

# Learner Corpora and the Acquisition of Word Order: A Study of the Production of Verb-Subject Structures in L2 English<sup>1</sup>

---

Cristóbal Lozano<sup>2</sup> and Amaya Mendikoetxea<sup>3</sup>

## 1 Introduction

Though there is a long tradition of research into word order phenomena in Second Language (L2) acquisition, this area of enquiry has recently been given a new impetus both from theoretical developments on the form-function interplay and, crucially, from the emergence of learner corpora. This paper focuses on a particular phenomenon which has received considerable attention in the literature: the production of postverbal subjects in L2 English by L1 speakers of languages characterised as allowing ‘free inversion’ of the subject in V(erb) S(ubject) structures, such as Spanish and Italian. In previous research emphasis has been placed on the learners’ production of ungrammatical VS order (see for instance Rutherford, 1989; Zobl, 1989 and, more recently, Oshita, 2004). Our approach, however, seeks to identify the syntactic, semantic and pragmatic conditions, as well as conditions deriving from processing mechanisms, under which learners produce inverted subjects, regardless of errors resulting from the syntactic encoding of those notions.

We analyse VS vs. SV structures in the Italian and Spanish subcorpus of ICLE (Grange *et al.*, 2002) and we compare our results with preliminary results obtained from a similar native English corpus (LOCNESS). Thus, we incorporate some of the fundamental tenets of what is known as *Contrastive Interlanguage Approach* (see, e.g. Granger, 1996 and Gilquin, 2001), which establishes comparisons between: (a) native and non-native data, and (b) different non-native data. Our main purpose is to see if the properties that govern the occurrence of postverbal Ss in native English, as currently analysed in the theoretical and descriptive literature, are the same as those operating in the non-native grammars of Spanish and Italian speakers.

We first examine the properties of VS order in English vs. Spanish/Italian (Section 2). In Section 3, we review previous L2 studies on postverbal Ss. Our hypotheses are presented in Section 4. Section 5 describes the method used to extract and code data from the corpus and their statistical treatment. Results are presented and

---

<sup>1</sup> This research has been partially funded by the Ministerio de Educación y Ciencia (HUM 2005-01728FILO) and by a co-financed project grant by the Comunidad Autónoma de Madrid and the Universidad Autónoma de Madrid (09/SHD/016). The Spanish Ministerio de Educación y Ciencia has also funded one of the authors to carry out research at Lancaster University in the academic year 2006-07 (grant PR2006-0054), which is gratefully acknowledged. The content of this paper has been partially presented at the seminar *Linking up contrastive and learner corpus research* at the 4<sup>th</sup> ICLC (International Contrastive Linguistics Conference), Universidad de Santiago de Compostela, at TALC7 (7th Conference on Teaching and Language Corpora), University Paris7 Denis Diderot – Bibliothèque Nationale de France, AESLA 2007, Universidad de Murcia and ICAME 28, Stratford upon Avon. We would like to thank participants at both events for their comments, as well as the editors of a forthcoming volume where part of the content of this research will be published (B. Díaz, G. Guilquin & S. Papp (eds.) *Linking up contrastive and learner corpus research*, Amsterdam: Rodopi).

<sup>2</sup> Universidad de Granada  
*e-mail:* cristoballozano@ugr.es

<sup>3</sup> Universidad Autónoma de Madrid  
*e-mail:* amaya.mendikoetxea@uam.es

discussed in Section 6. In Section 7, we compare learner data with data obtained from LOCNESS and Section 8 presents the conclusion

## 2 Theoretical background

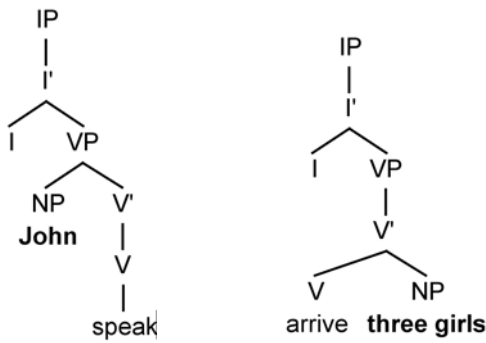
English has ‘fixed’ word order, as opposed to Italian and Spanish where word order is said to be ‘free’. VS order can, however, be found in English in very restricted contexts. Research into word order has shown that the properties of VS order have to be analysed at different levels: (a) the *lexicon-syntax interface*, to account for the lexico-semantic properties of Vs and their interaction with the grammatical properties of the structure; (b) the *syntax-discourse interface*, to account for the discourse status (topic or focus) of the preverbal and the postverbal elements and their interaction with the syntactic properties of the structure; and (c) the *syntax-phonology interface*, to account for the grammatical/phonological properties of the postverbal S along a ‘heaviness’ scale.

### 2.1 Postverbal subjects: syntax-lexicon interface

According to Perlmutter’s (1978) Unaccusative Hypothesis, which is widely accepted in the theoretical literature, it is assumed that there are two classes of intransitive Vs: unergative Vs and unaccusative Vs. Syntactically *unergatives* have an external argument but no internal argument and *unaccusatives* have an internal argument, but no external argument. Semantically, unergatives typically denote activities controlled by an agent, (*speak* in (1a) and also *cry, cough, sweat, jump, run, dance, work, play...*) while unaccusatives have themes (or patients) as their only argument (*arrive* in (1b) and also *blossom, appear, exist, deteriorate, come...*). Levin & Rappaport Hovav (1995) further distinguish between two semantic classes of unaccusative Vs: (a) change of state (*melt, break, open, rust, grow* etc.) and (b) existence and appearance (*arrive, arise, exist, emerge...*)

Theoretical frameworks differ in whether the semantic difference between unergatives and unaccusatives correlates with a syntactic difference or not. In the generative grammar literature, unergatives and unaccusatives have different (base) structures concerning the position occupied by the only argument of the verb. Themes and agents occupy different positions in the structure. Themes are internal arguments and occupy the position of complement of V, while agents are external arguments and are generated in the position of the specifier of the VP (<Spec, VP>) after the introduction of the *VP-internal subject hypothesis* (see Koopman & Sportiche, 1991). Thus, unergative Vs appear in initial structures (D(eep)-Structures) like (1a), with *John* as the external argument, while unaccusatives appear in initial structures like (1b), with *three girls* as the internal argument. In the course of the derivation, the NPs in (1) move to <Spec, IP> to satisfy their Case requirements (i.e., to be assigned nominative Case) and/or the requirement that <Spec, IP> in English must be occupied by an overt element (roughly, Chomsky’s (1981) Extended Projection Principle), as in (2):

- (1) a. *unergative* 'John spoke'                      b. *unaccusative* 'Three girls arrived'                      (D-Structure)



