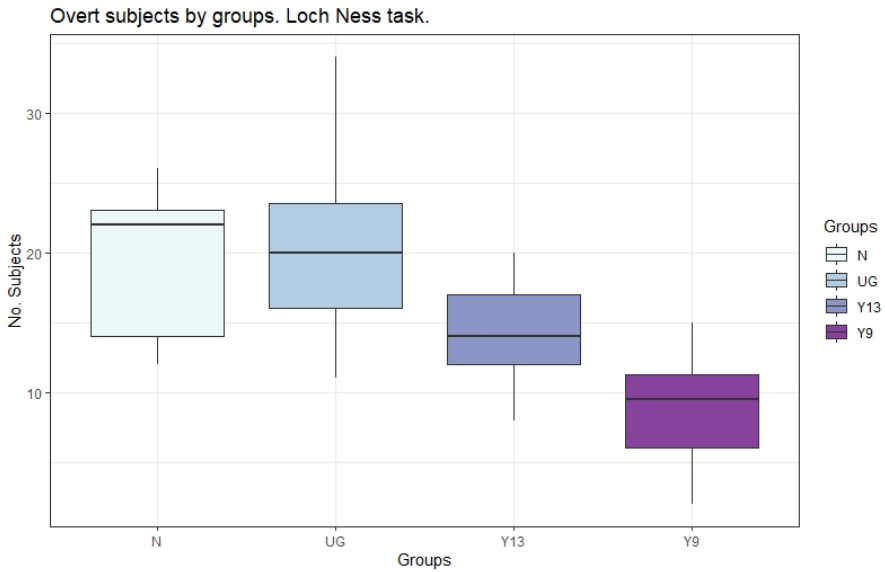Overt subjects by groups. Loch Ness task.

Figure 6: Overt subjects for each participant group (Loch Ness task)

We analysed the use of overt subjects in this task to investigate whether participants preferred to use a pronoun or a full DP. The Y9 speakers did not produce any pronouns and only 0.2% of the overt subjects produced by the Y13 group was a pronoun. This rate of use is 0.5% for the UG group and 0.6% for the controls. This indicates that these two tasks did not elicit high rates of pronominal subjects. This could be explained by the nature of the task as participants based their productions on what was depicted on a series of pictures. It was easy for the participants to move from picture to picture, introducing the third person subject in each picture as a new referent (which does not require the use of a pronoun).

An independent in-between groups ANOVA yielded a statistically significant effect for both null subjects ($F(3, 71) = 21.77, p < 0.001$) and overt subjects ($F(3, 71) = 25.79, p < 0.001$). Tukey multiple comparisons of means at 95% confidence level reveal no significant differences between UG and the native controls. On the other hand, Y9 and Y13 learners have a significantly different pattern of use of both null and overt subjects when compared to the advanced learners and the native controls. Y9 and Y13 are significantly different, too. These results are shown in Table 8 (null subjects) and Table 9 (overt subjects).

Table 8: Results of the Tukey multiple comparisons (null subjects)

Group	diff	lwr	upr	p adj
UG-N	−1.766667	−5.351916	1.8185825	0.5682405
Y13-N	−6.166667	−9.751916	−2.5814175	0.0001369
Y9-N	−9.666667	−13.251916	−6.0814175	0.0000000
Y13-UG	−4.400000	−7.719296	−1.0807042	0.0045702
Y9-UG	−7.900000	−11.219296	−4.5807042	0.0000002
Y9-Y13	−3.500000	−6.819296	−0.1807042	0.0348375

Table 9: Results of the Tukey multiple comparisons (overt subjects)

Group	diff	lwr	upr	p adj
UG-N	0.5	−3.589519	4.589519	0.9883962
Y13-N	−5.8	−9.889519	−1.710481	0.0021115
Y9-N	−10.7	−14.789519	−6.610481	0.0000000
Y13-UG	−6.3	−10.086159	−2.513841	0.0002329
Y9-UG	−11.2	−14.986159	−7.413841	0.0000000
Y9-Y13	−4.9	−8.686159	−1.113841	0.0058873

Next, we show a few samples of the oral productions from the corpus. The next three examples illustrate how three different Y9 learners (L1, L11 and L16) told the same part of the story using different amounts of null subjects. Participant L1 was able to use three null subjects (as indicated by pro) and participant L11 produced two. In contrast, participant L16 did not produce any null subject pronouns. L1 and L11 are able to produce some null subjects as they produced at least two sentences to describe the actions carried out by the same subject.

(12) L1: Hay mucha gente um [/] um al lado del lajo de Loch Ness y *pro* miran el monstruo. Hay mucho fotos y um [/] um *pro* hacen fotos um *pro* pensan que el monstruo es verdad.

