SELECTIVE DEFICITS AT THE SYNTAX-DISCOURSE INTERFACE: EVIDENCE FROM THE CEDEL2 CORPUS

Cristóbal Lozano

University of Granada
Campus de Cartuja
Facultad de Filosofía y Letras
Departamento de Filología Inglesa
Granada 18071
Spain

cristoballozano@ugr.es

TO APPEAR IN:
Snape, N., Y.-k. I. Leung & M. Sharwood Smith. (eds.). (in prep.) Representational Deficits in SLA. Amsterdam: John Benjamins.

ABSTRACT

Recent studies reveal that second language (L2) learners show persistent deficits at the interfaces were the language faculty meets other cognitive domains. In the context of the pro-drop parameter, it is well known that adult learners of L2 Spanish are sensitive to the formal syntactic mechanisms licensing overt and null pronominal subjects from early stages of acquisition, but they show residual deficits when their distribution is constrained by topic and focus at the syntax-discourse interface, even at advanced and near-native levels of proficiency. Importantly, it has been assumed in the L2 literature that all phi-features of the pronominal paradigm are equally vulnerable.

In the current paper I will argue that the observed deficits at the syntax-discourse interface are *selective*, i.e., they do not affect the whole set of phi-features in the pronominal paradigm but rather a subset, due to their representational nature. In order to test this, data were sampled from L1 English – L2 Spanish advanced and very advanced learners from the CEDEL2 corpus (*Corpus Escrito del Español L2* 'L2 Spanish written corpus'). Results confirm that third person singular animate (masculine and feminine) pronouns are more amenable to vulnerability at the syntax-discourse interface, while the rest of the paradigm remains rather stable and (in some cases) intact. This indicates that, while the general observation about vulnerability at the syntax-discourse interface in L2 acquisition is largely correct, certain structures are more resilient to vulnerability due to the representational nature of their features, which is constrained by Universal Grammar.

1 INTRODUCTION

In the second language literature over the past two decades, researchers have mainly focused on the role of formal (i.e., morphosyntactic) features in L2 acquisition. This is reflected in the publication of monographs on this issue (e.g., Liceras et al. 2007), textbooks (e.g., Hawkins 2001, White 1989, 2003) as well as innumerable articles. Several proposals have been put forth trying to account for how some features can lead to representational deficits but others lead to native-like knowledge (e.g., Hawkins & Chan's 1997 *Failed Functional Features Hypothesis*, Prévost & White's 2002 *Missing Surface Inflection Hypothesis*, Beck's 1998 *Local Impairment Hypothesis*, just to name a few). By contrast, relatively little is known about the role of features operating at the interfaces outside narrow syntax. Recent studies have started to address the issue of why features at the syntax-discourse interface can be problematic for L2 learners even at end-states (e.g., Sorace 2004, 2005, 2006). Further distinctions have been made about syntax-semantics vs. syntax-discourse features, with different predictions for vulnerability (Tsimpli & Sorace 2006).

In the context of the first property of the pro-drop parameter (or null-subject parameter) it is well known since the 80's that English-speaking learners of L2 Spanish acquire from early stages the *formal* features licensing null referential pronominal subjects in L2 Spanish (e.g., Liceras 1989, Lozano 2002a, Phinney 1987), yet studies from the late 90s report that such learners show deficits with the distribution of overt and null referential pronominals when constrained outside syntax, i.e., when regulated by *discursive* features like [Topic] and [Focus] (e.g., Lozano 2002b, Montrul & Rodríguez-Louro 2006, Pérez-Leroux & Glass 1997, 1999). These studies claim that the acquisition of pronominal subjects results in (i) native-like knowledge of formal features operating at narrow syntax from early states, yet (ii) divergent knowledge and deficits when features operate at the syntax-discourse interface, which appear to be persistently problematic even at end-states.

Crucially, this observed 'syntax-before-discourse' phenomenon (i.e., the claim that the *formal* features licensing null subjects are acquired early and effortlessly while the *discursive* features are persistently problematic) comes from studies presenting evidence mostly from third singular pronominal subjects, but the claim has been made about the *whole* pronominal paradigm (singular: 1st, 2nd and 3rd person; plural: 1st, 2nd and 3rd). In this study, evidence from the whole paradigm will be presented and it will be shown that deficits with pronominal subject at the syntax-discourse interface are *selective*, since not all persons in the paradigm are equally vulnerable. It will be proposed that such selectiveness is a result of how Universal Grammar organises pronominal features according to a pronominal 'Feature Geometry' (Harley & Ritter 2002).

This study is structured as follows. Section 2 presents the theoretical background on the distribution of pronominal subjects at the syntax-discourse interface in native Spanish and the so-called pronominal 'Feature Geometry Analysis'. Section 3 reports on previous L2 Spanish studies on the acquisition of the distribution of overt and null pronominal subjects. In section 4 the relevant predictions and hypotheses are presented. Section 5 describes the methodology used and section 6 presents the results. In section 7 the results are discussed and a conclusion is reached in section 8.

2 THEORETICAL BACKGROUND

In this section, I will discuss first how the distribution of overt and null pronominal subjects is constrained in native Spanish by discursive properties like Topic and Topic-Shift. Later, I will show an analysis of the hierarchy of pronominal subject features, which is constrained by Universal Grammar (UG).

2.1 Distribution of pronominal subjects at the syntax-discourse interface

Since the formulation of the null-subject parameter (Jaeggli 1982, Rizzi 1982), it is well known that in null-subject languages like Spanish, overt and null personal pronominal subjects can alternate. In (1) the overt pronoun *él* and the null pronoun *pro* are in free alternation (e.g., Fernández-Soriano 1989, 1993, 1999, Luján 1999, Picallo 1994, Rigau 1986). Notice that a full Noun Phrase (NP) subject like *Pedro* 'Peter' can also alternate in the same position.[1] By contrast, in non-null subject languages like English a null pronoun is not possible, (2). The licensing of null subjects in languages like Spanish has been attributed to formal syntactic features. In particular, the formal [+D] and [+AGR] features of the T(ense) head in Spanish is a proper licensor of *pro* (Rizzi 1997).[2]

(1) { Pedro / él / *pro* } tiene mucho dinero.
Peter/he/Ø have.3sg a lot of money
'Pedro/he has a lof of money'

(2) { Peter / he / * *pro* } has a lot of money.

The apparently free alternation of overt and null pronominal subjects is constrained by discursive factors in null-subject languages like Spanish (see, *inter alia*, Alonso-Ovalle et al. 2002, Fernández-Soriano 1989, Pérez-Leroux & Glass 1997, 1999). It is well known that at the syntax-discourse interface information packaging is crosslinguistically articulated into topic and focus and that different languages can use different mechanisms to encode it, such as morphology, syntax, prosody or a combination of these (e.g., Casielles-Suarez 2004, Rochemont 1998, Zubizarreta 1998, 1999, Vallduví 1992, Vallduví & Engdahl 1996). For the purposes of this study, I will focus on two different types of contexts that regulate the distribution of overt and null pronominal subjects: *topic* contexts and *topic-shift* contexts.

Topic information represents discourse-old, known information which has been previously evoked in the preceding discourse. Topic continuity is marked in Spanish via a null pronoun. Consider Spanish native data from the CEDEL2 corpus, (3), where the informant is talking about *el protagonista* 'the main character' of the film *Escondido* 'Hidden'. The first instance of *el protagonista* ([3] [sing] [masc]) is clearly focus (new information), as it has not been mentioned in the preceding context. The following references to *el protagonista* can, at least theoretically, be realised in three possible ways: as a full NP again (*el protagonista*), as an overt pronoun (*él* 'he') and as a null pronoun (*pro*). Due to economy reasons, a null pronoun (*pro*) is the pragmatically felicitous option since it marks topic continuity in the discourse, though the other options (full NP or overt pronoun) would be grammatically possible but pragmatically infelicitous in native Spanish.

(3) (*Previous context: the informant is talking about the main character of a film*)
En la película "Escondido" **el protagonista** tiene una familia y ***pro*** trabaja en un programa de televisión. Un día ***pro*** empieza a recibir videos anónimos... [RSZ, Spanish native, CEDEL2 corpus]
'In the film "Hidden", **the main character** has a family. **He** works on a TV program. One day **he** starts receiving anonymous videos...'

Obviously, *pro* can mark topic-continuity for any other person and number of the pronominal paradigm. For example, in (4) the informant is talking about herself and her holidays, so she uses the null pronoun *pro* (phi-features: [1] [sing] [±masc], discursive features: [Topic]).

(4) (*Previous context: the informant is talking about her holidays*)
A mediados de julio, ***pro*** estaba muy cansada y con mucha tensión emocional, porque ***pro*** tuve que dejar casi terminado un trabajo de cierta envergadura. ***pro*** pasé unos días en la playa de Guardamar... [MCL, Spanish native, CEDEL2 corpus]
'Around mid July, **I** was very tired and with a lot of emotional stress, because **I** had to finish off an important assignment. **I** spent a few days on Guardarmar beach...'

Topic-shift (also referred to in the literature as *contrastive focus*) requires a change of referent, which is typically realised via an overt pronoun. In (5), the native informant is talking about *los protagonistas* 'The main characters' of a film, a man and a woman. When the informant wants to refer to the man, the overt pronoun *él* 'he' (phi-features: [3] [sing] [masc], discursive features [Topic-shift]) is required. Likewise, a change of reference to the woman requires the overt pronoun *ella* 'she' (phi-features: [3] [sing] [fem], discursive features [Topic Shift]). While it would be grammatically possible to use a null pronoun *pro* to refer to either the man or the woman (since the

null pronoun is unspecified for gender and it could then refer to either of the referents), its use would cause ambiguity, hence an overt pronoun is pragmatically preferred.

(5) (*Previous context: the informant is talking about the main characters of the film 'The Illusionist': a man and a woman*)
La última película que he visto es la de "El Ilusionista"... **Los protagonistas** son dos jóvenes que se conocen y se enamoran. **Él** es de clase baja, mientras que **ella** es de familia noble... [SPH, Spanish native, CEDEL2 corpus]
'The last film I saw is "The illusionist" ... **The main characters** are two young people who meet and fall in love. **He** comes from a working-class family, whereas **she** comes from a noble family...'

Note that in topic-shift contexts it is also possible to use a full NP (instead of an overt pronoun). This fact has been often overlooked in the literature, where it has been standardly assumed that an overt pronoun is the default option. An analysis of the native data in the CEDEL2 corpus (see Table 4 and Figure 4, page15) reveals that in native Spanish (as well as in non-native Spanish) topic-shift is significantly realised via a full NP more frequently than via an overt pronoun. In (5), the informant is talking about the main characters of the film 'The Illusionist': *el príncipe* 'the prince' ([3] [sing] [masc]), *su prometida/la chica* 'his fiancé/the girl' ([3] [sing] [fem]) and *el ilusionista* 'the illusionist' ([3] [sing] [masc]). Given that two masculine referents are brought into the discourse (*el principe* and *el ilusionista*), the native informant chooses full NPs (which can be specified as [Topic-shift], as well as [Focus]) to mark topic shift, instead of overt pronouns (which would cause certain ambiguity, since *él* 'he' could refer to either masculine antecedent).

(6) (*Previous context: the informant is talking about the main characters of the film 'The Illusionist': the Prince, his fiancé, the illusionist*)
Un día **el príncipe** y **su prometida** acuden a ver el espectáculo... **El príncipe** sospecha de que **su prometida** le es infiel y ***pro*** manda a uno de sus secuaces a perseguirla... Al final, **el ilusionista** y **la chica** preparan su huida ... **El príncipe** termina suicidándose y, al final, **el ilusionista** y **la chica** consiguen estar juntos [SPH, Spanish native, CEDEL2 corpus]
'One day, **the Prince** and **his fiancé** go to see the show ... **The Prince** suspects that **his fiancé** is cheating on him and **he** orders one of his henchmen to chase her ... At the end, **the illusionist** and **the girl** prepare their escape ... **The Prince** ends up committing suicide and, at the end, **the illusionist** and **the girl** get together'.

