Postverbal subjects at the interfaces in Spanish and Italian learners of L2 English: a corpus analysis

Cristóbal Lozano
Universidad de Granada

Amaya Mendikoetxea
Universidad Autónoma de Madrid

Abstract

The purpose of this article is to characterise the production of postverbal subjects in two ICLE subcorpora (Italian and Spanish). The question has been dealt with before in the literature with emphasis on the production of ungrammatical inversion structures in L2 English of speakers from a variety of L1s, but in quite a scattered, unsystematic and rather intuitive fashion. Our approach seeks to identify the conditions under which learners produce inverted subjects. Based on previous research findings and our review of the theoretical literature, we hypothesise that for Spanish and Italian learners of L2 English, there is a tendency for subject inversion to occur when: the verb is unaccusative (H1), the subject is long or "heavy" (H2), and the subject is new (or relatively new) information or "focus" (H3). While H1 has found confirmation in the L2 literature, H2 and H3 have, to our knowledge, been untested and the facts they describe gone unnoticed in previous research. Our results show that the three conditions are met in the writing of Spanish and Italian L2 speakers of English, despite errors in the syntactic encoding of the structures concerned. Thus, a full account of the production of inverted subjects in L2 English must look at properties which operate at (i) the lexicon-syntax interface, (ii) the syntax-phonology interface, and (iii) the syntax-discourse interface.

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 two relevant ICLE subcorpora (*International Corpus of Learner English*; Granger et al. 2002). Our aim 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.[1]

This study is part of a larger project which seeks to gain an insight into the grammatical processes underlying word order in non-native grammars (L2), as compared to both the mother tongue (L1) and the target language. Thus, we incorporate some of the fundamental tenets of what is known as *Contrastive Interlanguage Analysis* (CIA) (see e.g. Granger 1996 and Gilquin 2000/2001), which establishes comparisons between: (a) native and non-native speakers (NS vs. NNS), and (b) different non-native speakers (NNS vs. NNS). The former involves 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. This includes errors, but it is conceptually wider as it seeks to identify overuse and underuse of certain linguistic features and patterns. As for (b), by comparing learner data from different L1 backgrounds, we can gain a better understanding of interlanguage processes and features, such as those which are the result of transfer or those which are developmental, common to learners with different L1s (see Granger 2002: 12-13 and references cited therein). Here, we will be comparing the production of inverted subjects in the writing of advanced Italian and Spanish learners of L2 English, as in (b); NS vs. NNS comparison is addressed in Lozano & Mendikoetxea (in preparation).

We first examine in detail the properties of VS order in English vs. Spanish/Italian as presented in the current theoretical literature (Section 2). In Section 3, we look critically at 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. The statistical analysis of the data is presented in Section 6, where we compare (i) production of postverbal Ss both with unaccusative and unergative Vs (6.1), and (ii) the nature of both postverbal and preverbal Ss in terms of "heaviness" (roughly, number of words) (6.2) and discourse status (topic or focus) (6.3). Results are discussed in Section 7, where it is shown that our data confirm previous findings in that postverbal Ss are only found with unaccusative Vs (with some significant differences between L1 Italian and Spanish speakers) and, more interestingly, they show that inversion is the result of two other factors not previously discussed in the L2 literature: S must be syntactically heavy, as well as display new information or focus. Section 8 is the conclusion, where we discuss suggestions for further research.

2.   VS order in English vs. Romance languages

English has been characterised as a "fixed" word order language, as opposed to Italian and Spanish where word order is said to be "free". In English the occurrence of VS order is highly restricted. It is found mostly in two types of constructions: *there*-constructions and so-called "inversion" constructions of the type XP-V-S, where (i) XP is an adverbial element, typically expressing time or place and linking the sentence to the prior discourse (in italics in [1]), (ii) V is an intransitive verb, typically expressing existence or appearance on the scene, and (iii) S is often syntactically/phonologically "heavy", consisting of a noun and a variety of pre- and/or postmodifiers, which introduce new information in the discourse (in bold in [1]). These properties are illustrated in (1) – a corpus example from Birner (1994: 254).

(1)   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)

Given the interplay of factors involved, 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 V 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. We deal with each of these issues in turn.

2.1 Postverbal Ss in English and Spanish/Italian at the lexicon-syntax interface: the Unaccusative Hypothesis

2.1.1 Postverbal Ss and the Null Subject Parameter

Both Modern English and Spanish/Italian show canonical SV(O(bject)) order as illustrated in (2), but while in English this order is "fixed", Spanish and Italian allow combinations like those in (3), where S appears in postverbal[2] position (VSO) without a special intonation pattern.[3]

(2) a. *Maria* has phoned the president.
     b. *María* ha telefoneado al presidente.   *Spanish*
     c. *Maria* ha telefonato al presidente.   *Italian*

(3) a. *Has phoned *Maria* the president.
     b. Ha telefoneado *María* al presidente.   *Spanish*
     c. Ha telefonato *Maria* al presidente.   *Italian*

Intransitive sentences in Spanish and Italian may also allow postverbal subjects as in (4), while the corresponding sentences in English are ungrammatical (5).

(4) a. i. Ha hablado *Juan*.   ii. Ha llegado *Juan*.   *Spanish*
     b. i. Ha parlato *Gianni*. ii. E' arrivato *Gianni*. *Italian*

(5) 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)* (see inter alia Chomsky 1981, Rizzi 1982, Burzio 1986, Fernández Soriano 1989, Jaeggli & Safir 1989, and more recently, Rizzi 1997, Luján 1999, Zagona 2002 and Eguren & Fernández Soriano 2004). "Free inversion" is among the cluster of properties

that distinguish languages that allow null Ss (i.e. are positively marked for the NSP), like Spanish and Italian (6), from languages which do not allow null Ss, like English (7).[4]

(6)   a.i. Ha hablado.   ii. Ha llegado.   *Spanish*
        b.i. Ha parlato.   ii. E' arrivato.   *Italian*
            '(He/She) has spoken.'   '(He/She) has arrived.'

(7)   i. *Has spoken.   ii. *Has arrived.

In the generative grammar literature, the possibility of having null Ss has been associated with "rich" (verb) agreement features (e.g. Rizzi 1982), which allows for the recovery of the "content" of a phonologically null pronominal element (*pro*) in the subject position in (6), while languages which do not allow null Ss have "poor" V agreement and lack *pro* as part of their lexical inventory (e.g. French [see note 4] and English).[5] As for sentences with postverbal Ss, a null element is also postulated for the preverbal subject position of structures like those in (3b, c), (4) and (8a), which we take to be the specifier of the Inflectional Phrase, <Spec, IP> (cf. note 2): *expletive pro* (pro_{expl}). This is the null equivalent of the overt expletives in languages negatively marked for the NSP, such as French *il* (8b) or English *there* (8c).

(8)   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**]].

Like its non-overt counterpart in Italian/Spanish, expletive *there* is involved in the licensing of the postverbal S in English: *pro$_{expl}$* and *there* enter a coindexing relation with the postverbal NP subject, which involves sharing of Case as well as (Person and Number) agreement features between the expletive and the associate NP (not for French, though, as *arrivé* in [8b] is singular).[6] Notice, however, that the presence of *there* in English does not render the structure in (3a) with a transitive V grammatical (as opposed to [3b, c] with *pro$_{expl}$*), as shown in (9a), and only certain intransitive Vs allow *there*-insertion: (9b), which is the result of adding *there* to (5i) with *speak*, is ungrammatical, as opposed to (8c) with *arrive*.
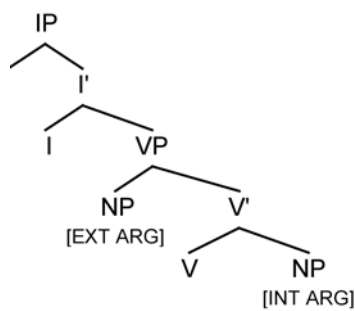
(9)   a. **There* has phoned **Maria** the president.
        b. **There* has spoken **John**.

This is so because *there*-constructions show certain constraints that are not found in constructions with *pro$_{expl}$* in Spanish and Italian. For a proper account of the distribution of *there* we have to refer to what is known as the *Unaccusative Hypothesis*.
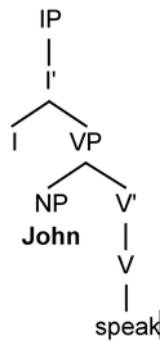
2.1.2   Postverbal Ss and the Unaccusative Hypothesis

Since Perlmutter (1978) first formulated the *Unaccusative Hypothesis*, it is commonly assumed in the generative grammar literature that there are two classes of intransitive Vs, which differ in the position occupied by their only argument: *unergative* Vs have an external argument but no internal argument, while *unaccusative* Vs have an internal argument, but no external argument. The internal argument occupies the position of complement of V, while the external argument is in the preverbal subject position (<Spec, IP>) (see Chomsky 1981, Burzio 1986). More recently, after the introduction of the *VP-internal subject hypothesis* (see Koopman & Sportiche 1991), by which all arguments are generated internal to the VP, external arguments are generated in the position of the specifier of the VP (<Spec, VP>), as in (10), and subsequently move to <Spec, IP>.[7] Thus, unergative Vs appear in initial structures (D[eep]-Structures) like (11a), while unaccusatives appear in initial structures like (11b).
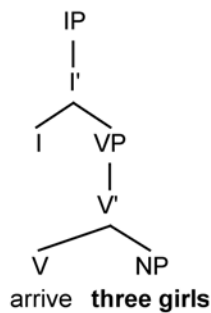
(10)

```
            IP
           /
          I'
         /  \
        I    VP
            /  \
          NP    V'
       [EXT ARG] / \
              V    NP
                 [INT ARG]
```
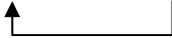
(11)   a. *unergative*   b. *unaccusative*   (D-Structure)
        'John spoke'   'Three girls arrived'

```
         IP                              IP
         |                               |
         I'                              I'
        /  \                            /  \
       I    VP                         I    VP
           /  \                             |
         NP    V'                           V'
        John   |                           /  \
               V                          V    NP
               |
             speak                     arrive  three girls
```

The syntactic distinction between the two classes of Vs is systematically related to a semantic distinction: roughly, unergatives typically denote activities controlled by an agent (*speak* in [11a] and also *cry*, *cough*, *sweat*, *jump*, *run*, *dance*, *work*, *play*, etc), while unaccusatives have themes (or patients) as their only argument (*arrive* in [11b] and also *blossom*, *appear, exist, deteriorate*, *arrive*, *come*, etc).