'There are many people, ehm ehm next to the Loch Ness and (*they*) look for the monster. There are many pictures and ehm ehm (*they*) take pictures and (*they*) think that the monster is real.'

(13) L11: Muchos personas ven eh [/] eh mucha yente pro ven eh [^ eng: I don't know] el monstruo y *pro* está en el tele eh. Un periodisto eh habla con la abuela y *pro* es en la tele xxx de verdad monstruo está en el lago.

'Many people see ehm ehm many people see eh [^ eng: I don't know] the monster and (*it*) is on tv eh. A journalist eh talks to the grandmother and (*she*) is on tv. Really, the monster is in the lake.'

(14) L16: Eh mucho periodista y eh mucho fotos y eh periodista hace fotos. Eh un chica y un chico eh parecer un Loch Ness monster. Eh un chico [//] no dos chicos y un chica eh un [/] un tele, Loch Ness Monster. eh [/] eh [/] eh abuela ehm nadan no Loch Ness monster [^ eng: it wasn't real] ehm [/] ehm periodista qui qui eh [/] eh un familia gone en un casa.

'Eh many journalists and eh many pictures and eh journalist take pictures. Eh a girl and a boy eh look like the Loch Ness monster. Eh a boy, no, two boys and one girl eh a tv, Loch Ness monster. Eh eh eh the grandmother eh swim, no Loch Ness monster [^ eng: it wasn't real]. Ehm Ehm journalist who who eh eh a family gone in a house.'

Participant L16 seems to avoid the use of null subjects by continuously introducing a new referent (in the form of a full DP) in every sentence. In contrast, example (15) shows data from one of the intermediate learners (L50) who manages to produce a null subject by using a subordinate clause with the same subject as the main clause:

(15)   L50: por la tarde muchas turistas y visitantes vienen ver el monstruo. Muchos [//] muchas de las personas sacar muchos fotos ehm porque *pro* pienso que el mons(truo) [//] el monstruo es real. Ehm los niños mira el monstruo en la tele. La abuelar ehm hablar con unar personar <der la> [/] de la televisión sobre ehm que el monstruo no es de verdad.

'In the afternoon many tourists and visitors come see the monster. Many many of the people take many pictures eh because (*they*) think that the monster is real. Ehm the children see the monster on tv. The grandmother ehm speaks with a person from tv about ehm the monster is not real.'

In the subordinate clause a null subject is preferred as there is no switch of referent from the referent introduced by the main clause. This learner produces the null pronoun because they are able to produce a complex structure which requires the subject to be null. The Y9 learners are not able to orally produce structures with such complexity which, in turn, reduces the chance of using a null subject in this task. Nevertheless, in this example we also see the same learner using shorter and simpler sentences to describe the actions in the pictures. This is a clear example of the mixed nature of the oral productions of learners at this intermediate level of proficiency. For comparison, example (16) shows how an advanced undergraduate student (L70) told the same part of the story using six null subjects:

(16)   L70: después por eso ehm llegan muchos periodistas y ehm muchas personas que *pro* tienen sorpresa> [//] que *pro* están sorprendientes de [/] de lo que ha pasado y eh ahí *pro* están y *pro* sacan muchísimas fotos ehm y [/] y *pro* sí ven [/] eh *pro* ven el monstruo en el lago ehm y por la noche o por la tarde las [//] los niños están en la casa y *pro* dicen ven [/] ven allí está el monstruo en [/] en el lago.

'After that ehm many journalists arrive and ehm many people (*who*) are surprised, (*who*) are surprised of what has happened and eh there (*they*) are and (*they*) take lots of pictures ehm and (*they*) do see the monster, (*they*) see the monster in the lake ehm and at night or in the evening the children are at home and (*they*) say come, come there is the monster in the lake.'

Participant L70 has used a null pronoun every time that the subject was not introducing a new referent. This is possible as the learner goes on to describe what a character does after they have been introduced in the discourse. This is a strategy which the less proficient learners hardly ever used, as we saw in

examples (12), (13) and (14) and which reduced the contexts in which pro would be preferred. In this respect, the type of task seems to have conditioned the use of null subjects especially for the beginner learners.

6.2 Paired-discussion task

This is the task that elicited the highest rate of null subjects for all groups. Data from the Y9 group are not available as this task was deemed too difficult for them to complete. Figures 7 and 8 show the use of null and overt subjects for the two learner groups and the controls. The Y13 learners show a lower rate of production of both forms, but particularly of null subjects, compared to the other two groups. Both forms are used at a similar rate for the control and advanced groups.

