**Interface conditions on postverbal subjects: a corpus study of L2 English***

Cristóbal Lozano, Universidad de Granada

Amaya Mendikoetxea, Universidad Autónoma de Madrid (Spain)

Running head: Interface conditions on postverbal subjects

Corresponding author:

Cristóbal Lozano

Universidad de Granada

Facultad de Filosofía y Letras

Departamento de Filología Inglesa

Granada 18071, Spain

cristoballozano@ugr.es

*This work has been supported by research grants from the Spanish Ministry of Education and Science (HUM2005-0127/FILO) and the Spanish Ministry of Science and Innovation (FFI2008-01584 and EDU2008-01268). Previous versions were presented in *Corpus Linguistics* 2007, *EUROSLA 17*, and *BUCLD 33*. We wish to thank the audiences at those conferences and all the members of the WOSLAC team (http://www.uam.es/woslac), with whom we have discussed many of the issues dealt with here. We are also grateful to three anonymous reviewers, whose comments have greatly improved this paper. All remaining errors are ours.
ABSTRACT

This paper investigates how syntactic knowledge interfaces with other cognitive systems by focusing on the production of postverbal subjects, V(erb)-S(ubject) order, in an L1 Spanish-L2 English corpus. VS order in both native and L2 English is shown to be constrained by properties operating at three interfaces: (i) LEXICON-SYNTAX: the verb is unaccusative (Unaccusative Hypothesis); (ii) SYNTAX-DISCOURSE: the subject is focus (End-Focus Principle) and (iii) SYNTAX-PHONOLOGY: the subject is heavy (End-Weight Principle). We show that learner produce postverbal subjects under the same interface conditions as in native English. Thus, unaccusativity is a necessary but not a sufficient condition for VS production. However, our learners overproduce VS and show persistent errors in their syntactic encoding. Our findings support recent proposals that these difficulties stem from problems at coordinating syntactic knowledge with knowledge from other external systems, but they suggest that the nature of the difficulties observed is not necessarily external to the syntax.
1 INTRODUCTION: THE MODEL

Over the last decade there has been a shift of perspective in L2 acquisition, along with a parallel shift in linguistic theory. Research has moved on from questions related to parameter resetting and U(niversal) G(rammar) accessibility to how syntactic knowledge interacts with other components of grammar and cognitive (sub)systems in the interlanguage of L2 learners. A number of studies have shown that linguistic phenomena at the interfaces are specially vulnerable in a wide range of acquisitional contexts: L2 acquisition, L1 attrition, child bilingualism; see Sorace & Serratrice (forthcoming) and White (2009) for overviews, as well as § 3 below. An area lending itself to this type of approach is the study of word order in L2 grammars.

In this paper, we analyse the production of Verb-Subject order (VS henceforth) in two corpora of L1 Spanish-L2 English and in a comparable native English corpus. English word order is said to be ‘fixed’, constrained by properties at the lexicon-syntax interface, which determine the syntactic position of verbal arguments. By contrast, Spanish ‘free’ word order is ruled by properties at the syntax-discourse interface (i.e., topic and focus).

The Minimalist Program (Chomsky 1995, 2000, 2005) with its emphasis on interface conditions is the general framework for our study. The language faculty has an ‘internal’ interface, connecting the lexicon, which specifies the elements that enter into computation, and the syntax or computational system, which generates derivations. Derivations are pairs of interface representations, understood as ‘instructions’ for the ‘external’ systems with which the language faculty interacts: Articulatory-Perceptual (i.e., Sensory-Motor (S-M)) and Conceptual-Intentional (C-I)
systems (Chomsky 1995: 169). Interaction with these systems takes place through the interface levels: Phonological Form (PF) and Logical Form (LF), respectively (see Figure 1 in § 4 below). A linguistic expression is then “nothing other than a formal object that satisfies the interface conditions in the optimal way.” (Chomsky 1995: 171). Recent research has focused on the properties which (external) interface conditions impose on the language faculty: e.g., data analysis for language acquisition and computational efficiency (e.g., Epstein 1999, Frampton and Gutmann 1999, Chomsky 2005). SLA research is now beginning to address questions related to the interaction between syntactic knowledge and knowledge from ‘external’ components in L2 grammars.

Two main results have been obtained in this study: (i) the parallel use of VS structures in both L1/L2 English (regarding verb class and the nature of the postverbal subject), indicating that VS appears to be constrained by the same conditions in native and non-native English, and (ii) Spanish learners of L2 English overuse the construction and produce mostly deviant structures. Transfer and, to a lesser extent, the role of input are explored as possible explanations. We also explore other factors, which are at the centre of debate in SLA interface studies: processing limitations (and their interaction with crosslinguistic influence) and the relative difficulty of acquiring lexicon-syntax and syntax-discourse properties.

The remaining of this paper is organized as follows: in § 2 we describe VS in the native grammars of English and Spanish. A brief review of previous research findings is provided in § 3. Our hypotheses are presented in § 4, followed by a description of the method (§ 5), our results (§ 6), a discussion (§ 7) and a conclusion (§ 8).
2 VS ORDER AT THE INTERFACES

Despite having ‘fixed’ word order, English allows subject inversion in certain contexts outside interrogatives. There are two main types: (i) ‘full inversion’ (Subject-Verb) in (1) and (ii) ‘partial inversion’ (subject-operator) (e.g., Only then did I realize what he meant). Though both are triggered by an element other than the subject in clause-initial position, their properties are quite different (Biber et al. 1999: 11.2.3). Our concern here is full inversion, which is found in English in certain, highly restricted, syntactic contexts, as in (1), an example of ‘locative’ inversion.\footnote{1}

(1) Somewhere between Shigeru Shibata and Silvio Berlusconi lies the ideal husband: demonstrative but sincere, flamboyant but faithful. [Jemima Lewis, ‘Why British men make good husbands’ The Independent 3/2/07, p. 14]

XP-V-S structures like (1) are much more commonly found in written registers than in conversation and constitute the main inversion type in fiction (see Biber et al. 1999: 11.2.3.8 for details). Typically, these structures have the following properties: (i) the initial XP contains familiar or given information (topic): thus, Shigeru Shibata, a member of the ‘Japan Doting Husbands Association’, has been mentioned in the preceding paragraph and Silvio Berlusconi is regarded as known information; (ii) the verb is intransitive, expressing existence, and (iii) the postverbal subject is a complex NP, with pre- and post-modifiers, which introduces the notion of the ideal husband as new information.\footnote{2}

In Spanish, in contrast, subjects ‘freely’ invert with all verb classes: transitives like comprar ‘to buy’ (2a) and both intransitive types: unergative (2b) and unaccusative (2c), with or without preverbal elements and requiring no special intonation.
An analysis of VS structures in the native grammars of English and Spanish must take into account: (i) the lexico-semantic properties of the verb and their interaction with structural properties (LEXICON-SYNTAX INTERFACE, § 2.1); (ii) the discourse status of the postverbal subject (SYNTAX-DISCOURSE INTERFACE, § 2.2); and, additionally, (iii) phonological heaviness or weight, a property responsible for non-canonical word order at the SYNTAX-PHONOLOGY (PF) INTERFACE (§ 2.3).

### 2.1 Postverbal subjects at the lexicon-syntax interface

Postverbal subjects and the Null Subject Parameter

‘Free inversion’ is among the cluster of properties that characterises languages positively marked for the Null Subject Parameter (e.g., Spanish) (see, *inter alia*, Rizzi 1982, Jaeggli & Safir 1989 and Fernández-Soriano 1993). Though there are counterexamples both ways (e.g., Cole 2000), there is a correlation between rich agreement and null subjects (see Rizzi 1982). Spanish rich agreement allows for the recovery of the ‘content’ of a null pronominal (*pro*) subject, (3). English, a non-Null Subject language, has ‘poor’ agreement and lacks *pro*, (4).³

(3) a. *pro* llegamos tarde
   pro-1pl arrived-1pl late
   ‘We arrived late.’

b. *pro* dormiste mucho
   pro-2sg slept-2sg a-lot
   ‘You slept a lot.’
(4)  a.  *pro arrived late  b.  *pro slept a lot

The element pro in (3) is to be distinguished from expletive pro (pro_{expl}) in the surface subject position (<Spec, IP>) of VS structures like those in (2), as represented in (5a) for (2c) above. This is the null counterpart of overt expletives like English there, (5b), which seems to have lost its original locative meaning and functions as a semantically empty grammatical subject like it in extraposition constructions. Expletives like those in (5) are commonly assumed to be involved in the licensing of a postverbal subject: they enter a co-indexing relation with the postverbal subject, with which they share agreement and Case features.

(5) a.  [IP pro_{expl} [VP llegaron tres chicas]].  (Spanish)
b.  [IP There [VP arrived three girls]].  (English)

But note that there seems to have a much more restricted distribution than pro_{expl}; there is compatible with verbs like arrive in (5b), but not with buy (transitive) or speak (intransitive: unergative) in (6).