In the course of the derivation, that is, in the mapping from D-Structure to S(urface) Structure, the NPs in (11) move to <Spec, IP> in order 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; see note 6), as in (12).

(12)   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 for external arguments like *John*, as in (12a), the internal argument of an unaccusative V may, however, remain in its initial position under certain conditions, as in *there*-constructions like (13) and inversion constructions like (1), now reproduced as (14), with *there* and PP in <Spec, IP>, respectively.

(13)   [$_{IP}$ [$_{NP}$ *There*] [$_{VP}$ [$_V$ arrived] [$_{NP}$ **three girls**]]]

(14)   [$_{IP}$ [$_{PP}$ *in the outside pocket*] [$_{VP}$ [$_V$ go] [$_{NP}$ **his schedule book, chap stick, gum, contact lens solution and hair brush**]]]

Characteristically, inversion constructions involve an opening XP adverbial in <Spec, IP>: typically a locative element as in (14) and (15a) below, which are examples of *locative inversion,* but also time adverbials, as in (15b), as well as other types like the *with*-PP in (15c).[8]

 (15)   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)

There is a considerable overlap between Vs appearing in *there*-constructions and typical locative inversion structures like (14) and (15a). For both types of constructions there is the requirement that the V be unaccusative, but this is not by itself sufficient to characterise the class of Vs appearing in these

structures.[9] Two main semantic classes of unaccusative Vs have been distinguished in approaches which seek to characterise unaccusativity at the lexicon-syntax interface in order to establish a relation between the semantic properties of the predicate and the syntactic properties of the constructions they enter: (i) Vs of change of state and (ii) Vs of existence and appearance (Levin & Rappaport Hovav 1995). Only the latter are found in this construction. That is, together with *be*, we find Vs such as *appear*, *arise*, *ascend*, *emerge*, *exist*, *come* and so on, but not unaccusative Vs of change of state like those in (16):

(16)   a. *There opened **the door**.   (= The door opened)
       b. *There broke **a window**.   (= A window broke)

It was with these considerations in mind that the Vs for this study were carefully chosen (see Section 5.2 and Table 2). In contrast, as we have seen, null expletives in Spanish and Italian may be found wherever we have a postverbal S, regardless of whether V is unaccusative or unergative. Their occurrence is not determined by properties operating at the lexicon-syntax interface. Rather, Romance "free" inversion appears to be governed by features that operate at the syntax-discourse interface, at least in informationally non-neutral contexts. This could, in principle, lead to the expectation that Spanish and Italian learners of English may overgeneralise and produce postverbal subjects with all verb classes. However, we will see that there are reasons to predict that VS structures in the NNS of our learners are found with exactly the same subset of Vs as in English NS. This issue will be addressed in our discussion of previous L2 research below (Section 3).

2.2   Postverbal Ss in English and Spanish at the syntax-discourse interface: inversion as focalisation

2.2.1   Inversion as focalisation

It has been argued recently that information structure notions, such as topic and focus, appear to play a crucial role in the position of S in free-order languages, with preverbal Ss usually analysed as topics or given information, and postverbal Ss as (presentational/informational) focus or new information (see, among others, Vallduví 1990, Fernández-Soriano 1993, Liceras et al. 1994, Picallo 1998, Zubizarreta 1998, 1999, Belletti 2001, 2004b, Domínguez 2004, Lozano 2006). The following question-answer pairs from Italian and Spanish clearly indicate that the postverbal S can be interpreted as (non-contrastive) focus, i.e. new information. When S is preverbal, it is interpreted as old information (topic), which explains why the (c) examples in (17) and (18) are pragmatically-odd (marked as #).[10]

(17)   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

(18)   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

Additionally, Belletti (2004b) points out that in Italian, (18b) is also an appropriate answer to neutral, out-of-the-blue questions like (19), i.e. when the whole clause is new information, as opposed to the sentence with the preverbal S in (18c) (examples from Belletti 2004b: 22).

(19)   Che cosa è sucesso?
        What has happened?

Empirical studies on Spanish native speakers' acceptability preferences show, however, that verb choice may determine whether the answer to questions like (19) may contain either a preverbal or a postverbal S. In particular, preverbal Ss are favoured for unergative Vs and postverbal Ss in unaccusative contexts (Hertel 2003, Lozano 2003, 2006) (see also Pinto 1997 for similar observations for Italian), as in (20).

(20)   a. i. *Juan* ha hablado.   ii. #Ha hablado *Juan*.
                Juan has spoken   has spoken *Juan*
        b. i. #*Juan* ha llegado.   ii. Ha llegado *Juan*.
                Juan has arrived   has arrived *Juan*

Though this issue requires further analysis, it is largely irrelevant for the present study as our data is extracted from written pieces of work in which we are not likely to find neutral or out-of-the-blue contexts (even the initial sentence is contextualised by the title of the essay). Thus, independently of whether the preverbal S can also be focus or new information in Romance Null S languages, what is crucial for our purpose is that in these languages inversion serves the purpose of focalising the subject. Notice also that in languages like Italian, Spanish and English, new information in clause initial position can be interpreted as contrastive focus (see Rizzi 1997), as shown in (21) in capitals.

(21)   a. IL TUO LIBRO ho comprato (non il suo).
        b. TU LIBRO he comprado (no el suyo).
        c. YOUR BOOK I have bought (not his).

From this, it follows that preverbal subjects are not necessarily topics or old information, but may also be analysed as contrastive focus. Non-contrastive (i.e. informational or presentational) focus, of the type we are looking at here, occurs postverbally (see also note 10).

Let us now turn to English, which makes a much more restricted use of inversion mechanisms, as we have seen. Postverbal Ss in English *there*-constructions and locative inversion structures are also often analysed as (presentational) focus (see, among others, Bolinger 1977, Rochemont 1986, Bresnan 1994). In Bresnan's (1994) analysis of locative inversion, the referent of the postverbal NP is introduced, or reintroduced, on the scene referred to by the preverbal PP: for instance in (15a) *on one long wall* provides the scene onto which *a row of Van Goghs* is introduced, which is characterised as a new discourse entity. A different approach is adopted by Birner (1994, 1995), who has argued on the basis of a large corpus of naturally-occurring tokens 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). Thus in (14), *in the outside pocket* is relatively more familiar than the material in postverbal position, as in the preceding context the speaker has mentioned the different pockets in which things are placed (see [1] above). The felicity of an inversion depends on the notion of relative discourse-familiarity: the preverbal constituent may not present "newer" information in the discourse than the postverbal constituent. This implies that the postverbal constituent need not always be discourse-new; it may, on the contrary, represent quite familiar information, given the appropriate discourse conditions, i.e. as long as it is relatively less familiar than the preverbal constituent. For Birner (1995), this cannot be accounted for by an analysis in which the postverbal NP is characterised as presentational focus and, therefore, discourse-new.

Given that both in English and in Italian/Spanish, inversion is used as some sort of focalisation device, we would expect the inverted Ss in the NNS of our study 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, whereas in NS English inversion appears to be restricted mostly to unaccusative Vs of existence and appearance, which in principle could lead to the expectation that our learners will use postverbal subjects in syntactic contexts in which it is not found in native English, as pointed out at the end of Section 2.1. We will return to this issue in Section 3 below.

2.2.2   Brief remarks on the definition of focus and topic

Focus and topic have been used here as common labels for new and old information, respectively, but these labels are just a first approximation to name

concepts and processes which may be adequate to express the relation between information structure and syntactic configurations (see Belletti 2004b: 42, note 1, for a similar comment). Regarding the old vs. new information dichotomy, 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*, defined by Prince (1992: 312) as "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". Compare the following examples used by Prince (1992: 305) to characterise inferrables:

(22)   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.

In (22a) *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 (22b) is not strictly speaking discourse-old, but it is treated as though it were already known to the hearer. Indeed, the hearer knows that the speaker is talking about the door of the Bastille, so it is not quite discourse-new. As Prince (1992) notices, inferrables have a double status as hearer-old (and discourse-old) entities, in that they rely on the earlier presence of another entity (e.g. *the Bastille*) triggering the inference, and as hearer-new (and discourse-new), in that the hearer is not expected to already have in his or her head the entity concerned. The question that arises is whether they can be collapsed with one or the other category, 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) in his corpus analysis of *it*-extraposition structures in English. He defines old (or given) information as "retrievable" from the preceding co(n)text and new information as not thus derivable (see Geluykens 1991 and Firbas 1992 on the concept of retrievability – or recoverability). Retrievability is a relative concept: some entities are more retrievable than others, depending on whether they can be derived directly (*evoked*) from the context (as in [22a]) or indirectly via inferences (*inferrable*), where context comprises both the verbal "co-text", as well as the situational context. This approach looks at the finished product of verbal interaction, the corpus text, rather than focusing on speakers' assumptions (referred to as the "psychologist's view") (see Kaltenböck 2005: Section 4.1).

In our study, both evoked and inferrable entities are considered to be topics, on the basis of Prince's (1992) analysis of Ss and Birner's (1994, 1995) findings that both entities are treated alike (as discourse-old) in inversion structures. As for new or "irretrievable" information, it can either be *brand-new* or *new-anchored* (an irretrievable state of affairs or entity, which is in some way linked to – "anchored in" – the previous context). We consider these as focus, together with what Prince (1992) refers to as *unused* (discourse-new, hearer-old

entities). An example of an *unused* entity is given in (23a) (from Prince 1992), where *the public* is discourse-new but the reader is expected to know of its existence. As for *new-anchored*, notice that the postverbal subject in the corpus example in (23b) is new in the sense that the idea it introduces has not been mentioned before, but it is anchored in that *this work* refers to what is mentioned in the preceding context.