An independent in-between groups ANOVA yielded a statistically significant effect for both null subjects ($F(2, 49) = 9.644, p < 0.001$) and overt subjects ($F(2, 49) = 15.83, p < 0.001$). Tukey multiple comparisons of means at 95% confidence level reveal no significant differences except for the UG-Native control comparison for both null and overt subjects (see Table 10 for null subjects and Table 11 for overt subjects).

Table 10: Null subjects

Group	diff	lwr	upr	p adj
UG-N	−1.872727	−13.41503	9.669580	0.9188738
Y13-N	−16.900000	−28.62127	−5.178731	0.0029598
Y13-UG	−15.027273	−24.37761	−5.676935	0.0008842

Table 11: Overt subjects

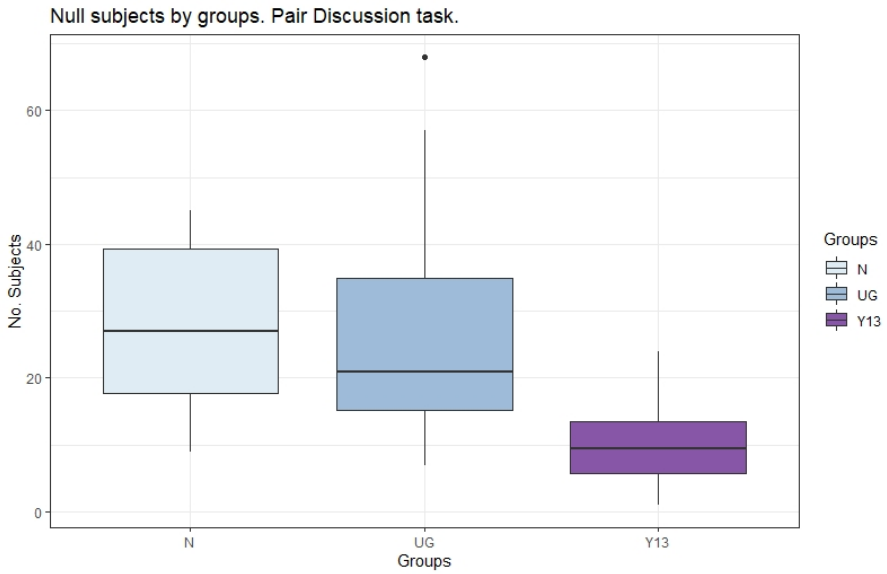| Group	diff	lwr	upr	p adj
UG-N	−3.854545	−10.16763	2.458537	0.3112893
V13-N	−13.250000	−19.66097	−6.839034	0.0000232
Y13-UG	−9.395455	−14.50964	−4.281273	0.0001492

Figure 7: Null subjects for each participant group (paired-discussion task)
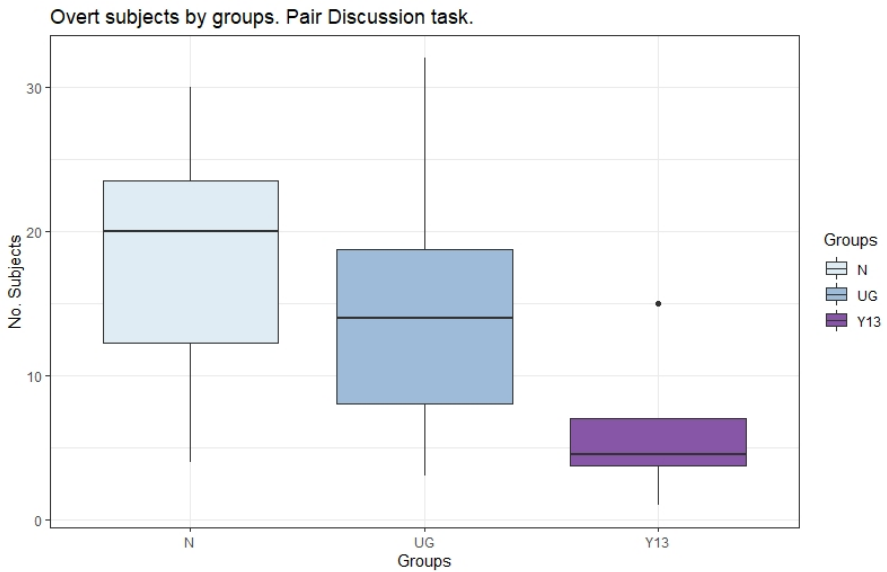

Figure 8: Overt subjects for each participant group (paired-discussion task)

Example (17) shows an exchange between two intermediate learners (D56 and D51) discussing some reasons why learning a foreign language is useful. Both learners show use of null pronouns. Learner D56 uses pro with *pienso* ('I think') and with *es importante* ('it is important') which in English requires a pleonastic *it*. Participant D51 shows various uses of the null subject form as shown in this example as well. It is clear that the structures this learner has chosen have the level of complexity which is appropriate for eliciting null pronouns, for instance by using subordinate clauses in which the subject is the same as in the main clause and does not need to be repeated.