From the data in (6) it appears that in topic-shift contexts the higher the number of potential antecedents, the higher the probability of ambiguity if overt pronouns are used. Full NPs are favoured in these contexts to avoid potential ambiguity, as (7) illustrates.

(7) (*Previous context: the informant is talking about the characters of the film 'Miss Sunshine', which consist of a family: the grandfather, the parents, two children and an uncle*)
En cuanto al argumento, trata de una familia de clase media estadounidense formada por **el abuelo paterno**, **los padres**, **dos niños** (**una niña de 7 años** y **otro de 16**) y **un tío materno**... En el viaje han de lidiar con problemas personales muy importantes, aquellos que han marcado su vida hasta el momento: **el tío** se encuentra con el amante que lo traicionó, **el padre** se da cuenta de que su socio lo ha engañado y lo ha dejado en la bancarrota, **el hijo mayor** descubre que es daltónico y ello le impedirá ser piloto, su sueño, y **el abuelo**, el peor parado de la historia, muere por sobredosis de cocaína... [CMM, Spanish native, CEDEL2 corpus]
'The plot is about a middle-class American family that consists of **the paternal grandfather**, **the parents**, **two kids** (**a 7 year-old girl** and **a 16 year-old boy**) and **a**

> **maternal uncle**... During their trip they face important personal problems that have marked their lives: **the uncle** meets the lover that betrayed him, **the father** realises that his partner has lied and bankrupted him, **the elder son** finds out he is colour-blind, which will prevent him from becoming a pilot, his dream, and **the grandfather**, the loser in the story, dies from a cocaine overdose...'

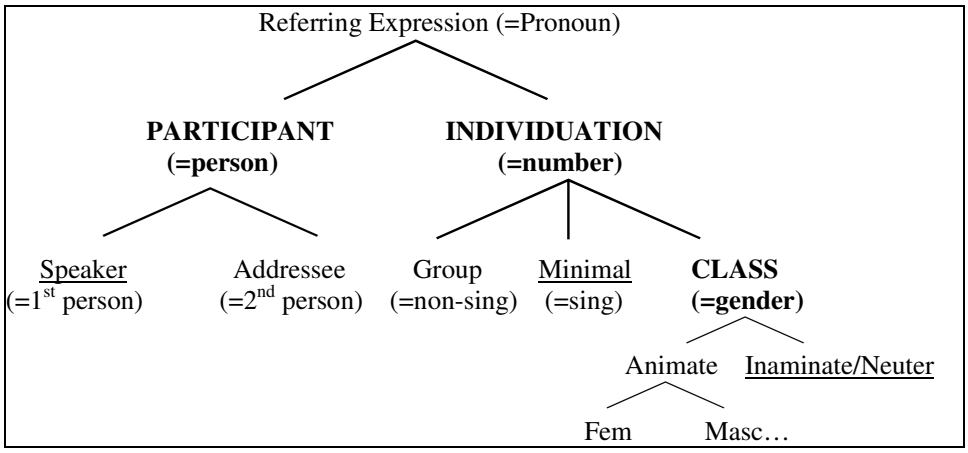
To sumarise, topic continuity is realised in native Spanish via a null pronoun, while topic shift is realised via overt material (either a full NP or an overt pronoun).

2.2 Pronominal feature geometry

In the generative second language acquisition (SLA) literature dealing with the acquisition of pronominal subjects, the inventory of pronominal features has been typically treated as an unordered 'bundle'. To illustrate the idea of the unordered bundle, consider two pronominal subjects, one specified for [3] [sing] and the other for [1] [sing]. In the literature there is no principled reason to assume that the first set of features is easier (or more difficult) to acquire than the second set. This entails that the relative order of acquisition of each pronominal feature would vary little between learners, who would therefore acquire pronominal features simultaneously in one 'bundle'. In other words, there has been no *systematic* prediction as to which pronominal feature(s), if any, are acquired first and which later. This is what has been implicitly assumed in the L2 literature since, as it was mentioned earlier, studies have focused mostly on third person singular pronouns and not on a systematic analysis of the full pronominal paradigm (e.g., Al-Kasey & Pérez-Leroux 1998, Davies 1996, Escutia 2002, Diaconescu & Goodluck 2002, Helland 2004, Kanno 1997, Lafond 2001, Liceras 1989, Liceras & Díaz 1999, López-Ortega 2006, Lozano 2002a, 2002b, Montrul 2004, Montrul & Rodríguez-Lourou 2006, Pérez-Leroux & Glass 1997, 1999, Pérez-Leroux & al. 1999, Phinney 1987, Polio 1995, Sorace & Filiaci 2006, White 1985).

Harley & Ritter (2002a, 2000b) have proposed the so-called 'Feature Geometry Analysis' for pronouns. Drawing from a wide range of typologically unrelated languages, they show that UG provides a constrained set of pronominal features which are systematically and hierarchically organized (Figure 1). The root node is termed *referring expression*, which corresponds to the traditional idea of a pronoun. There are three main sets of features: participant, individuation and class.[3] The PARTICIPANT node and its dependents, *Speaker* and *Addressee* represent 1st and 2nd person respectively. The INDIVIDUATION node and its dependents, *Group* and *Minimal* correspond to non-singular (plural and dual) and singular number respectively. The CLASS node encodes gender and other class information. Note that the underlined nodes *Speaker* (i.e., 1st person), *Minimal* (i.e., singular) and *Inanimate* (i.e., neuter) represent the default interpretation of the node.

Figure 1: Pronominal feature geometry analysis (Harley & Ritter 2002)



Importantly, the PARTICIPANT node encodes two features: *Speaker* (1st person) and *Addressee* (2nd person), while 3rd person is unmarked. It has been traditionally recognised in linguistics that there is a difference between 1st and 2nd person (which correspond to the speech-act participants, i.e., to a deictic use of the pronoun) vs. 3rd person (which does not correspond to the speech-act participants but rather to an anaphoric use of the pronoun) (e.g., Benveniste 1971, Bloomfield 1933, Forchheimer 1953, Jespersen 1924; see also Cornish 2006 and Saxena 2006 for more updated overviews). According to Harley & Ritter (2002a: 488) 'The geometry ... captures the intuition that so-called 3rd person is in fact not a true personal form ... When the Participant node is absent, the underspecified Referring Expression node receives a so-called 3rd person interpretation'. In other words, the authors claim that 3rd person is the absence of grammatical person (=participant) (see also Bianchi 2005, Kayne 2000).

Evidence that UG constrains the Feature Geometry analysis in language development comes from child L1 acquisition (e.g., Harley & Ritter 2002a, Hanson 2000). The uniformity and variability in the order of acquisition of pronominal subjects in different L1s shows that in the path of development 1st person singular and 3rd person singular neuter are normally the first pronouns to be acquired, while 3rd person animate comes later. Similar findings are reported for child Specific Language Impairment (Mastropavlou 2006), where 3rd person is selectively impaired or delayed, compared to 1st and 2nd person. Singular is also acquired before plural in normally developing children (Harley & Ritter 2002a, Hanson 2000). It seems then that in L1 acquisition defaults are acquired first (i.e., *Speaker*, *Minimal* and *Inanimate*). As we will see, this has implications for L2 acquisition since generative L2 research has not addressed the issue of emergence of the different pronominal features, i.e., all pronouns have been treated as a 'bundle' and no specific predictions have been made about the difficulty or emergence of each of them. Recent corpus SLA research from a functional perspective has shown that learners of L2 English (with L1 Swedish) treat 1st/2nd person pronouns differently from 3rd person (Ädel 2003). As we will see in the experimental section, the distinction will be crucial to account for the L1 English - L2 Spanish data.

3 PREVIOUS L2 STUDIES

In the context of the pro-drop parameter, the distribution of overt and null subjects in L2 Spanish has been extensively researched. It is well known since the late 80's that English-speaking learners of L2 Spanish acquire from early stages and with relatively little effort the *formal* properties that license a null subject, *pro* (*inter alia*, Lafond 2001, Liceras 1989, Lozano 2002a, Phinney 1987). In particular, learners know that an overt and a null pronoun are in free alternation in Spanish, as in (8).

(8) $\left\{ \begin{matrix} \text{Él} \\ \textit{pro} \end{matrix} \right\}$ come pasta.
He/*pro* eat.3sg pasta
'He eats pasta'

More recent studies have shown that, while the observation that *formal* licensing features are acquired early is correct, learners do show some deficits with the *discursive* features that constrain the distribution of overt and null pronouns in the discourse. Pérez-Leroux & Glass (1997) found that very advanced and near-native learners of L2 Spanish (with L1 English) showed native-like knowledge of formal constraints, yet non-native behaviour at the syntax-discourse interface. In one of the tasks, learners were shown topic-shift contexts like (9), where there are several referents (*Sampras, Edberg* and *Agassi*) and one of them has to be chosen for contrastive purposes. Learners were required to translate a sentence into Spanish. The context is manipulated such that the expected target translated sentence should contain an overt pronoun to express a shift of referent (i.e., to contrast *Sampras* against the rest of players). A null pronoun would be unpragmatic since it would cause ambiguity and could refer to any of them.[4] Learners showed a low percentage of overt pronouns, but a high percentage of null pronouns, which leads to an unpragmatic overgeneralisation of null pronouns for contrastive purposes. This deficit persists even at end-states.

(9) [Context] My friends are all excited about the US Open Tournament. The player that is most on their mind is Pete Sampras$_i$. They've barely mentioned Stefan Edberg$_j$ and Andre Agassi$_k$.
[To translate] Everybody thinks that he will win.
[Expected target sentence] Todo el mundo$_m$ cree que **él$_i$** ganará.
[Unexprected unpragmatic sentence] Todo el mundo$_m$ cree que ***#pro*$_m$** ganará.

In a later study testing several proficiency levels (elementary, intermediate and advanced), Pérez-Leroux & Glass (1999) found developmental evidence for the early mastery of formal constraints regulating the overt/null distribution, yet late and persistently problematic mastery of such distribution when constrained by topic and contrastive focus. Unpragmatic overuse of null pronouns in topic-shift contexts was found again in the advanced group. The results showed that 'knowledge of the marking of the topic/focus distinction is acquired over time and experience' (p. 242), though residual deficits remain. Similar results for contrastive focus contexts have been reported in other studies (e.g., Lozano 2002b, 2003).

In topic contexts, it has been reported that learners of Spanish overuse overt pronouns, which results in redundancy (e.g., Lozano 2006a, Montrul & Rodríguez-Louro 2006). In particular, in contexts when the topic is clearly set (*Profesor Antonio* 'Professor Antonio') and there are no potential referents that could lead to ambiguity, as in (10),

learners' tolerance of the (a) sentence, where the overt pronoun *él* 'he' is pragmatically redundant, is significantly higher than the Spanish natives' tolerance. This tolerance persists even at end-states (Lozano 2006a).

(10) Aunque el profesor Antonio$_i$ parece pobre...
(a) los estudiantes dicen que **#él**$_i$ tiene mucho dinero
(b) los estudiantes dicen que ***pro***$_i$ tiene mucho dinero
'Even though professor Antonio seems poor...
(a) students say that he has a lot of money
(b) students say that has a lot of money'

Other studies testing several aspects of the pronominal paradigm on L2 Spanish support the finding that formal constraints are in place early, yet discursive constrains are acquired over time and tend to result in residual deficits (e.g., Al-Kasey & Pérez-Leroux 1998, Liceras & Díaz 1999, Pérez-Leroux & al. 1999).