- (2) a. [IP [NP *John<sub>i</sub>*] [VP [NP *t<sub>i</sub>*] [V' spoke]]]                      *Unergative*
- ↑
- b. [IP [NP *Three girls<sub>i</sub>*] [VP [V arrived] [NP *t<sub>i</sub>*]]]                      *Unaccusative*
- ↑

The internal argument of an unaccusative V may, however, remain in its initial, postverbal position under certain conditions. This is the case for *there*-constructions like (3) and *inversion* constructions like (4), where the opening XP adverbial in <Spec, IP> is typically a locative element as in (4a) below (locative inversion structure), but it can also be a time adverbial, as in (4b), as well as other types like the *with*-PP in (4c). In both cases, VS order is restricted to a subclass of unaccusative Vs: those expressing existence or appearance and shows low frequency (see e.g. Biber *et al.*, 1999: 945 and our own results in Section 7)

- (3) [IP [<sub>Expl</sub> *There*] [VP [V arrived] [NP **three girls**]]]
- (4) a. *On one long wall hung a row of Van Goghs.*  
 b. *Then came the turning point of the match.*  
 c. *With incorporation, and the increased size of the normal establishment came **changes which revolutionized office administration.***  
 (corpus examples from Biber *et al.* 1999: 912-913)

Let us look now at subject inversion in 'free inversion' languages like Spanish and Italian, where word order is more flexible. Intransitive sentences in these languages freely allow postverbal Ss as in (5), while the corresponding sentences in English are ungrammatical (6).

- (5) a. i. Ha hablado *Juan*                      ii. Ha llegado *Juan*.                      *Spanish*  
 b. i. Ha parlato *Gianni*.                      ii. E' arrivato *Gianni*.                      *Italian*
- (6) i. \*Has spoken *John*                      ii. \*Has arrived *John*.

Subject inversion in languages like Spanish/Italian is associated with the fact that in these languages the subject need not be expressed by an explicit pronominal element. Both properties characterise languages which are positively marked for the *N(ull) S(ubject) P(arameter)*, as opposed to languages which are negatively marked for the NS Parameter (e.g., Chomsky, 1981; Rizzi, 1982; Burzio, 1986; Eguren & Fernández Soriano, 2004; Jaeggli & Safir, 1989; Luján, 1999; Rizzi, 199; Zagana, 2002).

In VS structures, a null element is postulated for the preverbal S position of structures like those in (5) (which we take to be the specifier of the IP, <Spec, IP>):

*expletive pro* ( $pro_{expl}$ ). This is the null equivalent of the overt expletive in non null-S languages, such as French *il* (7b) or English *there* (7c).

- |        |                              |  |                |
|--------|------------------------------|--|----------------|
| (7) a. | [ <sub>IP</sub> $pro_{expl}$ | [ <sub>VP</sub> llegaron <b>tres chicas</b> ]].    | <i>Spanish</i> |
| b.     | [ <sub>IP</sub> <i>Il</i>    | [ <sub>VP</sub> est arrivé <b>trois filles</b> ]]. | <i>French</i>  |
| c.     | [ <sub>IP</sub> <i>There</i> | [ <sub>VP</sub> arrived <b>three girls</b> ]].     |                |

The assumption is that the NPs in (7) surface in the postverbal subject position they occupy as internal arguments, as in (1b). Unergatives, however, are not allowed in the construction in English, hence the ungrammaticality of (8).

- |        |  |
|--------|--|
| (8) a. | * <i>There</i> has phoned <b>Maria</b> the president |
| b.     | * <i>There</i> has spoken <b>John</b>                |

To summarise, postverbal Ss are possible in English, but their occurrence is restricted to (a subset of) unaccusative Vs and highly constrained and their frequency is low, while Spanish and Italian show what appears to be ‘free inversion’; i.e. the occurrence of inverted subjects is not restricted to a lexico-syntactic class of intransitive verbs.

## 2.2 Postverbal subjects: syntax-discourse interface

It has been claimed recently that information structure notions such as topic and focus play a crucial role in the position of S in null-subject languages, with postverbal Ss usually analysed as (presentational/informational) focus, i.e., new information (e.g., Vallduví, 1990; Fernández-Soriano, 1993; Licerias *et al*, 1994; Into, 1997; Picallo, 1998; Zubizarreta, 1998, 1999; Belletti 2001, 2004b; Domínguez, 2004, Lozano 2006). Thus, in narrow focus constructions like those in (9) and (10) in which we are asking about subject, we expect the answer to contain a postverbal subject as in the (b) examples below (pragmatic anomaly in the (c) examples is marked as #).

- |        |   |                |
|--------|---|----------------|
| (9) a. | ¿Quién ha llegado/hablado?<br>who has arrived/spoken?                     | <i>Spanish</i> |
| b.     | Ha llegado/hablado <b>Juan</b><br>has arrived/spoken <b>Juan</b>          |                |
| c.     | # <b>Juan</b> ha llegado/hablado<br>Juan has arrived/spoken               |                |
| (10)a. | Chi è arrivato/ha parlato?<br>who has arrived/spoken?                     | <i>Italian</i> |
| b.     | É arrivato/Ha parlato <b>Gianni</b><br>has arrived/spoken <b>Gianni</b>   |                |
| c.     | # <b>Gianni</b> è arrivato/ha parlato<br><b>Gianni</b> has arrived/spoken |                |

In English *there*-constructions and locative inversion structures are also often analysed as involving (presentational) focus (see, among others, Bolinger, 1977; Rochemont, 1986; Bresnan, 1994). For Bresnan (1994) in locative inversion structures the referent of the postverbal NP is introduced (or reintroduced) on the scene referred to by the preverbal PP: for instance in (4a) *on one long wall* provides the scene onto which *a row of Van Goghs* is introduced, which is characterised as a new discourse entity. By contrast, Birner (1994, 1995) argues that the discourse function of all inversion constructions is that of “linking relatively unfamiliar information to the prior context

through the clause-initial placement of information that is relatively familiar in the discourse” (Birner, 1995: 238). This is the case in (11), where *in the outside pocket* is relatively more familiar (topic) than the material in postverbal position (focus).

- (11) 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 [<sub>PP</sub> *in the outside pocket*] [<sub>V</sub> *go*] [<sub>NP-SUBJECT</sub> **his schedule book, chap stick, gum, contact lens solution and hair brush**]. (Land’s End March 1989 catalog. p. 95, quoted in Birner, 1994: 254)

In this study, we use focus and topic as common labels for new vs. old information, respectively. Information structure is analysed along a continuum with topic and focus as concepts which encompass a variety of notions which are best analysed in terms of a gradience (see Prince, 1981 and Kaltenböck, 2005). Both evoked and inferrable entities are considered to be topics, on the basis of Prince’s (1992) study and Birner’s (1994, 1995) findings that both entities are treated alike (as discourse-old) in inversion structures. The notion focus encompasses a similar gradience: *brand-new* information (i.e., completely new, not previously mentioned in the discourse) is less retrievable than *new-anchored* information (i.e., an irretrievable state of affairs or entity, which is in some way linked to (‘anchored in’) the previous context).