(17)   D56: eh para mí ehm lo más importante es para poder ir a otro país y poder comunicarnos con los habitantes de allá [/] de allá porque *pro* pienso que *pro* es importante hablar con los extranjeros en su lengua.

   'For me ehm the most important thing is to be able to go to another country and be able to communicate with the speakers there because (I) think that (it) is important to speak to foreigners in their language.'

   D51: Sí *pro* tienes razón porque cuando *pro* visito un otro país *pro* lo odio cuando <los eh> [/] <los eh> [//] las turistas hablan más alto y eh más claro pero en su lengua eh normal con [//] como inglés porque *pro* piensan que es eh los extranjeros ehm conocerían los [//] conocerían.

   'Yes, you are right because when (I) visit another country (I) hate it when the tourists speak louder and clearer but in their own language like English because they think that the foreigners would know it.'

In this task, a large number of sentences contain the first person singular pronoun (*yo*) as the participants were giving their own reasons for defending their ranking of solutions to the problems. In contrast, most of the subjects elicited by the Loch Ness task were third person which may be a factor for explaining the lower use of null subjects produced by all the groups in that task. We discuss the implications of this distinction in §7.

7 Discussion

In this study, we have investigated the acquisition of Spanish null and overt subjects by three groups of English learners at beginner, intermediate and advanced proficiency levels. We examined the acquisition of these structures using oral production data as evidence, which have not been properly investigated in previous studies on this topic. Overall, our findings are in line with existing research

(mostly using comprehension/judgment data as evidence) which has shown that this is an area of Spanish which English speakers are able to acquire by the time they reach an advanced level of proficiency (Pérez-Leroux & Glass 1999, Liceras & Díaz 1999, Lozano 2002, 2006, Hertel 2003, Montrul 2004, Montrul & Rodríguez-Louro 2006, Belletti et al. 2007, Rothman & Iverson 2007, Domínguez 2013, Pladevall Ballester 2013). The oral data we have discussed clearly show an increase of use of both forms relative to proficiency and experience and towards target-like use.

Since both null and overt subjects show similar levels of pragmatic complexity, we predicted that these two forms would pose the same processing demands to these learners. According to this assumption, null subjects could potentially be difficult to acquire, particularly at early stages of acquisition. Following Cho & Slabakova (2014), we made two further predictions: that null and overt subjects would be ultimately acquired, and that overt subjects may be more difficult to acquire than null subjects as this is a structure which requires reassembly for English speakers.

Our first prediction was born out as beginner and intermediate learners consistently behaved differently to the advanced group for the use of both forms. We found no evidence in any of our analyses to suggest that null subjects are problem-free. In this sense, the analysis of these oral data complements comprehension data reported by previous research which also found null subjects to be somewhat difficult to acquire by English speakers (Montrul & Rodríguez-Louro 2006, Domínguez 2013, Pladevall Ballester 2013, Clements & Domínguez 2017). Our second prediction was also born out as advanced speakers behaved like the native controls in all of the tests which suggests that the advanced learners are able to master how to use these forms appropriately in different tasks. This finding supports Cho & Slabakova (2014)'s assumption that whether the L1 and the L2 use similar morphological means to express a particular feature or structure (as opposed to context) is relevant for the acquisition task.

The third prediction, however, was not completely supported. Since overt subjects require reassembly of existing form-meaning pairs, we predicted that learners may have more problems acquiring overt subjects than null subjects for this reason. The results we discussed for the two oral tasks revealed that although beginner and intermediate learners used both overt and null subjects at a lower rate than the controls, the intermediate group used overt subjects at a higher rate than null subjects in the paired-discussion task. That is, their rates of use of subjects were closer to the target for this form. It may be the case that the oral data that we have analysed are not able to provide us with the crucial evidence needed to conclude whether overt subjects are indeed more problematic as we

are not able to see, for instance, whether the learners would accept these forms in inappropriate contexts.

Crucially, the data show that although learners are aware that null subjects are available in Spanish, their use in oral production is sparse and does not show target-like levels until learners reach advanced levels of proficiency in Spanish. This is puzzling since, as Rothman & Iverson (2007) points out, the trigger for learning the underlying structure (or resetting the NSP parameter in Rothman's study) is salient and frequent. Pladevall Ballester (2013) concludes, after analysing advanced learners' justifications for the choices in a contextualised judgment task, that instruction seems to have a positive effect on the acquisition of the syntactic properties of Spanish null subjects but not for their distribution and use in context. In the case of our participants, it is also possible that linking pro to an existing referent in the discourse is a harder task in an oral production task in which learners, especially the least experienced ones, may feel more under pressure than when completing a written task.