(23)   a. … from reporters eager to tell the public about … (from Prince 1992: 312)
        b. So should scientists thus be condemned for allowing it to become possible for these women to have babies? Certainly there exists *a demand for this work to be done*. (from the LOCNESS corpus, file ICLE-ALEV-0014.8)

We thus consider topic and focus as concepts which encompass a variety of notions which are best analysed in terms of a gradience.[11, 12]

2.3   Postverbal Ss at the syntax-phonology interface

Together with discourse notions such as "newness" or "focus", choice of ordering is influenced by properties to do with the phonetic realisation of the strings of words generated by the grammar. Thus, at the syntax-phonology interface, there can be operations affecting the linear ordering of constituents which are not triggered by syntactic features. This is presumably what happens in structures involving "Heavy NP-Shift", where a "heavy" NP appears to have been "displaced" to the end of the sentence.[13] Thus, while (24a) shows canonical V-NP-PP, the non-canonical V-PP-NP order in (24b) is also possible with a "heavy" NP.

(24)   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 generalisation that "heavier" constituents should follow "lighter" constituents is essentially what 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* (see Arnold et al. 2000 for a review of these two approaches). In fact, the two concepts are difficult to separate neatly, as revealed by Wasow's (1997) corpus study, which shows high correlations among the various characterisations of heaviness. In sum, 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 (see Hawkins 1994). 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 (focalisation) appear to reinforce each other (see Biber et al. 1999: 898). As pointed out by Arnold et al. (2000: 34) "items that are new to the discourse tend to be complex, and items that are given tend to be simple".

The end-weight principle appears to be in operation also in structures involving postverbal Ss. Culicover & Levine (2001) claim that certain structures which appear to be the result of locative inversion become a good deal less natural with "lighter NPs", as shown in (25) (highlighting is ours), and analyse such structures as involving Heavy-NP shift.

(25)   a. *Into the room* walked carefully **Robin**.
        b. *Into the room* walked carefully **the students in the class who had heard about the social psych experiment that we were about to perpetrate**.
        (from Culicover & Levine 2001: 291)

Our own analysis of the corpus examples used by Levin & Rappaport Hovav (1995) (L&RH) in their study of locative inversion reveals, indeed, that overwhelmingly the postverbal S is heavy. When it is a proper noun or a relatively light NP, it is normally followed by material in apposition, as in (26) (highlighting is ours).

(26)   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 an NP, the more likely it is to be placed in clause-final position. This approach has the additional advantage that it can be used to measure "relative" weight, which has been shown to be relevant in studies involving two constituents which may appear in different orders, such as the direct and indirect object of ditransitive Vs (Arnold et al. 2000).

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 its general purpose appears to be related to easing the processing burden on the receiver, by placing long, complex (and new) elements towards the end of the clause, we will assume that this is a universal processing principle (Frazier 2004). Thus (27a), which shows canonical word order in Spanish with the adjunct PP following the NP complement, appears to be less "natural" than (27b), where the heavy object follows the PP adjunct.

(27)   a. Vi [_{NP} *los libros que tanto gustaban al padre de tu vecino*] [_{PP} en la librería].

I saw [_{NP} *the books that your neighbour's father liked so much*] [_{PP} in the bookshop]

b. Vi [_{PP} en la librería] [_{NP} *los libros que tanto gustaban al padre de tu vecino*].

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 subjects which are long and complex, as a reflection 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.[14]

3.   Previous L2 findings

Little is known about the production of postverbal Ss in L2 English with intransitive Vs, apart from the fact that they are found with unaccusative Vs. White's (1985, 1986) studies on subject-verb inversion, as a feature of the NSP, which are based on grammaticality judgements, are rather inconclusive, due mostly to verb choice, as pointed out by Oshita (2004: 106). Two production studies, Zobl (1989) and Rutherford (1989), support the hypothesis that learners with Spanish L1 produce VS only with unaccusatives, but are rather small in scale. Zobl (1989) reports from a small corpus that learners of L2 English with Null S L1s (mostly Japanese, Spanish and Arabic) occasionally produce postverbal Ss with unaccusatives, as shown in (28) (highlighting is ours).

(28)   a. I was just patient until dried **my clothes**. (L1 Japanese)

b. *Sometimes* comes **a good regular wave**. (L1 Japanese)

(Zobl 1989: 204)

Similarly, Rutherford (1989) reports that L1 Spanish and Arabic speakers produce postverbal Ss with unaccusatives in a written corpus of L2 English, as in (29) (highlighting is ours), but never with unergatives.

(29)   a. *In the town* lived **a small Indian**. (L1 Spanish)
b. *On this particular place called G…* happened ***a story which now appears on all Mexican history books*…**. (L1 Spanish)
c. After that, they'll be lead to their house, and *with that* comes **the end of the wedding**. (L1 Arabic)
d. The bride was very attractive, *on her face* appeared ***those two red cheeks*…** (L1 Arabic)
(Rutherford 1989: 178-179)

Zobl (1989) and Rutherford (1989), however, differ in their explanation of why VS order is found with unaccusative verbs. For Zobl (1989), this word order results from the syntactic S of an unaccusative sentence being actually the logical or semantic object of V or the *theme*,[15] what we have been referring to as its internal argument (see Section 2.1.2). This word order preserves at S-Structure the position occupied by the internal argument at D-Structure (see [12b] above) and therefore we do not expect unergatives to appear in VS structures since their only argument is an external argument (see [12a]). VS structures are just one type of solution to the problem of how learners map logical relations to surface structure configurations in languages like English, where Ss and Os are defined in terms of their structural position (see examples [1]-[6] in Zobl [1989: 204] for other solutions learners come up with). For Zobl (1989), the production of VS structures is developmental and precedes a stage in which learners are able to determine the canonical alignment between semantic roles and syntactic structure.

Rutherford (1989), however, argues that VS order in the learner English of Spanish and Arabic L1 speakers is the direct result of transfer. His main hypothesis is that "the tendency for canonical word order permutation in written ILs [interlanguages] will correlate directly with the propensity of the learner's native language to permute its own canonical constituents" (Rutherford 1989: 166). Crucially, Rutherford's prediction is that not all word order differences will lead to transfer. Transfer is expected when the learners' L1 has what Rutherford (1989) refers to as "Pragmatic Word Order" (PWO), but not when L1 and L2 show different "Grammatical Word Orders" (GWO). Thus, Spanish and Italian are basically SVO languages, like English, and, as was discussed in Section 2 above, variations on this word order are mostly due to properties operating at the syntax-discourse interface, showing what Rutherford refers to as PWO. On the other hand, Japanese and English differ in their canonical word order: SOV vs. SVO. While transfer is expected in the L2 English of Italian and Spanish (as well as Arabic) L1 speakers, this is not the case for Japanese L1 speakers. Rutherford's predictions are borne out by his results. While no violations of word order were found in the written production of 21 Japanese L1 learners of English (but see Zobl 1989), the 59 compositions produced by Spanish L1 learners of English contained 20 instances of violations of English canonical word order and in virtually all cases the resulting order was (X)VS (see Rutherford 1989: 168). For Rutherford (1989), this is the result of transfer from PWO L1s. However, he offers no explanation as to why (X)VS order in the IL of Spanish (and Arabic) L1

speakers is restricted to a definable class of lexical verbs: those which we have referred to as unaccusative verbs of existence and appearance, following Levin & Rappaport Hovav (1995).

It has to be mentioned that both Zobl's (1989) and Rutherford's (1989) conclusions on VS production by Spanish L1 learners of English are based on a relatively small number of VS instances. This is roughly 20 instances in Rutherford (1989). As for Zobl's (1989) study, 13 examples of VS order are found in the written work of 114 informants from different language backgrounds, of which 90 are Japanese and 10 Spanish L1 speakers; 10 of those examples were produced by Japanese speakers, which is used by Zobl (1989) to rule out transfer, contrary to Rutherford's (1989) results. Though these studies point out that the unergative-unaccusative distinction is psychologically real for learners, not enough data is provided to support their claims and very little information is provided about the learners, sample size and so on.

Oshita's (2004) is the first study to use a large electronic corpus: *Longman Learners' Corpus* (version 1.1, March 1993, LLC). His purpose is to investigate the psychological reality of null expletives, for which he extracts 941 token sentences on 10 common unaccusative verbs and 640 token sentences with 10 common unergative verbs from compositions written by speakers of Italian (684), Spanish (1,079), Japanese (1,363) and Korean (236). For VS order, he extracts sentences with preverbal overt expletives (i.e. *it*, *there*) and null expletives. L1 Spanish and Italian learners produced postverbal Ss only with unaccusatives (never with unergatives), as in (30) (highlighting is ours). Both groups' production ratios were similar: 14/238 (6%) Spanish; 14/346 (4%) Italian.

(30)   a. …*it* will happen **something exciting**... (L1 Spanish)
   b. …because *in our century* have appeared ***the car and the plane***… (L1 Spanish)
   c. …*it* happened ***a tragic event***… (L1 Italian)
   d. *One day* happened **a revolution**. (L1 Italian)
   (Oshita 2004: 119-120)

Additionally, in a study of existential *there*-V-S constructions, Palacios-Martínez & Martínez-Insua (2006) report that Spanish learners of English (ICLE corpus) use the verb *be*, except for 5 tokens with the unaccusative *exist*, which represents 1.1% of all existential *there*-constructions. While this study does not address the issue of unaccusativity and postverbal Ss, it reveals that the only verb, other than *be,* that appears in these postverbal S structures is an unaccusative verb.

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 L2 acquisition, as demonstrated by studies on learners' 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). This is despite the fact that English lacks overt marking for unaccusatives, rendering the unergative-unaccusative distinction inaccessible: unaccusative Vs overwhelmingly appear in SV constructions and there is no auxiliary or morphological marker for unaccusative Vs. That is, although inversion structures with unaccusatives are found in English, as discussed at length in Section 2 above, the rarity of the construction makes it unlikely that such VS order is sufficiently represented in the input data to learners to count as positive evidence (see, by way of illustration, the percentages given in Biber et al. [1999: 945] for existential clauses with Vs other than *be*).

Without entering the debate of whether VS order is developmental or the result of transfer, our study should be considered as part of the growing body of research which takes 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 (see Section 4 below). There are two aspects in which our study differs considerably from previous L2 studies: (i) we intend to show that unaccusativity is a necessary but not sufficient condition for VS order to occur, as it is not just the nature of V (unaccusativity), but crucially the features of S (heaviness and focus) that trigger inversion, and (ii) unlike previous L2 studies, we do not focus on errors, but rather on the conditions under which postverbal subjects are found.