Given that both in English and in Italian/Spanish, inversion is used as some sort of focalization device, we do expect the inverted Ss in our learners’ grammar to be discourse-new or focus. It has to be stressed, however, that Italian and Spanish make use of this device with *all* verb types, while in native English inversion appears to be restricted mostly to unaccusative Vs of existence and appearance. Despite this, previous studies have found that VS order in the L2 English of Spanish and Italian learners is only found in unaccusative contexts (see section 3 below).

### 2.3 Postverbal subjects: syntax-phonology interface

Linear ordering is also influenced by operations which alter the canonical word order of constituents post-syntactically (at the Phonological Form level) and which do not involve changes in meaning (in terms of truth conditions. An example of this is ‘Heavy NP-Shift’, where a heavy’ NP has been ‘displaced’ to the end of the sentence (12b), as opposed to (12a) with canonical V-NP-PP order:

- (12)a. I bought [<sub>NP</sub> *a book written by a specialist in environmental issues*] [<sub>PP</sub> *for my sister*].  
 b. I bought [<sub>PP</sub> *for my sister*] [<sub>NP</sub> *a book written by a specialist in environmental issues*].

The question is how we define ‘heavy’. Heaviness is often related to structural or grammatical complexity. In fact, heaviness can be defined simply as a matter of string length (number of words) or on the basis of more sophisticated criteria to do with grammatical complexity (see Arnold *et al.*, 2000 for a review of these two approaches to the notion of heaviness). In fact, the two concepts are difficult to tear apart, as revealed by Wasow’s (1997) corpus study which shows high correlations among the various characterizations of heaviness. What emerges out of these studies is that long and complex elements tend to be placed towards the end of the clause, an operation which reduces the processing burden and, thus, eases comprehension by the receiver. Since long and complex grammatical elements typically also carry new information, the end-weight principle and the discourse principle by which new information tends to be placed towards the end of the clause appear to reinforce each other (see Biber *et al.*, 1999:

11.13). As pointed out by Arnold *et al.* (2000) in a study designed to compare the influence of heaviness and newness in constituent ordering “items that are new to the discourse tend to be complex, and items that are given tend to be simple” (p. 34).

The operation of Heavy NP Shift illustrated is just one device, among others, to alter the canonical order of constituents to comply with the principle of end-weight. The end-weight principle appears to be in operation also in VS structures. Thus in the inversion structure in (11) above, the subject is clearly ‘heavy’. Our own analysis of the corpus examples used by Levin & Rappaport Hovav (1995) (LRH) in their study of locative inversion reveals, indeed, that the postverbal S is overwhelmingly heavy (see also Culicover & Levine, 2001): when it is a proper noun or a lighter NP, it is normally followed by material in apposition, as in (13) (highlighting is ours).

- (13)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, cited in L&RH: 221]
- 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, cited in L&RH: 257)

The gradience approach adopted for information status is also adopted in our study for ‘heaviness’: the heavier a NP is the more likely it is to be placed in clause-final position. The relatively ‘free’ word order of Spanish and Italian means that the principle of end-weight may be less noticeable in these languages. Given that it serves a general processing mechanism, we will assume, following Hawkins (1994) that this is a universal principle (see also Frazier, 2004).

The conclusion, then, is that long and complex information tends to be placed at the end in both English and Spanish/Italian. Therefore, we expect learners to produce postverbal Ss which are long and complex, as a reflex of this general processing mechanism. As we have seen, the principle of end-weight interacts with information structure principles which operate at the syntax-discourse interface, by which (discourse-) new information tends to be placed towards the end of the clause. Thus, Ss which are focus, long and complex tend to occur postverbally in those structures which allow them in both English, on the one hand, and Spanish and Italian, on the other hand. This is also the prediction made for the learners in our study.

### 3 Previous research on L2 postverbal subjects

Previous studies show a remarkably consistent pattern in which unaccusative and unergative verbs are treated differently by learners of English regarding the occurrence of postverbal Ss. Learners of L2 English with different background (Spanish, Italian, Arabic, Japanese) produce postverbal Ss with unaccusatives only (Rutherford, 1989; Zobl, 1989), where the postverbal S is shown in bold, (14) and (15).

- (14)a. On this particular place called G... happened **a story which now appears on all Mexican history books**... (L1 Spanish)
- b. The bride was very attractive, on her face appeared **those two red cheeks**... (L1 Arabic).  
(Source: Rutherford 1989)
- (15) a. ...because in our century have **appeared the car and the plane**... (L1 Spanish)
- c. ...it happened **a tragic event**... (L1 Italian)  
(Source: Oshita 2004)

This adds to other type of evidence, provided in Oshita (2004) which points towards the fact that the Unaccusative Hypothesis, that is, the unaccusative-unergative distinction, is psychologically real in SLA. This is demonstrated by studies on learners' preference of VS with unaccusatives but SV with unergatives (Hertel, 2003; Lozano 2006), auxiliary selection (Sorace, 1993, 1995), the production of 'passivised' unaccusative structures (Zobl, 1989, Oshita, 2000) and learners' reluctance to accept SV order with unaccusatives (Oshita ,2002) (see also Balcom, 1997; Hirakawa, 1999; Montrul, 2004 and Yusa, 2002).

We follow previous research in taking the Unaccusative Hypothesis to be psychologically real for L2 learners of English from speakers of Null Subject languages like Italian and Spanish, which is the reason for predicting that VS order will be found only with unaccusative verbs in our corpus

## 4 Hypotheses

While early studies clearly show that unaccusativity is a necessary condition for the production (and/or acceptance) of postverbal Ss, they overlook the fact that unaccusativity is a necessary *but not sufficient* condition for the production of postverbal subjects. As the discussion in the previous section has emphasised, there is a clear tendency for postverbal S to be heavy (i.e., phonologically long) and focus (i.e., new information). Thus, we postulate the three hypotheses in (16). While H<sub>1</sub> has found support in the literature, H<sub>2</sub> and H<sub>3</sub>, to our knowledge, have not been tested before in the L2 literature.

- (16) **H<sub>1</sub>: Lexicon (*Lexicon – Syntax interface*):** Both Spanish and Italian learners of L2 English will produce postverbal Subjects only with unaccusatives, but never with unergatives.  
**H<sub>2</sub>: Weight (*Syntax – Phonology interface*):** In those contexts where inversion is allowed, both groups of learners will tend to place subjects (i) in postverbal position when S is heavy but (ii) in preverbal position they are light.  
**H<sub>3</sub>: Focus (*Syntax – Discourse interface*):** In those contexts where inversion is allowed, both groups of learners will tend to place subjects (i) in postverbal position when S is focus but (ii) in preverbal position when they are topic.