To further investigate this possibility, the preliminary qualitative analysis we conducted on the data showed important differences in the overall ability to successfully communicate orally across the groups. We argue that the low production of null subjects observed for the beginner learners may be (partially) due to their limited knowledge of the type of complex structures which require the use of a null subject, such as a subordinate clause which adds extra information about a subject referent previously introduced in the discourse. Some of the intermediate learners are starting to use some of these more complex structures and are also able to provide more details to describe what the characters in the Loch Ness story were doing. Using coordinating sentences to describe a character's actions would elicit null subjects, a strategy which is rare for the beginner group. We see some of these examples in the data of some intermediate learners, but it is not until later on in the acquisition process that its use is widespread. Thus, it is likely that the overall linguistic ability and capability for oral communication of the learners also play a role in the rate of production of the forms we are investigating.

These results support the view that type of task used to elicit the data seems to be a very important factor when investigating the use of target forms in oral production (see Tracy-Ventura & Myles 2015, Domínguez 2019). In our results, the native controls' use of null and overt subjects varied according to the task. This was also the case for all the learners. The paired-discussion task elicited more null subjects than the Loch Ness task, perhaps because the referent used in this task was often the speakers themselves. It is easy to assume that sentences with first person pronouns *yo* 'I' would not often require an overt subject in

Spanish. In contrast, most of the subjects in the Loch Ness task were third person as participants had to describe the actions carried out by fictional characters. The participants may have used more overt subjects in this task as there were more opportunities to introduce new referents in every scene. We leave it to further research to clarify the effect of the type of referent (first vs third person) for the elicitation of null subjects.

It is interesting to point out that similar data discussed in Domínguez (2013) for the same forms and for the same group of learners also corroborates the finding that the type of task can influence the rate of use of null and overt subjects in oral tasks. This study describes the results of the data elicited using a semi-spontaneous interview with one of the investigators. In this task, which is the least constrained in terms of giving participants the freedom to discuss topics they were happy with, the native speakers used null subjects at a high rate (71%) when compared to the 29.4% rate of use of overt subjects. This is the task which elicited the highest number of null subjects for all groups (74% for Y9, 71% for Y13 and UG). This is also the task in which participants chose to speak mostly about themselves (same referent which is salient in the discourse), so many of the subjects produced were used in [−topic shift] contexts, the context in which pro is more likely to be used. When the participants used an overt pronoun, *yo* was the preferred choice as shown in Table 12.

Table 12: Average use of pronouns (from Domínguez 2013)

	Yo 'I' (%)	*El/Ella* 'he/she' (%)	Other (%)
Y9	83.3	16.6	0.0
Y13	77.0	22.9	0.0
UG	61.3	34.9	3.7
NS	83.5	10.0	6.4

Lozano (2009) reports overproduction and underproduction of third person animate singular pronoun (*él/ella* 'he/she') in a written corpus of L2 Spanish. The interesting result in this study is that the third person pronoun was the only pronoun that was problematic for learners: some learners used this overt pronoun redundantly in [−topic shift] contexts, while other learners used a null pronoun when an overt third person pronoun would be pragmatically felicitous.[3]

---

[3]One relevant finding from child language acquisition is that monolingual Spanish children seem to master null subjects corresponding to 1st and 2nd person before those corresponding to a 3rd person (see e.g. Forsythe et al. 2021).

Altogether, this seems to suggest that when learners have to describe actions carried out by a third person referent, they may find it harder to produce the correct form (null and overt). Since the Loch Ness task was the task with the highest number of third person referents, the lower use of null pronouns could be explained by the demands of the task on the less proficient learners.

Overall, the results of the oral data analysed in this study corroborate some of the previous findings using other methodologies: that overt and null subjects can be acquired in Spanish, but that their acquisition is not completely problem free. The analysis of the oral data has also shown that the type of task and the low proficiency of some of the learners may be obstacles to producing null subjects. We conclude by pointing out the benefits of using L2 oral data to investigate the acquisition of morphosyntactic phenomena and to test predictions which are theoretically inspired. This is very much in the spirit of Myles's pioneering work promoting the use of L2 corpora in SLA research (Myles 2004, 2005, Domínguez et al. 2013).

References

Alexiadou, Artemis & Elena Anagnostopoulou. 1998. Parameterizing AGR: Word order, V-movement and EPP-checking. *Natural Language and Linguistic Theory* 16. 491–539.

Almoguera, Rosa & Conchita Lagunas. 1993. El parámetro pro-drop en el español no nativo. In Juana M Liceras (ed.), *La lingüística y el análisis de los sistemas no nativos*, 140–149. Ottawa: Dovehouse.

