# VERB-SUBJECT ORDER IN L2 ENGLISH: NEW EVIDENCE FROM THE ICLE CORPUS

CRISTÓBAL LOZANO KOETXEA
*Universidad de Granada*
AMAYA MENDI
*Universidad Autónoma de Madrid*

ABSTRACT. *In this paper, we examine the production of Verb-Subject order in the Spanish and Italian subcorpora of ICLE. Our findings confirm that learners produce postverbal subjects only with unaccusative verbs (never with unergatives or with transitives) as reported in previous research. However, we show that unaccusativity is a necessary but not sufficient condition for postverbal subject production, since, crucially, there is a tendency for the postverbal subject to be heavy (i.e., phonologically long) and focus (i.e., discourse-new information). Thus, a full account of the production of inverted subjects in L2 English must look at properties which operate at the interfaces: (i) the lexicon-syntax interface (i.e., unaccusativity), (ii) the syntax-phonology interface (i.e., weight), and (iii) the syntax-discourse interface (i.e., focus).*
KEYWORDS: *L2 acquisition, learner corpora, word order, unaccusatives, unergatives, postverbal subjects*

RESUMEN. *En este estudio examinamos la producción del orden Verbo-Sujeto (sujetos postverbales) en inglés L2 en el corpus ICLE (subcorpus de Español e Italiano). Tal como se ha observado en estudios previos, los aprendices producen sujetos postverbales sólo con verbos inacusativos (pero nunca con inergativos ni con transitivos). A diferencia de los estudios previos, en este estudio demostramos que la inacusatividad es una condición necesaria pero no suficiente para la producción de sujetos postverbales ya que, crucialmente, el sujeto postverbal tiende a ser pesado (es decir, fonológicamente largo) y foco (es decir, información discursiva nueva). Por tanto, una descripción completa de la producción de sujetos postverbales en inglés L2 debe de investigar las propiedades que operan en las interfaces: (i) la interfaz léxico-sintaxis (inacusatividad), (ii) la interfaz sintaxis-fonología (peso) y (iii) la interfaz sintaxis-discurso (foco).*
PALABRAS CLAVE: *adquisición de L2, corpus de aprendices, orden de palabras, inacusativos, inergativos, sujetos postverbales*

## 1. INTRODUCTION

In this paper we analyse in detail the production of postverbal subjects (V(erb)-S(ubject) order) in the writing of Spanish and Italian speakers of English as represented in the ICLE corpus (Granger et al. 2002). While previous research has focused on errors, our approach seeks to identify the conditions under which learners produce inverted S, regardless of problems to do with the syntactic encoding of the structures. 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 discussed in Section 6 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. In what follows, we show 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.

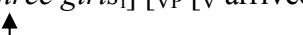### 2.1. *Postverbal subjects: syntax-lexicon interface*

In the generative literature it is commonly assumed that there are two classes of intransitive Vs *unergatives* have an external argument but no internal argument, while *unaccusatives* have an internal argument, but no external argument (Perlmutter's (1978) Unaccusative Hypothesis). The internal argument occupies the position of complement of V, while the external argument is 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.

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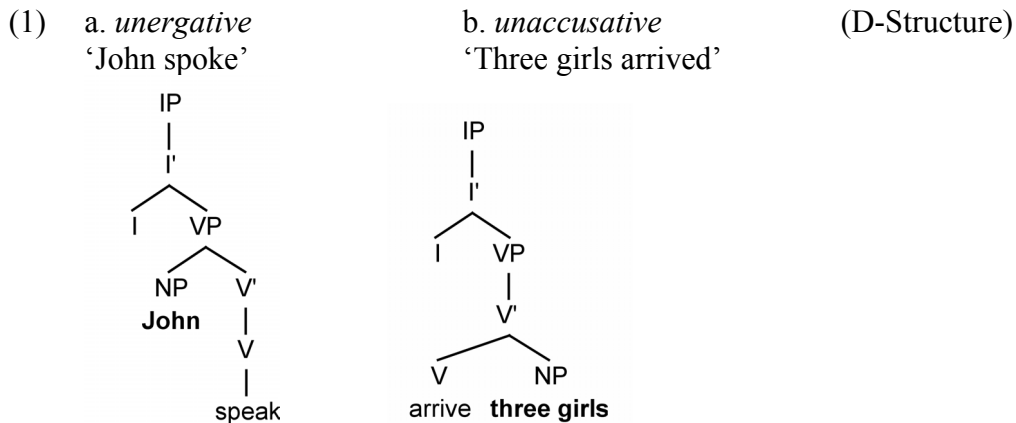


Since Perlmutter (1978), the syntactic distinction between the two classes of Vs is systematically related with a semantic distinction; roughly, unergatives typically denote activities controlled by an agent, (*speak* in (1a) and also *cry, cough, sweat, jump, run, dance, work, play*...) while unaccusatives 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, come, appear,* etc.)