This 'syntax-before-discourse' observation on pronominal subjects does not represent an isolated phenomenon in Spanish L2 acquisition, as it has been attested in other acquisition contexts, as in L1 English–L2 Chinese (Polio 1995), L1 Croatian–L2 Italian (Kras 2006), L1 English–L2 Italian (Sorace & Filiaci 2006), L2 Italian with learners of several L1s (Belletti & Leonini 2004), English-Italian bilingual children (Serratrice 2004, Serratrice et al. 2004), Italian-Dutch bilingual children (Pinto 2006), L1 Greek and Italian attrition under the influence of L2 English (Tsimpli et al. 2004), L1 Spanish attrition under the influence of L2 English (Satterfield 2003), L1 Spanish heritage speakers with dominant English (Montrul 2004) and Spanish L1 acquisition (Grinstead 2004). Additionally, it has been also reported that for the second property of the pro-drop parameter (Subject-Verb inversion), learners of L2 Spanish also show early knowledge of the formal properties licensing inversion, but persistent problems with the discursive properties that constrain inversion in the discourse (Hertel 2003, Lozano 2006b, 2006c). Similar results are reported in Spanish heritage speakers with dominant English (Valenzuela & McIlwraith 2007), L2 Portuguese (Fruit 2007) and attrition in L1 Catalan with dominant L2 English (Helland 2004).

In this context, Sorace (2004) observes that 'aspects of grammar at the syntax-discourse interface are more vulnerable ... than purely syntactic ones' (p. 143) and that 'interfaces, because they are more complex than narrow syntax, are inherently more difficult to acquire.' (p. 144). So, while the 'syntax-before-discourse' phenomenon is beyond dispute, its causes are still unclear. Two main proposals have appeared recently in the literature. First, the *representational deficit account* postulates that underspecification of [+interpretable] features like [Topic] and [Topic-Shift] become underspecified at the syntax-discourse interface, thus triggering the observed deficits (e.g., Montrul 2004, Sorace 2004), but it may be also the case that it is [–interpretable] features that are responsible for such deficits (Lozano 2006b, 2006c). Second, the *processing deficit account* postulates that the language processor cannot process efficiently properties at the syntax-discourse interface, which results in shallow processing (Sorace 2005, 2006, Sorace & Filiaci 2006).

Importantly, in the context of pronominal subjects, most studies mentioned above present evidence from 3rd person singular, but the claims are made (implicitly or explicitly) about the whole pronominal paradigm. In other words, the 'syntax-before-discourse' phenomenon has been assumed to affect the whole pronominal paradigm

(all three persons and the two numbers). In this study, I will present evidence from the whole pronominal paradigm. As will be shown in the next section (Predictions), deficits at the syntax-discourse interface are *selective*, since not all persons are equally vulnerable at the syntax-discourse interface.

4 PREDICTIONS

Based on existing previous L2 studies (section 3) it was predicted that advanced and highly advanced learners of L2 Spanish (with L1 English) would show deficits at the syntax-discourse interface, in particular, overproduction of overt pronouns in topic contexts where a null pronoun is required. Additionally, following the theoretical assumptions of Harley & Ritter's Feature Geometry (section 2), it was predicted that not all pronominal features are equally complex/simple to acquire. The hypothesis in (11) was then formulated..

(11) H_1: Deficits at the syntax-discourse interface do not affect the whole pronominal paradigm, but are rather *selective*.
In particular, we expect (i) robustness and native-like knowledge with speech-act participants (1st and 2nd person) and with neutrals (3rd person inanimate), but (ii) vulnerability with 3rd person animate.

In short, syntax-discourse deficits are selective as they affect certain persons in the pronominal paradigm and not all persons (as previously assumed in the literature).

5 METHOD

In this section I will detail the properties of the corpus used in the study, as well as the subjects of the corpus. Then, I describe the concordance software used to analyse the corpus and how the corpus raw data were treated statistically.

5.1 Corpus

CEDEL2 (*Corpus Escrito Del Español L2* 'L2 Spanish Written Corpus') is a learner corpus that is being developed at the Universidad Autónoma de Madrid (Spain).[5] Currently the corpus consists of an L1 English–L2 Spanish subcorpus plus a native Spanish subcorpus used for comparative purposes. At present, the corpus has reached approximately 250,000 words, though it is envisaged that by the end of the project data collection will have reached over half a million words. Data consist of written compositions in Spanish. In particular, learners must fill in several online forms via the internet at: `http://www.uam.es/woslac/start.htm`

i. a *learning background questionnaire* which provides crucial information such as the learners' chronological age, age of first exposure to Spanish, years of instruction in Spanish, length of stay in Spanish-speaking countries, their father's and mother's native language, language(s) spoken at home, etc. As we will see in section 5.2 (Table 1, page 11), this type of data are useful to filter out learners that do not meet the researcher's requirements.
ii. a *composition in Spanish* form, where participants can choose between twelve different topics that vary in difficulty and can elicit different types of linguistic structures.

iii. a *Spanish placement test* (only for learners), which is an independent and standardized measure of grammatical proficiency in Spanish (University of Winconsin College-level Placement Test: Spanish grammar, form 96M).

Spanish natives must fill in two online forms only: (i) a *formación académica* form, which is similar to the learners' learning background form, but it obviously excludes questions that are not applicable to natives, such as length of instruction in Spanish, age of first exposure, etc; (ii) a *redacción en español*, which is identical to the learners' composition in Spanish form.

Given the nature of the data for this study, I selected only those compositions where topic continuity and topic shift would be more likely, i.e., compositions where one character would be activated during the narration (topic continuity) and those where several characters appeared and a change of character would be likely (topic shift). The composition titles were mostly of the type *Resume una película que hayas visto recientemente* 'Summarise a film you have seen recently', *Describe un viaje que hayas hecho recientemente* 'Describe a trip you have made recently', *¿Qué hiciste el año pasado durante tus vacaciones?* 'What did you do last year during your holidays?'

5.2 Subjects

For this study I selected a small group of CEDEL2 texts that met certain proficiency criteria (see below). Learners were classified into two learner groups (N=10 each group) according to proficiency level. An independent proficiency measure was administered to check learners' grammatical competence (University of Winconsin College-Level Placement Test – Spanish, 1998). A summary of the learners' and natives' bio-data in shown in Table 1 (full bio-data details of each participant are presented in the tables in the Appendix section).

Table 1: Summary of subjects' bio-data (see Appendix for full details)

Group	N	Mean proficiency	Mean chronol age	Mean age first exposure	Mean years instruct	Mean stay (months)
Upper-advanced	10	99%	35	14	8	29
Lower-advanced	10	93%	32	15	6	27
Spanish natives	12	n/a	37	n/a	n/a	n/a

As can be seen in Table 1, the proficiency score was higher in the upper-advanced group (mean=99%; range=98%-100%) than in the lower-advanced group (mean=93%; range=95%-91%). Overall, the upper-advanced group was first exposed to Spanish earlier (mean=14 years; median=14; range=3-20 years) than the lower-advanced group (mean=15 years; median=14; range=6-26). Overall, the upper-advanced group received more years of instruction (mean=8; range=5-11) than the lower-advanced (mean=6; range=3-15). Regarding the learners' stay in a Spanish-speaking country, the upper-advanced group stayed overall longer (mean=29 months; range=3-18 months; outlier=204 months) than the lower-advanced group (mean=27 months; range=0-12 months; outlier=228 months). These variables confirm that the

upper-advanced group can be safely regarded as a highly proficient group, reaching near-native grammatical competence, while the lower-advanced group can be considered a very proficient group. The rationale behind the creation of two advanced groups instead of only one was the following: given that the SLA debate on deficits at the syntax-discourse interface has focused on whether learners can overcome discursive deficits at end-states (see sections 1 and 3), it was necessary to discriminate between those learners who had indeed achieved a near-native degree of competence (upper-advanced group) and those who had not but still showed a high level of competence (lower-advanced group).

5.3 Software (concordancer)

The *UAM Corpus Tool* (version 1.0) is a stand-off XML, freely-downloadable software for the annotation of linguistic corpora developed by Michael O'Donnell.[6] Some of its main features, which were used in the current study, are as follows:

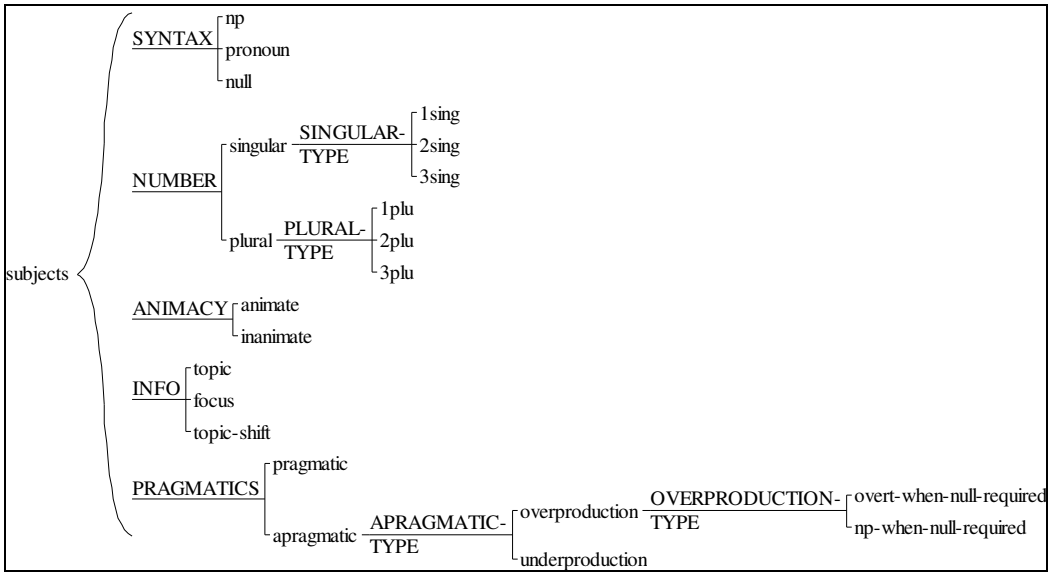
i. Annotation of segment using an annotation scheme of your design. The annotation of each segment can be made at multiple levels (e.g., NP, Clause, Sentence, whole document). In this study, I annotated each grammatical subject in finite sentences (null pronoun, overt pronoun or full NP) according to the annotation scheme in section 5.4, Figure 2.
ii. Searching for instances of a feature (or combination of features), e.g., any instance of the feature topic-shift, or any instances containing an overt pronoun which is a topic-shift, or any instances of third-person singular subjects realised as a full NP which is topic-shift.
iii. Comparative statistics across subsets, e.g., contrasting pragmatic and unpragmatic production of third person singular subjects.

5.4 Data analysis

Using *UAM Corpus Tool* I designed an annotation scheme (Figure 2). The annotation scheme was implemented taking into account Harley & Ritter's (2002) pronominal Feature Geometry and previous literature on pronominal subjects. Each subject (whether a full NP, an overt pronoun or a null pronoun) in finite sentences was tagged for the following properties:

i. *Syntax*: each subject was coded as a full NP, an overt pronoun or a null pronoun.
ii. *Number*: singular (1^{st}, 2^{nd} or 3^{rd} person) or plural (1^{st}, 2^{nd} or 3^{rd} person).
iii. *Animacy*: animate or inanimate.
iv. *Information status*: topic (which encodes topic-continuity in the discourse), focus (which encodes discourse-new information), and topic-shift (which encodes a change of topic in the discourse). In the current study, only topic and topic-shift will be analysed.
v. *Pragmaticality*: whether the subject was pragmatically appropriate or rather inappropriate, in which case I coded the type of lack of pragmaticality. This can be of two types: *underproduction* (i.e., using a null pronoun in topic-shift contexts which require the use of overt material) or *overproduction* (i.e., using an overt pronoun or a full NP when a null pronoun is required in topic-continuity contexts).