## 5 Method

### 5.1 Corpora

We used the Spanish and Italian subcorpora of International Corpus of Learner English, ICLE (Granger et al. 2002), which consists of 11 subcorpora of academic essays written by advanced L2 English learners of 11 different L1s. In total, 427,461 words were used in our analyses (Table 1).

<b>Corpus</b>	<b>Number of essays</b>	<b>Number of words</b>
ICLE Spanish	251	200,376
ICLE Italian	392	227,085
<b>TOTAL</b>	<b>643</b>	<b>427,461</b>

**Table 1:** Corpora

## 5.2 Data analysis

Following Levin (1993) and Levin & Rappaport-Hovav (1995), we constructed an inventory of unaccusative (n=32) and unergative (n=41) lemmas in English (see Table 6 in the appendix), which were searched in the concordancer WordSmith Tools 4.0 (Scott 2002). All possible forms of the lemma (both native English and possible misspelt learner forms) were queried, e.g., for the lemma APPEAR: *appear, appears, appearing, appeared, appeared, apear, apears, appearing, apeared, apeard*. The concordances output by WordSmith were filtered manually according to 51 criteria to discard those structural contexts in which inversion in English is not possible, regardless of the nature of V. Approximately  $\frac{3}{4}$  of the concordances turned out to be unusable since they did not meet the filtering criteria, like those in (20) (see Lozano & Mendikoetxea, forthcoming and in preparation, for further details). The filtering process resulted in 1510 usable concordances, as shown in Table 2).

- (17) a. The V must be intransitive (unaccusative or unergative) (4e discarded (un)grammatical uses of transitive unaccusatives like e.g., *parents grew their children*)  
 b. The V must be finite  
 c. The V must be in the active voice.  
 c. The subject must be a Noun Phrase.

Subcorpus	V type	# usable concordances
Spanish	Unergative	153
	Unaccusative	640
Italian	Unergative	143
	Unaccusative	574
TOTAL		1510

**Table 2:** Usable concordances

Regarding the *weight* of the S, most authors use length in number of words as a simple measure of weight. While there are certainly more sophisticated measures such as syntactic complexity, it is well known that length and complexity are highly correlated (e.g., Wasow, 1997; Wasow & Arnold, 2003). In this study we report only on length (in number of words), though we have previously used both syntactic complexity and number of words, which produced very similar results (e.g., Lozano & Mendikoetxea, forthcoming 2007).

As for the analysis of the *discursive status* of the S, Ss were coded as either topic or focus, according to our earlier definition of these terms, by which topic and focus are concepts encompassing a variety of notions which are best analyzed in terms of a gradient scale such as the retrievability scale in Kaltenböck (2004, 2005). Coding was performed manually, taking into consideration the preceding discourse and context to determine whether each S was topic or focus.

The issue of grammaticality/acceptability arises whenever one is dealing with learner data. Since our focus is on the conditions under which learners produce postverbal subjects, and not on errors, we coded in both grammatical and ungrammatical sentences containing postverbal subjects. The term ‘ungrammatical’ for learner data is commonly used in the L2 acquisition literature as synonymous with deviant in relation with the

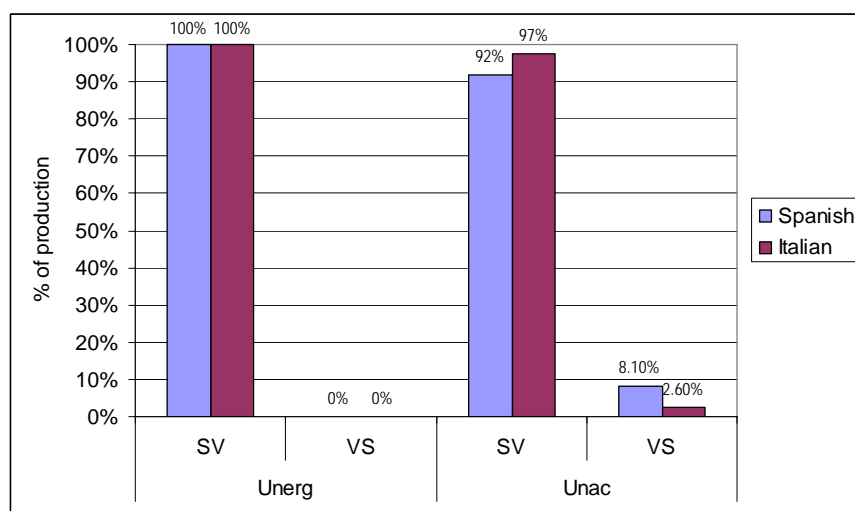


native form. Here, however, we are abstracting away from standard ungrammaticalities such as S-V agreement or wrong past tense form. If a learner produces a sentence conforming to the conditions under which native speakers produce postverbal subjects (initial XP or expletive *there*, unaccusative V, subject focus and/or long), the sentence is considered to be grammatical. Thus, a sentence like *Then come the necessity to earn more* [spm07023] is coded in as grammatical as the postverbal-subject structure is possible in English, though lack of S-V agreement would render it ungrammatical in native English. Conversely, *it*-insertion renders a postverbal subject structure ungrammatical as in *\*I do believe that it will not exist a machine or something able to imitate the human imagination* [spm01007] (see (21a) below). Our results show that most unaccusative postverbal Ss produced by our learners are ungrammatical in the sense in which we use this term, the difference being more marked in the Spanish corpus (65.4% ungrammatical vs. 34.6% grammatical) than in the Italian corpus (53.3% vs. 46.7%), but the difference is not significant [ $\chi^2=0.723$ ,  $df=1$ ,  $p=0.395$ ].

## 6 Results and discussion

### 6.1 Results for H<sub>1</sub>: syntax-lexicon interface

There is no difference between both groups in their production rates with unergatives (100% of SV), yet their production with unaccusatives differs significantly: 8.1% vs. 2.6% of VS [ $\chi^2=17.630$ ,  $df=1$ ,  $p<0.001$ ], (Figure 1). While this between-group difference is statistically significant, it is important to note that both groups produce VS only with unaccusatives (see also Table 3), as expected, and that their relative rates of production are not significantly different, as we will see below (for an explanation of this difference, see Lozano & Mendikoetxea, in press, in preparation).