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):

(2)    a.    [IP [NP *John*i]   [VP [NP *t*i] [V' spoke]]]        *Unergative*

b.    [IP [NP *Three girls*i] [VP [V arrived] [NP *t*i]]]        *Unaccusative*

While this movement is obligatory in English for external arguments like *John*, as in (2a), 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*,* 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

(3)    [IP [Expl *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)

Consider now VS structures in the so-called 'free inversion' languages like Spanish and Italian, where word order is more flexible. Intransitive sentences in these languages 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*.

The grammaticality of postverbal Ss in languages like Spanish/Italian vs. English, has often been attributed to the *N(ull) S(ubject) P(arameter)* (e.g., Chomsky 1981, Rizzi 1982, Burzio 1986, Eguren & Fernández Soriano 2004, Jaeggli & Safir 1989, Luján 1999, Rizzi 1997, Zagona 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.    [IP *pro*expl     [VP llegaron **tres chicas**]].    *Spanish*
        b.    [IP *Il*          [VP est arrivé **trois filles**]].    *French*
        c.    [IP *There*     [VP arrived **three girls**]].

Unergatives, as expected, are not allowed in the construction, hence the ungrammaticality of (8).

(8)    a.    *There* has phoned ***Maria*** the president
        b.    *There* has spoken ***John***

To summarise, postverbal Ss are possible in English, yet their frequency is low and their occurrence (see Biber *et al.* 1999) is highly restricted to unaccusative Vs in *inversion* constructions and a subset of unaccusatives in *there*-constructions, while in Spanish and Italian postverbal Ss are preferred with unaccusatives in neutral, out-of-the-blue contexts where focus does not play a role.

## 2.2. *Postverbal subjects: syntax-discourse interface*

As the Unaccusative Hypothesis would predict, empirical studies show that verb choice determines word order in Spanish and Italian: in out-of-the-blue replies headed by questions like *¿Qué pasó?* 'What happened?' (i.e., those where none of the constituents in the reply is focused in particular) Spanish natives favour preverbal Ss with unergative V, yet postverbal Ss with unaccusatives (Hertel 2003, Lozano 2003, 2006) (see also Pinto 1997 for similar data for Italian), as in (10).

(9)     ¿Qué ha pasado? (Spa) / Che cosa è sucesso? (Ital)
        'What has happened?'

(10)  a.   i.    *Juan* ha hablado          ii.   #Ha hablado *Juan*
                 *Gianni* ha parlato               #Ha parlato *Gianni*
                 *Juan* has spoken                 has spoken *Juan*
      b.   i.    #*Juan* ha llegado         ii.   Ha llegado *Juan*
                 #*Gianni* è arrivato              È arrivato *Gianni*
                 *Juan* has arrived                has arrived *Juan*

It has, indeed, 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, Liceras et al. 1994, Picallo 1998, Zubizarreta 1998, 1999, Belletti 2001, 2004b, Domínguez 2004, Lozano 2006; see Pinto 1997 for Italian), as in (11) and (12) (pragmatic anomaly is shown as #). In these contexts, information structure overrides the lexical bias, i.e., focused constituents (e.g., subjects) are sentence-final, irrespective of verb type, unaccusative, unergative or transitive.

(11)  a.   ¿Quién ha llegado/hablado?            *Spanish*
           who has arrived/spoken?
      b.   Ha llegado/hablado **Juan**
           has arrived/spoken **Juan**
      c.   #**Juan** ha llegado/hablado
           Juan has arrived/spoken

(12)  a.   Chi è arrivato/ha parlato?            *Italian*
           who has arrived/spoken?
      b.   É arrivato/Ha parlato **Gianni**
           has arrived/spoken **Gianni**
      c.   #**Gianni** è arrivato/ha parlato
           **Gianni** has arrived/spoken

In English *there*-constructions and locative inversion structures are also often analysed as (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 (13), where *in the outside pocket* is relatively more familiar (topic) than the material in postverbal position (focus).

(13)   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 [PP *in the outside pocket*] [V go] [NP-SUBJECT **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)

Thus, the felicity of an inversion depends on the notion of relative discourse-familiarity: the preverbal constituent cannot present 'newer' information in the discourse than the postverbal constituent. This implies that the postverbal constituent needs not always be discourse-new, as it may represent quite familiar information when it is relatively less familiar than the preverbal constituent.