Figure 2: Annotation scheme used in the software *UAM Corpus Tool*



The annotation scheme was applied to the following corpora (Table 2). The upper-advanced corpus consisted of 10 texts (i.e., 10 learners, 1 text per learner) containing a total of 8,188 words and 453 tags (i.e., 453 annotated subjects which were statically analysed at a later stage). The ratio of pronominal subjects used in this group was 5.5% (i.e., on average there were 5.5 tagged pronominal subjects per 100 words). The lower-advanced corpus also consisted of 10 texts with a total of 8,521 words and 528 tags, corresponding to a 6.2% words/tags ratio. Finally, the Spanish native corpus consisted of 12 texts with a total of 5,954 words and 299 tags, the percentage of pronominal subjects being 5.1%. It is important to highlight that, although the total number of words is lower in the Spanish corpus than in the learner corpora (due to the natives' smaller amount of words per composition), the crucial fact is that the percentage of pronominal subjects used is similar in all three corpora, ranging from 5.1% to 6.1%, which indicates that the three corpora are comparable in terms of the ratio of syntactic subjects produced.

Table 2: Summary of linguistic data (see Appendix for full details)

Corpus	N texts	Total # words	Total # tags	Mean % pronominal subjects
Upper-advanced	10	8188	453	5.5%
Lower-advanced	10	8521	528	6.2%
Spanish natives	12	5954	299	5.1%

As explained briefly above, *UAM Corpus Tool* outputs raw frequency statistics for each feature (i.e., for each tag type) and compares them via inferential statistics based on the t-test. Since most statistical analyses presented in this study are based on the raw frequencies produced by our participants, I coded the output of raw frequencies into SPSS and treated them to chi-square analyses to check for significance, as is the standard practice in learner corpus studies.

6 RESULTS

In this section we will see three types of results regarding the syntactic subject: (i) its syntax and information status, (ii) its person and number, and (iii) its information status and pragmaticality. While result (i) is rather descriptive, it gives us an idea of how learners use subjects from a formal and functional point of view, a fact that has been often overlooked in the literature. Results (ii-iii) address our hypothesis and will reveal learners' deficits at the syntax-discourse interface.

6.1 Syntax and information status of the subject

In this section we will explore whether, according to the literature, a null pronoun is preferred over an overt pronoun in topic-continuity contexts, yet an overt pronoun is preferred to a null pronoun in topic-shift contexts.

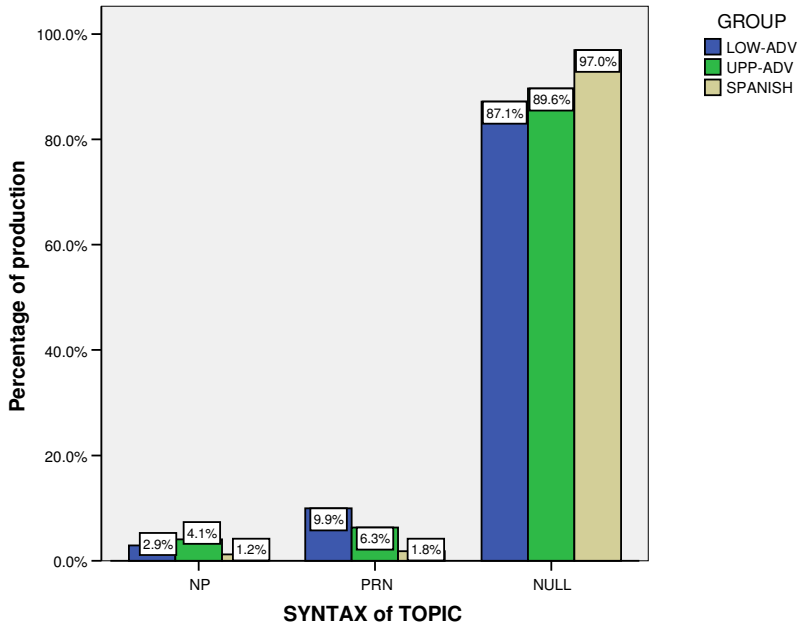
In topic contexts, most subjects are null pronouns in both the Spanish native corpus (97%) and the learner corpora (98.6% and 97.1% for the upper and lower advanced groups respectively), as expected (see Table 3, Figure 3). Overt pronouns are marginally used in topic contexts by natives (1.8%) and, to a larger extent, by learners (6.3% and 9.9%). Full NPs are clearly disfavoured to encode topic-continuity (1.2% natives; 4.1% and 2.9% learners).

Table 3: Group * Syntax of topic

			SYNTAX OF TOPIC			Total
			NP	PRN	NULL	
GROUP	LOW-ADV	Count	10	34	298	342
		% within GROUP	**2.9%**	**9.9%**	**87.1%**	100.0%
	UPP-ADV	Count	13	20	285	318
		% within GROUP	**4.1%**	**6.3%**	**89.6%**	100.0%
	SPANISH	Count	2	3	159	164
		% within GROUP	**1.2%**	**1.8%**	**97.0%**	100.0%

In short, in topic contexts all groups prefer mostly a null pronoun to encode topic-continuity, though overt pronouns and full NPs are marginally used in these cases, as can be visually appreciated in Figure 3. The learners' slightly higher (and pragmatically inapropriate) production of overt pronouns and full NPs in these contexts is significantly different from the Spanish natives' production (upper-advanced vs. natives: χ^2=8.002, df=2, p=0.018; lower-advanced vs. natives: χ^2=12.516, df=2, p=0.002). In section 6.3 we will come back to the issue of incorrect production of overt pronouns in topic contexts.

Figure 3: Syntax of topic



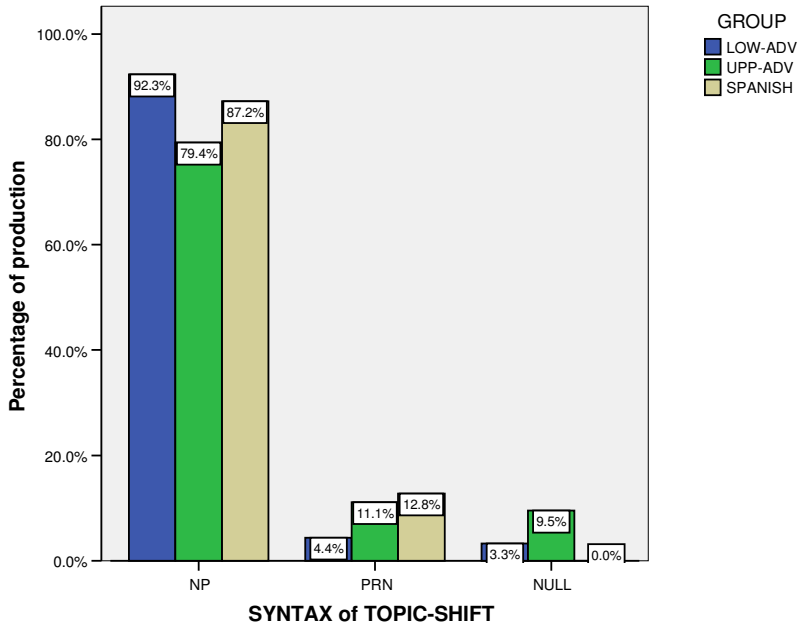
Regarding topic-shift, Table 4 shows that a change of topic is (i) normally encoded via a full NP (87.2% Spanish, 79.4% upper-advanced, 92.3% lower-advanced), (ii) occasionally encoded via an overt pronoun (12.8% Spanish, 11.1% upper-advanced and 4.4% lower-advanced), and (iii) never by a null pronoun in the native group (0%) and occasionally or seldom by a null pronoun (9.5% upper-advanced and 3.3% lower advanced).

Table 4: Group * Syntax of topic-shift

			SYNTAX OF TOPICSHIFT			Total
			NP	PRN	NULL	
GROUP	LOW-ADV	Count	84	4	3	91
		% within GROUP	**92.3%**	**4.4%**	**3.3%**	100.0%
	UPP-ADV	Count	50	7	6	63
		% within GROUP	**79.4%**	**11.1%**	**9.5%**	100.0%
	SPANISH	Count	41	6	0	47
		% within GROUP	**87.2%**	**12.8%**	**0.0%**	100.0%

The fact that topic-shift is encoded typically via a full NP can be better appreciated visually in Figure 4. The corpus data in this study show that natives as well as learners prefer producing an NP (rather than an overt pronoun) to encode topic-shift. There are no significant differences between the learner groups and the natives (upper-advanced vs. natives: χ^2=4.740, df=2, p=0.93; lower-advanced vs. natives: χ^2=4.634, df=2, p=0.99). This is a fact that has been largely overlooked both in the theoretical and L2 literature, where it has been assumed that in null-subject languages topic-shift is normally encoded via an overt pronoun. While the causes for this syntactic preference need serious investigation, they are beyond the scope of this study (see Blackwell 1998 and Reinhart 1995).

Figure 4: Syntax of topic-shift



6.2 Person and number of the subject

In this section I analyse each person in the pronominal paradigm and check whether its production is pragmatically (un)acceptable.

Regarding *1st person singular* (*yo*/*pro* 'I') results are shown in Table 5. All groups' (natives and learners) production of 1st singular was pragmatically correct (98.1% natives, 100% upper-advanced and 98.2% lower-advanced), there being no significant differences between any of the groups (χ^2=2.234, df=2, p=0.327). These results clearly show that learners show a native-like production of 1st singular, i.e., 1st singular does not lead to deficits at the syntax-discourse interface. This lends support to H_1 in (11) above.

Table 5: Group * Pragmaticality of 1st singular

			PRAGMATICALITY 1st sing		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	108	2	110
		% within GROUP	**98.2%**	**1.8%**	100.0%
	UPP-ADV	Count	120	0	120
		% within GROUP	**100.0%**	**0.0%**	100.0%
	SPANISH	Count	52	1	53
		% within GROUP	**98.1%**	**1.9%**	100.0%

As for *1st person plural* (*nosotros*/*nosotras*/*pro* 'we'), the results are similar to those of 1st singular. Production of 1st plural was pragmatically correct (100% for natives and learners), as Table 6 shows. There are obviously no significant differences between groups, since their production rates are identical (hence a χ^2 cannot be performed because there is no variation in the production rates). These results confirm

again hypothesis H_1, since learners' native-like production rates clearly indicate that they do not show any pragmatic deficits with 1st plural.

Table 6: Group * Pragmaticality of 1st plural

			PRAGMATICALITY 1st plural		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	42	0	42
		% within GROUP	**100.0%**	**0.0%**	100.0%
	UPP-ADV	Count	27	0	27
		% within GROUP	**100.0%**	**0.0%**	100.0%
	SPANISH	Count	22	0	22
		% within GROUP	**100.0%**	**0.0%**	100.0%

Regarding *2nd person singular* (*tú/pro* 'you') and *2nd person plural* (*vosotros/vosotras/pro* 'you all') all groups' production is pragmatically correct, though the frequencies are so low in our corpus that little else can be said about 2nd person. In particular, there were only two productions of 2nd singular (n=1 upper-advanced group and n=1 lower-advanced group) and only two of 2nd plural (n=1 upper-advanced group and n=1 lower-advanced group). Spanish natives did not produce any 2nd person at all.

