Interface conditions on the production of Verb-Subject structures in L2 English

EUROSLA 2007
University of Newcastle

Cristobal Lozano
Universidad de Granada

Amaya Mendikoetxea
Universidad Autónoma de Madrid

http://www.uam.es/woslac

AIM OF THE PRESENTATION

- To inform on the results of a study on the production of postverbal subjects (VS order) in non-native English.
- Purpose of the study: to characterize the interlanguage of advanced non-native speakers of English L2 (Spanish/Italian L1) by examining their production of both grammatical and ungrammatical VS structures, as represented in the relevant ICLE subcorpora (Granger et al. 2002), :
  - (1) ...and here emerges the problem.
  - (2) *In the name of religion it had occurred many important events

FRAMEWORK

Main question:
What are the conditions under which learners produce inverted subjects (VS structures), regardless of problems to do with syntactic encoding (grammaticality)?

Comparative Framework: to determine the role of L1 in L2 acquisition (transfer) in the areas under study
Learner corpora vs. native corpora (LOCNESS)

Word Order in L1 English (1)

(a) XP V S (Inversion structures with an opening adverbial)

(3) a. [On one long wall] hung a row of Van Goghs. [FICT]
   b. [Then] came the turning point of the match. [NEWS]
   c. [Within the general waste type shown in these figures] exists a wide variation. [ACAD]

(b) There-constructions

(4) a. Somewhere deep inside [there] arose a desperate hope that he would embrace her [FICT]
   b. In all such relations [there] exists a set of mutual obligations in the instrumental and economic fields [ACAD]
   c. [There] came a roar of pure delight as... [FICT]

ENGLISH and SPANISH/ITALIAN differ in devices employed for constituent ordering:
English ‘fixed’ order is determined by lexico-syntactic properties and Spanish/Italian ‘free’ order is determined by information structure, syntax-discourse properties.
Word order in L1 English (VS order)

- **Lexicon-syntax interface** (Levin & Rappaport-Hovav, etc):
  - Unaccussative Hypothesis (Burzio 1986, etc) [existence and appearance]
    - There sang four girls at the opera. [unergative verb]
    - There arrived four girls at the station. [unaccusative verb]
- **Syntax-discourse interface** (Biber et al, Birner 1994, etc):
  - Postverbal material tends to be focus/relatively unfamiliar information, while preverbal material links S to previous discourse.
  - We have complimentary soft drinks and coffee. Also complimentary is red and white wine.
- **Syntax-Phonological Form (PF) interface** (Arnold et al 2000, etc)
  - Heavy material is sentence-final (Principle of End-Weight, Quirk et al 1972) – general processing mechanism (reducing processing burden)

Subjects which are focus, long and complex tend to occur postverbally in those structures which allow them (unaccusative Vs).

Word Order L1 Spanish/Italian (1)

- Postverbal subjects are produced ‘freely’ with all verb classes:

  (9) a. Ha telefoneado María al presidente. (transitive)
  - Has phoned Mary the president
  b. Ha hablado Juan. (unergative) c. Ha llegado Juan (unaccusative)
  - has spoken Juan
  - has arrived Juan

- Inversion as ‘focalisation’:
  - Preverbal subjects are topics (given information)
  - and postverbal subjects are focus (new information) (Belletti 2001, 2004, Zubizarreta 1998)

  (10) ¿Quién ha llegado/hablado? (Sp)
  Who has arrived/spoken?
  i. Ha llegado/hablado Juan
  ii. # Juan ha llegado/hablado

  (11) Chi è arrivato/parlato? (It)
  i. É arrivato/ A parlato Gianni
  ii. # Gianni è arrivato/a parlato

The phenomenon in SLA


- L1 Spanish/Italian/Arabic – L2 English:
  (12) ...it arrived the day of his departure...
  (13) And then at last comes the great day.
  (14) In every country exist criminals
  (15) ...after a few minutes arrive the girlfriend with his family too.

- Only with unaccusative verbs (never with unergatives).
  - Unaccusatives: arrive, happen, exist, come, appear, live...
  - Unergatives: cry, speak, sing, walk ...

- **Explanation**: lexicon-syntax interface (Unaccusative Hypothesis)

Subjects which are focus, long and complex tend to occur postverbally, with no restrictions at the lexicon-syntax interface.
The psychological reality of the Unaccusative Hypothesis

- A number of studies have found that L2 learners are aware of the argument structure distinction between unaccusative and unergative Vs and that they use this as a guiding principle to construct L2 mental grammars.