**Figure 1:** Percentage of subjects produced (group x verb type)

Subcorpus	V type	# postverbal S	# usable concordances	Rate (%)
Spanish	Unergative	0	153	0/153 (0%)
	Unaccusative	52	640	52/640 (8.1%)
Italian	Unergative	0	143	0/143 (0%)
	Unaccusative	15	574	15/574 (2.6%)

**Table 3:** Postverbal subjects produced

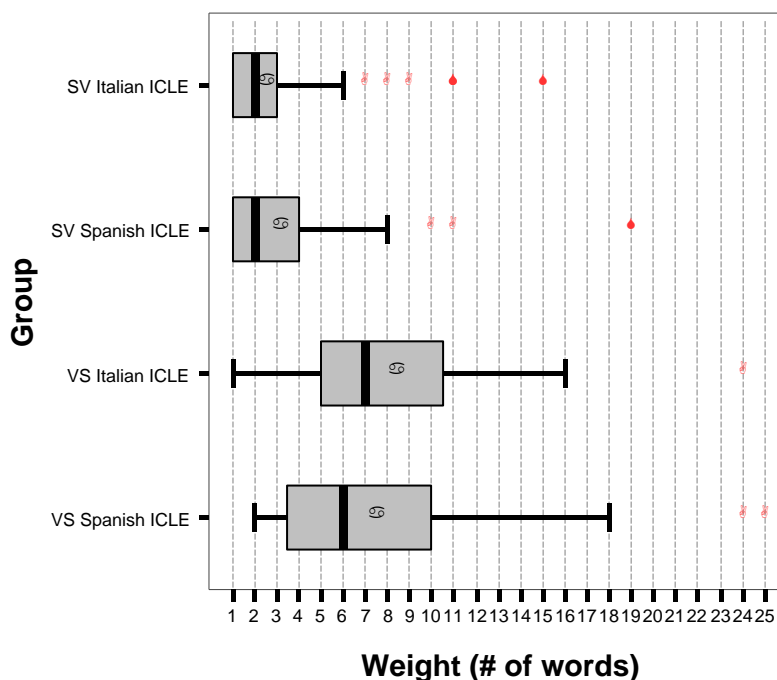
The production of unaccusative VS in the two groups contained both grammatical and ungrammatical constructions, though, as mentioned above, most constructions were ungrammatical [65.4% Spanish group; 53.3% Italian group]. Additionally, VS production rates are higher with unaccusatives of existence (*exist*) and appearance (*appear*). Finally, the type of VS constructions produced were of different types, often involving preverbal material, i.e., XP-V-S structures (see Lozano & Mendikoetxea, in press, in preparation, for further details), as in (18)-(23), where we also show the corpus file names (those beginning with ‘s’ belong to the Spanish subcorpus and those with ‘i’ to the Italian one).

- (18) Ungrammatical *it*-insertion:  
a. I do believe that *it* will not exist a machine or something able to imitate the human imagination. (spm01007)  
b. ...and *it* still live some farmers who have field and farmhouses. (itb07001)
- (19) Grammatical locative inversion:  
a. *In the main plot* appear **the main characters: Volpone and Mosca** ... (spal1002)  
b. Cesare Lombroso (1835/1909) criminological, asserted that *on the earth* lived **people which were born-criminal**. (itrl1005)
- (20) Insertion of any other type of phrase (*XP*-insertion), which is typically (but not exclusively) a PP:  
a. *There* exists a whole range of occ[a]sions in which we have had to be witness of how people from other nations usually fight abroad for foreign causes. (spm10015)  
b. ..., *there* still remains **a predominance of men over women**. (itto4006)
- (21) Ungrammatical  $\emptyset$ -insertion:  
a. Nevertheless exist other means of obtaining it [i.e., money] which are not so honourable, but quicker. (spm01013)  
b. Instead I think that exist factors which, on long term, can predispose human mind to that crime ... (itrl1010)
- (22) *AdvP* insertion:  
a. ..., and *here* emerges **the problem**. (spm01001)  
b. *Later* came **a world of disorder**, during and after the First World War ... (itrs1010)
- (23) Grammatical existential *there*-insertion:  
a. ...and *from this moment* begins **the avarice**. (spm04048)  
b. [No instances of *XP*-insertion were found in the Italian corpus]

To summaries, results show that Spanish and Italian learners of English produce postverbal Ss only with unaccusative verbs (and never with unergatives), as H<sub>1</sub> predicts and as shown in previous L2 studies.

## 6.2 Results for H<sub>2</sub>: syntax-phonology interface

The boxplot (Figure 2) represents the spread of weight (in number of words) for unaccusative pre- and post-verbal Ss in each subcorpus (Spanish and Italian), with circles representing outliers and asterisks representing extreme cases. While both heavy and light Ss appear in both preverbal and postverbal positions, both groups behave statistically alike: preverbal Ss are light for both groups [mean=3.2 (Spanish) and 2.6 (Italian),  $t=1.430$ ,  $df=175$ ,  $p=0.155$ ] as in (24a,b), while postverbal Ss are heavy (long) for both groups [mean=7.0 words (Spanish) and 7.5 (Italian),  $t=-0.554$ ,  $df=65$ ,  $p=0.581$ ], as in (25a',b').



**Figure 2:** Boxplot (with median and mean) of subject weight in number of words

- (24) Preverbal unaccusative subjects: light vs. heavy
- a. ...for the first time, *beggars* appeared. (spm02003)
  - a'. .... it was in that time when the *utopian societies created by the [e]arly socialists* appeared. (spm04019)
  - b. *Violence* does exist ... (itto2034)
  - b'. Nowadays, *the differences between men and women* should not exist any more,... (itto4006)
- (25) Postverbal unaccusative subjects: light vs. heavy
- a. ....and from there began **a fire**, ... (spm04011)
  - a'. .... and thus began the period known as Restoration, which in literature ended in 1707 on the death of George Farquhar, the last mahor writer of the "Comedy of Manners". (spm08005)
  - b. We could call it the body language and through it, emerges **the protagonists' personality**. (itrs1064)
  - b'. This is conveyed in line 25 where by the expression, emerges **the people's ignorance in having prejudices**. (itrs1065)

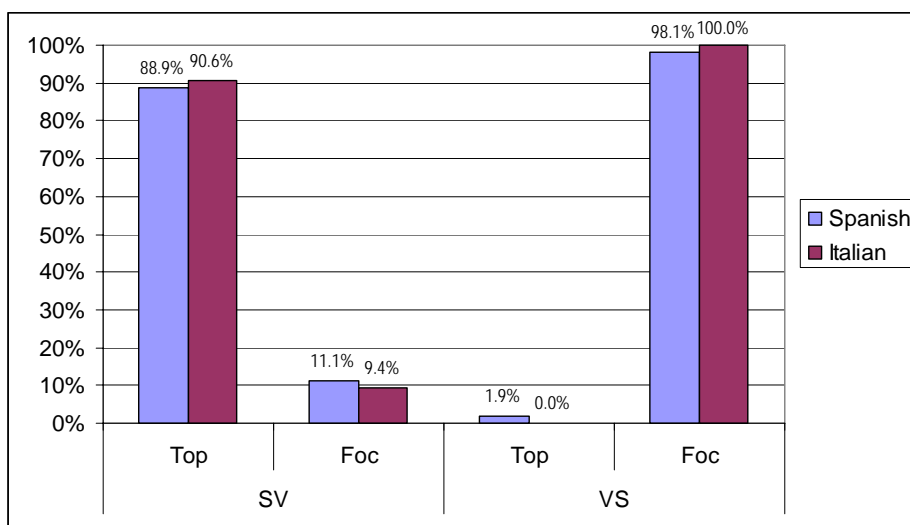
To summaries, Spanish and Italian learners of L2 English produce unaccusative Ss in postverbal position when they are heavy (long), yet in preverbal position when they are light (short). This finding confirms our H<sub>2</sub>.