(6) a.  * There has bought a book Mary  b.  * There spoke John

**Postverbal subjects and the Unaccusative Hypothesis**

Since Perlmutter’s (1978) *Unaccusative Hypothesis*, it is commonly accepted in generative grammar that there are two classes of intransitive verbs, which differ in the position occupied by their only argument at the initial level or D-Structure: unergative verbs in (7a) have an external argument, represented in <Spec, VP> (after Koopman & Sportiche 1991), but lack an internal argument. Unaccusative verbs in (7b), on the other hand, have an internal argument, but no external argument.
The syntactic distinction between the two classes of verbs is systematically related with a semantic distinction. Unergative verbs typically denote activities controlled by an agent (*speak* in (7a) and also *cry, cough, run*, etc.), while unaccusatives are associated with themes (*arrive* in (7b) and also *appear, exist, arrive*, etc.). Under the assumption that semantic relations are syntactically represented in a systematic way (Baker’s 1988 Uniformity of Theta-Assignment Hypothesis), agents are mapped as external arguments and themes as internal arguments. NP-movement applies to promote the NPs in (7) to <Spec, IP> at S-Structure (see arrowed lines) in order to satisfy their Case requirements, or, more recently, the requirement that <Spec, IP> must be filled by an overt element (roughly Chomsky’s EPP, see note 5).

The internal argument of an unaccusative verb may, however, remain in its base position under certain conditions: e.g., in *there*-constructions, (5b). Both *there*-insertion and NP-movement satisfy the requirement that English must have overt subjects in <Spec, IP>. Unergatives like *speak* in (7a) are, on the other hand, commonly assumed to be banned from *there*-constructions (but cf. Culicover and Levine 2001). Additionally, not all unaccusative verbs may appear *there-
constructions. Perlmutter’s (1978) intuition that unaccusativity is syntactically encoded but semantically determined has been the starting point for a number of recent studies on unaccusativity (and related phenomena) seeking to establish a relation between the lexical semantics of verbs and the syntactic properties of the constructions they enter. These studies have revealed that the class of unaccusatives is neither semantically, nor syntactically, homogeneous and two major (sub)classes have been distinguished: (i) change of state and (ii) existence and appearance (Levin & Rappaport-Hovav 1995). Only the latter appear in there-constructions in English, (8a,b,c), but not the former, (8d,e,f).

(8)  

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>There appeared a grotesque figure</td>
<td>(= A grotesque figure appeared)</td>
</tr>
<tr>
<td>b.</td>
<td>There exist a unique theory</td>
<td>(= A unique theory)</td>
</tr>
<tr>
<td>c.</td>
<td>There arrived three girls</td>
<td>(= Three girls arrived)</td>
</tr>
<tr>
<td>d.</td>
<td>*There opened the door</td>
<td>(= The door opened)</td>
</tr>
<tr>
<td>e.</td>
<td>*There broke a window</td>
<td>(= A window broke)</td>
</tr>
<tr>
<td>f.</td>
<td>*There melted the butter</td>
<td>(= The butter melted)</td>
</tr>
</tbody>
</table>

The class of verbs in there-constructions overlaps with the class in XP-V-S constructions like (1) above and also in (9) (corpus examples from Biber et al. 1999: 912-913, highlighting is ours). (9) shows non-canonical PP-V-NP and is descriptively analysed as a variant of equivalent canonical sentences. (9a), an example of locative inversion like (1), seems to be the result of the NP and the PP ‘switching positions’, when compared to (10) (canonical NP-V-PP). In (10) NP-movement applies, (11a), while inversion in (9a) results from a movement rule placing the PP preverbally, (11b).

(9)  

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>On one long wall hung a row of Van Goghs.</td>
</tr>
<tr>
<td>b.</td>
<td>Then came the turning point of the match.</td>
</tr>
<tr>
<td>c.</td>
<td>With incorporation, and the increased size of the normal establishment came changes which revolutionized office administration.</td>
</tr>
</tbody>
</table>
(10)  [NP A row of Van Goghs] [vp hung [pp on one long wall]].

(11)  a. [ip NPi [vp V [NP ti] PP]]  b. [ip PPi [vp V NP [pp ti]]]

In Spanish, however, inversion is possible with all verb classes, as shown in (2) above, so there are no constraints at the lexicon-syntax interface. Rather, Spanish (and Romance) ‘free’ inversion appear to be governed by features operating at the syntax-discourse interface, which also operate in English. This could, in principle, lead to the expectation that Spanish learners of L2 English may produce postverbal subjects with all verb classes, but several studies show that they produce VS with exactly the same subset of verbs as native speakers (see § 3, below).

### 2.2 Postverbal subjects at the syntax-discourse interface

Information structure plays a crucial role in the position of the subject in ‘free’ word order languages like Spanish. In preverbal position, subjects (and other elements) are interpreted as discourse topics, i.e., presupposed, given or (relatively) familiar information, while postverbal subjects are interpreted as (presentational/information) focus, i.e., non-presupposed, new or (relatively) unfamiliar information (e.g., Belletti 2001, 2004, Domínguez 2004, Liceras et al. 1994, Lozano 2006a, Fernández-Soriano 1993, Vallduví 1993 and Zubizarreta 1998). In question-answer sequences like (12), the focused subject appears in postverbal position, where it is interpreted as presentation/information focus, (12Bi), while a preverbal subject is pragmatically odd, (12Bii).
(12) A: ¿Quién ha llegado/hablado?
‘Who has arrived/spoken?’
B: i. Ha llegado/hablado Juan
has-3sg arrived/spoken Juan
(ii. #Juan ha llegado/hablado
Juan has-3sg arrived/spoken
‘John has arrived/spoken.’

Unlike what happens in focus contexts in (12) above, verb type does seem to play a role in ‘out-of-the-blue’ contexts, i.e., with neutral questions such as ¿Qué pasó? ‘What happened’: unergatives pattern like transitives, with postverbal subjects as pragmatically odd, (13B), while inversion with unaccusatives is preferred even if the whole clause is new, not just the subject (13B’), as well as in focused contexts (12B).

(13) A: ¿Qué pasó anoche en la fiesta?
‘What happened last night at the party?’
B: i. #Lloró Juan.
cried Juan
ii. Juan lloró.
Juan cried
‘John cried’
B’: i. Vino Juan.
arrived Juan
ii. #Juan vino
Juan arrived
‘John arrived’

Thus, inversion in Spanish, as in other Romance languages like Italian, serves the purpose of focalising the subject (Zubizarreta’s 1998 ‘nuclear stress’): a postverbal subject is interpreted as presentational/information focus in transitive/unergative structures but it is ambiguous in unaccusative structures between a focus and a neutral interpretation (cf. also contrastive focus in note 9).

As for English, postverbal subjects in there-constructions, (8a-c), and inversion structures, (9) and (1), are often characterized as functionally equivalent (but cf. Birner & Ward 1993), with the postverbal subjects as focus. The tendency for
elements with higher informational content to appear towards the end of the clause is descriptively known as the principle of End-Focus (Quirk et al. 1985: § 18.3). Thus, locative inversion constructions like (9a) are commonly analysed as involving PRESENTATIONAL FOCUS (e.g., Bresnan 1994, Bolinger 1977, Rochemont 1986): the referent of the postverbal NP is introduced, or reintroduced, on the scene referred to by the preverbal PP.

The terms ‘focus’ and ‘topic’ are used here as common labels for new and given (old) information, respectively (Belletti 2004: 42, note 1). Prince (1981, 1992) distinguishes between hearer status (Hearer-old vs. Hearer-new) and discourse status (Discourse-old vs. Discourse-new). Only the latter distinction is relevant for subjecthood. There is also a third status for an entity in discourse: inferrable, defined by Prince (1992: 312) as NPs evoking entities that have not been mentioned, and which the reader has no prior knowledge of, but whose existence can be inferred “on the basis of some entity that was previously evoked and some belief I have about such entities”. Thus, in (14a) the door is discourse-old. In contrast the door in (14b) has a double status: as hearer-old (and discourse-old), in that it relies on the earlier presence of another entity (e.g., the Bastille) triggering the inference, and as hearer-new (and discourse-new) in that the hearer is not expected to already have in his/her head a mental representation of the entity concerned.

(14) a. He passed by the door of the Bastille and the door was painted purple.
b. He passed by the Bastille and the door was painted purple.

We consider topic and focus as concepts which encompass a variety of notions which are best analysed in terms of a gradience (see Kaltenböck 2005). In our study, both evoked and inferrable entities are considered to be topics, on the basis of
Prince’s (1992) analysis of subjects and Birner’s (1994, 1995) findings that both entities are treated alike (as discourse-old) in inversion structures. The term focus refers to ‘irretrievable’ information, whether brand-new or new-anchored (linked in some way to the previous context), together with what Prince (1992) refers to as unused (discourse-new, hearer-old entities).\(^{11}\)

Since inversion is used as a focalisation device in English and Spanish, we expect the inverted subjects produced by our learners to be discourse-new/focus or relatively less familiar than the preverbal XP (Birner, 1994). Importantly, Spanish makes use of this device with all verb types, while English inversion is restricted to unaccusative verbs of existence and appearance, as discussed above.

### 2.3 Postverbal subjects at the syntax-phonology (PF) interface and processing components

Choice of ordering seems to be also influenced by notions related to the phonetic realization of the strings generated by the grammar, which operate in the interface with the articulatory-perceptual or sensory-motor (S-M) systems. At Phonological-Form (PF), there can be operations affecting linear ordering which are not triggered by syntactic features and do not affect meaning: e.g., ‘Heavy NP-Shift’ in (15b) (vs. (15a) with canonical V-NP-PP).\(^{12}\)

\[(15)\]
\[
\begin{align*}
\text{a.} & \quad \text{I bought [}_{\text{NP}} \text{a book written by a specialist in environmental issues}] \text{ for my sister.}
\text{b.} & \quad \text{I bought for my sister [}_{\text{NP}} \text{a book written by a specialist in environmental issues}.}
\end{align*}
\]

Heavy NP Shift is but one manifestation of a general tendency observed in many languages for constituents to occur in order of increasing size or complexity
(Wasow 1997). This is descriptively known as the Principle of End-Weight (Quirk et al. 1985: § 18.9).