- However, learners have difficulty in determining the range of appropriate syntactic realizations of the distinction, and this can persist into near-native levels of proficiency (see R. Hawkins 2001: 5.4).

CRUCIAL DIFFERENCE: these previous studies focused on ERRORS, thus emphasising the differences between native and non-native structures. By contrast, our study emphasises the similarities between native and non-native structures.

Hypotheses

GENERAL HYPOTHESIS:
Conditions licensing VS in L2 Eng are the same as those in Native Eng, DESPITE differences in syntactic encoding.

- **H1 [LEXICON]:** Lexicon-syntact interface:
  - Postverbal subjects with **unaccusatives** (never with unergatives)

- **H2 [WEIGHT]:** Syntax-PF interface:
  - Postverbal subjects: **heavy** (but preverbal light)

- **H3 [FOCUS]:** Syntax-Discourse interface:
  - Postverbal subjects: **focus** (but preverbal topic)

METHOD (1)

- Based on Levin (1993) and Levin & Rappaport-Hovav (1995):
  - **Unergatives:** cough, cry, shout, speak, walk, dance…
  - **TOTAL:** 41
  - **Unaccusatives:** exist, live, appear, emerge, happen, arrive…
  - **TOTAL:** 32

METHOD (2)

- Learner corpus: L1 Spa – L2 Eng; L1 Ital – L2 Eng
- ICLE (Granger et al. 2002)

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of essays</th>
<th>Number of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE Spanish</td>
<td>251</td>
<td>200,176</td>
</tr>
<tr>
<td>ICLE Italian</td>
<td>392</td>
<td>227,085</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>643</td>
<td><strong>427,461</strong></td>
</tr>
</tbody>
</table>

(Problem: proficiency level?)

- **WordSmith** v. 4.0 (Scott 2004)
  - Concordance queries can be performed automatically with WordSmith, by targeting specific verbs BUT there is a lot of manual work (filtering out unusable data, coding data in Excel, analysing data in SPSS, etc).

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>V type</th>
<th># usable concordances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>Unergative</td>
<td>153</td>
</tr>
<tr>
<td>Italian</td>
<td>Unergative</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Unaccusative</td>
<td>640</td>
</tr>
<tr>
<td></td>
<td>Unaccusative</td>
<td>572</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td><strong>1510</strong></td>
</tr>
</tbody>
</table>
**H1 results: syntax-lexicon**

![Graph showing percentage of production for Spanish and Italian in Unergative and Unaccusative subcorpora.]

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>V type</th>
<th>Postverbal S</th>
<th>Usable concordances</th>
<th>Rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>Unergative</td>
<td>0</td>
<td>153</td>
<td>0/153 (0%)</td>
</tr>
<tr>
<td>Italian</td>
<td>Unergative</td>
<td>0</td>
<td>143</td>
<td>0/143 (0%)</td>
</tr>
<tr>
<td></td>
<td>Unaccusative</td>
<td>52</td>
<td>640</td>
<td>52/640 (8.1%)</td>
</tr>
<tr>
<td>Italian</td>
<td>Unaccusative</td>
<td>15</td>
<td>574</td>
<td>15/574 (2.6%)</td>
</tr>
</tbody>
</table>

**H1: Unaccusative:**
- **grammatical vs ungrammatical VS**
  - **GRAMM.**
    - 35% Spa
    - 47% Ital
  - **UNGRAM.**
    - 65% Spa
    - 53% Ital

**Examples H2: syntax-phonology**

**SV: typically LIGHT**

(23) a. …but they may appear everywhere.
   b. …since the day eventually came …
   c. …these people should exist, …

**VS: typically HEAVY**

(24) a. Against this society drama emerged an opposition headed by Oscar Wilde and Bernard Shaw.
   b. …exists yet in Spain a group of people who are supposed to be professional soldiers.
   c. It is almost disappearing the use of writing nice letters to friends.
Discourse status (topic/focus) has to be measured manually by establishing theoretical criteria and then by checking the context (or even the essay) manually.

Examples H3: syntax-discourse

**VS: FOCUS**

(25) a. ...there also exists a wide variety of optional channels which have to be paid.
   b. So arised the Saint Inquisition.
   c. In 1880 it begun the experiments whose result was the appearance of the television some years later.