4.   Hypotheses

As a general hypothesis, we expect Spanish and Italian advanced learners of L2 English to produce postverbal subjects in the same contexts in which inversion takes place in native English, given that, except for the condition on unaccusativity, the native English contexts are substantially the same as those for Spanish and Italian, as we have seen. We do not expect significant differences between the groups of Spanish and Italian learners. In particular, in accordance with previous L2 findings and the theoretical analysis in Section 2, we can hypothesise the following, (31-33).

(31)   H1. *Lexicon-Syntax interface*: As reported in previous research, both
          groups of learners will produce postverbal Ss with unaccusatives only, but
          never with unergatives, as in L1 English.
(32)   H2. *Syntax-Phonology interface*: Given our theoretical analysis, we
          predict that, in those contexts where inversion is allowed, both groups of
          learners will tend to place Ss in postverbal position when S is heavy.
(33)   H3. *Syntax-Discourse interface*: Given our theoretical analysis, we predict
          that, in those contexts where inversion is allowed, both groups of learners
          will tend to place Ss in postverbal position when S is the focus.

5.   Method

5.1   Corpora

We used the Spanish and Italian subcorpora of the *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, nearly half a million words were used in our analyses (Table 1).

Table 1. Corpora

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

5.2   Analysis of concordances

Following Levin (1993) and Levin & Rappaport Hovav (1995), we constructed an inventory of unaccusative (n = 32) and unergative (n = 41) lemmas in English (Table 2). The unaccusative Vs selected belong to those semantic classes that can be found in inversion structures in English (see Section 2.1.2).

Searches were then performed with WordSmith Tools 4.0 (Scott 2002) on all verbal forms of the lemmas. 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, appeard, apear, apears, apearing, apeared, apeard.* Additionally, the sentences under analysis were both grammatical and ungrammatical in native English since, as stated previously, we are not interested in ungrammaticality but rather in the conditions under which postverbal Ss appear.

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 three quarters of the concordances output by the software turned out to be unusable since they did not meet the filtering criteria. For conciseness, we present here only the main filtering criteria, cf. (34) (see Lozano & Mendikoetxea in preparation for details on the filtering criteria).

(34)   a. *V must be intransitive (unaccusative or unergative, with either preverbal or postverbal S)*. We discarded (un)grammatical uses of transitive unaccusatives (e.g. *parents grew their children*) and *let-*constructions (e.g. *let him develop*).

b. *V must be finite*. We discarded instances of *to*-infinitive clauses (e.g. *He would like to leave*), infinitival clauses (e.g. *to avoid this happen again*), gerundive clauses (e.g. *Returning to the title of this paper*).

c. *V must be in the active voice with(out) auxiliary or modal*. We discarded case of passivised unaccusatives (e.g. *This situation has already been happened*).

d. *S must be a Noun Phrase*. We discarded cases of the NP being a clause, as in extraposition (e.g. *it happened that the countries which make the weapons are…*).

Table 2. Inventory of unaccusative and unergative verbs

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

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

Consider now the analysis of weight for the (pre- and postverbal) S of the concordances. Length in number of words has been typically used to measure heaviness (see Section 2.3). This measuring method is legitimate, yet it is unclear in previous studies where the dividing line between heavy/light lies, i.e. it is not known how long (in words) a constituent must be for it to be classified as heavy. Hence, we used three types of measurement to arrive at a satisfactory nominal heavy/light scale (Table 3): (i) a *numeric scale* (length in words, as in earlier research), (ii) an *ordinal scale* (ranging from 0 [lightest] to 3 [heaviest]) based on syntactic structure, as shown in the table, and (iii) a dichotomous *nominal scale* (light vs. heavy) based on the ordinal scale where the "light" category corresponds to ranks 0 and 1 in the ordinal scale, while "heavy" corresponds to 2 and 3. To our knowledge, there is no standardised scale linking syntactic structure to weight. We followed a basic (yet intuitive) principle to link syntactic structure to weight. The syntactic structure corresponding to "light" implies a basic structure with a pronoun (PRN) or a noun (N) as the head, with basic premodification and postmodification: either Ø, or a determiner (D) and/or an adjective (ADJ). The rest of the combinations were regarded as "heavy", as shown in the table.

Table 3. Scales for measuring weight

Nominal scale	Ordinal scale	Syntactic structure			
LIGHT	0	(D)		N	
				PRN	
	1	(D)	ADJ	N	
HEAVY	2	(D)		N	PP
		(D)	ADJ*	N	
		(D)	(ADJ)	N*	
	3	(D)		N	PP*
		(D)	(ADJ)	N	AdjP*
		(D)	ADJ	N	PP
		(D)		N	IP/CP
		(D)	(ADJ)	N*	PP*
		(D)	ADJ	N*	(PP*)

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

To ensure that the nominal scale is a reliable measurement of the numeric scale, we performed a correlation test between the numeric and the ordinal scale. Results reveal that they are highly correlated [$\rho = 0.914$, $p<0.001$]. Hence, it can be safely assumed that the division between ranks 0 and 1 (light) vs. ranks 2 and 3 (heavy) is a reliable dividing line between what we considered to be light vs. heavy subjects.

As for the analysis of the discursive status of S, each (pre- and postverbal) S in the selected concordances was coded as either topic or focus, according to our definition of these terms in Section 2.2.2, by which topic and focus are concepts encompassing a variety of notions which are best analysed in terms of a 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 S was topic or focus. Both evoked and inferable entities were coded as focus.

Finally, consider the analysis of preverbal unaccusative subjects. The total number of unaccusative SV concordances was 588 for the Spanish corpus and 599 for the Italian corpus. Given that coding manually these 1,147 concordances in terms of weight and discursive status is time-consuming and requires a considerable amount of human resources, we randomly sampled a relatively high number of concordances (minimum 80 for each corpus). The Vs pooled in the sample were the top 4 unaccusative inversion Vs: *exist, appear, begin* and *come* (Spanish corpus) and *exist, emerge, live* and *come* (Italian corpus), as we will see in Figure 4. The sampling process resulted in the following unaccusative SV concordances: n = 81 Spanish and n = 96 Italian. These amounts will be contrasted against the unaccusative VS concordances (n = 52 Spanish and n = 15 Italian) in terms of weight and discourse status (Sections 6.2 and 6.3).

6.   Results

6.1   H1: Postverbal Ss and unaccusativity

As Table 4 shows, the total raw frequency of usable concordances produced was: 153 unergatives plus 640 unaccusatives for the Spanish corpus, and 143 unergatives plus 574 unaccusatives for the Italian corpus. Crucially, postverbal Ss are produced only with unaccusatives (8.1% for the Spanish corpus vs. 2.6% for the Italian corpus), but never with unergatives (0% for both the Spanish and Italian corpora). This finding supports H1 and previous research.

Table 4. Rate of postverbal Ss produced

Subcorpus	V type	# postverbal Ss	# 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%)

Given that learners never produce VS with unergatives, in what follows we will leave aside unergative Vs and will focus on unaccusatives with pre- and postverbal Ss, i.e. unaccusative SV and VS orders.

Consider now the different rates of unaccusative SV vs. VS orders. The production of postverbal Ss (i.e. VS) is rather low (Figure 1), since the majority of structures produced are SV structures: 91.9% in the Spanish corpus and 97.4% in the Italian corpus. While the VS rates are relatively low, a statistical analysis reveals that the difference in VS production between the two corpora is significantly different: 8.1% in the Spanish corpus and 2.6% in the Italian corpus [χ^2 = 17.630, df = 1, *p*<0.001]. This result is unexpected since, according to our general hypothesis, we expect both groups to show similar rates, given that L1 Spanish and Italian behave similarly regarding the conditions under which VS order is produced. This is discussed below, but crucially note that, (i) as we have just seen, both groups behave similarly at the lexicon-syntax interface by producing VS with unaccusatives only and, (ii) as we will see in the next sections, both groups' behaviour is similar for the rest of the interfaces under investigation (weight at the syntax-phonological form interface and information status at the syntax-discourse interface).
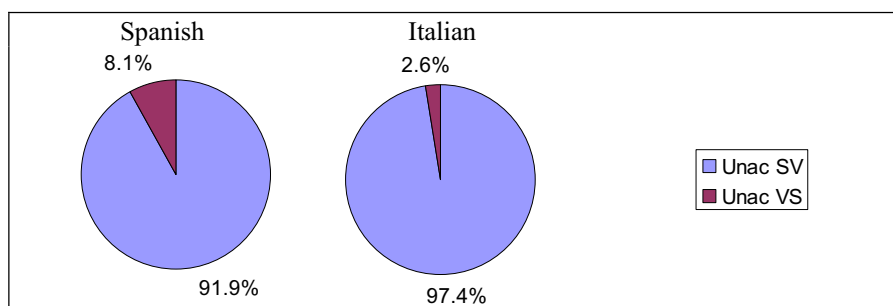
Figure 1. Preverbal vs. postverbal subjects produced with unaccusatives

We should bear in mind that, despite the proportion of postverbal subjects produced by our learners, unaccusative SV concordances represent the majority of productions with unaccusatives, namely 588 (640–52, i.e. 91.9%) in the Spanish corpus and 559 (574–15, i.e. 97.4%) in the Italian corpus, as just seen in Figure 1.

The issue of grammaticality/acceptability inevitably arises when dealing with learner data. Recall that we are interested in postverbal subject constructions with unaccusative Vs, regardless of their grammatical status in native English.[16] Overall, Figure 2 shows that most unaccusative postverbal Ss produced by our learners are ungrammatical ([35] vs. [36] below), 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 [χ^2 = 0.723, df = 1, *p* = 0.395]. In contrast, 100% of preverbal-subject structures with both unergatives and unaccusatives are grammatical in both corpora, as in (36) (codes in round brackets represent the ICLE filename: those filenames starting with an "s" belong to the Spanish subcorpus and those starting with an "i", to the Italian subcorpus).
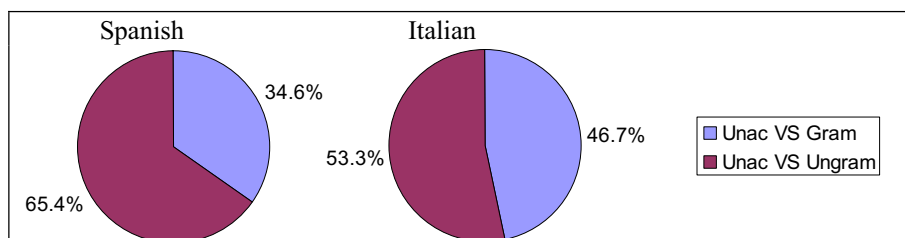
Figure 2. Grammatical vs. ungrammatical postverbal subjects (unaccusative VS)

(35)   a. […] in the evolution of the human specie *it* would disappear ***the capacity of thought* in a near future**. (spm04006)
        b. *It* is almost disappearing *the* **use of writing nice letters to friends [**…]
        (itrs1018)
(36)   a. *Then* come **the necessity to earn more** […] (spm07023)
        b. *Then* came **psychoanalysis**. (itrs1010)

The production of a postverbal S here often implies the presence of preverbal material. Overall, six different types of XP-V-NP$_{SUBJECT}$ were produced, cf. (37)-(42).