There is considerable debate regarding (i) the correct characterization of ‘heaviness’, and (ii) the reasons for end-weight. Concerning (i), two characterisations are often found in the literature: structural/grammatical complexity and string length (number of words). The two concepts are difficult to tear apart and studies reveal high correlations among the various characterizations of heaviness (e.g., Wasow 1997 and Arnold et al. 2000). As for (ii), explanations of the end-weight principle take either the listener’s perspective – putting long and complex elements towards the end of the clause facilitates parsing, as it reduces the processing burden (e.g., Hawkins 1994) – or the speaker’s perspective – weight effects exist mostly to facilitate planning and production (Wasow 1997, 2002: ch. 2: § 5). What is important for our purposes is that processing constraints in the performance systems, like the end-weight principle, influence syntax, leading to non-canonical word order structures.

Since constituents expressing new information tend to occur postverbally and heavy/complex grammatical elements tend to appear towards the end of the clause, the end-weight principle and the end-focus principle appear to reinforce each other (see Biber et al. 1999: 11.13 and Birner & Ward 1998). Thus, postverbal subjects are usually focus and tend to be heavy, as in the corpus examples used by Levin & Rappaport Hovav (1995: 221, 257) in (16) (highlighting is ours).^{13}

(16)  

a. And when it is over, off will go Clay, smugly smirking all the way to the box office, the only person better off for all the fuss. [R. Kogan, “Andrew Dice Clay Isn’t Worth ‘SNL’ Flap” 4]  
b. Above it flew a flock of butterflies, the soft blues and the spring azures complemented by the gold and black of the tiger swallowtails. [M. L’Engle A Swiftly Tilting Planet, 197]
Little attention has been paid to the effects of end-weight in Spanish (with the exception of Bolinger 1954 and Taboada 1995). While the relatively ‘free’ word order of Spanish means that weight effects may be less noticeable, (17a) shows that canonical word order with the adjunct PP (en el parque) following the NP object (a los chicos...historias) appears to be less ‘natural’ than (17b), where the heavy object is in sentence-final position following the adjunct.

(17) a. #Vi [a los chicos de los que quería haberte contado varias historias] [en el parque].
   ‘I saw [the boys I would have liked to tell you stories about] [in the park]’

   b. Vi [en el parque] [a los chicos de los quería haberte contado varias historias].
   ‘I saw [in the park] [the boys I would have liked to tell you stories about]’

Given that weight effects serve general processing and planning mechanisms, and that (end) weight appears to be a universal phenomenon, a characteristic of the human parser (see Hawkins 1994), we will assume that weight is the linguistic manifestation of extralinguistic properties which interact in language design (e.g., principles of structural architecture for computational efficiency; see Chomsky 2005). For SVO languages, the effects of weight involve placing long and complex elements towards the end of the clause.

The conclusion is that postverbal subjects which are focus, long and complex tend to occur postverbally in those structures which allow them. This is also the prediction for the learners in our study. The gradience approach adopted for information status is also adopted for ‘heaviness’: the heavier an NP is, the more likely it is to be placed in clause-final position.
3 PREVIOUS L2 FINDINGS

Previous research has focused mostly on the VS production in L2 English by L1 speakers of Null Subject languages (e.g., White 1986, Rutherford 1989, Zobl 1989, Oshita 2004). The focus of these studies is on parameter-resetting, transfer, UG accessibility and the psychological reality of theoretical hypotheses. Recently, the focus has shifted to how syntactic knowledge interacts with knowledge from other components (e.g., discourse) in the interpretation and use of null subjects and VS structures (see White, forthcoming, and references cited therein).

Following White’s (1986) study, two production studies (Zobl 1989 and Rutherford 1989), support the hypothesis that L1 speakers of Null Subject languages produce postverbal subjects in L2 English with unaccusative verbs only, (18).

(18) a. Sometimes comes a good regular wave. (L1 Japanese; source: Zobl 1989: 204)
   b. On this particular place called G… happened a story which now appears on all Mexican history books…. (L1 Spanish; source: Rutherford 1989: 178-179)
   c. The bride was very attractive, on her face appeared those two red cheeks…. (L1 Arabic; source: Rutherford 1989: 178-179)

These two studies, however, differ in their explanation of why VS order is found with unaccusatives. For Zobl (1989), it is developmental: it precedes a stage when learners are able to determine the canonical alignment between semantic roles and syntactic structure. For Rutherford (1989), VS production is the result of transfer, but no explanation is offered as to why (XP)VS order in the learners’ grammar is restricted to unaccusatives. The problem with these studies is that their conclusions are based on a relatively small number of learners and a very small number of VS instances; additionally, not enough information is provided about learners’ sample, proficiency level and so on.
Oshita (2004) was the first study to use a relatively large electronic corpus: *Longman Learners' Corpus* (version 1.1). He extracted 941 token sentences (concordances) on 10 common unaccusative verbs and 640 token sentences with 10 common unergative verbs from L2 English compositions written by L1 speakers of Italian, Spanish, Japanese and Korean. Since his objective was to investigate the psychological reality of null expletives, he extracted VS sentences with preverbal overt expletives (i.e., *it*, *there*) and null expletives. His results corroborate the role of unaccusativity for L1 Spanish and Italian learners, as in (19) (Oshita 2004: 119-120), whose production ratios are similar (14/238 (6%) Spanish; 14/346 (4%) Italian), but his conclusions are also based on a relatively small number of tokens.¹⁴

(19)  

a. …it will happen **something exciting**…. (L1 Spanish)  
b. …because in our century have appeared **the car and the plane**… (L1 Spanish)

Previous studies show a remarkably consistent pattern in which unaccusative and unergative verbs are treated differently by L2 learners of English regarding VS order. This adds to other types of (morpho)syntactic evidence which point towards what is referred to in the literature as the ‘psychological reality’ of the Unaccusative Hypothesis in SLA (e.g., Balcom 1997, de Miguel 1993, Hertel 2003, Hirakawa 1999, Ju 2000, Lozano 2003, 2006a, 2006b, Montrul 1999, 2004, Oshita 2000, Sorace 1993, Yuan 1999).¹⁵ This is so despite the fact that English lacks overt marking for unaccusatives, rendering the unergative-unaccusative distinction inaccessible, and that unaccusatives overwhelmingly appear in SV constructions. In other words, although inversion is found with unaccusatives in English (see § 2 above), the rarity of the construction (Biber et al. 1999: 945) makes it unlikely that VS order is sufficiently represented in the input to count as positive evidence for learners.
As it has become clear in the previous discussion, choice of VS order is determined by interface conditions. Additionally, conceptualizing interfaces involves mastering the processing mechanisms for integrating information from different domains in the non-native grammar. These are issues that have barely begun to be explored in SLA research. Regarding postverbal subjects, research has focused on VS order in the L2 grammars of languages like Italian and Spanish by speakers of languages with fixed word order (e.g., English) (see, amongst others, Belletti and Leonini 2004, Hertel 2003, Lozano 2006a, Belletti et al. 2007). The results of these studies show consistently that learners experience problems with linguistic phenomena at the syntax-discourse interface. There are disagreements, however, concerning the nature of these problems: are they grammar external (processing deficits) or internal (representational deficits)? As for the source, at least four explanations have been given: feature underspecification, crosslinguistic influence, processing limitations and quantity/quality of input (see Sorace and Serratrice, forthcoming, for a review).

Two aspects distinguish our study from previous studies: (i) for us unaccusativity is a necessary but not a sufficient condition for VS production: a full account of VS order in L2 (and L1) grammars must necessarily look at the properties of both the verb (lexicon-syntax interface) and the subject (syntax-discourse and syntax-phonology interfaces); (ii) while previous studies focus mostly on ‘errors’, we focus on the interface conditions for VS production in both L1 and L2 English.

In Lozano & Mendikoetxea (2008) we lay the basis for the current study and set out to investigate the interface conditions under which postverbal subjects were produced in L2 English by analysing VS/SV structures in the Italian and Spanish subcorpora of ICLE (International Corpus of Learner English, Granger et al. 2002)
(1,510 usable concordances). Postverbal subjects were produced with unaccusative verbs like *disappear*, as predicted, and were shown to be heavy as well as new information (focus), (20), as opposed to preverbal subjects with the same verbs.

(20) a. [...] in the evolution of the human specie it would disappear *the capacity of thought in a near future.* (L1 Spanish)
b. It is almost disappearing *the use of writing nice letters to friends.* (L1 Italian)

In the current study, we compare our results on VS production by L1 Spanish learners of English with those obtained from a comparable English native corpus in order to throw some light on issues to do with, among other things, transfer, processing constraints and the role of input. Given the focus of the present study, our results are also relevant for the current debate in SLA regarding the nature of deficits at the interfaces. Within the framework of the so-called ‘interface hypothesis’, Sorace (2004, 2005) has argued that advanced learners are more likely to show deficits at the ‘external’ interfaces (syntax-discourse) than at the ‘internal’ interfaces (lexicon-syntax) (but see Ivanov’s 2009 for native-like behaviour at the interfaces with L2 Bulgarian clitics). Our results, however, show that learners display native-like behaviour regarding interface conditions (lexicon-syntax, syntax-discourse and syntax-phonology), but overuse the VS construction and show persistent problems in its syntactic encoding.

4 HYPOTHESES

As a general hypothesis, we expect no substantial differences between intermediate/advanced L1 Spanish learners of L2 English and English native speakers
regarding the *interface conditions* for VS structures production. We saw that both in English and Spanish, postverbal subjects represent new information, while preverbal subjects are given information (the *Principle of End-Focus*) and that postverbal subjects are typically heavier than preverbal subjects (the *Principle of End-Weight*).