**SV: typically TOPIC**

(26) a. I use the Internet ... I find windows ... if they press on any of these windows ... these windows cannot appear because a child could enter easily...
   b. ...the world of drugs; mafias ... problems with mafias finished ... dangerous people making money ... no reason why these people should exist.

Conclusion

Our results and CIA (1)

- **Contrastive Interlanguage Analysis** (Granger 1996, Gilquin 2001)

(a) **NNS vs. NNS:**

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 L1.

- No significant differences between Italian and Spanish learners, as expected, except for frequency of inversion [8.1% (Spanish) vs. 2.6% (Italian)]
  - Possible explanation: lexical bias.
Our results and CIA (2)

(b) NNS vs. NS.

It 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 (Granger 2002: 12-13).

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of essays</th>
<th>Number of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE Spanish (L1 Sp L2 Eng)</td>
<td>251</td>
<td>200,376</td>
</tr>
<tr>
<td>UAM WriCLE (L1 Sp L2 Eng)</td>
<td>85</td>
<td>63,836</td>
</tr>
<tr>
<td>TOTAL</td>
<td>336</td>
<td>264,212</td>
</tr>
</tbody>
</table>

vs. LOCNESS (L1 Eng): 436 324, 304

Corpus Number of essays Number of words
ICLE Spanish (L1 Sp L2 Eng) 251 200,376
UAM WriCLE (L1 Sp L2 Eng) 85 63,836
TOTAL 336 264,212

H1: Syntax-Lexicon
Ns vs. NNs
Percentage of VS

H2: Syntax-Phonology: NS vs. NNS
Heavy-Light subjects with Unacc Vs

- Subjects in SV structures with unacc Vs tend to be 'light' in NS (67.7% in ICLE+WriCLE and 68.1% in LOCNESS).
- Subjects in VS structures with unacc Vs are overwhelmingly 'heavy' in NS (81.0% in ICLE+WriCLE and 81.3% in LOCNESS).
H3: Syntax-Discourse: NS vs. NNS
Focus and Topic subjects with Unacc Vs

- Most subjects in SV structures with unacc Vs in NS are topic (89.9% in ICLE+WriCLE and 83.5% in LOCNESS); just a few are focus (10.5% in ICLE+WriCLE and 16.5% in LOCNESS; p=0.223).
- Subjects in VS structures with unacc Vs are overwhelmingly focus in NS (98.3% in ICLE+WriCLE 100% in LOCNESS; p=0.784).

Some preliminary conclusions: Ns vs NNS

- These results confirm that Spanish (and, presumably, Italian) learners of English produce postverbal subjects under exactly the same interface conditions as in L1 English (unaccusativity being a necessary but not a sufficient condition).
- Spanish and Italian learners show persistent problems in the syntactic encoding of the construction, producing mostly ungrammatical examples (it-insertion, 0-insertion, wrong XP).
- Spanish learners overuse the construction and show a lexical bias for the V exist.

REFERENCES (1)

Unaccusativity Hypothesis

(1) a. unergative
i. John spoke

b. unaccusative
i. Three girls arrived

VS in native Spanish

Unaccusatives: VS
A: ¿Qué pasó anoche en la reunión?
B: # Un hombre llegó.
Llegó un hombre.

Inergatives: SV
A: ¿Qué pasó anoche en la reunión?
B: Un hombre gritó.
# Gritó un hombre.

ADDITIONAL SLIDES TO FOLLOW:
VS in native Spanish (2)

A: ¿Quién gritó anoche en la reunión?
B: # Un hombre gritó.

Gritó un hombre.

A: ¿Quién llegó anoche a la reunión?
B: Un hombre llegó.

Llegó un hombre.

Data analysis (cont’d)

- MAIN FILTERING CRITERIA:
  - The verb must be intransitive (unergative or unaccusative).
  - The verb must be finite, active voice.
  - The subject can appear either postverbally (VS) or preverbally (SV).
  - The subject must be an NP.
  - The sentence can be either grammatical or ungrammatical in native English.