### 6.3 Results for H<sub>3</sub>: syntax-discourse interface

For both groups the unaccusative S is produced (i) postverbally to mark focus but (ii) preverbally to express topic (Figure 3 and Table 4). While the vast majority of unaccusative preverbal Ss are topic [88.9% Spanish; 90.6% Italian], all postverbal Ss are focus [98.1% Spanish; 100% Italian]. Importantly, both groups behave similarly: preverbal Ss [ $\chi^2=0.480$  df=1,  $p=0.488$ ] and postverbal Ss [ $\chi^2=0.293$  df=1,  $p=0.588$ ].

Subcorpus	Word order	Topic	Focus
Spanish	Unac SV	72/81 (88.9%)	9/81 (11.1%)
	Unac VS	1/52 (1.9%)	51/52 (98.1%)
Italian	Unac SV	87/96 (90.6%)	9/96 (9.4%)
	Unac VS	0/15 (0%)	15/15 (100%)

**Table 4:** Production of subjects (topic vs. focus) with unaccusatives



**Figure 3:** Production of subjects (topic vs. focus) with unaccusatives

Example (26) illustrates unaccusative postverbal Ss being new information (focus), i.e., the S has not been mentioned previously in the discourse. By contrast, (27) illustrates unaccusative preverbal Ss which are topics (shown in *italics*) since they have been mentioned in the prior discourse (shown in underlined typeface).

- (26) a. In the world, dominated by science, technology and industrialisation, there is no a place for dreaming and imagination. Thanks to science and its consequences, technology and insdustrialisation, appeared the big factories and the capitalism system. (spm03007)

- b. It seems impossible, but although we have now reached through technology a high standard of life, we are very pessimists. It seems as progress has stolen our imagination and therefore the love for small things. I can give few examples that such a fact: television is becoming lately the killer of conversation between parents and children; it is almost disappearing the use of writing nice letters to friends, since there is the telephone. (itrs1018)
- (27)
- a. The approval of acting of women were something essential. Women started to perform female characters and this contribute to give a sexual and realistic atmosphere. [...] *Female characters* appear with a stronger personality they really love these men. (spm08014)
  - b. The idea of Europe doesn't ignore these differences, but inglobes them, accept them and upon them construct its identity.[...] If I think of the concept of Europe I cannot think of anything else that of a whole of different countries, but that all together produce the European identity. *The differences* have always existed in the Europe and for ages its peoples fought one against the other. (itrs1008)

To summarise, Spanish and Italian learners of L2 English produce unaccusative Ss in postverbal position when they are focus, yet in preverbal position when they are topic. This finding confirms our H<sub>3</sub>.

## 7 Native grammars vs. non-native grammars: preliminary observations

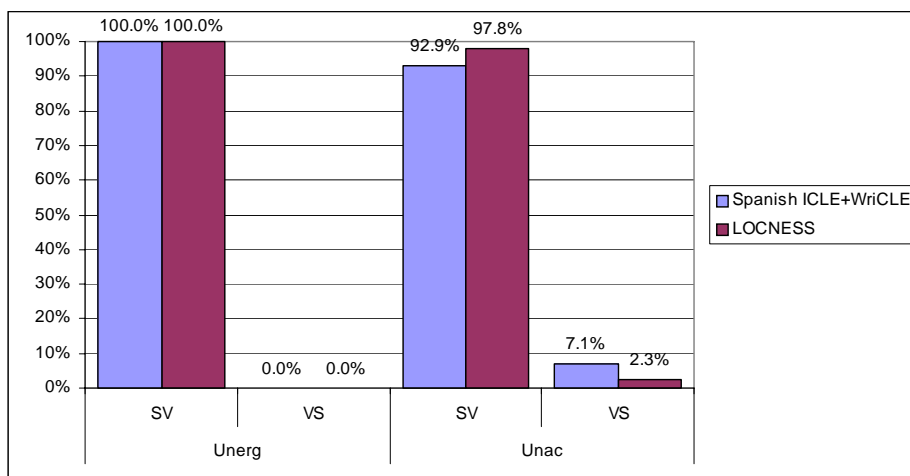
As mentioned in the Introduction *Contrastive Interlanguage Approach* (CIA) involves crucially comparing native and non-native data (NS vs. NNS): a detailed analysis of linguistic features in native and non-native corpora to uncover and study non-native features in the speech and writing of (advanced) non-native speakers (see Granger 2002). Though a full comparison of our results against those obtained from native Spanish/Italian and English corpora is yet to be accomplished, we are in a position to offer some preliminary results of a comparison between the performance of Spanish learners of English and English native speakers regarding the occurrence of postverbal subjects. In order to do so, we analyzed the occurrence of postverbal subjects in the Spanish subcorpus of ICLE, as presented throughout this paper, as well as in 85 essays collected at the Universidad Autonoma de Madrid, following the same parameters as ICLE, as part of the WriCLE (Written corpus of learner English) currently being collected (see <http://www.uam.es/woslac>) and we compared our results with those obtained from a parallel analysis of an equivalent native English corpus: LOCNESS (Louvain corpus of native English Essays, UCL, Louvain-la-Neuve), as shown in Table 5.

<b>Corpus</b>	<b>Number of essays</b>	<b>Number of words</b>
ICLE Spanish + WriCLE	336	264,211
LOCNESS	436	324,304

**Table 5:** NS vs. NNS Corpora

As expected, native speakers did not produce inversion structures with unergative Vs. As for unaccusative Vs, these structures are significantly less frequent in the native speakers' writing than in the writing of the Spanish learners: 58 out of 820 concordances

for ICLE+WriCLE (7.1%) and 16 out of 702 for LOCNESS (2.2%), as represented in Figure 4:



**Figure 4:** Production of postverbal subjects in NS vs. NNS

These results indicate that though Spanish learners are sensitive to the unergative-unaccusative distinction, they ‘overuse’ inversion of the subject in unaccusative contexts. The results are not surprising, given that, as shown in section 2, inversion in English is much more highly constrained both syntactically and pragmatically, than it is in Spanish.