Regarding *3rd person singular animate* (*NP/él/ella/pro* 'NP/he/she'), learners produce a considerable amount of pragmatically incorrect forms (14.9% upper-advanced, 16.7% lower-advanced), compared to the negligible pragmatically incorrect production of the Spanish native group (1%, which corresponds to just 1 token), as Table 7 shows. Given that the observed frequencies in the Spanish group are smaller than 5, I performed Fisher's Exact test (instead of Chi-square), which reveals that each learner group significantly differs from the Spanish group (upper-advanced vs. natives: $p \leq 0.001$; lower-advanced vs. natives: $p \leq 0.001$). By contrast, the upper-advanced group does not differ from the lower-advanced group ($p=0.386$). These results support H_1, since learners clearly show deficits with 3rd singular animate (i.e., they produce a considerably high number of non-pragmatic instances), while the learner groups do not differ from each other.

Table 7: Group * Pragmaticality of 3rd singular animate

			PRAGMATICALITY 3rd sing animate		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	175	35	210
		% within GROUP	**83.3%**	**16.7%**	100.0%
	UPP-ADV	Count	120	21	141
		% within GROUP	**85.1%**	**14.9%**	100.0%
	SPANISH	Count	97	1	98
		% within GROUP	**99.0%**	**1.0%**	100.0%

Examples (12) and (13) illustrate overproduction deficits with 3rd person singular animate. In (12) the upper-advanced learner produces an overt pronoun (*él* 'he', [3] [sing] [masc] [anim]) to refer to the previous antecedent *mi novio de EEUU* 'my boyfriend from USA'. The overt pronoun is pragmatically redundant, since there is no

ambiguity of referent in this situation. Similarly, in (13) another upper-advanced learner produces a full NP (*Penelope*, [3] [sing] [fem] [anim]) to refer to the previous antecedent (*Penelope*). The second NP is pragmatically redundant since there is no ambiguity in the discourse. In both (12) and (13) a null pronoun would have been the pragmatically desirable option.

(12) [Context: The informant is talking about a recent trip: she went to Seville (Spain) on holiday and her boyfriend came over from the USA to visit her]
Cuando me faltaban dos semanas, **mi novio de EEUU** me visitó unos días para ver la ciudad que me encantaba tanto. Era la primera vez que **#él** salió de su pais, por eso era un viaje importante. [CPB, upper-advanced, CEDEL2 corpus]
'Two weeks before I went back, **my boyfriend from the USA** paid me a visit for a few days to see the city I liked so much. It was the first time that **he** left his country, that is why it was an important trip.'

(13) [Context: The informant is talking about the film 'Volver': Penelope Cruz kills her father and she and her mum hide the corpse in a fridge in their restaurant]
...**Penelope** limpia todo el sangre y lo ocultan en la nevera del restaurante abajo, que esta cerrado y por venta. #**Penélope** tiene las llaves por si acaso alguien quiera echar un vistazo para comprarlo. [JEL, upper-advanced, CEDEL2 corpus]
'**Penelope** cleans all the blood and they hide it [the corpose] in the fridge in the restaurant below, which is shut down and for sale. **Penelope** has the restaurant keys in case anyone wants to have a look and buy it.'

Similarly, deficits with 3rd singular animate can be of the underproduction-type. In (14), the production of a null pronoun (*pro*) is pragmatically ambiguous, as it could refer to two potential antecedents in the immediately preceding context: either *mi amiga* 'my (girl) friend' or *su novio-prometido* 'her fiancé'. As it stands, the sentence can have two possible readings, (i) it is the girl friend who wanted to check whether her fiance would like Canada, or (ii) it is the fiancé who wanted to check whether he himself would like Canada. In this context, an overt pronoun would be pragmatically desirable (either *él* 'he' or *ella* 'she', depending on the informant's choice).

(14) [Context: The informant is talking about her best girl friend and her fiancé]
Mi amiga me contó que iban a visitar a Canadá este mes para que **su novio-prometido** conociera a su familia aquí y sus amigos también. **Su novio** tampoco había estado en Canadá así que ***#pro*** quería ver si a él le gustara y cosas así para ver si querían volver a Canadá después de casarse. [KEM, upper-advanced, CEDEL2 corpus]
'**Mi best friend** told me they were going to visit Canada this month so that **her fiancé** could meet her family here [in Canada] and her friends too. **Her fiancé** hadn't been to Canada either so that ***pro*** [=he/she] wanted to check whether he like it and stuff, to check whether they wanted to come back to Canada after their marriage.'

By contrast, learners show native-like behaviour with *3rd person singular inanimate* (*NP/ello/pro* 'NP/it'), Table 8. Learners' production is pragmatic (97.7% upper-advanced, 100% lower-advanced), similar to Spanish natives' production (98.7%, except for 1 residual token which represent 1.3% of lack of pragmaticality). As the raw data reveal, the learner groups are not significantly different from the Spanish group. Fisher's exact test indeed confirms this: upper-advanced vs. natives, p=0.589; lower-advanced vs. natives, p=0.467). These results confirm H_1, since learners do not show any pragmatic deficits with 3rd singular inanimate.

Table 8: Group * Pragmaticality of 3rd singular inanimate

			PRAGMATICALITY of 3rd sing inanim		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	90	0	90
		% within GROUP	**100.0%**	**0.0%**	100.0%
	UPP-ADV	Count	43	1	44
		% within GROUP	**97.7%**	**2.3%**	100.0%
	SPANISH	Count	78	1	79
		% within GROUP	**98.7%**	**1.3%**	100.0%

Sentence (15) illustrates how an upper-advanced learner produces a pragmatically felicitous sentence with 3rd singular inanimate. There are two potential inanimate antecedents in the preceding context (*Sevilla* 'Seville' and *autobús* 'bus'). The informant wants to say that the city was impressive, hence he/she uses the overt NP *la ciudad* 'the city' to avoid ambiguity (otherwise, the use of a null pronoun could be interpreted as (i) it is Seville that was impressive, or (ii) it is the bus that was impressive). Sentence (16) shows how the null pronoun (*pro*) is used pragmatically to refer to the 3rd person inanimate antecedent *El paseo de los ingleses* 'The English avenue'.

(15) [Context: The informant is talking about her arrival to Seville (Spain)]
Cuando llegué a **Sevilla**, supe que había un **autobús** que iba al centro de la ciudad. No pensaba que hubiera ninguno, y por eso, había planeado ir en taxi. **La ciudad** era muy impresionante. [AK, upper-advanced, CEDEL2 corpus]
'When I arrived in **Seville**, I know that there was a **bus** going to the city centre. I didn't think there was going to be any, that's why I had planned on going by taxi. **The city** was very impressive.'

(16) [Context: The informant is talking about her trip to a university course in France and how she went from her apartment to the university every day]
Cada día caminaba de mi apartamento a la universidad por "**El paseo de los ingleses**". ***pro*** era un camino muy lindo con vistas de hoteles y también el mar azul y claro del Mediterráneo. [ARGL, upper-advanced, CEDEL2 corpus]
'Every day I walked from my apartment to the university via the "**The English avenue**". **It** was a beautiful walk with sights overlooking the hotels and the sky-blue Mediterranean'.

Regarding *3rd person plural animate* (*NP/ellos/ellas/pro* 'NP/they'), Table 9 shows that learners' production of unpragmatic forms (9.7% upper-advanced, 9.6% lower advanced) is higher than natives' production (5.6%), though these differences are non-significant (Fisher's Exact Test: upper-advanced vs. natives: p=0.348; lower-advanced vs. natives: p=0.394). These results indicate that, while production of 3rd plural animate is largely pragmatic for all groups but learners show a higher (but non-significant) percentage of unpragmatic production than natives do. While differences between learners and natives are non-significant, results lends support to H_1 since learners do show some pragmatic deficits with 3rd plural animate.

Table 9: Group * Pragmaticality of 3rd plural animate

			PRAGMATICALITY of 3rd plu animate		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	47	5	52
		% within GROUP	**90.4%**	**9.6%**	100.0%
	UPP-ADV	Count	102	11	113
		% within GROUP	**90.3%**	**9.7%**	100.0%
	SPANISH	Count	34	2	36
		% within GROUP	**94.4%**	**5.6%**	100.0%

Finally, regarding *3rd plural inanimate* (*NP/pro* 'they'), learners show native-like behaviour (Table 10) since they do not produce unpragmatic tokens (0% upper-advanced, 4.5% lower-advanced, which corresponds to only one residual token, which inflates the percentage due to the low number of productions, i.e., 22). Spanish natives do not produce any unpragmatic tokens either (0%). These results confirm that with 3rd plural inanimate, learners' behaviour is native-like as their production is pragmatic and they do not significantly differ from Spanish natives (since both the Spanish native group and the upper-advanced group produced 100% of pragmatic cases, there are no significant differences; regarding the lower-advanced group, they produced only 1 token which corresponds to 4.5%, but this difference is non-significant when compared to the natives, p=0.710 with Fisher's Exact Test). These results support H_1 as learners do not show any pragmatic deficits with 3rd plural inanimate.

Table 10: Group * Pragmaticality of 3rd plural inanimate

			PRAGMATICALITY of 3rd plu inanim		Total
			PRAGMATIC	UNPRAGMATIC	
GROUP	LOW-ADV	Count	21	1	22
		% within GROUP	**95.5%**	**4.5%**	100.0%
	UPP-ADV	Count	9	0	9
		% within GROUP	**100.0%**	**0.0%**	100.0%
	SPANISH	Count	14	0	14
		% within GROUP	**100.0%**	**0.0%**	100.0%

To summarise (Table 11), learners show deficits only with 3rd person animate (in particular in the singular number), as predicted by H_1. By contrast, the rest of the pronominal paradigm is intact in our learners' representation. This suggests that deficits at the syntax-discourse interface are selective and do not affect the whole pronominal paradigm, as stipulated in H_1.

Table 11: Summary of learners' results on person and number

	SINGULAR	PLURAL
1st	**No deficits:** Learners behave statistically like Spanish.	**No deficits:** Learners behave statistically like Spanish.
2nd	**No deficits:** Learners behave statistically like Spanish. [BUT more research needed due to low frequencies]	**No deficits:** Learners behave statistically like Spanish. [BUT more research needed due to low frequencies]
3rd	☞**Deficits: Animate** Learners behave differently from Spanish.	**No deficits: Animate** Learners behave statistically like Spanish. ☞**BUT** learners produce more unpragmatic tokens than Spanish (non sig.)
	No deficits: Inanimate Learners behave statistically like Spanish.	**No deficits: Inanimate** Learners behave statistically like Spanish.

6.3 Information status and pragmaticality of the subject

In this section I focus on the (a)pragmaticality of the information of status of the subject, i.e., whether the use of the full NP/overt pronoun/null pronoun is pragmatically correct.

First, consider *topic* contexts, where a null pronoun is expected to express topic continuity (Table 12, shown graphically in Figure 5A). Both learner groups (12.3% lower-advanced, 8.8% upper-advanced) produce a considerable amount of unpragmatic topic, which is significantly higher than those of Spanish natives (3%): upper-advanced vs. natives: χ^2=5.621, df=2, p=0.018; lower-advanced vs. natives: χ^2=11.269, df=2, p=0.001. Notice that the learner groups do not significantly differ from each other (χ^2=2.137, df=2, p=0.144). Example (17) illustrates unpragmatic topic: the upper-advanced learner produces an overt pronoun (*ellos* 'they') to refer to the immediate antecedent (*los chicos* 'the boys') in a topic-continuity context which would require a null pronoun, as there is no possible ambiguity. Notice that the learner subsequently uses a pragmatic null pronoun. Similarly, in (18) the lower-advanced learner is talking about *la madre* 'the mother'. A null pronoun is expected, as used in the first instance (*...pero **pro** es muy trabajadora* 'but she is hard-working'), but the learner later uses two overt pronouns (*ella* 'she'), which are pragmatically redundant.