Alonso-Ovalle, Luis & Francesco D'Introno. 2001. Full and null pronouns in Spanish: The zero pronoun hypothesis. In Hector Campos, Elena Herburger, Alfonso Morales-Front & Thomas Walsh (eds.), *Hispanic linguistics at the turn of the millennium*, 189–210. Somerville, MA: Cascadilla Press.

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

Barbosa, Pilar. 2009. Two kinds of subject *pro. Studia Linguistica* 63. 2–58.

Belletti, Adriana, Elisa Bennati & Antonella Sorace. 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory* 25(4). 657–689.

Bini, Milena. 1993. La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro pro-drop. In Juana M Liceras (ed.), *La lingüística y el análisis de los sistemas no nativos*, 126–139. Ottawa: Dovehouse.

Blackwell, Sarah E. 2003. *Implicatures in discourse: The case of Spanish NP anaphora.* Amsterdam: John Benjamins.

Cho, Jacee & Roumyana Slabakova. 2014. Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second Language Research* 30(2). 159–190.

Chomsky, Noam. 1981. *Lectures on ggovernment and binding*. Dordrecht: Foris.

Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Marti, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on minimalism in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Clements, Maria & Laura Domínguez. 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.

Díaz, Lourdes & Juana Liceras. 1990. Formulación de parámetros y adquisición de lenguas extranjeras. In Carlos Martín Vide (ed.), *Proceedings of the V congress on Natural and Formal Languages*, 465–480. Barcelona: Universidad de Barcelona.

Dippold, Doris. 2006. Face in L2 argumentative discourse: Psycholinguistic constraints on the construction of identity. *Language, Culture and Identity in Applied Linguistics* 21(10). 163–180.

Domínguez, Laura. 2013. *Understanding Interfaces: Second language acquisition and first language attrition of Spanish subject realization and word order variation*. Amsterdam: John Benjamins.

Domínguez, Laura. 2019. A "mixed methods" approach for investigating Aspect in a second language: Evidence from the SPLLOC project. *Dutch Journal of Applied Linguistics* 8(1). 41–66.

Domínguez, Laura & María J. Arche. 2014. Subject inversion in non-native Spanish. *Lingua* 145. 243–265.

Domínguez, Laura & María J. Arche. 2008. Optionality in L2 grammars: The acquisition of SV/VS contrast in Spanish. In Harvey Chan, Heather Jacob & Enkeleida Kapia (eds.), *Proceedings of the 32nd annual Boston University conference on language development*, 96–107. Somerville, MA: Cascadilla Press.

Domínguez, Laura, Nicole Tracy-Ventura, Maria J. Arche, Rosamond Mitchell & Florence Myles. 2013. The role of dynamic contrasts in the L2 acquisition of Spanish past tense morphology. *Bilingualism: Language and Cognition* 16(3). 558–577.

Fernández Soriano, Olga. 1989. Strong pronouns in null subject languages and the Avoid Pronoun Principle. *MIT Working Papers in Linguistics* 11. 228–240.

Forsythe, Hannah, Daniel Greeson & Cristina Schmitt. 2021. After the null subject parameter: Acquisition of the null-overt contrast in Spanish. *Language Learning and Development* 18(2). 171–200.

Goodall, Grant. 2001. The EPP in Spanish. In William Davies & Stanley Dubinsky (eds.), *Objects and other subjects*, 193–223. Dordrecht: Kluwer.

Grinstead, John. 1998. *Subjects, sentential negation and imperatives in child Spanish and Catalan*. UCLA. (Doctoral dissertation).

Hertel, Tammy J. 2003. Lexical and discourse factors in the second language acquisition of Spanish word order. *Second Language Research* 19(4). 273–304.

Holmberg, Anders. 2005. Is there a little *pro*? Evidence from Finnish. *Linguistic Inquiry* 36. 533–564.

Hwang, Sun Hee & Donna Lardiere. 2013. Plural-marking in L2 Korean: A feature-based approach. *Second Language Research* 29(1). 57–86.

Jaeggli, Osvaldo A. 1982. *Topics in Romance syntax*. Dordrecht: Foris.

Jaeggli, Osvaldo A. & Ken J. Safir. 1989. The null subject parameter and parametric theory. In Osvaldo A. Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*, 1–44. Dordrecht: Springer. DOI: 10.1007/978-94-009-2540-3_1.

Al-Kasey, Tamara & Ana T. Pérez-Leroux. 1998. Second language acquisition of Spanish null subject. In Suzanne Flynn, Gita Martohardjono & Wayne O'Neil (eds.), *Second language acquisition of Spanish null subjects*, 161–185. Mahwah: Lawrence Erlbaum Associates.