Focus and topic can be used as common labels for new vs. old information, respectively. Prince (1981, 1992) distinguishes between *hearer status* (hearer-old. vs. hearer-new) and *discourse status* (discourse-old vs. discourse-new) and argues that only the latter distinction is relevant for subjecthood. There is also a third status for an entity in discourse: *inferrables*, "NPs evoking entities which were not previously mentioned and which I as the reader had no prior knowledge of, but whose existence I could infer on the basis of some entity that was previously evoked and some belief I have about such entities." (Prince 1992: 312). In (14a) *the door* is discourse-old, as it has been mentioned before, the hearer is assumed to have a mental representation of it. In contrast *the door* in (14b) is not strictly speaking discourse-new, but it is treated as though it were already known by the hearer since *the Bastille* acts as a trigger which implies the existence of *the door*.

(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.
               (Examples from Prince 1992:305)

The question that arises is whether inferrables can be collapsed with topic or focus, whether they form a separate category, or whether we have some sort of information continuum, with inferrables somewhere in the middle. The continuum or 'gradience' approach is adopted by Kaltenböck (2005), who defines old (or given) information as 'retrievable' from the preceding context and new information as not retrievable (see Firbas (1992) and Geluykens (1991) on the concept of retrievability (or recoverability)).  Some entities are more retrievable than others, depending on whether they can be derived directly *(evoked)* from the context (as in (14a)) or indirectly via inferences *(inferrable)* (see Kaltenböck's 2005: 4.1 discussion of these issues).

This is the approach we adopt in our study.  We consider topic and focus as concepts which encompass a variety of notions which are best analysed in terms of a gradience. 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).

Finally, note that, 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 4 below).

### 2.3. *Postverbal subjects: syntax-phonology interface*

Choice of linear ordering is also influenced by properties to do with the phonetic realization of the strings of words generated by the grammar at the syntax-phonology interface: there are operations which alter the canonical word order of constituents post-syntactically (at the Phonological Form level) and which do not involve changes in meaning. An example of this is 'Heavy NP-Shift', where a heavy' NP has been 'displaced' to the end of the sentence (15b), as opposed to (15a) with canonical V-NP-PP order:

(15) a.    I bought [NP *a book written by a specialist in environmental issues*] [PP for my sister].
b.    I bought [PP for my sister] [NP *a book written by a specialist in environmental issues*].

The generalization that 'heavier' constituents should follow 'lighter' constituents is known as the 'end-weight' principle (Quirk et al. 1972). 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*, e.g., number of nodes (see Arnold et al. 2000 for a review). In fact, the two concepts are highly correlated (see Wasow 1997): 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 (Hawkins 1994). Since these 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).

The end-weight principle appears to be in operation also in VS structures. Thus in the inversion structure in (13) above, the subject is clearly 'heavy'. Our own analysis of the corpus examples used by Levin & Rappaport Hovav (1995) 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 light*er* NP, it is normally followed by material in apposition, as in (16) (highlighting is ours).

(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, 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 L2 STUDIES

Early studies found that 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, (17) and (18).

(17) 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)

(18) 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)

All these 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. 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, as demonstrated by studies on learners' preference of VS with unaccusatives but SV with unergatives (Hertel 2003, Lozano 2006a, 2006b), 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).

However, 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 since, as we will show, the postverbal S needs to be heavy (i.e., phonologically long) and focus (i.e., new information).

4. Hypotheses

Following previous theoretical and acquisitional research, we predicted $H_1$. Given that unaccusativity is a necessary but not sufficient condition for VS production, we also hypothesized $H_2$ and $H_3$, which, to our knowledge, have not been tested before in the L2 literature.

(19) **$H_1$: Lexicon (*Lexicon – Syntax interface*)**: Both Spanish and Italian learners of L2 English will produce postverbal Subjects only with unaccusatives, but never with unergatives.
**$H_2$: 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.
**$H_3$: 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).

| Corpus | Number of essays | Number of words |
|---|---|---|
| ICLE Spanish | 251 | 200,376 |
| ICLE Italian | 392 | 227,085 |
| TOTAL | 643 | 427,461 |

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 4 in the appendix), which were searched in the concordancer WordSmith Tools 4.0 (Scott 2002). In order to filter out irrelevant concordances, we employed 51 filtering criteria, though for conciseness reasons we will mention only the crucial ones, (20) (see Lozano & Mendikoetxea, forthcoming 2007 and in preparation, for further details). The filtering process resulted in 1510 usable concordances, which were analysed (Table ).

 (20) a. TheV must be intransitive (unaccusative or unergative)
    b. TheV must be finite and 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, recall that 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.