(37)   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)
(38)   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)
(39)   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)
(40)   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)

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

(42) 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]

Ungrammatical *it*-insertion, as in (37), and grammatical locative inversion, as in (38), were the most frequent structures in both groups (Figure 3). Due to space limitations, we will leave aside the study of these different preverbal structures and will focus on the conditions under which post- and preverbal unaccusative Ss are produced.
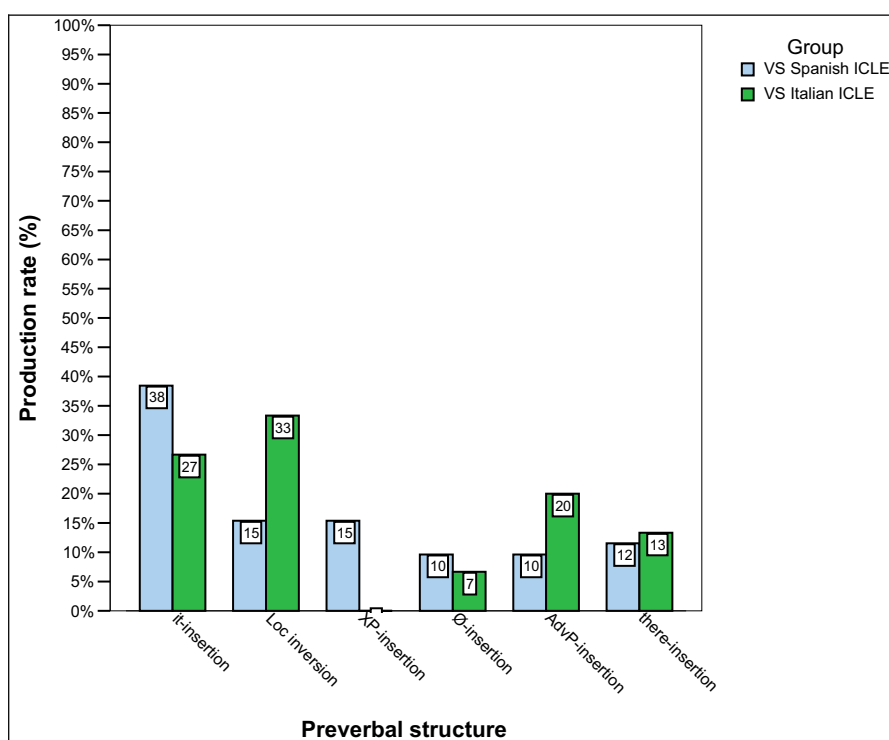
Figure 3. Preverbal material in unaccusative VS structures (XP–V– NP)

Consider, finally, the different types of unaccusative lemmas that trigger VS order. Figure 4 represents, for each unaccusative lemma, its frequency of inversion (in %) out of the grand total number of concordances. We can observe that in the Spanish corpus, out of the 8.1% of VS structures produced (recall Figure 1), the top four inversion unaccusatives are: *exist* (3.4%), *appear* (1.7%),
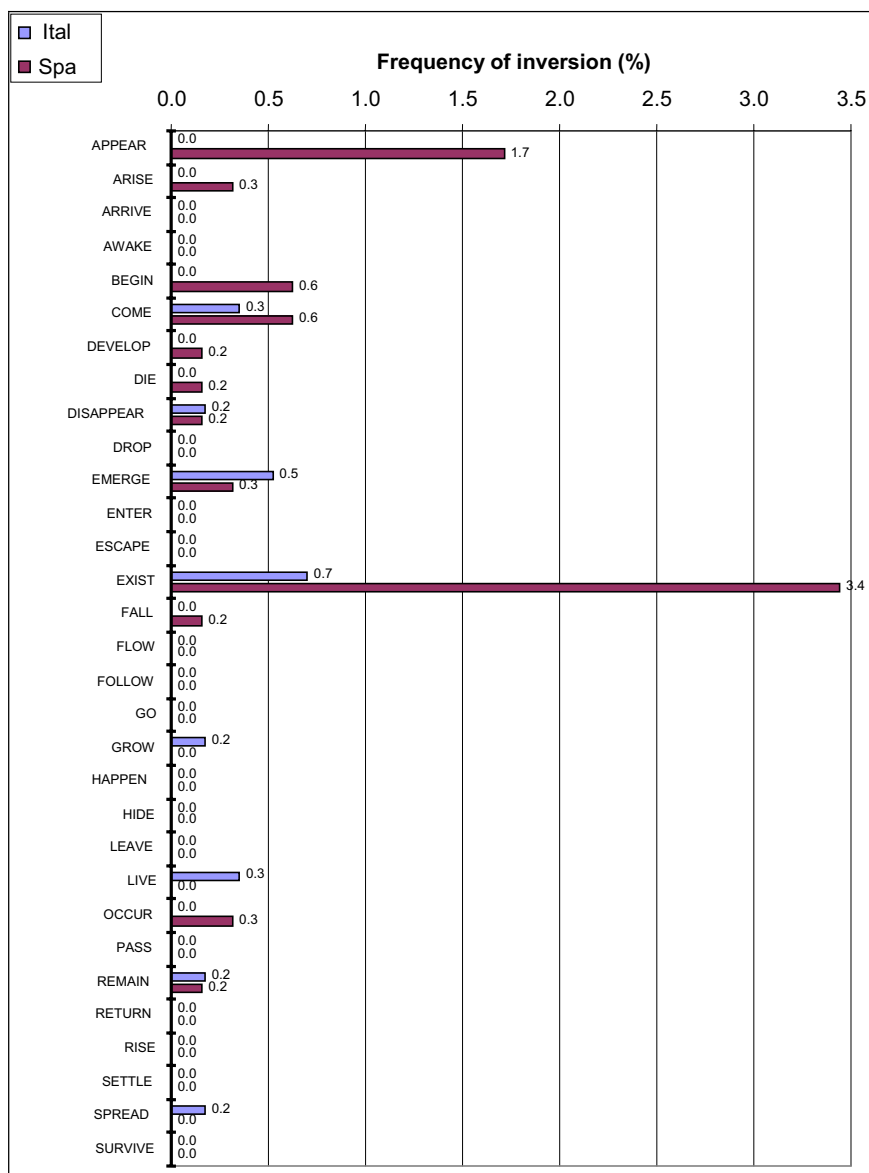
Figure 4. VS structures produced with each unaccusative lemma

begin (0.6%) and *come* (0.6%), while the rest contribute a small percentage to inversion. In the Italian corpus, the top four inversion unaccusatives are: *exist* (0.7%), *emerge* (0.5%), *come* (0.3%) and *live* (0.3%). We will return to these

results (Section 7), as they are crucial to interpret the difference in VS rates between the two corpora, Spanish (8.1%) and Italian (2.6%).

6.2 H2: Postverbal Ss and weight

Consider first the results on weight in the numeric scale (as measured in number of words). This will be crucial to see the differences in the spread of scores between both groups and the actual length of pre- vs. postverbal Ss, which will be the basis to later interpret the results on the nominal scale (heavy/light).

The boxplot (Figure 5) represents the spread of weight of unaccusative pre- and postverbal Ss for both groups (Spanish and Italian), with circles representing outliers and asterisks representing extreme cases. While both heavy and light subjects appear in both preverbal and postverbal positions, as the contrast between (43) and (44) shows, a statistical analysis reveals that, for both groups, preverbal Ss are typically light (short), as in (43a, b), while postverbal Ss are typically heavy (long), as in (44a', b').
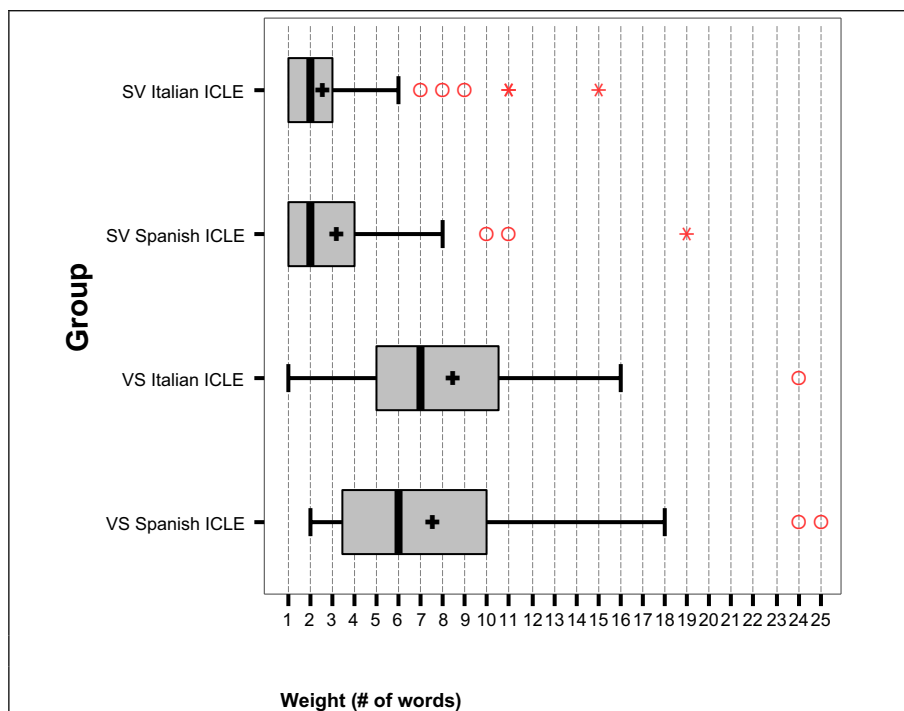

Figure 5. Boxplot (with median and mean) of subject weight in number of words

(43)   *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)

(44)   *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)

In particular, in SV structures, both the Spanish and Italian groups behave similarly: the median for both groups is 2, i.e. preverbal Ss containing 2 words divide the spread of weight scores. The mean (which is marked by a plus sign in Figure 5) is 3.2 in the Spanish corpus and 2.6 in the Italian corpus. This difference between both corpora is not significant [t = 1.430, df = 175, p = 0.155]. By contrast, the spread of postverbal Ss is different from preverbal Ss in both corpora. In particular, the median for postverbal Ss is 6 in the Spanish corpus and 7 in the Italian corpus, and the means are 7.5 in the Spanish corpus and 8.4 in the Italian corpus. The difference in means between both groups is not significant [t = -0.554, df = 65, p = 0.581].