LaFond, Larry, Rachel Hayes & Rakesh M. Bhatt. 2001. Constraint demotion and null-subjects in Spanish L2 acquisition. In Joaquim Camps & Caroline R. Wiltshire (eds.), *Romance syntax, semantics and their L2 acquisition*, 121–135. Amsterdam: John Benjamins.

Lardiere, Donna. 2005. On morphological competence. In Laurent Dekydtspotter, Rex A. Sprouse & Audrey Liljestrand (eds.), *Proceedings of the 7th Generative Approaches to Second Language Acquisition conference (GASLA 2004)*, 178–192. Somerville, MA: Cascadilla Press.

Lardiere, Donna. 2008. Feature-assembly in second language acquisition. In Juana Liceras, Helmut Zobl & Helen Goodluck (eds.), *The role of features in second language acquisition*, 106–140. Mahwah, NJ: Lawrence Erlbaum.

Lardiere, Donna. 2009. Some thoughts on a contrastive analysis of features in second language acquisition. *Second Language Research* 25(2). 173–227.

Liceras, Juana. 1988. Syntax and stylistics: More on the pro-drop parameter. In James Pankhurst, Mike Sharwood Smith & Paul Van Buren (eds.), *Learnability and second languages: A book of readings*, 71–93. Dordrecht: Foris. DOI: 10.1515/9783110874150-005.

Liceras, Juana. 1989. On some Properties of the "pro-drop" parameter: Looking for missing subjects in non-native Spanish. In Susan M. Gass & Jacquelyn Schachter (eds.), *Linguistic perspectives on second language acquisition*, 109–133. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139524544.009.

Liceras, Juana, Anahi Alba de la Fuente & Christina Martínez Sanz. 2010. The distribution of null subjects in non-native grammars: Syntactic markedness and interface vulnerability. In 84–95. Somerville, MA: Cascadilla Proceedings Project.

Liceras, Juana & Lourdes Díaz. 1999. Topic-drop versus pro-drop: Null subjects and pronominal subjects in the Spanish L2 of Chinese, English, French, German, Japanese and Korean speakers. *Second Language Research* 15. 1–40.

Liceras, Juana, Elena Valenzuela & Lourdes Díaz. 1999. L1 and L2 developing Spanish grammars and the "pragmatic deficit hypohtesis". *Second Language Research* 15(2). 161–190.

Lozano, Cristóbal. 2009. Selective deficits at the syntax. In Neal Snape, Yan-kit Ingrid Leung & Michael Sharwood Smith (eds.), *Representational deficits in SLA: Studies in honor of Roger Hawkins*, 127–166. Amsterdam: John Benjamins.

Lozano, Cristóbal. 2002. The interpretation of overt and null pronouns in non-native Spanish. *Durham Working Papers in Linguistics* 8. 53–66.

Lozano, Cristóbal. 2006. The development of the syntax-information structure interface: Greek learners of Spanish. In Vincent Torrens & Linda Escobar (eds.), *The acquisition of syntax in Romance languages*, 371–399. Amsterdam: John Benjamins. DOI: 10.1075/lald.41.18loz.

Lubbers Quesada, Margaret & Sarah E. Blackwell. 2009. The L2 Acquisition of null and overt Spanish subject pronouns: A pragmatic approach. In *Selected Proceedings of the 11th Hispanic Linguistics Symposium*, 117–130. Somerville, MA: Cascadilla Proceedings Project.

Luján, Marta. 1985. Binding properties of overt pronouns in null pronominal languages. *CLS* 21. 123–143.

Luján, Marta. 1986. Stress and binding of Pronouns. *CLS* 22. 69–84.

MacWhinney, Brian. 1991. *The CHILDES project: Tools for analyzing talk.* Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk.* 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

Margaza, Panagiota & Aurora Bel. 2006. Null subjects at the syntax–pragmatics Interface: Evidence from Spanish interlanguage of Greek speakers. In Mary Grantham O'Brien, Christine Shea & John Archibald (eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition conference (GASLA 2006)*, 88–97. Somerville, MA: Cascadilla Press.

Mitchell, Rosamond, Laura Domínguez, María J. Arche, Florence Myles & Emma Marsden. 2008. SPLLOC: A new database for Spanish second language acquisition research. *EuroSLA Yearbook* 8(1). 287–304.

Montrul, Silvina. 2004. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7. 125–42.

Montrul, Silvina & Celeste Rodríguez-Louro. 2006. Beyond the syntax of the null subject parameter. In Vincent Torrens & Linda Escobar (eds.), *The acquisition of syntax in Romance languages*, 401–418. Amsterdam: John Benjamins. DOI: 10.1075/lald.41.19mon.