(17) [Context: The informant is talking about a group of teenagers in Ecuador. She was her teacher during her stay there]
Cuando me integré en el grupo, en realidad **los chicos** no podían cantar ni tocar muy bien. Sin embargo, poco a poco a lo largo del año, **#ellos** se mejoraron bastante y no sólo ***pro*** desarrollaron su grupo y sus talentos musicales, sino también ***pro*** crecieron como individuos. [ELS, Upper-Advanced, CEDEL2 corpus]
'When I got into the group, **the boys** could not really sing or play well. However, little by little during that year, **they** improved a lot, ***pro*** developed their group and their musical skills, and ***pro*** grew up as individuals.'

(18) [Context: The informant is talking about the main character of the film "Spanglish"]
La madre no puede hablar inglés pero ***pro*** es muy trabajadora. **#Ella** empieza a trabajar ... **#Ella** no puede comunicar[se] con esta familia …. [SMM, Low-Advanced, CEDEL2 corpus]
'**The mother** cannot speak English but ***pro*** is very hard-working. **She** starts working ... **She** cannot communicate with the family...'

Table 12: Group * Pragmaticality of Topic

			PRAGMATICALITY		Total
			PRAG TOPIC	UNPRAG TOPIC	
GROUP	LOW-ADV	Count	299	42	341
		% within GROUP	**87.7%**	**12.3%**	100.0%
	UPP-ADV	Count	290	28	318
		% within GROUP	**91.2%**	**8.8%**	100.0%
	SPANISH	Count	159	5	164
		% within GROUP	**97.0%**	**3.0%**	100.0%

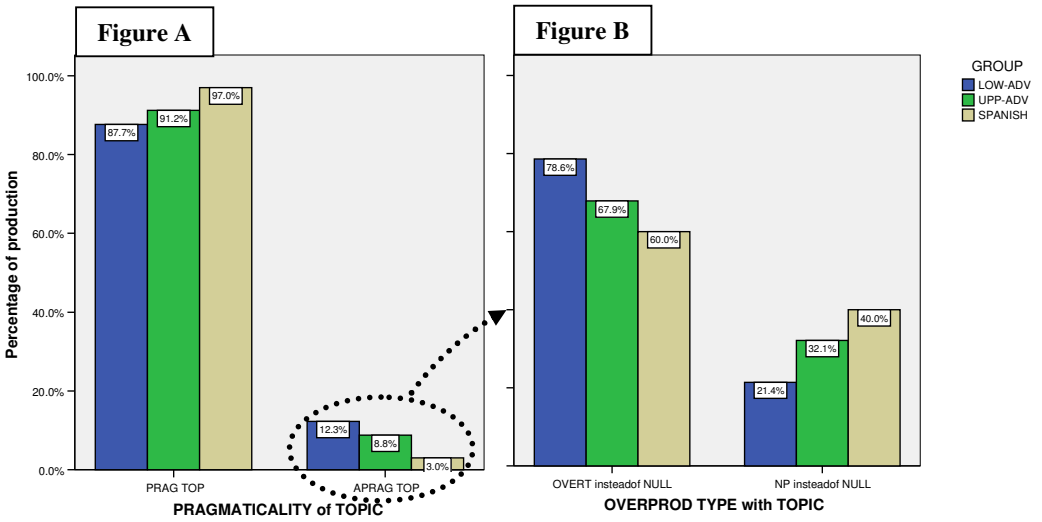
To summarise, while the production of topic subjects is largely correct for all groups, the two learner groups produced a large proportion of unpragmatic topic subjects, which is significantly higher than the Spanish natives' production. This corroborates H_1 in that learners at advanced and very advanced levels of competence show deficits with discursive features like [Topic].

Importantly, recall that the unpragmatic errors with topic can be of two types, namely, (i) the production of an overt pronoun when a null pronoun is required, and (ii) the production of a full NP when a null pronoun is required. Figure 5B (which corresponds to Table 13), shows the percentage of overproduction types with topic, out of the percentages of unpragmatic topic (Figure 5A). As can be appreciated in Figure 5B, all groups overproduce more overt pronouns than full NPs when a null pronoun is required, although the trend of overproduction of overt pronouns decreases towards the native norm, but the trend of overproduction of full NPs increases towards the native norm.

Table 13: Group * Overproduction type with Topic

			OVERPRODUCTION TYPE		Total
			OVERT instead of NULL	NP instead of NULL	
GROUP	LOW-ADV	Count	33	9	42
		% within GROUP	**78.6%**	**21.4%**	100.0%
	UPP-ADV	Count	19	9	28
		% within GROUP	**67.9%**	**32.1%**	100.0%
	SPANISH	Count	3	2	5
		% within GROUP	**60.0%**	**40.0%**	100.0%

Figure 5: Pragmaticality of Topic



Instances of overproduction of overt pronouns are shown in (19), where the use of an overt pronoun (*ellos* 'they') to refer to the antecedent *los chicos* 'the boys' is pragmatically redundant. A null pronoun would be more adequate, as the learner correctly produces in the next clauses. Additionally, overproduction of full NPs is shown in (19) as well, where the final instance of *los chicos* 'the boys' (in *los chicos habían decidido...*) is redundant, since *los chicos* is the topic in the discourse and, therefore, requires a null pronoun (*pro*), as the learner correctly produces in the previous lines.

(19) [Context: The informant is talking about a group of teenagers in Ecuador. She was her teacher during her stay there]
Cuando me integré en el grupo, en realidad **los chicos** no podían cantar ni tocar muy bien. Sin embargo, poco a poco a lo largo del año, **#ellos** se mejoraron bastante y no sólo ***pro*** desarrollaron su grupo y sus talentos musicales, sino también ***pro*** crecieron como individuos. Tuvimos un retiro en que hablamos sobre las razones por las cuales **#los chicos** habían decidido participar en el grupo...[ELS, Upper-Advanced, CEDEL2 corpus]
'When I got into the group, **the boys** could not really sing or play well. However, little by little during that year, **they** improved a lot, ***pro*** developed their group and their musical skills, and ***pro*** grew up as individuals. We retreated to a quiet place where we talked about the reasons why **the boys** had decided to participate in the group...'

To summarise, the data on the pragmaticality of topic and the types of overproduction errors with topic indicate that learners significantly overproduce more overt material (mainly overt pronouns but also full NPs) than Spanish natives do in topic continuity contexts. This indicates that advanced and even end-state learners show overproduction deficits at the syntax-discourse interface, as predicted by H_1.

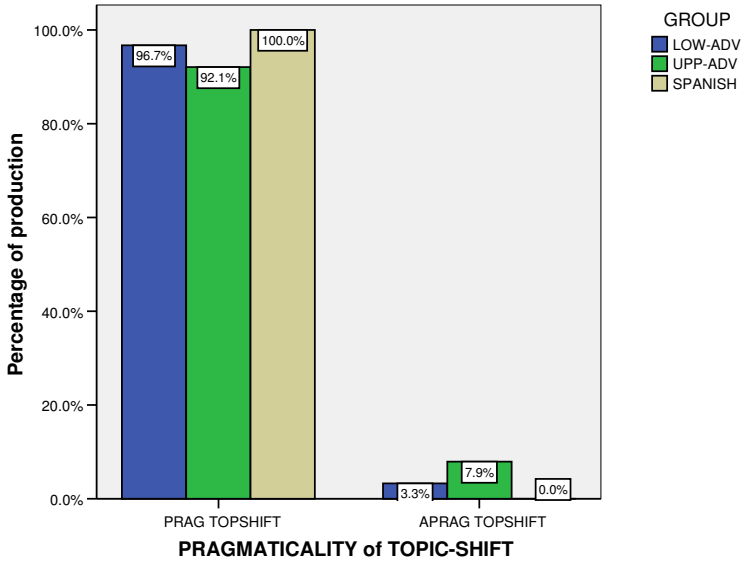
Consider now the (a)pragmaticality of *topic-shift*. Recall that to mark a change of topic, overt material is required (either an overt pronoun or a full NP) in native Spanish, which prevents ambiguity between referents. The type of unpragmatic errors with topic-shift relate to underproduction, i.e., production of a null pronoun when overt material is required (cf. sentence (14)). Table 14 (shown visually in Figure 6) indicates that learners produce relatively low percentages of underproduction (7.9% upper-advanced, 3.3% lower-advanced), which corresponds to a few tokens, while Spanish natives never underproduce (0%). Technically, learners behave like natives in

topic-shift contexts, as there are no significant differences, but learners do still produce some residual null subjects when an overt pronoun is required (Fisher's Exact Test: upper-advanced vs. natives p=0.57, just non-significant; lower-advanced vs. natives p=0.283). This lends support to our H_1, as learners show underproduction deficits at the syntax-discourse interface in topic-shift contexts, though these deficits are not significantly different from Spanish natives' zero underproduction rate.

Table 14: Group * Pragmaticality of Topic-shift

			PRAGMATICALITY		Total
			PRAG TOPICSHIFT	UNPRAG TOPICSHIFT	
GROUP	LOW-ADV	Count	88	3	91
		% within GROUP	**96.7%**	**3.3%**	100.0%
	UPP-ADV	Count	58	5	63
		% within GROUP	**92.1%**	**7.9%**	100.0%
	SPANISH	Count	47	0	47
		% within GROUP	**100.0%**	**0.0%**	100.0%

Figure 6: Pragmaticality of Topic-shift



7 DISCUSSION

Results on the non-pragmaticality of topic and topic-shift confirm the general hypothesis that advanced and end-state L2 learners show deficits at the syntax-discourse interface with pronominal subjects. In particular, English-speaking learners of L2 Spanish produce a considerable proportion of unpragmatic subjects in topic contexts, i.e., they produce overt material (mostly overt pronouns, but also full NPs) in topic-continuity contexts where a null pronoun would be pragmatically acceptable. They also show a residual and non-significant amount of underproduction, i.e., production of a null subject in topic-shift contexts requiring overt material.

It may be argued that these results are based on *written* production data from the CEDEL2 corpus and therefore do not reflect directly learners' spontaneous intuitions as in *oral* corpora or in experimental work. However, the general findings on overproduction and underproduction with pronominal subjects clearly replicate those of previous research. This indicates that written corpora are a legitimate source for exploring learners' competence (see Granger 2002a, 2002b, 2003).

Most L2 Spanish studies reviewed above report on *unidirectionality* in the type of production, i.e., overproduction (and not underproduction) is the norm, though *bidirectionality* has been also attested (Montrul & Rodríguez-Louro 2006). This entails that errors of both overproduction and underproduction are found, but there is an asymmetry, since overproduction errors are significantly greater than underproduction errors. The results of the current study show bidirectionality, though significant differences between learners and natives were found only with overproduction (and not with underproduction). Sorace (2006) argues that deficits at the syntax-discourse interface in Italian L1 attrition are unidirectional (overproduction only) as a result of processing deficits at the interfaces. In particular, overt pronominal subjects can be used as the 'default' processing option when the language processor is overloaded as it tries to process both syntactic and interface information. Sorace & Filiaci (2006) also report unidirectionality in L2 Italian, which is claimed to be a result of representation deficits. In particular, [+interpretable] features like [Topic] at the syntax-discourse interface are vulnerable. Whatever the causes (whether processing deficits or incomplete representations), unidirectionality has been also observed in native Spanish speakers, who tend to produce some overt pronominal subjects in topic-continuity contexts (Alonso-Ovalle et al 2002). Our data also show that Spanish natives do produce some unpragmatic overt material in topic-continuity contexts (see Table 12 and Table 13). This is surprising, as the theory would predict that a null pronoun is the pragmatically preferred choice to mark topic continuity. Further research will need to elucidate whether L2 learners overproduce overt subjects as a consequence of a supposedly universal default mechanism.