More interesting for our present purposes are the results concerning our H<sub>2</sub> and H<sub>3</sub> in section 4, which show no difference between NS and NNS regarding the weight and information status of the postverbal subject. As in NNS, preverbal subjects with unaccusative Vs tend to be ‘light’ in LOCNESS (67.7% in ICLE+WriCLE and 68.1% in LOCNESS), while postverbal subjects are overwhelmingly ‘heavy’ (81.0% in ICLE+WriCLE and 81.3% in LOCNESS). As for information status, most preverbal subjects with unaccusative Vs in LOCNESS are topic (83.5%) and just a few are focus (16.5%). The same pattern is observed for our learners (89.9% topic, 10.5% focus), with no significant differences between NS and NNS ( $p=0.223$ ). Postverbal subjects, on the contrary, are overwhelmingly focus: 100% in LOCNESS and 98.3% in ICLE+WriCLE, with no significant differences between NS and NNS ( $p=0.784$ ).

These results confirm that Spanish (and, presumably, Italian) learners of English produce postverbal subjects under exactly the same interface conditions. However, our Spanish learners overuse the construction and show persisting problems in the syntactic encoding of VS structures, producing mostly ungrammatical examples.

## 8 Conclusion

In the present study we have used a large-scale learner corpus (ICLE) to show that, similarly to what has been found in previous research, Spanish and Italian learners of English produce postverbal subjects (VS order) only with unaccusative verbs. Unlike previous research, we have also shown that unaccusativity is a necessary yet not sufficient condition for the production of VS, since the postverbal subject needs to be phonologically heavy (i.e., long) and discursively focus (i.e., new information).

These findings show that a full account of postverbal subjects in L2 English needs encompass factors at the interfaces: (i) syntax-lexicon interface (unaccusativity), (ii) syntax-phonology interface (weight) and (iii) syntax-discourse interface (focus).

## References

- Arnold, J.E., T. Wasow, A. Losongco, and R. Ginstrom (2000) 'Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering' *Language*, 76, 28–55.
- Balcom, P. (1997) 'Why is this happened? Passive morphology and unaccusativity'. *Second Language Research* 13, 1–9.
- Belletti, A. (2001) "Inversion" as focalization, in A. C. Hulk, and J-Y Pollock, (eds), *Subject Inversion in Romance and the Theory of Universal Grammar*, pp 60–90. Oxford: Oxford University Press.
- Belletti, A. (2004a) (ed.) *Structures and Beyond. The Cartography of Syntactic Structures. Vol 3.* New York: Oxford University Press.
- Belletti, A. (2004b) Aspects of the low IP area, in L. Rizzi (ed.) *The Structure of CP and IP. The Cartography of Syntactic Structures. Vol 2*, pp 16–51. New York: Oxford University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan (1999) *The Longman Grammar of Spoken and Written English*. London: Longman.
- Birner, B. (1994) 'Information status and English inversion'. *Language* 70, 233–59.
- Birner, B. (1995) 'Pragmatic constraints on the verb in English inversion'. *Lingua* 97, 223–56.
- Bolinger, D. (1977) *Form and Meaning*, London: Longman.
- Bresnan, J. (1994) 'Locative inversion and the architecture of Universal Grammar'. *Language* 70, 72–131.
- Burzio, L. (1986) *Italian Syntax*. Dordrecht: Kluwer.
- Chomsky, N. (1981) *Lectures on Government and Binding*. Dordrecht: Foris.
- Chomsky, N. (1995) *The Minimalist Program*, Cambridge, MA: MIT Press.
- Culicover, P. W. and R. D. Levine (2001) 'Stylistic inversion in English: a reconsideration'. *Natural Language and Linguistics Theory* 19, 2, 283–310.
- Domínguez, L. (2004) *Mapping Focus: The Syntax and Prosody of Focus in Spanish*. Boston University: Unpublished PhD dissertation.
- Eguren, L. and O. Fernández-Soriano (2004) *Introducción a una sintaxis minimista*. Madrid: Gredos.
- Fernández-Soriano, O. (1993) 'Sobre el orden de palabras en español'. *Cuadernos de Filología Hispánica* 11, 113–51.
- Firbas, J. (1992) *Functional Sentence Perspective in Written and Spoken Communication*, Cambridge: CUP.
- Frazier, L. (2004) '(Default) Focus structure in sentence processing'. Paper presented at the workshop "*Information structure in language processing and language acquisition*", University of Potsdam, October 2004.
- Geluykens, R. (1991) Information flow in English conversation: a new approach to the given-new distinction, in E. Ventola (ed.) *Functional and Systemic Linguistics: Approaches and Uses*, pp. 141-167. Berlin: Mouton de Gruyter.
- Gilquin, G. (2001) 'The Integrated Contrastive Model. Spicing up your data'. *Languages in Contrast* 3, 1, 95–123.

- Granger, S. (1996) From CA to CIA and back: An integrated approach o computerized bilingual and learner, in K. Aijmer, B. Altenberg . and M. Johansson . (eds) *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English* 88. pp. 37–51. Lund: Lund University Press.
- Granger, S. (2002) A bird's-eye view of computer learner corpus research, in: S. Granger, J. Hung and S. Petch-Tyson (eds), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Language Learning and Language Teaching* 6. pp 3–33. Amsterdam & Philadelphia: Benjamins,
- Granger S., E. Dagneaux and F. Meunier (2002) *The International Corpus of Learner English. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Hawkins, J. (1994) *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.
- Hertel, T.J. (2003) 'Lexical and discourse factors in the second language acquisition of Spanish word order'. *Second Language Research*, 19, 273–304.
- Hirakawa, M. (1999) L2 acquisition of Japanese unaccusative verbs by speakers of English and Chinese, in K. Kanno (ed.) *The Acquisition of Japanese as a Second Language*, pp. 292–312 Amsterdam: John Benjamins.
- Jaeggli, O. and K. Safir (eds) (1989) *The Null Subject Parameter*. Dordrecht: Kluwer.
- Kaltenböck, G. (2004) *It-extraposition and non-extraposition in English: A study of syntax in spoken and written texts*. Wien: Braumüller.
- Kaltenböck, G. (2005) 'It-extraposition in English: A functional view'. *International Journal of Corpus Linguistics*, 10, 2, 119–59.
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London & New York: Longman.
- Koopman, H. & D. Sportiche (1991) 'The position of subjects'. *Lingua* 85: 211–58.
- Levin, B. (1993) *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Levin, B. and M. Rappaport-Hovav (1995) *Unaccusativity at the Lexical Semantics-Syntax Interface*. Cambridge, MA.: MIT Press.
- Liceras, J., B. Soloaga and A. Carballo (1994) 'Los conceptos de tema y rema: problemas sintácticos y estilísticos de la adquisición del español'. *Hispanic Linguistics* 5: 43–88.
- Lozano, C. (2003). *Universal Grammar and focus constraints: The acquisition of pronouns and word order in non-native Spanish*. Unpublished PhD dissertation. University of Essex.
- Lozano, C. (2006) 'Focus and split intransitivity: The acquisition of word order alternations in non-native Spanish', *Second Language Research* 22: 1–43.
- Lozano, C. and A. Mendikoetxea (forthcoming) 'Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: a corpus study', in G. Gilquin, B. Díez and S. Papp (eds). *Linking up Contrastive and Learner Corpus Research*. Amsterdam: Rodopi.
- Lozano, C. and A. Mendikoetxea (in preparation) *Interface conditions on postverbal subjects: a corpus study of 'inversion' in non-native grammars*, (ms.), Universidad de Granada/Universidad Autónoma de Madrid.
- Luján, M. (1999) Expresión y omisión del pronombre personal, in I. Bosque and V. Demonte (eds), *Gramática descriptiva de la lengua española*, pp. 1275–1315. Madrid: Espasa-Calpe.
- Montrul, S. (2004) 'Psycholinguistic evidence for split intransitivity in Spanish second language acquisition'. *Applied Psycholinguistics* 25, 239–67.