To summarise, the results on the numeric scale reveal that for both learner groups preverbal Ss are short (around 2~3 words long), while postverbal Ss are long (around 7~8 words long).

Consider now the same results in terms of the nominal (heavy/light) scale (Figure 6). The majority of unaccusative postverbal Ss are heavy: 78.8% in the Spanish corpus vs. 86.7% in the Italian corpus. The difference between the two corpora is non-significant [χ^2 = 0.455, df = 1, p = 0.50].

By contrast, as Figure 7 reveals, the majority of unaccusative preverbal Ss are light: 67.9% in the Spanish corpus vs. 78.1% in the Italian corpus. The difference is non-significant [χ^2 = 2.355, df = 1, p = 0.125].

To summarise, for both groups postverbal Ss tend to be heavy, while preverbal Ss tend to be light. Since the results of the nominal scale confirm the results of the numeric scale, these results support hypothesis 2.
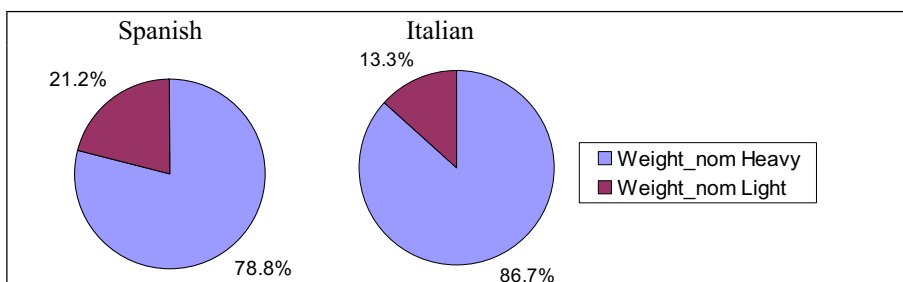
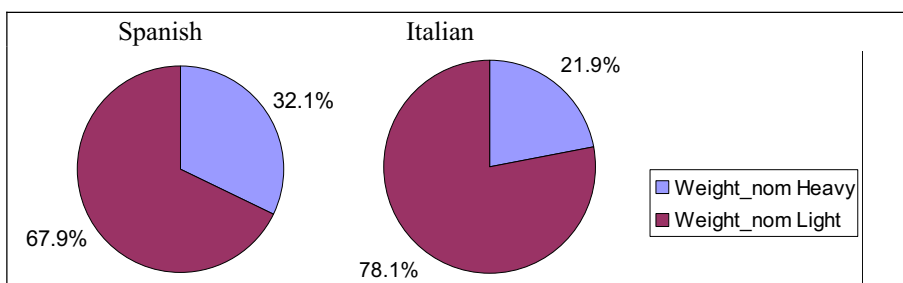**Figure 6.** Heavy vs. light postverbal subjects (unaccusative VS)

Figure 7. Heavy vs. light preverbal subjects (unaccusative SV)

6.3 H3: Postverbal Ss and information status

The last property of unaccusative Ss under investigation is information status.[17] Figure 8 clearly shows that postverbal Ss are focus (evoked or inferable from the co(n)text): 98.1% in the Spanish corpus and 100% in the Italian corpus, the difference between corpora being non-significant [$\chi^2 = 0.293$, df = 1, $p = 0.588$].[18]
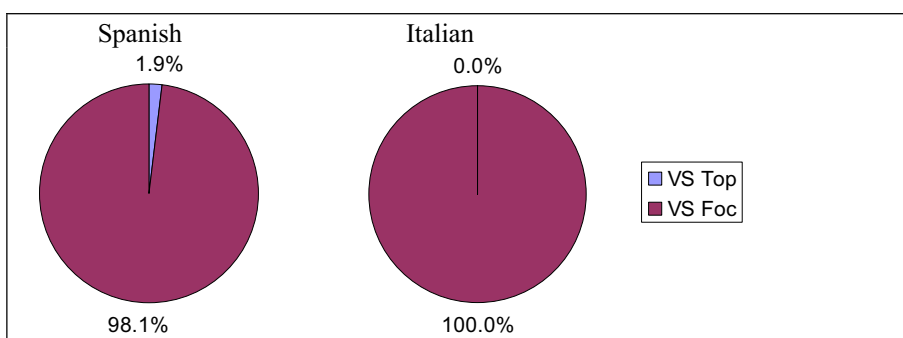

Figure 8. Topic vs. focus postverbal subjects (unaccusative VS)

The contexts in (45) illustrate this finding: postverbal subjects are brand-new information (focus), i.e. the subject has not been mentioned previously in the discourse. These examples are taken directly from the first paragraph of the ICLE essay to illustrate the point that the subject is a new entity in the discourse.

(45)   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)

There exist other contexts where the subject is *new-anchored* (to use Kaltenböck's [2005] terminology), i.e. while part of the subject has been mentioned in the preceding context, typically a postmodifier PP, e.g. *theatre* in (46a) and *men* and *women* in (46b), the keyword in the NP is brand-new information, e.g. *the decline* and *a predominance* respectively. These have also been coded as focus, as has been done previously in the literature (see Kaltenböck 2004, 2005 and references therein).

(46)   a. In the 2nd half of the 17th century we've got the Racionalism and so the comedy was the best mean to express its ideas. The most important for the restoration <u>theatre</u> was to know the human; so in its comedies appears the satire to know human behaviours. The theatre was highly professional because it was in the hands of actors, and so what they wanted was to make money the quickliest possible, so they represented what was more acceptable, *so* came **the decline of the theatre** because they repeated a lot of time the same themes. (spm06010)
b. Nowadays, the difference between <u>men</u> and <u>women</u> should not exist any more, at least in our mind. It is true that, from the physical point of view, excluding particular situations (for example women who practice sports or do works which implies physical strength), *there* still remains **a predominance of men over women**. (itto4006)

Consider now the information status of preverbal material (XP) with unaccusative VS structures, i.e. unaccusative structures like XP-V-NP$_{SUBJECT}$. Figure 9 shows that preverbal material is typically topic in both the Spanish and Italian corpora: 80% and 90% respectively. This difference is not significant between groups [$\chi^2 = 0.480$, df $= 1$, $p = 0.488$]. The examples in (47) show preverbal XP$_{TOPIC}$ while those in (48) represent preverbal XP$_{FOCUS}$.
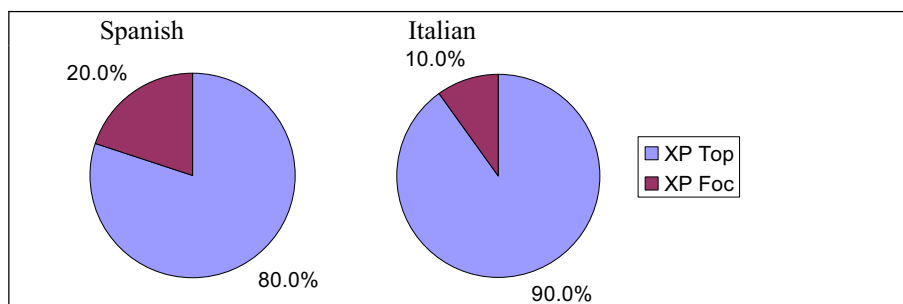
Figure 9. Topic vs. focus preverbal material (unaccusative XP-V-S)

(47) a. The Victorian society drama is one of the 3 manifestations that emerged in the XVIII century. This kind of theatre represents the victorian values, which is the conservatism. It was developed in a drawing-room or an isolated set. That is why it is called the "drawing roam" drama. *Against this society drama* emerged **an oposition headed by Oscar Wilde and Bernard Shaw**. (spm08007)

b. In the passage there are many points in which the so called [deleted word here] emerges, and this is conveyed by the many characters' gestures. We could call it the body language and *through it* emerges **the protagonists' personality**. (itrs1064)

(48) a. I think we have been created for to think and also for to dream and to imagine. Sometimes, when you have got a bad moment in your live, when you become depressed, to dream that maybe things will be better soon, to imagine that it can exist a better world can help to keep you alive in those hard days that everybody use to have. If some day we are not going to need dreaming nor imagining because everything will be done, then we won't need either to be sad or hungry or to have problems because science will make something for not to exist problems, for our mind not to work about nothing. If this became so (and I say this arriving to a very extremed and exaggerated point of view) might someday babies will born without need to use their minds which means that *in the evolution of the human specie* it would disappear **the capacity of thought in a near future**. (spm04006)

b. I think that the fashion of being vegetarian is lunked to this general going towards something more spiritual in an era in which we have got everything. Even if it could seem add after this introduction I gave up eating meat at about) years ago. The reason of my choice concerned especially my personal tastes (I have never liked meat very much) but something strange that I can't exactly explain happened to me one day 9 years ago. I refused to eat meat because it remind me about the animal alive. […] In my opinion a lot of vegetarian people have made this choice first to follow the fashion and not because of their own belief. *In the U.S.A. for example, the country of contradictions, where there is the*

highest number of obese persons but where the fitness mania involves every level of the population has recently spread out **the fashion fo[r] natural and healthy food that of course doesn't include meat**. (itb13001)