However, while VS order is restricted to a subset of unaccusative contexts in English, no such restrictions apply in Spanish. This difference could lead us to predict that Spanish learners of English may invert more freely than English native speakers (as a result of transfer), but previous L2 research consistently shows that the Unaccusative Hypothesis is a guiding principle in building learners’ mental grammars (irrespective of their L1-L2 pairings). Thus, based on previous research findings (in § 3 above) and theoretical studies (in § 2 above), we formulated three hypotheses, (21)-(23).

(21)  \( H_1 \): **LEXICON-SYNTAX INTERFACE**: Unaccusativity and the postverbal subject.  
As in native English, postverbal subjects will be produced with unaccusative verbs only (never with unergatives) by Spanish learners of L2 English.

(22)  \( H_2 \): **SYNTAX-DISCOURSE INTERFACE**: Information status of the postverbal subject.  
As in native English, learners will place unaccusative subjects in postverbal position when they are focus (i.e., discourse and/or hearer new), yet in preverbal position when they are topic (i.e., discourse and/or hearer old, inferable, containing inferable).

(23)  \( H_3 \): **SYNTAX-PHONOLOGY (PF) INTERFACE**: Weight of the postverbal subject.  
As in native English, learners’ unaccusative postverbal subjects will (typically) be heavier than their preverbal counterparts.

The shift of perspective brought about by Chomsky’s Minimalist Program involves the elimination of theory-internal levels such as D-Structure and S-Structure so that the language faculty minimally contains an ‘internal’ interface (lexicon-syntax) and two ‘external’ interfaces, connecting the language faculty with external
systems: PF and LF (see § 1 above, based on Chomsky 1995: ch 4). Figure 1, representing the minimalist reinterpretation of the classic generativist Y-model, shows where our three hypotheses are located within that model.

**Figure 1: The architecture of language and our hypotheses**

5 **METHOD**

5.1 *Instrument: corpora and concordancer*

We used two learner corpora of Spanish learners of L2 English (Table 1): the Spanish subcorpus of *ICLE* (International Corpus of Learner English, Granger et al., 2002), with a total of 200,376 words of essays written by L1 Spanish learners of L2 English at two Spanish universities: Universidad Complutense de Madrid and Universidad de Alcalá de Henares; and (ii) *WriCLE* (Written Corpus of Learner English, version 1.0), which is being developed at the Universidad Autónoma de Madrid, following the ICLE guidelines: it contains essays written by upper-intermediate Spanish learners of L2 English (63,836 words).
The learner corpora were contrasted against a comparable corpus of native English: LOCNESS (Louvain Corpus of Native English Essays), consisting of argumentative essays written by British and American students at pre-university and university level (288,177 words).

Table 1: Corpora details

<table>
<thead>
<tr>
<th>Learner corpus</th>
<th>Words</th>
<th>Native corpus</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE-Spanish</td>
<td>200,376</td>
<td>LOCNESS USarg</td>
<td>149,574</td>
</tr>
<tr>
<td>WriCLE</td>
<td>63,836</td>
<td>LOCNESS USmixed</td>
<td>18,826</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOCNESS Alevels</td>
<td>60,209</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOCNESS BRsur</td>
<td>59,568</td>
</tr>
<tr>
<td><strong>Total no. of words:</strong></td>
<td><strong>264,212</strong></td>
<td><strong>LOCNESS</strong></td>
<td><strong>288,177</strong></td>
</tr>
</tbody>
</table>

5.2 Data coding and analysis

Regarding H$_2$ (LEXICON-SYNTAX INTERFACE), we chose 73 lemmas, or word types, (32 unaccusative and 41 unergative verbs) from the inventory of intransitives (Table 2) proposed by Levin (1993) and Levin & Rappaport-Hovav (1995). For each lemma, we searched for all possible L1 and L2 forms (misspellings and overregularisations, e.g., appeared, arised, etc), using the concordancer WordSmith Tools version 4.0 (Scott 2004).
Table 2: Inventory of unaccusatives (based on Levin 1993, Levin & Rappaport-Hovav 1995)

<table>
<thead>
<tr>
<th>Unaccusatives: Semantic class</th>
<th>Unergatives: Semantic class and subclass</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Existence:</strong></td>
<td>Emission:</td>
</tr>
<tr>
<td>exist, flow, grow, hide, live, remain, rise, settle, spread, survive</td>
<td>Light emission: beam, burn, flame, flash.</td>
</tr>
<tr>
<td><strong>Appearance:</strong></td>
<td>Sound emission: bang, beat, blast, boom, clash, crack, crash, cry, known, ring, roll, sing.</td>
</tr>
<tr>
<td>appear, arise, awake, begin, develop, emerge, flow, follow, happen, occur, rise</td>
<td>Smell emission: smell.</td>
</tr>
<tr>
<td><strong>Disappearance:</strong></td>
<td>Substance emission: pour, sweat</td>
</tr>
<tr>
<td>die, disappear</td>
<td>Communication:</td>
</tr>
<tr>
<td>Inherently directed motion:</td>
<td>Manner of speaking: cry*, shout, sing*.</td>
</tr>
<tr>
<td>arrive, come, drop, enter,</td>
<td>Talk verbs: speak, talk</td>
</tr>
<tr>
<td>escape, fall, go, leave, pass, rise*, return</td>
<td>Bodily processes:</td>
</tr>
<tr>
<td><strong>Emission:</strong></td>
<td>Breathe verbs: breath, cough, cry*, sweat**.</td>
</tr>
<tr>
<td>Light emission: beam, burn, flame, flash.</td>
<td>Nonverbal expressions: laugh, sigh, smile</td>
</tr>
<tr>
<td>Sound emission: bang, beat, blast, boom, clash, crack, crash, cry, known, ring, roll, sing.</td>
<td>Manner of motion:</td>
</tr>
<tr>
<td>Smell emission: smell.</td>
<td>Run verbs: fly, jump, run, swim, walk, ride, travel, slide</td>
</tr>
<tr>
<td>Substance emission: pour, sweat</td>
<td>Performance:</td>
</tr>
<tr>
<td>Communication:</td>
<td>Monadic agentives: dance, phone, play, sing, work</td>
</tr>
<tr>
<td>Manner of speaking: cry*, shout, sing*.</td>
<td>Snooze: sleep</td>
</tr>
</tbody>
</table>

**TOTAL UNACCUSAT.: 32**  **TOTAL UNERGATIVES: 41**

Notes: (*) see also sound emission. (**) see also substance emission. (***) see also existence.

Concordances (i.e., VS and SV sentences containing those verbs) were retrieved automatically. Concordances were then filtered manually according to 51 filtering criteria were used (due to space limitations, we present only the key filters, (24)) so as to discard structural contexts in which inversion in English is not possible, because our interest is in contexts were the choice of SV/VS is not restricted by structural factors. Only ¼ of the concordances output by the concordancer were usable: n=58 in the learner corpora and n=16 in the native corpus.

(24) **FILTERING CRITERIA:**

a. The verb must be intransitive (either unaccusative or unergative from Table 3 below).\(^{19}\)
b. The verb must be finite.\(^{20}\)
c. The verb must be in the active voice.\(^{21}\)
d. The subject must be an NP.\(^{22}\)
e. The sentence must be declarative.
f. Set expressions are excluded.\(^{23}\)

Preverbal and postverbal subjects were coded manually for weight and discursive status. While VS concordances were all coded due to their small number
(n=58+16), we selected a random sample of all SV concordances with unaccusative verbs (n=762 for learners and n=703 for natives) from the only four inverted unaccusatives (come, exist, begin and remain) in the native corpus (n=91 unaccusative SV concordances) and the top four inversion unaccusatives (exist, appear, begin and come, as Figure 6 shows) in the learner corpora (n=96 unaccusative SV concordances).

As for H$_2$ (SYNTAX-DISOURSE INTERFACE), for each concordance we coded the discursive status of the (preverbal/postverbal) subject. The entire composition was inspected to determine the nature of each subject as topic or focus, where these terms encompass a variety of notions (see § 2.2).

Regarding the analysis used for H$_3$ (SYNTAX-PF INTERFACE), there is no agreement in the literature as to the most appropriate measuring instrument for syntactic weight, as seen in § 2.3. While word length, as measured by number of words, is standardly used (e.g., Kaltenböck 2004, Arnold et al. 2000), it cannot tell us how long (or short) a constituent must be to be considered ‘heavy’ (or ‘light’). Thus, while the histogram in Figure 2 shows that the most frequent lengths for the subject constituent in the learner corpora were 2 and 5 words, and the mean of all lengths was 7.5 words, it is difficult to establish the cut-off point between heavy vs. light.
To solve this dilemma, we started off from a syntactic scale of complexity in order to arrive at a reliable \textit{nominal} dichotomous scale (heavy/light), which was then correlated with a \textit{numeric} scale (no. of words). The rationale is that, if the nominal scale is a good predictor of weight, then it should correlate significantly with the numeric scale, Table 3.

Table 3: A syntactic scale for measuring syntactic weight

<table>
<thead>
<tr>
<th>SYNTACTIC WEIGHT</th>
<th>SYNTACTIC STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nominal scale</td>
</tr>
<tr>
<td>LIGHT</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>HEAVY</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: (i) The asterisk (*) represents a complex (i.e., recursive) categorical or phrasal structure. (ii) Parentheses indicate the optional realization of the bracketed category or phrase.