## 6. RESULTS AND DISCUSSION

### 6.1. *Results for $H_1$: 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 2), 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 2007 and in preparation).
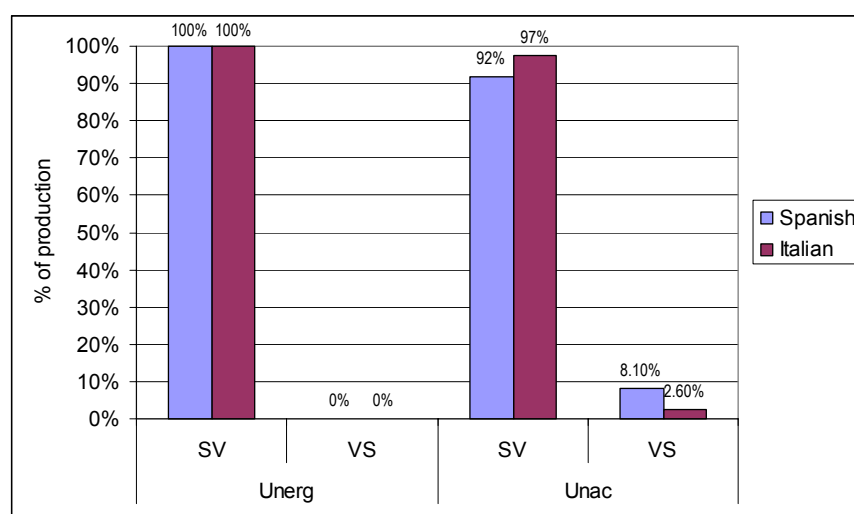


Figure 1. *Percentage of subjects produced (group x verb type)*

| 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 2. *Postverbal subjects produced*

The production of unaccusative VS in the two groups contained both grammatical and ungrammatical constructions, though 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, 2007 and in preparation, for further details), as in (21)-(26), 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).

(21) 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)

(22) Grammatical locative inversion:
a. *In the main plot* appear **the main characters: Volpone and Mosca** … (spal1002)
b. Cesare Lombroso (1835/1909) criminologal, asserted that *on the earth* lived **people which were born-criminal.** (itrl1005)

(23) 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)

(24) Ungrammatical Ø-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)

(25) *AdvP* insertion:
a. …, and *here* emerges **the problem**. (spm01001)
b. *Later* came **a world of disorder**, during and after the First World War … (itrs1010)

(26) 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 $H_1$ predicts and as shown in previous L2 studies.

## 6.2. *Results for $H_2$: 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 (27a,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 (28a',b').
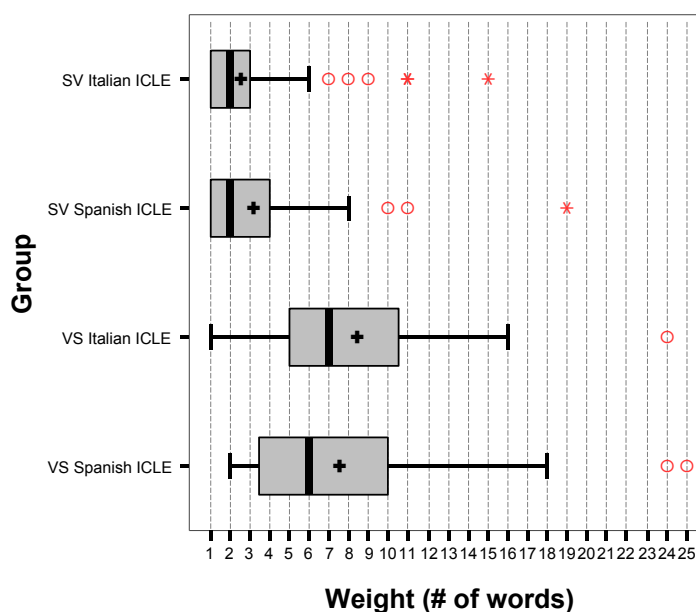


Figure 2. *Boxplot (with median and mean) of subject weight in number of words*

(27)  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)

(28)  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)

107

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 $H_2$.