Results on **phi-features** have shown that learners' behaviour is statistically similar to Spanish natives regarding 1st person (singular and plural), 2nd person (singular and plural) and 3rd person inanimate. By contrast, learners behave significantly different from natives as regards 3rd person animate singular (and, to a lesser extent, plural). In other words, it seems that learners show robust knowledge at the syntax-discourse interface with the deictic use of pronouns (i.e., when these refer to speech-act participants: 1st and 2nd person) and with inanimacy (3rd neutral). Vulnerability at the interface is observed only with 3rd person animate, i.e., with the anaphoric use of the pronouns. This indicates that deficits with discursive features like [Topic] and [Topic-shift] at the syntax-discourse interface are *selective* and do not affect the whole pronominal paradigm, as certain features (3rd animate) are vulnerable, while the rest appear to be relatively robust. Montrul (2006) compares incomplete acquisition in heritage speakers with other cases of language change in progress in Brazilian Portuguese (BP). Interestingly, BP is shifting towards the English value of the null subject parameter (i.e., a non-null subject language) and the most affected person is 3rd singular, as reported in this study.

Additionally, the observed syntax-discourse deficits are mainly a matter of overproduction, since overt pronouns, which are specified as [Topic-shift], can be

occasionally specified for [Topic]. Such deficit interacts with participant features of the Feature Geometry, as deficits affect the [non-participant] feature. Table 15 shows a likely mental representation of discursive pronominal features for both Spanish natives and L2 Spanish learners based on the results. The discursive feature [Topic] is realised in Spanish via a null pronoun (*pro*) unspecified for person, number and animacy, i.e., the null pronoun can refer to any person, number and animate or inanimate entities. English-speaking learners of L2 Spanish also realise the [Topic] feature mainly via a null pronoun for the whole pronominal paradigm, but they show a residual deficit in the sense that discursive [Topic] can be realised via an overt pronoun when it is specified for [3] and [+animate]. In topic-shift contexts, Spanish natives as well as learners realise the discursive feature [Topic-shift] as an overt pronoun for the whole pronominal paradigm. As indicated by the round brackets, learners may exceptionally encode [Topic-shift] via a null pronoun with a certain combination of phi-features: [3] and [+animate]. While this last unpragmatic choice shows a exceptionally low frequency in the learners, it is not significantly different from Spanish natives (who never produce it).

Table 15: Likely mental representation of discursive pronominal features

	Topic contexts		**Topic-shift contexts**	
	Discurs. Feature	**Realisation**	**Discurs. Feature**	**Realisation**
Spanish natives	[Topic]	***pro*** [***u***pers][***u***num][***u***anim]	[Topic-shift]	**overt** [***u***pers][***u***num][***u***anim]
Spanish L2ers	[Topic]	{***pro*** [***u***pers][***u***num][***u***anim] / **overt** [3][+anim]}	[Topic-shift]	{**overt** [***u***pers][***u***num][***u***anim] / (***pro*** [3][+anim])}

Recall that in L1 acquisition the relative order of acquisition is typically 1st and 3rd inanimate, then 3rd animate. This path of development is claimed to be guided by the Feature Geometry, which is ultimately constrained by UG. Our L2 data suggest that English-speaking adult learners' knowledge of L2 Spanish may be also guided by such Feature Geometry in the sense that (i) 1st and 3rd inanimate must have been acquired before end-states as they do not appear to lead to deficits at the syntax-discourse interface in our learners, while (ii) 3rd animate is acquired later and can lead to residual deficits at the interface, as observed. As it stands, this proposal needs further empirical corroboration by means of L2 developmental studies testing (i) how the Feature Geometry interacts with discursive features like topic and topic-shift in L2, and (ii) whether the observed path of development in child L1 acquisition is also observed in adult L2 acquisition.

In the context of general theories of SLA, recall from the introduction that most proposals have focused on formal (morphosyntactic) features operating at narrow syntax, and not on discursive features like [Topic] and [Focus] operating at the syntax-discourse interface. Despite their original aim, these theories centre around two opposing views, as the source of deficits can be of two general types: *representational* vs. *mapping* deficits. Let us see in turn how these theories can account for our data.

The *representational* approach stipulates that learners' linguistic representation of the relevant features show deficits of various kinds. Beck's (1998) *Local Impairment*

Hypothesis states that strong (i.e., uninterpretable) features become unspecified, which leads to a permanent state of unconstrained optionality. If this were the case, we would expect our learners to show a more random behaviour and produce optionally both overt and null pronominals in topic contexts and topic-shift contexts, which is contrary to our results. Hawkins & Chan's (1997) *Failed Functional Feature Hypothesis*) stipulates that L2 uninterpretable formal features become defective if they have not been acquired before the Critical Period. This proposal is difficult to test in the current study since the learner sample contains both pre- and post-pubescent learners (see Table 16 and Table 17)

The *mapping* approach claims that learners' linguistic representation of features is intact and that the observed deficits are a result of performance when mapping linguistic knowledge onto the relevant morphophonological forms, i.e., the learner cannot retrieve the required morphological item (Prévost & White 2002). If this were the case, we would expect our learners to show the inverse deficit, i.e., they would underproduce null subjects in topic-shift contexts significantly more than they do now since they would supposedly have surface problems when mapping the relevant features onto the correct pronominal form and thus retrieve a null form (*pro*). Additionally, we would expect our learners not to overproduce overt material in topic-continuity contexts, which is contrary to fact.

Current approaches to the problem of the interfaces in language acquisition (e.g., Sorace 2004, 2005, 2006) postulate that it is *interpretable* discursive features like [Topic] that are vulnerable since they operate at the interface between the computational system and other systems of thought. This is a plausible explanation for the current data but (i) it cannot account for why deficits are selective, i.e., they are observed only with 3rd person animate subjects and not with the rest of the paradigm, and (ii) it would entail that the problem is discursive (i.e., learners are unable to acquire the discourse-related features [Topic] and [Topic-shift] to a native-like extent) rather than a problem with the realisation of these discursive features onto the correct pronominal form. There is no principled reason to believe that advanced learners are unable to interpret the information status of [Topic] as representing old information and [Topic-shift] as a change of discourse referent, since information status is a cross-linguistic universal (e.g., Casielles-Suarez 2004, Vallduví 1992, Vallduví & Engdahl 1996). It seems more reasonable to assume that the learners' deficits lie in the realization of the relevant discursive feature onto the correct pronominal form. Further research will need to investigate more closely the locus of deficits with discursive features and how these interact with the Feature Geometry.

Finally, the conclusion that deficits at the syntax-discourse interface with pronominal subjects are *selective* in L2 Spanish must be taken cautiously since the evidence presented here comes from **corpus data**. Two of the well-known limitations of data-driven approaches to language acquisition is that (i) performance data does not guarantee a genuine reflection of learners' competence and (ii) biases in the corpus sample can skew the data (as is the case in this study with 2nd person). Corpus results must be complemented with detailed experimental work. Future research needs to address whether the observed selective impairment in production data is also attested in comprehension data and, if so, what is the ultimate source of the observed deficits.

8 CONCLUSION

This study has shown that English-speaking learners of L2 Spanish at advanced and very advanced levels of competence show deficits with the properties that constrain the distribution of overt and null pronominal subjects at the syntax-discourse interface, as previous SLA research has shown. Unlike previous SLA research, it has been shown that such deficits are *selective*, as they do not affect equally all phi-features in the pronominal paradigm, but rather a subset of them. It is concluded that the observed deficits stems from the way Universal Grammar constrains pronominal features.

9 REFERENCES

Ädel, A., 2006. Metadiscourse in L1 and L2 English. Amsterdam: John Benjamins.

Al-Kasey, T., Pérez-Leroux, A.-T., 1998. Second language acquisition of Spanish null subjects. In: Flynn, S., Matohardjono, G., O'Neil, W. (eds.), The Generative Study of Second Language Acquisition, 161-185. Hillsdale, N.J.: Lawrence Erlbaum.

Alonso-Ovalle, L., Fernández-Solera, S., Frazier, L., Clifton, C., 2002. Null vs. overt pronouns and the topic-focus articulation in Spanish. Journal of Italian Linguistics 14, 151-169.

Beck, M.-L., 1997. Regular verbs, past tense and frequency: tracking down a potential source of NS/NNS competence differences. Second Language Research 13, 93-115.

Belletti, A., Leonini, C., 2004. Subject inversion in L2 Italian. In: Foster-Cohen, S., Sharwood Smith, M., Sorace, A., Mitsuhiko, O. (eds.), EUROSLA yearbook: Volume 4, 95-118. Amsterdam: John Benjamins.

Benveniste, E., 1971. Problems in General Linguistics. In: Benveniste, E. (ed.), The nature of pronouns. Coral Gables, FL: University of Miami Press.

Bianchi, V., 2005. The person asymmetry: underspecification of person and number features? Paper presented at the workshop 'Underspecification in morphology and syntax', University of Cologne (Germany).

Blackwell, S.E., 1998. Constraints on Spanish NP anaphora: The syntactic versus the pragmatic domain. Hispania 81, 606-618.

Bloomfield, L., 1933. Language. New York: Holt, Rinehart and Winston.

Casielles-Suarez, E., 2004. The Syntax-Information Structure Interface: Evidence from Spanish and English. Oxford: Routledge.

Davies, W.D., 1996. Morphological uniformity and the null subject parameter in adult SLA. Studies in Second Language Acquisition 18, 475-493.

Diaconescu, C.R., Goodluck, H., 2002. Structural and discourse factors in Romanian L2 English Learners' interpretation of pronouns. In: Liceras, J.M., Zobl, H., Goodluck, H. (eds.), Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002), 71-75. Somerville, MA: Cascadilla Press.

Escutia, M., 2002. Universal Grammar, transfer and variability: a case study. Estudios Ingleses de la Universidad Complutense 10, 67-85.

Forchheimer, P., 1953. The Category of Person in Language. Berlin: Walter de Gruyter.

Fruit, M.N., 2007. Divergence at the syntax-discourse interface: Evidence from the L2 acquisition of contrastive focus in European Portuguese. Paper presented at ISB6 (6th International Symposium on Bilingualism), University of Hamburg.

Grinstead, J., 2004. Subjects and interface delay in child Spanish and Catalan. Language 80, 40-72.

Granger, S., Dagneaux, E., and Meunier, F., eds. 2002a. *International Corpus of Learner English [inc. CD ver 1.1]*. Louvain: UCL Presses Universitaires de Louvain.

Granger, S., Hung, J. and Petch-Tyson, S., eds. 2002b. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Granger, S. and Petch-Tyson, S., eds. 2003. Extending the Scope of Corpus-Based Research: New Applications, New Challenges. Amsterdam: Rodopi.

Gundel, J.K., 1998. On different kinds of focus. In: Bosch, P. (ed.), Focus: Linguistic, Cognitive and Computational Perspectives, 293-205. Cambridge: Cambridge University Press.

Hanson, R., 2000. Pronoun acquisition and the morphological feature geometry. Calgary Working Papers in Linguistics 22, 1-14.

Harley, H., Ritter, E., 2002a. Person and number in pronouns: A feature-geometric analysis. Language 78, 482-526.

Harley, H., Ritter, E., 2002b. Structuring the bundle: A universal morphosyntactic feature geometry. In: Simon, H.J., Weise, H. (eds.), Pronouns: Grammar and Representation.

Hawkins, R., 2001. Second Language Syntax: A Generative Introduction. Oxford: Blackwell.

Hawkins, R., Chan, C.Y., 1997. The partial availability of Universal Grammar in second language acquisition: The 'failed functional features hypothesis'. Second Language Research 13, 187-226.

Helland, C., 2004. Attrition and syntactic subjects in Catalan. Paper presented at The Romance Turn, Universidad Nacional de Educación a Distancia (Madrid, September).

Hertel, T.J., 2003. Lexical and discourse factors in the second language acquisition of Spanish word order. Second Language Research 19, 273-304.