Therefore, bare N(ouns) that optionally take a D(eterminer), (25a), were regarded as light and ranked as 0 in an \textit{ordinal scale}, which ranges from 0
(syntactically simple) to 3 (syntactically complex). This is a more fine-grained scale mediating between the syntactic-complexity scale and the nominal scale. Its usefulness can be seen in (25b): NPs with an optional D, as in (D)+ADJ+N, are slightly more complex than (D)+N and thus can be ranked as 1, though they are still light. Heavy structures can be syntactically less elaborated (ranked as 2), as (25c) with D+N+PP, or more complex (ranked as 3), as with a full NP+Finite-clause.

(25) a. However, with the awareness also came discrimination. [usarg.txt]
b. …that it can exist a better world … [spm04006]
c. With this theory also came the area of quantum mechanics. [usarg.txt]
d. …for there exists a general judicial presumption that Parliament does not intend to legislate contrary to European Community Law. [brsur3.txt]

With this procedure, we obtained three weight measures for each postverbal subject: number of words, degree of syntactic complexity and weight. A highly significant correlation was found in the learner corpus between (i) word length and complexity (\(\rho=0.73, n=58, p<0.001\) with Spearman’s rho test), (ii) length and weight (nominal scale) (\(\rho=0.61, n=58, p<0.001\)) and (iii) complexity and weight (\(\rho=0.83, n=58, p<0.001\)). Similar significant correlations were also found in the native LOCNESS corpus, which indicates that our nominal scale (heavy/light) is a reliable indicator of weight. Hence we decided to adopt this nominal scale together with the widely-used numeric (word-length) scale.
6 RESULTS

6.1 Results for H₁: Unaccusativity (lexicon-syntax)

The difference between preverbal vs. postverbal subject production is highly remarkable (Figure 3). While the proportion of postverbal subjects with unaccusatives was 7.1% for learners (i.e., 58 out of a total of 820 concordances), there were no instances of postverbal subjects with unergatives (0/181 ratio, i.e., 0%). This SV/VS difference is statistically significant ($\chi^2=12.65$, df=1, $p<0.001$; since the observed frequency in one of the cells is smaller than 5, Fisher’s exact test, rather than the standard Pearson’s test, was performed). The difference is also remarkable for natives (2.2% for unaccusatives and 0% for unergatives). Similarly, preverbal subjects are the norm in both corpora (97.8% for natives and 92.9% for learners). However, learners produce postverbal subjects with unaccusatives significantly more often than natives do: learners (7.1%) and natives (2.2%) ($\chi^2=19.67$, df=1, $p<0.001$ with Pearson’s test).

Figure 3: Pre-verbal (SV) and post-verbal (VS) subjects produced with intransitives
Regarding the type of VS structures (Figure 4), the most frequent structure produced by learners is *IT-INSERTION*, (26a,b), which accounts for 41.4% of all VS structures and is ungrammatical in native English.\(^{24}\) This is followed by: grammatical *LOCATIVE INVERSION* (15.5%) in (27a,b); *XP-INSERTION* (13.8%) in (28a,b), which are ungrammatical in the learner data due to the absence of expletive *there*; grammatical *THERE-INSERTION* (10.3%) in (29a,b); *AdvP-INSERTION* (10.3%) in (30a,b), which can be classed as a type of loco/temporal inversion and is a possible grammatical structure in native English; and, finally, ungrammatical *Ø-INSERTION* (8.6%) (31a,b).

Natives, on the other hand, produce the following VS structures: *XP-insertion* (43.8%) in (28c,d), *there-insertion* (37.5%) in (29c,d) and *AdvP-insertion* (18.8%) in (30c) (the only example produced). Surprisingly, no instances of locative inversion are found, (27c), which can be taken simply as a gap or perhaps a bias in the native corpus, since locative inversion is a relatively common VS structure in written English (see Biber et al. 1999: § 11.2.3). As expected, natives did not produce any instances of ungrammatical *it-insertion* (26c) and *Ø-insertion* (31c).
Figure 4: Frequency of the types of postverbal-subject structures produced

![Graph showing frequency of postverbal-subject structures]

(26) **It-Insertion**
Learners:
- *In the name of religion it had occurred many important events.* [spm02009]
- *I do believe that it will not exist a machine or something able to imitate the human imagination.* [spm01007]

(27) **Locative Inversion**
Learners:
- *In some places still exist popularly supported death penalty.* [spm05001]
- *In the main plot appear the main characters: Volpone and Mosca.* [spal1002]

(28) **XP-Insertion**
Learners:
- *[…] and from this moment begins the avarice.* [spm04048]
- *In 1760 occurs the restoration of Charles II in England.* [spm08007]
Natives:
- *After De Gaulle came George Pompidou in 1969.* [brsur1.txt]
- *With this competitiveness comes the desire to stand out from the crowd and be the best.* [usarg.txt]

(29) **There-Insertion**
Learners:
- *There exist about two hundred organizations such as Greenpeace, which have increased the number of its members laterly.* [spm04052]
- *Furthermore there also exists a wide variety of optional channels which have to be paid.* [spm03018]
Natives:
- *[…] there existed social ills as serious as the ones that exist today.* [usarg.txt]
- *Certainly there exists a demand for this work to be done, […]* [alevels8.txt]
(30) **AdvP-insertion**

*Learners*

a. [...] and *thus* began the **period known as the Restoration** ... [spm08005]  
b. *So arised the Saint Inquisition* in case someone blasphemed about God. [spm05012]  

*Natives*

c. *Thus* began the **campaign to educate the public** [...] [usarg.txt]

(31) **Ø-insertion**

*Learners*

a. It is difficult that exist **volunteers with such a feeling against it**. [spm04026]  
b. [...] because exist **the science technology and the industrialisation**. [spm04006]

Figure 5 represents the learners’ proportion of structures according to their grammaticality: two thirds (65.5%) of the VS structures produced by our learners are ungrammatical (i.e., structurally impossible in native English).²⁵

![Figure 5: Postverbal-subjects structures according to their grammaticality](image)

In short, learners produce mostly ungrammatical *it*-insertion (a construction that is neither in their L2 input nor directly a result of L1 transfer) and grammatical locative inversion. Natives produce mostly *XP*-insertion and *there*-insertion. We will come back to these results in the following section.

Let us focus now on the unaccusative verb itself. Figure 6 shows the percentage each particular unaccusative verb contributes to the grand VS total (7.1%
for learners and 2.3% for natives, as shown in Figure 3 above). For learners, the top inversion verbs are: exist (2.9%), appear (1.7%), begin (0.6%), come (0.5%) and arise/emerge/occur (0.2% each). Natives inverted with only four unaccusatives: come (1.3%), exist (0.7%), and begin/remain (0.1% each). In § 7 we discuss the production of exist in the learner vs. the native corpora.

Figure 6: Production of postverbal subjects (VS) according to verb: VS/TotalConcordances ratio
6.2 Results for $H_2$: Information status (syntax-discourse)

As Figure 7 shows, postverbal subjects are always focus or new information (100% for natives and 98.3% for learners). By contrast, preverbal subjects tend to be topic or evoked information (83.5% for natives and 89.6% for learners). This implies that natives and learners place unaccusative subjects in postverbal position if and only if the subject represents new information, yet in preverbal position when it typically represents old information. The topic vs. focus difference is significantly different for each word order in each corpus (p<0.001 for all comparisons).

Both natives and learners behave alike, as there are no significant differences between them in terms of preverbal subjects ($\chi^2=1.485$, df=1, p=0.223) and postverbal subjects ($\chi^2=0.280$, df=1, p=0.597). These findings are exemplified in (32) and (33), where we provide the context preceding the target concordance to see the focus status of the postverbal subject since (i) it has not been mentioned in the prior discourse, (ii) is not known to the reader and (iii) it cannot be inferred by the reader.

Figure 7: Information status of pre-verbal and post-verbal subjects with unaccusatives
(32) **Natives:** postverbal focus subject

[...] Albert Einstein came onto the scene. Although his theory (his and his wife's) was basically scientific in nature, it can and has been applied to all areas of human existence. The theory I'm speaking of is relativity. *With this theory* also came *the area of quantum mechanics.* [usarg.txt]

(33) **Learners:** postverbal focus subject

Principally, in the middle ages all the cities were filled with churches, monasteries, convents, priests and nuns. Hence, religion was part of the population. [...] God was the most important, the center their own existence. *So arose the Saint Inquisition* in case someone blasphemed about God. [spm05012]

By contrast, in (34) and (35) the information status of the preverbal is known (topic), as it has been previously mentioned in the discourse (shown by underlined lettering).

(34) **Natives:** preverbal topic subjects

[...] In order to examine why Louis took such exception to Hugo it is necessary to reflect on the way in which the Party regarded intellectuals. *Hugo* came from a bourgeoisie background, [...] [brsur1.txt]

(35) **Learners:** preverbal topic subjects

*The feminism* is the movement that try to obtain that women have the same rights than men, [...] *The feminism* begun with the French Revolution and the Industrial Revolution, [...] [spm07020]

Additionally, the preverbal XP in XP-V-S structures such as (28a) above shows a strong tendency to represent known/old information (topic) in both corpora: 80% (16/20) for learners and 100% (8/8) for natives (p=0.237 with Fishers’ exact test).