### 6.3. *Results for $H_3$: 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 3). 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 3. *Production of subjects (topic vs. focus) with unaccusatives*
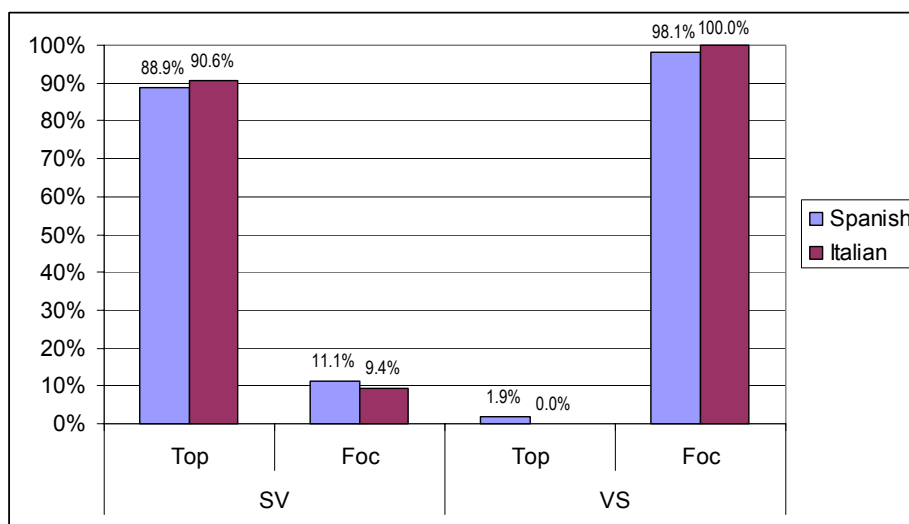


Figure 3. *Production of subjects (topic vs. focus) with unaccusatives*

Example (29) illustrates unaccusative postverbal Ss being new information (focus), i.e., the S has not been mentioned previously in the discourse. By contrast, (30) illustrates unaccusative preverbal Ss which are topics (shown in *italics*) since they have been mentioned in the prior discourse (shown in underlined typeface).

(29) a.    In the world, dominated by science, technology and industrialisation, there is no a place for dreaming and imagination. Thanks to science and its consecuences, 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)

(30) a.    The approval of acting of women were something essential. Women started to perform    <u>female characters</u> 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 <u>these differences</u>, 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 <u>different</u> 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 summaries, 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$_3$.


7. 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).


NOTES

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: Hulk, A.C. and Pollock, J.-Y. (eds), *Subject Inversion in Romance and the Theory of Universal Grammar*. Oxford: Oxford University Press, 60-90.

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. 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: 233-259.

Birner, B. (1995), Pragmatic constraints on the verb in English inversion, *Lingua* 97: 223-256.

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.

Chocano, G., C. Lozano, R. Jiménez, A. Mendikoetxea and S. Murcia (in preparation), 'Retrievability, recency and the Topic-Focus distinction: a study of subjects', (ms.), Universidad Autónoma de Madrid.

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: 113-151.

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,* Berlin: Mouton de Gruyter, 141-167.

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 B. and M.Johansson M. (eds), *Languages in Contrast. Text-based cross-linguistic studies. Lund Studies in English 88.* Lund: Lund University Press. 37-51.

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. Amsterdam & Philadelphia: Benjamins. 3-33.

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: Kanno, K. (ed.), *The Acquisition of Japanese as a Second Language.* 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): 119-159.

Kennedy, G. (1998), *An Introduction to Corpus Linguistics.*  London & New York: Longman.

Koopman, H. & D. Sportiche (1991), 'The position of subjects', *Lingua* 85: 211-258.

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.*  University of Essex: Unpublished PhD dissertation.

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 2007), 'Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: a corpus study', in Gilquin, G., B. Díez and S. Papp (eds). *Linking up Contrastive and Learner Corpus Research.* 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*, Madrid: Espasa-Calpe. 1275-1315.

Montrul, S. (2004), Psycholinguistic evidence for split intransitivity in Spanish second language acquisition. *Applied Psycholinguistics* 25: 239-267.

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.  The Japan Second Language Association

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: 157-189.

Picallo, C. (1998), 'On the Extended Projection Principle and null expletive subjects', *Probus,* 10: 219-241.

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*, London: Academic Press. 223-255.

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*, Amsterdam: John Benjamins. 295-325.

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. Dordrecht: Kluwer Academic Publishers.

Rizzi, L. (2004), 'Locality and left periphery',  in: A. Belletti (ed.). 223-251.

Rochemont, 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*, Cambridge: Cambridge University Press. 163-182.

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.*  Amsterdam: John Benjamins, 153-175.

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: 347-361.

Wasow, T. & Arnold, J. (2003). 'Postverbal constituent orderin in English'. In Rohdenburg, G. and B. Mondorf (eds). *Determinants of Grammatical Variation*. 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*. Lisbon: Associaçao Portuguesa de Linguistica. 289-296.

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*, Cambridge: Cambridge University Press. 203-221.

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.

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

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

Table 4. *Inventory of searchable lemmas*