Jaeggli, O., 1982. Topics in Romance Syntax. Dordrecht: Foris.

Jespersen, O., 1924. The Philosophy of Grammar. London: George Allen & Unwin.

Kayne, R.S. (2000) Person morphemes and reflexives in Italian, French, and related languages.
In Kayne. R.S. (2000) Parameters and universals, 131-162. Oxford: Oxford University Press.

Kanno, K., 1997. The acquisition of null and overt pronominals in Japanese by English speakers. Second Language Research 13, 265-287.

Kras, T., 2006. Interface instability in bilingual language acquisition: in search of the causes. Paper to be presented in EUROSLA 2006 (13-16 September, Turkey).

Lafond, L., Hayes, R., Bhatt, R., 2001. Constraint demotion and null-subjects in Spanish L2 acquisition. In: Camps, J., Wiltshire, C. (eds.), Romance Syntax, Semantics and L2 Acquisition Amsterdam: John Benjamins.

Liceras, J.M., 1989. On some properties of the "pro-drop" parameter: looking for missing subjects in non-native Spanish. In: Gass, S.M., Schachter, J. (eds.), Linguistic Perspectives on Second Language Acquisition, 109-133. Cambridge: CUP.

Liceras, J.M., Díaz, L., 1999. Topic drop versus pro-drop: null subjects and pronominal subjects in the Spanish L2 of Chinese, English, French, German and Japanese speakers. Second Language Research 15, 1-40.

Liceras, Juana M, Zobl, Helmut, and Goodluck, Helen. The Role of Formal Features in Second Language Acquisition. 2007. Mahwaw, NJ, Lawrence Erlbaum Associates.

López Ortega, N.R., 2006. Presencia y ausencia de sujeto en el discurso oral de español no-nativo: un estudio cualitativo longitudinal en contexto de inmersión. In: Carrió Pastor, M.L. (ed.), Perspectivas interdisciplinares de la lingüística aplicada (tomo 1), 175-186. Valencia: Universidad Politécnica de Valencia.

Lozano, C., 2002a. Knowledge of expletive and pronominal subjects by learners of Spanish. ITL Review of Applied Linguistics 135/6, 37-60.

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

Lozano, C., 2006a. Explaining the "syntax-before-discourse" phenomenon: Pronominal subject distribution in L1 Greek-L2 Spanish. Paper presented at "The Romance Turn II", University of Utrecht, 7th September.

Lozano, C., 2006b. Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish. Second Language Research 22, 1-43.

Lozano, C., 2006c. The development of the syntax-discourse interface: Greek learners of Spanish. In: Torrens, V., Escobar, L. (eds.), The Acquisition of Syntax in Romance Languages, 371-399. Amsterdam: John Benjamins.

Mastropavlou, M., 2006. *The Role of Phonological Salience and Feature Interpretability in the Grammar of Typically Developing and Language Impaired Children.* Unpublished PhD dissertation: Aristotle University of Thessaloniki.

Montrul, S., 2004. Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. Bilingualism: Language and Cognition 7, 125-142.

Montrul, S., 2006. Bilingualism, incomplete acquisition and language change. In: Lefebvre, C., White, L., Jourdens, C. (eds.), *L2 Acquisition and Creole Genesis. Dialogues*, 379-400. Amsterdam: John Benjamins.

Montrul, S., Rodríguez-Louro, C., 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, V., Escobar, L. (eds.), The Acquisition of Syntax in Romance Languages Amsterdam: John Benjamins.

Pérez-Leroux, A.T., Glass, W.R., 1999. Null anaphora in Spanish second language acquisition: probabilistic versus generative approaches. Second Language Research 15, 220-249.

Pérez-Leroux, A.T., Glass, W.R., 1997. OPC effects on the L2 acquisition of Spanish. In: Pérez-Leroux, A.T., Glass, W.R. (eds.), Contemporary Perspectives on the Acquisition of Spanish, volume I: Developing Grammars, 149-165. Sommerville, MA: Cascadilla Press.

Pérez-Leroux, A.T., Scott, A.M., Hertel, T., Kellar, V., Glass, W.R., 1999. Sources of knowledge in second language acquisition. Spanish Applied Linguistics 3, 33-63.

Phinney, M., 1987. The pro-drop parameter in second language acquisition. In: Roeper, T., Williams, E. (eds.), Parameter Setting, 221-238. Dordrecht: Reidel.

Pinto, M., 2006. Subject pronouns in biliguals: interference or maturation? In: Torrens, V., Escobar, L. (eds.), The Acquisition of Syntax in Romance Languages, 331-350. Amsterdam: John Benjamins.

Polio, C., 1995. Acquiring nothing? The use of zero pronouns by nonnative speakers of Chinese and the implications for the acquisition of nominal reference. Studies in Second Language Acquisition 17, 353-377.

Reinhart, T., 1995. Interface Strategies. OTS Working Papers, TL-95-002.

Rizzi, L., 1982. Issues in Italian Syntax. Dordrecht: Foris.

Rizzi, L., 1997. A parametric approach to comparative syntax: properties of the pronominal system. In: Haegeman, L. (ed.), The New Comparative Syntax, 268-285. London: Longman.

Rochemont, M.S., 1998. Phonological focus and structural focus. In: Culicover, P.W., McNally, L. (eds.), Syntax and Semantics (vol. 29): The Limits of Syntax, 337-363. London: Academic Press.

Satterfield, T., 2003. Economy of interpretation: Patterns of pronoun selection in transitional bilinguals. In: Cook, V. (ed.), Effects of the Second Language on the First, 214-233. Clevedon: Multilingual Matters.

Serratrice, L., 2004. Anaphoric interpretation of null and overt pronominal subjects in Italian. Paper presented at The Romance Turn, Universidad Nacional de Educación a Distancia, Madrid, September.

Serratrice, L., Sorace, A., Paoli, S., 2004. Transfer at teh syntax-pragmatics interface: subjects and objects in Italian-English bilingual and monolingual acquisition. Bilingualism: Language and Cognition 2, 169-186.

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

Sorace, A., 2006. Possible manifestations of shallow processing in advanced second language speakers. Applied Psycholinguistics 27, 88-91.

Sorace, A., 2005. Selective optionality in language development. In: Cornips, L., Corrigan, K.P. (eds.), Syntax and variation: Reconciling the biological and the social, 55-80. Amsterdam: John Benjamins.

Sorace, A., Filiaci, F., 2006. Anaphora resolution in near-native speakers of Italian. Second Language Research 22, 339-368.

Tsimpli, I.-M., Sorace, A., 2006. Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. *BUCLD Proceedings* 30, 653-664.

Tsimpli, I.M., Sorace, A., Heycock, C., Filiaci, F., 2004. First language attrition and syntactic subjects: a study of Greek and Italian near-native speakers of English. International Journal of Bilingualism 8, 257-277.

University of Wisconsin, 1998. The University of Wisconsin College-Level Placement Test: Spanish (Grammar) Form 96M. Madison, WI: University of Wisconsin Press.

Valenzuela, E., McIlwraith, E., 2007. Topic and focus in early bilingual grammars. Paper presented at ISB6 (6th International Symposium on Bilingualism), University of Hamburg.

Vallduví, E., 1992. The Informational Component. Garland: New York.

Vallduví, E., Engdahl, E., 1996. The linguistic realization of information packaging. Linguistics 34, 459-519.

White, L., 1985. The "pro-drop" parameter in adult second language acquisition. Language Learning 35, 47-62.

White, L., 2003. Second Language Acquisition and Universal Grammar. Cambridge: Cambridge University Press.

White, L., 1989. Universal Grammar and Second Language Acquisition. Amsterdam: John Benjamins.

Zubizarreta, M.L., 1998. Prosody, Focus, and Word Order. Cambridge, MASS: MIT Press.

Zubizarreta, M.L., 1999. Las funciones informativas: tema y foco. In: Bosque, I., Demonte, V. (eds.), Gramática descriptiva de la lengua española (vol. 3) Madrid: Espasa.

APPENDIX

Table 16: Upper-advanced group's data

Initials	Profic	Age	Age first exposure	Years instruct	Stay (yrs;mths)	# words	# tags	% pronom subjects
ELS	100%	25	15	5	18	647	40	6.2%
JSD	100%	20	12	9	13	819	41	5.0%
KEM	100%	28	20	8	10	1032	65	6.3%
KEM2	100%	28	20	8	10	866	50	5.8%
OPE	98%	48	15	11	17	798	43	5.4%
CPB	98%	21	11	11	3	1010	45	4.5%
MEA	98%	52	3	?	204	865	38	4.4%
JGP	98%	58	13	7	6	554	27	4.9%
LP	98%	51	16	6	4	781	46	5.9%
JEL	98%	20	13	8	7	816	58	7.1%
	99% *(mean)*	**35** *(mean)*	**14** *(mean)*	**8** *(mean)*	**29** *(mean)*	**8188** *(total)*	**453** *(total)*	**5.5%** *(mean)*

Table 17: Lower-advanced group's data

Initials	Profic	Age	Age first exposure	Years instruct	Stay (yrs;mths)	# words	# tags	% pronom subjects
ARGL	95%	49	16	?	9	981	63	6.4%
SAR	95%	58	20	3	12	858	46	5.4%
DKH	95%	28	6	15	3	870	65	7.5%
AK	95%	23	13	10	1	907	63	6.9%
SMG	93%	19	16	3	1	840	65	7.7%
KMH	93%	21	10	11	4	834	47	5.6%
CCO	93%	20	14	5	0	849	44	5.2%
SMM	91%	18	13	6	0	705	50	7.1%
CN	91%	55	26	3	228	871	42	4.8%
ACC	91%	25	12	2	7	806	43	5.3%
	93% *(mean)*	**32** *(mean)*	**15** *(mean)*	**6** *(mean)*	**27** *(mean)*	**8521** *(total)*	**528** *(total)*	**6.2%** *(mean)*

Table 18: Spanish native group's data

Initials	Age	# words	# tags	% pronom subjects
MSCL	33	642	17	2.6%
GDC	51	392	15	3.8%
CAC	33	377	17	4.5%
MCL	40	667	38	5.7%
AHN	33	470	30	6.4%
ENB	39	442	21	4.8%
MHMR	56	335	17	5.1%
CLB	32	456	17	3.7%
RSZ	26	538	27	5.0%
MDD	32	422	28	6.6%
SPH	30	372	32	8.6%
CMM	33	841	40	4.8%
	37 *(mean)*	**5954** *(total)*	**299** *(total)*	**5.1%** *(mean)*

FOOTNOTES

[1] In current generative work, what was traditionally analysed as Noun Phrase (NP) is now analysed as Determiner Phrase (DP). For simplicity reasons, I will choose the terminology NP throughout, as the precise syntactic analysis (whether NP or DP) is irrelevant in this work.

[2] In the generative literature, there are many theoretical explanations of why null-subject languages allow *pro*, while non-null subject languages cannot. While the precise technical details of the mechanisms licensing *pro* vary in these studies, what is common to all of them is that *formal* features are responsible for the licensing of *pro*. For different views, see, e.g., Alexiadou & Anagnostopoulou (1989), Rizzi (1997).

[3] In the Feature Geometry structure shown in Figure 1, the expressions in brackets have been added for clarification purposes, e.g., '*participant (=person)*', since what Harley & Ritter term *participant* has been traditionally termed *person*.

[4] Unpragmatic constructions are indicated with the hash mark (#). Note constructions can be pragmatically infelicitous while being grammatically correct, i.e., lack of pragmaticality does not entail agrammaticality.

[5] More information on CEDEL2 is available at: http://www.uam.es/woslac/cedel2.htm

[6] UAM Corpus Tools can be freely downloaded. More information at: http://www.wagsoft.com/CorpusTool/index.html