To summarise, discourse clearly constrains both natives and learners’ distribution of the subject in SV/V-S structures. When subjects are focus, they appear postverbally (with the preverbal material as topic). By contrast, when subjects are topic, they occur preverbally.
6.3 Results for H₃: Weight (syntax-phonology)

The weight rates (Figure 8) are similar in both corpora: postverbal subjects are more frequently heavy (81% for learners and 81.3% for natives) than light (19% and 18.8% respectively). Thus, learners’ behaviour does not significantly differ from natives’ ($\chi^2=0.000$, df=1, $p=0.647$ with Fisher’s exact test). By contrast, preverbal subjects are more frequently light (67.7% for learners and 68.1% for natives) than heavy (32.3% and 31.9% respectively). Once again, learners’ rates do not differ from natives’ ($\chi^2=0.004$, df=1, $p=0.951$ with Pearson’s test).

Figure 8: Weight of pre-verbal and post-verbal subjects with unaccusatives (nominal scale)

![Graph showing frequency of production for learners and natives in post- and preverbal positions](image)

On the nominal scale, then, learners and natives behave alike in placing subjects in postverbal position when they are heavy, (36) (though a small portion of them can be light, (37)), but in preverbal position when they are light, (38) (though some of them can be heavy, (39)).
Postverbal heavy subjects
a. [...] in 1880, it begun the experiments whose result was the appearance of television some years later. [learner, spm03051]
b. Thus began the campaign to educate the public on how one contracts aids. [native, usarg.txt]

Postverbal light subjects
a. [...] and from there began a fire. [...] [learner, spm04011]
b. Along with the traffic congestion, comes pollution. [native, alevels1.txt]

Preverbal light subjects
a. [...] but they may appear everywhere. [learner, 0002.cor]
b. These debates began over two decades ago. [native, usarg.txt]

Preverbal heavy subjects
a. However, I strongly believe the cases of men mistreated do not appear in the media because [...] [learner, 0006.cor]
b. However, my curiosities about sexual relationships still existed. [native, usarg.txt]

Consider now these production rates when measured numerically (word length). The boxplot in Figure 9 shows the dispersion of word-length of the subject. For SV in both the native and the learner corpus, the median (represented by the vertical bar inside the box) is 2 words long (i.e., half of the preverbal subjects produced are two words long or less, while the other half are two words long or more) and the word-length mean (indicated by the cross) is 3.14 words (natives) and 3.24 (learners), the difference being non-significant (t=0.252, df=185, p=0.801, with an independent-samples t test). As for VS, in the native corpus, the median is 11 words and the word-length mean is 11.06, while in the learner corpus the median is 6 words and the mean is 7.52, the difference being statistically significant (t=-2.229, df=72, p=0.029). Additionally, preverbal subjects are significantly shorter than postverbal subjects in each corpus (t=8.221, df=105, p<0.001 for natives; t=6.715, df=152, p<0.001 for learners), as Figure 9 clearly shows.
In short, both natives and learners produce significantly longer subjects in postverbal position than in preverbal position, though the natives’ postverbal subjects are significantly longer than the learners’. This finding confirms our previous results on the nominal (heavy/light) scale.

7 DISCUSSION

The main result of our study is the parallel use of VS structures in L1 and L2 English (L1 Spanish): postverbal subjects are produced with essentially the same (sub)class of unaccusative verbs and under the same interface conditions (i.e., when the subject is focus and/or heavy). Yet, learners produce a high proportion of ungrammatical structures, due mostly to the wrong choice of preverbal element.

The fact that VS occurs only with (a class of) unaccusatives but never with unergatives, confirms our hypothesis H₁ (LEXICON-SYNTAX INTERFACE) and is in line
with previous L2 research on VS structures, as well as with studies showing that L2 learners are sensitive to the (morpho-)syntactic and semantic manifestations of unaccusativity (see § 3). Finally, within the context of L2 research into the interfaces, our results could be taken as confirmation of previous findings suggesting that the internal interface (lexicon-syntax) is not problematic for advanced learners (see e.g. Tsimpli & Sorace 2006; Sorace & Serratice forthcoming & White forthcoming).

Regarding the two classes of unaccusative verbs (see § 2.1), our results confirm the possibility of VS structures with verbs of existence and appearance in L1 Spanish-L2 English (see also Lozano & Mendikoetxea 2008 for L1 Italian-L2 English), but further research is needed to show that unaccusative verbs of change of state do not trigger inversion, as in L1 English.

It still needs to be explained why our learners do not produce VS structures with unergatives, which are possible in their L1 Spanish (see also Oshita 2004: § VI). Since inversion structures, though rare, are found in English with a subclass of unaccusative verbs and never with unergatives (but see note 13), it may be the case that learners have internalised the fact that postverbal subjects are possible only with unaccusatives on the basis of the input received. Given the rarity of the construction, however, we have to resort to the Unaccusative Hypothesis: learners are aware of the unergative/unaccusative distinction and use it to build mental grammars in the process of L2 acquisition, as previous research has shown (see note 15). Thus, while INPUT (and, certainly, instruction) are crucial for the development of interlanguage grammars, our results cannot be accounted for by input, or at least by input alone. The fact that natives’ top inversion unaccusatives (as shown in Figure 6 and also as reported by Birner 1995) do not coincide with learners’ (and the same applies to low
inversion unaccusatives) could be taken as some evidence for this (though it may also
be a bias in the sample).

Both learners and natives display a very strong and clear-cut pattern in terms
of **INFORMATION STATUS**, by producing (i) subjects in postverbal position only if they
represent focus (new) information, yet (ii) in preverbal position only if they represent
known (topic) information. Additionally, the preverbal material in XP-V-S structures
is topic, which ‘anchors’ the information to the preceding discourse (see Birner 1995
on anchoring). The syntactic distribution of VS/SV is thus regulated discursively,
which supports \( H_2 \) (**SYNTAX-DISCOURSE INTERFACE**).

Additionally, the results on **WEIGHT** clearly reveal that learners behaved like
natives by producing (i) VS when the subject is heavy, yet (ii) SV when the subject is
light, which confirms hypothesis \( H_3 \) (**SYNTAX-PHONOLOGY INTERFACE**). To our
knowledge, this is the first L2 corpus-based study that shows that L2 English learners
appear to be observing the *Principle of End-Weight* (see § 2.3, above) irrespective of
the (un)grammaticality of the structures produced.

Much theoretical research from a functional perspective points out that **END-
WEIGHT** and **END-FOCUS** are related (Arnold et al. 2000, Hawkins 1994, Wasow
1997, Wasow 2002 and § 2.3 above). When speakers need a significant number of
words to develop a new idea (i.e., a heavy subject), it tends to be in a position
typically reserved for new information (i.e., postverbally in sentence-final position),
whereas short subjects appear in a position typically reserved for old or familiar
information (i.e., preverbally in sentence-initial position).

The end-focus and end-weight principles stem from a more general
mechanism that facilitates language processing, since it lessens the processing load on
the listener/reader (Hawkins 1994) and also on the speaker/writer (Wasow 1997) by
leaving longer and heavier (and, consequently, new) elements to be processed at the end. Thus, there is a gradual rise in information load as the sentence unfolds (Biber et al. 1999). It then seems that learners of L2 English are overriding grammatical criteria in favour of more general (perhaps universal) processing criteria in their production of subjects with unaccusatives.

While the correlation between DISCOURSE STATUS and WEIGHT (as summarised in Table 4) is by no means perfect in all cases, it is certainly the case that (i) all postverbal subjects are focus yet all preverbal subjects are topic and (ii) most postverbal subjects are heavy (81%) yet most preverbal subjects are light (68%). It then seems that if a subject is heavy (long), it tends to be placed in the same position as new information constituents, namely, postverbally. The fact that natives produce longer subjects than learners has to do with the correlation between complexity and proficiency: as it is well known, phrasal composition increases in complexity with developmental level (Klein & Perdue 1992), with length of production as one of the measures for quantifying complexity or linguistic maturity (see Ortega 2000, 2003).

Table 4: Tendency of syntactic distribution with unaccusative subjects: weight x info status

<table>
<thead>
<tr>
<th>Weight</th>
<th>Information status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Light</td>
<td>Preverbal</td>
</tr>
<tr>
<td>Heavy</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>Topic</td>
</tr>
<tr>
<td></td>
<td>Focus</td>
</tr>
</tbody>
</table>

Regarding the combination of the three interfaces, previous L2 research (Oshita 2004, Rutherford 1989, Zobl 1989) has clearly shown that learners produce VS structures with unaccusatives only, but they have failed to notice that
unaccusativity is a necessary but not sufficient condition for the production of postverbal subjects: VS order in both native and non-native English is constrained by properties operating at three interfaces, (40).

(40) Conditions for postverbal-subject production
   a. Lexicon-syntax interface: The verb must be unaccusative (and not unergative).
   b. Syntax-phonology (PF) interface: The postverbal subject will tend to be heavy (long).
   c. Syntax-discourse interface: The postverbal subject must be focus (new info).

There are two main aspects in which learners’ production deviates from natives’ use of VS structures. Natives’ production of VS structures (2.2%) is significantly lower than learners’ (7.1%). It is important here to look at the verbs triggering inversion. Almost half of learner VS structures (2.9%) contain exist, as opposed to 0.7% by natives. This clearly inflates the learners’ inversion rates (Figure 6) (see also Palacios-Martínez & Martínez-Insua’s 2006). Since in L1 Spanish, existir ‘exist’ typically requires a postverbal subject, this may, in principle, indicate transfer. But if this was a result of L1 transfer alone, learners would be expected to produce mostly VS with exist, which is contrary to fact, since they produced 46 SV sentences out of 70 concordances. Hence, our learners’ higher overproduction rates must stem from a lexical bias (overuse of exist) rather than from L1 transfer alone.