Myles, Florence. 2004. From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society* 102(2). 139–168.

Myles, Florence. 2005. Interlanguage corpora and second language acquisition research. *Second Language Research* 21(4). 373–391.

Myles, Florence. 2007. Using electronic corpora in SLA research. In Dalila Ayoun (ed.), *French applied linguistics*, 377–400. Amsterdam/Philadelphia: John Benjamins. DOI: 10.1075/lllt.16.18myl.

Papp, Silvia. 2000. Stable and developmental optionality in native and non-native Hungarian grammars. *Second Language Research* 16(2). 173–200.

Pérez-Leroux, Ana T. & William R. Glass. 1999. Null anaphora in Spanish second language acquisition: Probabilistic versus generative approaches. *Second Language Research* 15. 220–249.

Phinney, Marianne. 1987. The pro-drop parameter in second language acquisition. In Thomas Roeper & Edwin Williams (eds.), *Parameter setting*, 221–238. Dordrecht: Reidel. DOI: 10.1007/978-94-009-3727-7_10.

Pladevall Ballester, Elisabet. 2013. Adult L2 Spanish development of syntactic and discourse subject properties in an instructional setting. *RAEL: Revista electrónica de lingüística aplicada* 12. 111–129.

Ramchand, Gillian & Peter Svenonius. 2008. Mapping a parochial lexicon onto a universal semantics. In Theresa Biberauer (ed.), *Limits of syntactic variation*, 219–245. Amsterdam: John Benjamins.

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

Rizzi, Luigi. 1986. Null objects in Italian and the theory of *pro. Linguistic Inquiry* 17. 501–557.

Roberts, Ian. 2010. A deletion analysis of null subjects. In Theresa Biberauer, Anders Holmberg, Ian Roberts & Michelle Sheehan (eds.), *Parametric variation: Null subjects in Minimalist theory*, 58–87. Cambridge: CUP. DOI: 10.1017/CBO9780511770784.002.

Rothman, Jason & Michael Iverson. 2007. On L2 clustering and resetting the null subject parameter in L2 Spanish: Implications and observations. *Hispania* 90(2). 329–342.

Sheehan, Michelle. 2016. Subjects, null subjects and expletives. In Susann Fischer & Christoph. Gabriel (eds.), *Manual of grammatical interfaces in Romance*, 329–362. Berlin: De Gruyter. DOI: 10.1515/9783110311860-014.

Silva-Corvalán, Carmen. 2001. *Sociolingüística y pragmática del español.* Washington, DC: Georgetown University Press.

Slabakova, Roumyana. 2009. Features or parameters: Which one makes second language acquisition easier, and more interesting to study? *Second Language Research* 25(2). 313–324.

Slabakova, Roumyana. 2013. What is easy and what is hard to acquire in a second language: A generative perspective. In María del Pilar García Mayo, María Juncal Gutiérrez Mangado & María Martínez-Adrián (eds.), *Contemporary approaches to second language acquisition* (AILA Applied Linguistics Series 9), 5–28. Amsterdam: John Benjamins. DOI: 10.1075/aals.9.04ch1.

Sorace, Antonella. 2000. Gradients in auxiliary selection with intransitive verbs. *Language* 76(4). 859–890.

Sorace, Antonella. 2004. Native language attrition and developmental instability at the syntax-discourse interface: Data, interpretations and methods. *Bilingualism: Language and Cognition* 7. 143–145.

Sorace, Antonella. 2011. Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism* 1(1). 1–33.

Sorace, Antonella & Francesca Filiaci. 2006. Anaphora resolution in near-native speakers of Italian. *Second Language Research* 22(3). 339–368.

Torrego, Esther & Itziar Laka. 2015. The syntax of φ-features: Agreement with plural DPs in Basque and Spanish. In Beatriz Fernández & Et Salaburu (eds.), *Ibon Sarasola, Gorazarre. Homenatge, Homenaje*, 633–646. Bilbao: UPV/EHU.

Tracy-Ventura, Nicole & Florence Myles. 2015. The importance of task variability in the design of learner corpora for SLA research. *International Journal of Learner Corpus Research* 1(1). 58–95.

Villa-García, Julio. 2018. Properties of the extended verb phrase: Agreement, the structure of INFL, and subjects in Spanish. In Kimberly L. Geeslin (ed.), *The Cambridge handbook of Hispanic linguistics*, 329–350. Cambridge: Cambridge University Press.

Villa-García, Julio. 2013. On the role of children's deterministic learning in the "no-overt-subject" stage in the L1 acquisition of Spanish. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol. 37, 375–388. DOI: 10.3765/bls.v37i1.843.

White, Lydia. 1989. *Universal Grammar and Second Language Acquisition.* Philadelphia: John Benjamins.