- Oshita, H. (2000) 'What is happened may not be what appears to be happening: a corpus study of 'passive' unaccusatives in L2 English'. *Second Language Research* 16, 293–324.
- Oshita, H. (2002) 'Uneasiness with the easiest: on the subject-verb order in L2 English', *Second Language* 1, 45–61.
- Oshita, H. (2004) 'Is there anything there when *there* is not there? Null expletives in second language data'. *Second Language Research* 20, 95–130.
- Perlmutter, D. (1978) 'Impersonal passives and the unaccusative hypothesis'. *Papers from the Annual Meeting of the Berkeley Linguistics Society* 4, 157–89.
- Picallo, C. (1998) 'On the Extended Projection Principle and null expletive subjects'. *Probus*, 10: 219–41.
- Pinto, M. (1997) *Licensing and Interpretation of Inverted Subjects in Italian*. Utrecht: LEd.
- Prince, E. F. (1981) 'Toward a Taxonomy of Given-New Information' in P. Cole (ed.). *Radical Pragmatics*, pp 223–55. London: Academic Press.
- Prince, E. F. (1992) The ZPG letter: Subjects, definiteness and information status, in S. Thompson and W. Mann, (eds) *Discourse Description: Diverse Analyses of a Fund Raising Text*, pp. 295–325. Amsterdam: John Benjamins.
- Quirk, R., S. Greenbaum, G. Leech & J. Svartvik (1972) *A Comprehensive Grammar of the English Language*, London: Longman.
- Rizzi L. (1982) *Issues in Italian Syntax*. Dordrecht: Foris.
- Rizzi, L. (1997) 'The fine structure of the left periphery', in L. Haegeman (ed.) *Elements of Grammar. Handbook of Generative Syntax Vol 1*, pp. 281–337. Dordrecht: Kluwer Academic Publishers.
- Rizzi, L. (2004) *Locality and left periphery*, in A. Belletti (ed.), pp. 223–51.
- Rochmont, M. S. (1986) *Focus in Generative Grammar*. Amsterdam: John Benjamins.
- Rutherford, W. (1989) *Interlanguage and pragmatic word order*, in S. Gass and J. Schachter (eds.) *Linguistic Perspectives on Second Language Acquisition*, pp. 163–82 Cambridge: Cambridge University Press.
- Scott, M. (2002) *Oxford WordSmith Tools (version 4.0)*. Oxford: Oxford University Press.
- Sorace, A. (1993) 'Incomplete vs. divergent representations of unaccusativity in nonnative grammars of Italian'. *Second Language Research* 9, 22–47.
- Sorace, A. (1995) *Acquiring linking rules and argument structures in a second language*, in L. Eubank, L. Selinker, M. Sharwood-Smith (eds.) *The current State of Interlanguage: Studies in Honor of William E. Rutherford*, pp.153–75. Amsterdam: John Benjamins.
- Vallduví, E. (1990) *The Informational Component*. PhD dissertation, University of Pennsylvania.
- Wasow, T. (1997) 'End-weight from the speaker's perspective'. *Journal of Psycholinguistic Research*, 26, 347–61.
- Wasow, T. & Arnold, J. (2003) *Postverbal constituent ordering in English*, in G. Rohdenburg and B. Mondorf (eds.) *Determinants of Grammatical Variation*, pp. 119–54. Cambridge: Cambridge University Press.
- Yusa, N. (2002) *The implications of unaccusative errors in L2 acquisition*, in J. Costa and M. J. Freitas (eds.) *Proceedings of the GALA 2001 Conference on Language Acquisition*, pp. 289–96. Lisbon: Associação Portuguesa de Linguística.
- Zagona, K. (2002) *The Syntax of Spanish*. Cambridge: Cambridge University Press.

- Zobl, H. (1989) Canonical typological structures and ergativity in English L2 acquisition, in S. Gass and J. Schachter (eds.) *Linguistic Perspectives on Second Language Acquisition*, pp. 203–221. Cambridge: Cambridge University Press.
- Zubizarreta, M. L. (1998) *Prosody, Focus and Word Order*. Cambridge, MA: MIT Press.
- Zubizarreta, M. L. (1999) ‘Las funciones informativas: tema y foco’, in I. Bosque and V. Demonte (eds), *Gramática descriptiva de la lengua española* (vol. 3). Madrid: Espasa.

## Appendix

Unaccusatives	Unergatives	
<p><i>Semantic class:</i></p> <p><i>Existence:</i> exist, flow, grow, hide, live, remain, rise, settle, spread, survive</p> <p><i>Appearance:</i> appear, arise, awake, begin, develop, emerge, flow<sup>***</sup>, follow, happen, occur, rise<sup>***</sup></p> <p><i>Disappearance:</i> die, disappear</p> <p><i>Inherently directed motion:</i> arrive, come, drop, enter, escape, fall, go, leave, pass, rise<sup>***</sup>, return</p>	<p><i>Semantic class:</i></p> <p><i>Emission:</i></p> <p><i>Communication:</i></p> <p><i>Bodily processes:</i></p> <p><i>Manner of motion:</i></p> <p><i>Performance:</i></p> <p><i>Snooze:</i></p>	<p><i>Semantic subclass:</i></p> <p><i>Light emission:</i> beam, burn, flame, flash</p> <p><i>Sound emission:</i> bang, beat, blast, boom, clash, crack, crash, cry, known, ring, roll, sing</p> <p><i>Smell emission:</i> smell</p> <p><i>Substance emission:</i> pour, sweat</p> <p><i>Manner of speaking:</i> cry<sup>*</sup>, shout, sing<sup>*</sup></p> <p><i>Talk verbs:</i> speak, talk</p> <p><i>Breathe verbs:</i> breath, cough, cry<sup>*</sup>, sweat<sup>**</sup></p> <p><i>Nonverbal expressions:</i> laugh, sigh, smile</p> <p><i>Run verbs:</i> fly, jump, run, swim, walk, ride, travel, slide</p> <p><i>Monadic agentives:</i> dance, phone, play, sing, work</p> <p>Sleep</p>
TOTAL UNACCUSATIVES: 32	TOTAL UNERGATIVES: 41	

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

**Table 6:** Inventory of searchable lemmas