Natives and learners also differ in the types of (XP)VS structures they produce, as shown in § 6.1. Differences concern the relative production rates of the various constructions, as well as the nature of the preverbal element: (i) an ‘expletive’-type element (there, it or Ø) or (ii) a ‘lexical’ element (locative PP, AdvP or another XP). Regarding (ii), though learners are proficient in their use of preverbal locative PPs, they show difficulties in precisely figuring out what counts as an appropriate XP in other contexts. This is not surprising as, from a grammatical
perspective, it is difficult to describe what can and what cannot appear as a preverbal XP in VS structures in L1 English: grammars simply provide lists of examples with locative and time adverbials, as well as ‘other types of’ adverbials (e.g., Biber et al. 1999: 11.2.3.1). For native speakers, frequency and patterns of use may be playing a part in their choice of preverbal XP in a particular context. Even advanced learners, however, have a relatively limited exposure to the language and without clear grammatical rules they are left to make their own choices. This is a matter that needs to be explored in further studies.

More interesting for our purposes is the use of expletive-type constructions: grammatical There-V-S and ungrammatical *It-V-S and *Ø-V-S. In Lozano & Mendikoetxea (2009a, b) we compare these three structures in the Spanish, Italian and French subcorpora of ICLE. The most common ungrammatical structure for all three learner groups is *It-V-S (of all (XP)-V-S structures: 41.4% for L1 Spanish: 27% for L1 Italian and 9.1% for L1). This is followed by Ø-V-S, produced by Spanish (8.6%) and Italian (7.2%) speakers, but not by French speakers.

Production of -V-S structures is often attributed to crosslinguistic influence (transfer). Equivalent constructions in Spanish and Italian contain a null expletive (see § 2.1 above). There is evidence that for L1 speakers of Null Subject languages null expletives are harder to expunge from their L2 English than null referential pronouns: learners omit expletives in sentences like In winter, snows a lot in Canada (L1 Spanish, White, 1986) (see also Hannay & Martínez Caro, 2008) and non-use of overt expletives persists longer than non-use of overt referential pronouns (Phinney, 1987, Tsimpli & Roussou, 1991). That is, even at advanced levels, learners may still produce sentences like those in (31) above, as an instance of what Sorace (2005) calls ‘residual’ optionality, but not sentences with a missing referential pronoun, equivalent
to *I, you, he, etc. The transfer account would also explain why Ø-V-S is not produced by L1 French learners of L2 English.

Production of *It-V-S shows Spanish learners of L2 English are aware that the subject position must be filled, but despite positive evidence (e.g., there-insertion), it is the preferred expletive for that position. Presentational there-constructions, as those in (29) above (with verbs other than be), are rarely used by L1 Null Subject languages learners (and those with L1 topic-drop languages, as noticed by Oshita 2004). By contrast, a corpus study by Palacios-Martínez and Martínez-Insúa (2006) shows that L1 Spanish learners of L2 English ‘overuse’ existential there-constructions (there+be). It may well be that ‘underuse’ of presentational there and ‘overuse’ of existential there could be related and input could be part of the explanation. Existential there-constructions are introduced at an early stage in the learning process, and (one suspects) are high frequency structures in the input. They are learned as formulaic or prefabricated chunks with the V be (Palacios Martínez and Martínez-Insúa, 2006). Thus, there may not be used as an independent expletive until advanced levels of proficiency (Oshita 2004: note 21).30 Learners’ (lack of) exposure to presentational there constructions may also be a factor. This is, of course, highly speculative, both empirically and theoretically: once we break down VS structures into different types, we are dealing with a relatively small number of tokens, and the role of input in SLA is not well understood. Further research, possibly experimental, needs to be carried out to see whether learners actually favour some VS structures and are reluctant to accept/use others.

It is clear from the previous discussion that our results cannot be accounted by input alone, nor can they be accounted for by L1 TRANSFER alone either. Firstly, if there was transfer, we would expect VS production with both unaccusatives and
unergatives, since inversion structures can be also found in native Spanish with unergatives, e.g., *En el parque jugaban niños* ‘In the park played children’ (Ortega-Santos 2005 and Torrego 1989, but see Mendikoetxea 2006 for a different analysis of these constructions). Secondly, our learners’ postverbal subject rates are relatively low (7.1%), as they mainly produced grammatical SV (92.9%). Experimental work shows that Spanish natives significantly (and drastically) prefer VS to SV with unaccusatives, yet SV to VS with unergatives (Hertel 2003, Lozano 2003, 2006a). Figure 3. Hence, if L1 transfer were taking place, we would expect our learners to show higher VS rates.31

However, there are other ways in which transfer or crosslinguistic influence may be affecting learner production. The fact that our learners produce mostly deviant VS structures may be interpreted in the light of current SLA research, according to which linguistic phenomena at the (external) interfaces are particularly vulnerable in non-native grammars and that failure to acquire a fully native grammar may be largely attributed to problems at integrating different types of knowledge at the interfaces: interfaces are targets for deficits such as fossilization and optionality, even at end-states (e.g., Sorace 2004, 2005) (see § 3 above). This may be because (i) structures requiring the integration of syntactic knowledge and knowledge from other domains require more processing resources than aspects of grammar requiring only syntactic knowledge, and/or (ii) learners may be less efficient at integrating multiple types of information in on-line comprehension and production of structures at the (syntax-discourse) interface (see Sorace & Serratrice, forthcoming and references therein).

Though the precise nature of processing limitations is not well understood, they could be in part explain why our learners produce mostly ungrammatical VS structures. As pointed our by Sorace & Serratrice (forthcoming), there is evidence
from bilinguals that complete de-activation of one of the two languages when hearing/speaking the other is rarely possible: they are always simultaneously active and in competition with one another (Dijkstra and van Heuven 2002; Green 1998). However, several factors affect their relative activation levels and the strength of competing structures: task, proficiency in each language, and frequency of use, among others. Thus, while not directly the result of transfer, some of the difficulties learners experience in VS production may be the result of difficulties at integrating different types of knowledge due to competition from the L1 VS form: e.g., use of $\emptyset$-V-S and It-V-S (see Lozano & Mendikoetxea 2009a, b). Our results suggest that these deficits are not external to the grammar, as learners have no difficulties in identifying topic/focus, for instance, but rather they belong to the computational system and/or the failure to map this information into appropriate syntactic structures: learners cannot encode the end-weight and end-focus principles onto the correct grammatical constructions and overuse the construction, possibly due to processing difficulties and crosslinguistic influence. Future research (cross-sectional or longitudinal) is needed to clarify the precise role of interfaces (syntax-discourse and syntax-phonology) in SLA and to determine the source of deficits in those areas where the computational system interacts with external (sub)systems.

8 CONCLUSION

Previous research has shown that learners of English produce postverbal subjects with a (sub) type of unaccusative verbs, but never with unergatives. In SLA research the Unaccusativity Hypothesis has been invoked to account for this phenomenon. Crucially, our study shows that unaccusativity is a necessary but not a sufficient
condition for VS order to occur not only in both L1 and L2 English. A full account of VS order must look not only at the nature of the verb, but also at the characteristics of the postverbal subject, a fact that has gone unnoticed in previous studies. Given the appropriate structural conditions (e.g., unaccusative verb), there is a strong tendency for VS order to be produced when the subject is syntactically heavy, as well as new information or focus. Interestingly, there is an interrelation between these two factors, as both conditions are designed to ease the processing burden and thus they can be considered (external) interface conditions imposed on the language faculty.

Our study does not only have a wider scope than previous studies but also has a different focus: while previous studies have mostly focused on errors, we have sought to identify the interface conditions under which learners produce VS structures: the verb is unaccusative (lexicon-syntax interface) and the subject is heavy (syntax-phonology interface) and focus (syntax-discourse interface). These interface conditions constrain VS production, regardless of learners’ problems with the syntactic encoding and their overuse of the construction. Our study shows that the production of English natives and L1 Spanish learners of L2 English do not differ significantly in the interface conditions that constrain VS order, but rather on the (un)grammaticality of its syntactic encoding.

REFERENCES


1 Other types of full inversion in English, e.g. in reporting clauses with verbs such as *say, think, matter*, etc., fall outside the scope of this paper, as they have very different properties (see Biber et al. 1999: 11.2.3.6).

2 “Postverbal” refers here to a subject that occurs after the verb, and, “preverbal” to a subject which occurs before the verb. We are not concerned with the specific structural position occupied by these subjects (see Hulk & Pollock 2001 for Romance subject inversion).

3 Null-Subject or *pro*-drop languages like Spanish and Italian are to be distinguished from *topic*-drop languages like Japanese and Korean, which lack ‘rich’ subject agreement morphology, as well as *pro* (Huang 1984).

4 The two expletives, *there* and *it*, have different origins and distributions. *There* is the subject of presentational (5b) and existential constructions (*There is a problem in this analysis*), while *it* is used as the subject of weather predicates (*It is freezing cold!*), extraposition constructions (*It is a pity that he is not fond of dogs*) and one-argument (and passive) verbs with clausal objects (*It seems that everybody agrees on this matter*). We will see in § 6 and 7 that L2 learners of English often use *it* in contexts where *there* is required.

5 No such agreement relation is observed in the French equivalent to these sentences (*Il est arrivé trois filles*), with a singular verb but a plural postverbal subject. Expletive *il* appears to be in the structure simply to satisfy the Extended Projection Principle (EPP), i.e., the requirement that clauses have subjects (Chomsky 1981).

6 A further restriction concerns (in)definiteness (Safir 1985). However, definite NPs (and proper nouns) are not uncommon in *there*-constructions, e.g. when the postverbal element is a list as in (i) below (see Biber *et al*. 1999: 947; Prince 1992: 299). We thank an anonymous reviewer for bringing this fact to our attention.