Contrary to what happens with postverbal Ss (cf. Figure 8), Figure 10 shows that preverbal unaccusative Ss are typically topic: 88.9% in the Spanish corpus and 90.6% in the Italian corpus. The difference is non-significant [$\chi^2 = 0.145$, df = 1, $p = 0.703$]. The examples in (49) illustrate preverbal subjects which are topics (shown in *italics*), since they have previously been mentioned in the discourse (shown in underlined typeface). Note that personal pronouns are also topics, since by necessity their referent must have been mentioned in the prior discourse, cf. (50).
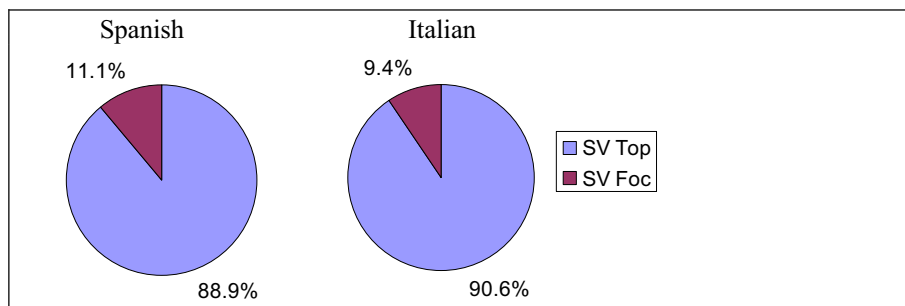

Figure 10. Topic vs. focus preverbal subjects (unaccusative SV)

(49)   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 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)

(50)   a. Also the <u>records</u>, <u>movies</u> have more success if *they* appear several days on the little screen. (spm07018)
b. <u>Violence</u> does exist and *it* will always exist. (itto2034)

To summarise, in unaccusative VS structures, the postverbal S is typically focus and the preverbal material (in those cases where it appears) is topic. By contrast, in unaccusative SV structures, the preverbal S is typically topic. This confirms hypothesis 3.

7. Discussion

As H1 predicted, both groups produce VS with unaccusatives only, yet their rates (8.1% in the Spanish corpus vs. 2.6% in the Italian corpus) differ statistically. Recall that this quantitative difference is unexpected given the similarities between Spanish and Italian (see Section 2). Additionally, Oshita (2004) did not find remarkable differences either (Spanish 6% vs. Italian 4%). A possible explanation for this may lie in the ratios of inversion with certain unaccusative Vs. As Figure 4 revealed, the Spanish results are somewhat inflated by the excessively high ratio of inversion with two of the unaccusatives, viz. *exist* (3.4%) and *appear* (1.7%). Both account for the majority of inversions (5.1% out of the total 8.1%). Additionally, recall that Palacios-Martínez & Martínez-Insua (2006) found that *exist* is the only unaccusative V in the Spanish ICLE corpus that appears with postverbal Ss in existential *there*-constructions. By contrast, the inversion rates with those two Vs in the Italian group are unexpectedly lower than in the Spanish group: 0.7% with *exist* and 0% with *appear*. A possible explanation for these discrepancies may come from the number of frequencies produced. We observed that in the Spanish group the raw frequency of concordances produced with those two Vs was very high: 58 concordances for *exist* (i.e. 9.1% of all concordances) and 60 concordances for *appear* (i.e. 9.4%). By contrast, the Italian counterparts are less frequent: 33 concordances for *exist* (i.e. 5.7%) and 13 concordances for *appear* (i.e. 2.3%). Follow-up analyses revealed that, for all verbs, there is a significant correlation between the frequency of concordances and the frequency of inversions in both corpora, Spanish (r = 0.397, p = 0.027) and Italian (r = 0.364, p = 0.044). This implies that the higher the number of concordances produced with unaccusatives, the higher the probability of inversion. So, it could be assumed that the unusually high proportion of inversions in the Spanish group with *exist* and *appear* is a consequence of their unusually high proportion of concordances.

It could then be asked why there is such a high number of concordances for *exist* and *appear* in the Spanish corpus. Factors like composition topic and/or proficiency level could be the source of such differences. Regarding proficiency, it may be the case that the Italian group's proficiency is higher than the Spanish group's. Kaszubski (2001) notes that the ICLE Spanish subcorpus belongs to the upper-intermediate level, as opposed to the advanced level of the rest of the subcorpora. Additionally, the frequencies of grammatical vs. ungrammatical unaccusative VS produced suggest that this is the case (see Figure 2): the Italian group produces more grammatical structures (46.7%) than the Spanish group (34.6%). We should take this assumption provisionally, since, crucially, the ICLE does not provide an independent measure of proficiency for each participant.

To summarise H1 (lexicon-syntax interface), the results support the hypothesis since both Spanish and Italian learners of English produced postverbal Ss with unaccusatives only, never with unergatives. This finding is in line with previous research (Rutherford 1989, Zobl 1989, Oshita 2004).

Regarding H2 (syntax-phonology interface), the results confirm the hypothesis that both Spanish and Italian learners of English tend to produce heavy Ss in postverbal position, but light Ss in preverbal position. To our knowledge, this finding has not been previously reported in the L2 literature.

Consider now H3 (syntax-discourse interface). The results confirm the hypothesis that both Spanish and Italian learners of English consistently produce focus Ss in postverbal position, but topic Ss in preverbal position. To our knowledge, this finding has not been previously reported in the L2 literature either.

To summarise all the results with unaccusative VS structures, the postverbal S is focus and it tends to be heavy, while preverbal material (in those cases where it appears) is topic. By contrast, with unaccusative SV structures, the preverbal S is topic and tends to be light. This is schematised in Table 5.

Table 5. Summary of weight and information status with unaccusatives

Construction	Structure		
Unaccusative postverbal S	$\left(\dfrac{XP}{Top} \right)$	$\dfrac{V}{Unac}$	$\dfrac{NP_{SUBJECT}}{Foc \quad Heavy}$
Unaccusative preverbal S	$\dfrac{NP_{SUBJECT}}{Top \quad Light}$	$\dfrac{V}{Unac}$	

In conclusion, for learners of L2 English, unaccusativity is a necessary but not sufficient condition for Ss to be produced in postverbal position. In particular, other conditions must be met for S to appear postverbally. These are (in order of strictness): unaccusative verb, focus S, heavy S, cf. (51).

(51)   *Conditions on the production of postverbal Ss:*
        (i) Unaccusative V
        (ii) Focus S
        (iii) Heavy S

8.   Conclusion

Inverted Ss are considered to be a feature of the interlanguage of L1 Spanish and Italian learners of L2 English. The Unaccusative Hypothesis (Perlmutter 1978) has been used to account for why inverted Ss appear to be produced only with unaccusative Vs. In recent years, the issue of unaccusativity has become part of a research agenda which seeks to establish a relation between the lexical semantics of Vs and the syntactic properties of the constructions they enter, in accordance with Perlmutter's (1978) original intuition that unaccusativity is encoded in the syntax but is semantically determined. Within this framework, Levin &

Rappapport Hovav (1995) distinguish two major classes of unaccusative Vs: (i) Vs of change of state and (ii) Vs of existence and appearance. Only the latter appear in constructions with inverted Ss, such as locative inversion structures and *there*-constructions in native English. Our results from L1 Spanish/Italian learners confirm the possibility of inversion with unaccusative Vs of existence and appearance. While these results confirm that learners are sensitive to the unaccusative-unergative distinction, and thus provide support for the psychological reality of the Unaccusative Hypothesis, further research should show that unaccusative Vs of change of state do not trigger inversion, as in L1 English. Thus, learner data could be used to provide further evidence for the existence of these two verb classes, following the tradition of research which takes learner data "to make a unique and potentially significant contribution to theoretical linguistics" (Oshita 2004: 95-96) and the investigation of interlanguage "a valuable endeavour for any researcher interested in extending our knowledge about what language is" (Rutherford 1989: 164) (see Rutherford 1986 and Oshita 2004: Section 1 on the contribution of language acquisition studies to theoretical linguistics).

The same can be said about the other two conditions identified for subject inversion (heaviness and discourse status), which support findings in the native English literature that inversion of S occurs, given the appropriate structural conditions, when S is syntactically heavy, as well as new (or *newer*) information or focus. Interestingly, there is an interrelation between these two factors (the newer an entity is, the more words are used to refer to it). Both conditions are designed to ease the processing burden on the receiver, by placing new and long constituents towards the end of a clause. The gradience approach adopted in some L1 studies to information status and heaviness turned out to be the most appropriate to code and explain our results concerning these two conditions, which, to our knowledge, have never been discussed in L2 studies of inversion. These studies have mostly focused on errors, while we have adopted a wider approach seeking to identify the conditions under which learners produce inverted Ss, regardless of problems to do with syntactic encoding.

No significant differences were found between Italian and Spanish learners regarding the conditions under which inversion occurs, though, as was pointed out, Italian speakers produced significantly fewer inversions with unaccusative Vs. To provide a proper explanation of this fact, we would require precise information about the Italian and Spanish learners' proficiency level. An issue that has not been mentioned but which could also prove extremely revealing would be the use of native Spanish and Italian corpora to determine to what extent the differences observed between the two groups of learners could be attributed to L1 properties (and thus to transfer). In fact, though we have been assuming throughout, on the basis of analyses provided in the theoretical literature, that Spanish and Italian are identical regarding the phenomenon known as "free inversion" (but see note 3), corpus studies may well reveal differences between the two languages. We are not aware of the existence of any such studies in either Italian or Spanish.

Similarly, a comparison between our L2 corpora and L1 English corpora would be crucial to identify overuse or underuse of inversion constructions, as part of the CIA approach mentioned in the introduction to this paper. Intuitively, it seems that at least Spanish learners appear to overuse postverbal subject structures. This intuition is confirmed by a preliminary comparison of the results obtained from the Spanish ICLE subcorpus with LOCNESS (*Louvain Corpus of Native English Essays*, CECL, Université catholique de Louvain, see http://cecl.fltr.ucl.ac.be), which does indeed reveal that this is the case. We leave a detailed analysis of these facts for further research (see Lozano & Mendikoetxea in preparation). Also for further research is the comparison of our results with Italian and Spanish NS data, as mentioned above, which would be crucial to make meaningful statements regarding transfer, as well as the analysis of L2 data from speakers of languages which do not allow free inversion. A preliminary analysis of the French ICLE subcorpus reveals a total of 2.3% of inversions with unaccusative verbs. This figure is similar to the rate of postverbal subjects found in the Italian subcorpus (2.6%) (see Figure 1 above). However, the inversion structures in the French subcorpus are mostly grammatical, as opposed to those found in the Italian and the Spanish subcorpora (see Figure 2 above). These facts require an in-depth examination.