(i) Michael puts loose papers like class outlines in the large file-size pocket. He keeps his checkbook handy in one of the three compact pockets. The six pen and pencil pockets are always full and in the outside pocket go his schedule book, *chap stick, gum, contact lens*. 

---

1 Other types of full inversion in English, e.g. in reporting clauses with verbs such as *say, think, matter*, etc., fall outside the scope of this paper, as they have very different properties (see Biber et al. 1999: 11.2.3.6).

2 “Postverbal” refers here to a subject that occurs after the verb, and, “preverbal” to a subject which occurs before the verb. We are not concerned with the specific structural position occupied by these subjects (see Hulk & Pollock 2001 for Romance subject inversion).

3 Null-Subject or *pro*-drop languages like Spanish and Italian are to be distinguished from *topic*-drop languages like Japanese and Korean, which lack ‘rich’ subject agreement morphology, as well as *pro* (Huang 1984).

4 The two expletives, *there* and *it*, have different origins and distributions. *There* is the subject of presentational (5b) and existential constructions (*There is a problem in this analysis*), while *it* is used as the subject of weather predicates (*It is freezing cold!*), extraposition constructions (*It is a pity that he is not fond of dogs*) and one-argument (and passive) verbs with clausal objects (*It seems that everybody agrees on this matter*). We will see in § 6 and 7 that L2 learners of English often use *it* in contexts where *there* is required.

5 No such agreement relation is observed in the French equivalent to these sentences (*Il est arrivé trois filles*), with a singular verb but a plural postverbal subject. Expletive *il* appears to be in the structure simply to satisfy the Extended Projection Principle (EPP), i.e., the requirement that clauses have subjects (Chomsky 1981).

6 A further restriction concerns (in)definiteness (Safir 1985). However, definite NPs (and proper nouns) are not uncommon in *there*-constructions, e.g. when the postverbal element is a list as in (i) below (see Biber *et al*. 1999: 947; Prince 1992: 299). We thank an anonymous reviewer for bringing this fact to our attention.

(i) Michael puts loose papers like class outlines in the large file-size pocket. He keeps his checkbook handy in one of the three compact pockets. The six pen and pencil pockets are always full and in the outside pocket go his schedule book, *chap stick, gum, contact lens*. 

---
solution and hair brush. [Land’s End March 1989 catalog. p. 95] (Birner 1994: 254, our underlining)

7 See Biber et al. (1999: 954) for differences in use between there-constructions and locative inversion. Structurally, the two constructions are not mutually exclusive (e.g. Out of the house there came a tiny old lady.)

8 As pointed out by an anonymous reviewer, Zubizarreta’s (1998) nuclear stress assignment may account for the differences between English and Spanish. While in English nuclear stress may be assigned either (i) to the lowest argument of a verb (e.g. the subject with unergative and unaccusative verbs or the object with transitive verbs) or (ii) to the lowest constituent in a structure, in Spanish only (ii) applies. This means that assignment of nuclear stress to the subject necessarily involves placing it postverbally, as the lowest constituent, with all verb classes. This, in turn, is possible as Spanish does not require the subject to be in <Spec, IP> for nominative Case assignment.

9 As is well known, (non-contrastive) presentational/information focus is to be distinguished from contrastive focus (e.g., Kiss 1998). The two types of focus constructions are the result of different rules or processes (Kiss 1998, but also Zubizarreta 1998). Our concern here is non-contrastive focus.

10 As pointed out by Belletti (2004, sec. 5), with the appropriate intonation pattern and the right pragmatic conditions, a postverbal subject can also be interpreted as topic, as illustrated in (i) for Italian (Belletti 2004: 22) (parallel examples can be given for Spanish).

(i) A: Che cosa ha poi fatto Giovanni? B: Ha parlato, Gianni.
‘What has Gianni finally done?’ has spoken, Gianni

11 An example of an unused entity is …from reporters eager to tell the public about… (Prince 1992: 312), where the public is discourse-new but the reader is expected to know of its existence.

12 We use the term Heavy NP-shift descriptively, to denote an NP ‘displaced’ to sentence-final position. This should not be taken as a commitment to any transformational analysis.

13 Note that the verb fly in (16b) is unergative. On the occurrence of unergative verbs in these constructions, see Levin & Rappaport Hovav (1995: ch 6), Mendikoetxea (2006), and Culicover & Levine (2001).
Oshita’s (2004) study is based on 346 usable unaccusatives concordances (L1 Spanish-L2 English corpus), while ours contains 820 (after filtering criteria) (see § 5.1 below). Additionally, he analysed only 10 unaccusative verbs, while we analysed 32 (see § 5.2 below). Finally, the ratio of inversion in Oshita’s study, 14/346 (4%), is lower than ours, 58/820 (7.1%), though the difference is not statistically significant ($\chi^2=3.44$, df=1, $p=0.64$). These facts imply that our results are clearly in line with Oshita’s work, yet our methodology is stricter and our sample larger.

The psychological reality of theoretical constructs has always been a central concern in the generative tradition. A common assumption is that L1 transfer may be taken as evidence for the psychological reality of the transferred item (e.g., Oshita 2004). While our results may be interpreted in that light, we believe that the notion of what is psychologically real remains, at best, elusive.

The current version of WriCLE contains approximately 700,000 words and will soon be available at: http://www.uam.es/proyectosinv/woslac/Wricle (see Mendikoetxea, O’Donnell and Rollinson, forthcoming).

See http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/localness1.htm for further details about LOCNESS.

The source file in the corpus is indicated as follows: *spm* (Spanish-ICLE); *cor (WRICLE); usarg, usmixed, alevels and brsur for LOCNESS.

As pointed out by an anonymous reviewer, L2 inversion may also occur with verbs other than those in the study (e.g., transitives). Indeed, there are more instances of inversion in the corpus than those we have found (e.g., inversion with be). However, it is not our aim to provide a full account of inversion in L2 English, but to see whether unaccusative/unergative VS structures are produced under the same interface conditions in L1/L2 English. Since in L1 English unaccusativity is a condition for VS order, we focus on the Unaccusative Hypothesis, which distinguishes between unergative and unaccusative verbs. Undoubtedly, a fuller picture can be obtained if all cases of inversion are analysed, as we intend to do in future research.

An anonymous reviewer points out that learners of L2 French often produce inverted subjects in questions with non-finite verbs. But note that we are focusing only on those contexts in which VS is possible in native English, hence excluding non-finite contexts, as our aim is to see if learners and
natives produce VS structures under the same interface conditions. Inevitably, this means that there are instances of inversion that may have escaped our notice (see note 19).

21 For further details on the well-attested phenomenon of L2 unaccusative passivisation (*My mother was died when I was young; This situation has already been happened*), see e.g. Balcom (1997), Oshita (2000), Rutherford (1989), and Zobl (19899.

22 Constructions with clausal postverbal subjects, as in (i), fall outside the scope of this paper:

(i) […] it only remains to add that nowadays we live in a world in which […] [spm03050]

23 Since it is well-known that set expressions and formulaic constructions are learnt as unanalysed chunks, we excluded them from our analysis (e.g., *as sometimes happen; They fall in love*). Likewise, we excluded set expressions that are clearly transferred from L1 Spanish (*It can happens the same; It happens the same*, from Spanish *Puede pasar lo mismo; Pasa lo mismo*).

24 Note that in sentence-initial position learners sometimes produce extra material before the expletive, as in (26a) (XP-it-V-S), where XP is typically realised via a PP and, less frequently, via an AdvP.

25 The term ‘grammatical’ must be understood here as ‘structurally possible’ in native English. For example, the sentence *Furthermore there also exist a wide variety of optional channels which have to be paid* [spm03018] is technically ungrammatical (subject-verb agreement mismatch), but it is still regarded as grammatical for our analysis in the sense that there-insertion is a grammatically possible structure in native English in this context.

26 There is a residual 1.7% for learners that corresponds to topic (1 doubtful case out of 58).

27 Interestingly, one fifth of preverbal subjects in the learner corpora (i.e., 19/96 = 19.8%) are realised as personal pronouns (e.g., *he*), which clearly encode a topic.

28 The rest of the concordances contained preverbal material which could not be coded for information-status: overt and null expletives (*there, it, Ø*), and adverbial expressions like *usually, later, etc.*

29 Measures of syntactic complexity can be taken as indicators for linguistic maturity (e.g., mean sentence length and subordination ratios), but as Meunier (2000: 188-189) points out there are no precise definitions of syntactic complexity, as the debate regarding whether subordination is an indicator of linguistic maturity illustrates (e.g., Beaman, 1984). In some cases, condensation and
simplification are better indicators of maturity, than, for instance, subordination. We are grateful to F. Meunier (P.C.) for bringing these facts to our attention.

30 Remarkably, *There-V-S* structures amount to 50% of VS structures produced by L1 French learners in ICLE, showing a higher production rate than that of natives (37.5%) (see Lozano & Mendikoetxea 2009a, b).

31 In relation to transfer, an anonymous reviewer points out that, since in Spanish VS is found in both oral and written registers, it would be useful to analyse the learners’ oral production to see whether VS structures are transferred to their ‘oral’ L2 English. Indeed, oral production provides a very valuable insight into interlanguage grammar, but we believe that if learners produce the structure, whether in their written or in their oral L2, it is a feature of their interlanguage grammars, in the same way as VS is a feature of L1 English, despite being mostly found in written contexts. As Biber (1999: 11.2.3.8) the fact that inversion is much more frequent in the written texts can be explained because writers of fiction make more use of the resources of the language.