It is also our intention to broaden the empirical scope of this research into the production of inverted Ss to include the V *be*, which, according to Biber et al.'s (1999: 954) native English corpus findings appears in 95% of *there*-constructions and about half of all locative inversion structures. Of the 1,778 tokens of the study reported in Birner (1994), 654 were instances of inversion around *be*, making it the most common V in inversion constructions. We also intend to combine corpus data with experimental data (e.g. acceptability judgements) in order to get a bigger picture of the mental processes and structures which underlie learners' production of inversion structures (see e.g. Kennedy 1998 on the need to combine corpus data with other types of linguistic evidence).

In sum, though the results obtained in this study are highly significant in that they reveal conditions for the occurrence of postverbal Ss in L2 English that have never been uncovered before, extending the scope of the research by comparing our results with native corpora and other learner corpora (different L2) and broadening the empirical scope by looking at inversion with the V *be*, as well as by using experimental language data, will no doubt allow us to gain a better understanding of the processes underlying this phenomenon in both native and non-native grammars.

Notes

2   The term "postverbal" as used in this paper is purely descriptive: an S that occurs *after* V. Equally, "preverbal" refers to an S which occurs *before* V. We do not address here the question of what are the specific structural positions occupied by both the postverbal and the preverbal S, an issue which has recently received a lot of attention in the theoretical literature (see e.g. Hulk & Pollock 2001 for Romance subject inversion).

3   There are well-known differences between Spanish and Italian concerning VSO order: when S is an NP, as opposed to the PP in (2c), VSO is considered to be less acceptable in Italian. Italian does not allow VOS either. These restrictions are not found in Spanish, which seems to have a "freer" word order (see Belletti 2004b). Since we will be dealing with intransitive sentences like those in (4), in which Spanish and Italian behave alike, the differences observed in transitive sentences are not relevant here.

4   Among the Romance languages, French has a special status in that it is negatively marked for the NSP: it does not allow for NSs, nor for the type of "free inversion" found in other Romance languages. Thus, the French counterpart of the examples in (3b, c) and (4) is ungrammatical.

5   Languages like Chinese, Korean and Japanese may also have null Ss, but do not display "rich" agreement and appear to lack *pro* (see Huang 1984). For an account of these languages along different lines, within the context of L2 acquisition, see Oshita (2004).

6   Within the Minimalist Program, Chomsky (1995: Ch. 4) has attributed the presence of *there* in constructions like (8c) (and, presumably, pro_{expl}) with the satisfaction of the Extended Projection Principle (EPP), by which clauses must have Ss (see Chomsky 1981). There are, however, proposals which seek to eliminate null expletives from the theory (for instance Picallo 1998 and Yusa 2002).

7   We are actually adopting a simplified version of the VP-internal subject hypothesis. In recent years more sophisticated proposals regarding the internal structure of the VP have been put forward in the literature (Chomsky 1995: Ch. 4 and references cited therein).

8   The inversion types in (15) are to be distinguished from subject-operator inversion, a construction found following negative opening elements (*not, not only, hardly…*), after *so*, etc.

9   There are, of course, additional requirements. For instance, for *there*-constructions it has been claimed that the postverbal NP must be indefinite (see, for example, Safir 1985), although it is well known that this does not account for all the data and definite NPs can be found in these constructions (see e.g. Prince 1992: 299 and Biber et al. 1999: 967). Likewise, some indefinites are banned. In particular, as noted by Prince (1992: 299), plural generics are incompatible with *there*-constructions. Though Prince (1992) does not give reasons for this, the explanation has to be found in the semantics of the construction which requires indefinites to have existential meanings and not universal or generic meanings.

10   As pointed out by Belletti (2004b, Section 5), with the appropriate intonation pattern and the right pragmatic conditions, a postverbal S can also be interpreted as topic. This is illustrated in (i) for Italian (Belletti 2004b: 22) and the same can be said of parallel examples in Spanish.

       (i)   a. Che cosa ha poi fatto Giovanni?
                     What has Gianni finally done?
                  b. Ha parlato, Gianni.
                     Has spoken, Gianni.

       Examples like these are largely irrelevant for our purposes, since we are dealing with written corpora, in which examples like these are unlikely to occur. The other side of the coin concerns preverbal elements which are focus. These often receive a contrastive interpretation, which is marked by special intonation (see Rizzi 1997, 2004 for Italian and Domínguez 2004 for Spanish) (see also the examples in [21]):

       (ii)   a. GIANNI ha comprato il libro (non Maria).
                  b. JUAN ha comprado el libro (no María).
                  c. JOHN has bought the book (not Mary).

11   This is the approach taken in Chocano et al. (in preparation), in which the binary topic-focus distinction comprises both a retrievability scale, as well as a recency scale.

12   One attempt to introduce information structure notions such as topic and focus as part of syntactic structure is what is known as the "cartographic approach" to syntactic structure (see e.g. Rizzi 1997 and the papers contained in Belletti 2004a). Focalisation (and topicalisation) are analysed as involving movement of a phrase to a designated position either in the left periphery or in the right periphery. The interpretation of an element as focus or topic is derived automatically from the position occupied by that

element in clause structure: the interpretation is read off the structural configuration.

13   The term "Heavy NP Shift" is used here descriptively. We are not postulating a rightward movement rule of the type found in the transformational analyses of the 70s.

14   Though all the examples in this section involve instances of locative inversion structures, a parallel situation is observed for *there*-constructions which are often found with "heavy" postverbal elements.

15   As is commonly assumed in the Generative Grammar literature, we take *theme* to be the semantic role associated with the internal argument of a V. In other frameworks this is referred to as the *patient*.

16   It is important to note that we classify as "ungrammatical" only those postverbal Ss that are not possible in English. We are abstracting away from standard ungrammaticalities such as S-V agreement. For example, in the sentence *Then come the necessity to earn more* (spm07023), the postverbal-subject structure is possible in English, yet lack of S-V agreement renders it ungrammatical sensus stricto.

17   Apart from the weight and information status of S, we also analysed its definiteness as it has been traditionally claimed that postverbal focused Ss are indefinite, while preverbal topic Ss are definite. This has been recently confirmed in corpus studies (e.g. Prince 1992 and Biber et al. 1999; see also note 9 here). In order to test whether there was a relationship between the information status and definiteness of S, we coded every S as being definite or indefinite. With unaccusative VS structures, learners typically mark S as indefinite (59.6% Spanish vs. 53.3% Italian, non-significant [χ^2 = 0.189, df = 1, p = 0.664]), while in unaccusative SV structures, S is marked as definite (70.4% Spanish vs. 72.99% Italian, non-significant [χ^2 = 0.141, df = 1, p = 0.708]). In other words, our results follow the typical native English pattern, i.e. preverbal (topic) Ss tend to be definite, while postverbal (focus) Ss tend to be indefinite.

18   The 1.9% corresponds to only one case (out of 52) where the unaccusative postverbal S is a topic.

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.

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*. 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. 16-51.

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*. Unpublished PhD dissertation, Boston University.

Eguren, L. and O. Fernández Soriano (2004), *Introducción a una sintaxis minimista*. Madrid: Gredos.

Fernández-Soriano, O. (1989), 'Strong pronouns in null-subject languages and the Avoid Pronoun Principle', *MIT Working Papers in Linguistics,* 11: 228-239.

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: Cambridge University Press.

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. (2000/2001), '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 to computerized bilingual and learner corpora', in: K. Aijmer, B. Altenberg and M. Johansson (eds), *Languages in Contrast. Papers from a*

Symposium on Text-Based Cross-linguistic Studies. Lund 4-5 March 1994. 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*. Amsterdam & Philadelphia: John Benjamins Publishing Company. 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.

Huang, C.-T.J. (1984), 'On the distribution and reference of empty pronouns', *Linguistic Inquiry*, 15: 531-574.

Hulk, A.C. and J.-Y. Pollock (eds) (2001), *Subject Inversion in Romance and the Theory of Universal Grammar*. Oxford: Oxford University Press.

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.

Kaszubski, P. (2001), 'Tracing idiomaticity in learner language: the case of BE', in: P. Rayson, A. Wilson, T. McEnery, A. Hardie and S. Khoja (eds) *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster: University Centre for Computer Corpus Research on Language. 312-322.

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

Koopman, H. and 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*. 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 (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.

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.

Palacios-Martínez, I. and A. Martínez-Insua (2006), 'Connecting linguistic description and language teaching: native and learner use of existential *there*', *International Journal of Applied Linguistics*, 16(2): 213-231.

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: UiL OTS Dissertation Series.

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 Publishing Company. 295-325.

Quirk, R., S. Greenbaum, G. Leech and 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. 281-337.

Rizzi, L. (2004), 'Locality and left periphery', in: A. Belletti (ed.) *Structures and Beyond. The Cartography of Syntactic Structures*. Vol 3. New York: Oxford University Press. 223-251.

Rochemont, M.S. (1986), *Focus in Generative Grammar*. Amsterdam: John Benjamins Publishing Company.

Rutherford, W. (1986), 'Grammatical Theory and L2 Acquisition: a brief overview', *Second Language Research*, 2: 1-15.

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.

Safir, K. (1985), *Syntactic Chains*. 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 and M. Sharwood-Smith (eds) *The current State of Interlanguage: Studies in Honor of William E. Rutherford*. Amsterdam: John Benjamins Publishing Company. 153-175.

Vallduví, E. (1990), *The informational component*. Unpublished PhD dissertation, University of Pennsylvania.

Wasow, T. (1997), 'End-weight from the speaker's perspective', *Journal of Psycholinguistic Research*, 26: 347-361.

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

White, L. (1986), 'Implications of parametric variation for adult second language acquisition: an investigation of the Pro-drop parameter', in: V.J. Cook (ed.) *Experimental Approaches to Second Language Acquisition*. Oxford: Pergamon Press. 55-72.

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. 4215-4244.