- OTHER FILTERING CRITERIA (TOTAL = 28)

<table>
<thead>
<tr>
<th>Subcorpus</th>
<th>V type</th>
<th># usable concordances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>Unergative</td>
<td>153</td>
</tr>
<tr>
<td></td>
<td>Unaccusative</td>
<td>640</td>
</tr>
<tr>
<td>Italian</td>
<td>Unergative</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Unaccusative</td>
<td>574</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>1510</td>
</tr>
</tbody>
</table>
WordSmith: query searches:

- For every lemma (e.g., APPEAR, ARISE), we searched for:
  - All possible native forms:
    - appear, appears, appearing, appeared
  - All possible overregularised and overgeneralised learner forms:
    - arised, arosed, arisen, arosened ("So arised the Saint Inquisition")
  - All possible forms with probable L1 transfer of spelling:
    - apear, apears, apearing, apeared
  - All other possible misspelled forms:
    - appeard, apeard

Query searches (cont'd)------

- CONCORDANCES: 6 BASIC FILTERING CRITERIA:
  1. The verb must be intransitive (unergative or unaccusative).
     - ✓ In the screen of the television one or two “rombos” should appear. [unac]
     - ✓ Leontes cries and the statue talks. [unerg]
     - ✓ This government’s movement has created several opinions. [trans]
  2. The verb must be finite, with(out) aux.
     - ✓ ...also it exists the psychological aggressions... [finite no aux]
     - ✓ ...the cases of men mistreated do not appear in the media. [finite aux]
     - ✓ This contradiction could disappear [finite modal]
     - ✓ There’s no reason for it to exist. [for clause + to inf]
     - ✓ Poor people cross borders to escape from poverty. [to-inf clause]
     - ✓ ...let time pass... ['let constructions]
     - ✓ ...make everyone’s life go ahead [causative + infinitive]
     - ✓ ...Returning to the title of this paper,... [gerundive clauses]
     - ✓ ...they go away in order to escape to France. ['in order to' clauses]
     - ✓ ...women have to live with the agressor [have to/ought to/able to]
     - ✓ ...prudence was beginning to disappear. [verbal/aspectual periphrases]
     - ✓ Before entering the argumentation,... [small clauses]
     - ✓ ...instead of following... [complement of P]
     - ✓ ...likely to happen... [complement of A]
     - ✓ The tests to enter the army are quite difficult now. [complement of N]

- OTHER FILTERING CRITERIA
  - Target V + V (verbal coordination)
    - ✓ Families without father exist and work well.
  - Coordinator + target V
    - ✓ ...we can manage to obtain it and live in a better world.
  - Interrogatives (only if V is the target)
    - ✓ How could they live?
  - Formulaic & Set expressions in English
    - ✓ ...it happens the same.
  - Set expressions transferred from the L1
    - ✓ ...they fall into account that they have treated very badly Mr Hardcastle.
  - Phrasal verbs:
    - ✓ ...a scientist come up with an intention...
  - Quotes (literary or other):
    - ✓ "To what purpose, April, do you return again?"
    - ✓ "Feminism has to evolved or die", Friedan said in 1982...
Extraposition was discarded

- NOTE: extraposition discarded:
  - It only remains [to add that nowadays we live in a world...]
  - It happened [that the countries which make the weapons are...]

H1: Examples of VS in LOCNESS

- AdvP insertion (3 tokens)
  (27) Thus began the campaign to educate the public on how one contracts Aids.

- PP insertion (7 tokens)
  (28) ...along with this development has come opposition from both the medical field and a proportion of the general public, who...

- There-insertion (6 tokens)
  (29) Certainly there exists a demand for this work to be done.

Result: Unaccusative: Type of VS structures

Results H1: Unaccusative: grammatical vs. ungrammatical VS

Figure 1. Proportion (in %) of grammatical vs. ungrammatical unaccusative VS
Result H1: VS and specific unaccusative verbs

---

Scale (syntactic weight/complexity)

<table>
<thead>
<tr>
<th>NOMINAL SCALE</th>
<th>ORDINAL SCALE</th>
<th>SYNTACTIC STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIGHT</td>
<td>0</td>
<td>(D) N</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>(D) ADJ N</td>
</tr>
<tr>
<td>HEAVY</td>
<td>2</td>
<td>(D) ADJ PP</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(D) ADJ N PP*</td>
</tr>
</tbody>
</table>

---

Topic vs. Focus: Retrievability scale

<table>
<thead>
<tr>
<th>Discourse old (TOPIC)</th>
<th>Discourse new (FOCUS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ retrievable</td>
<td>+ retrievable</td>
</tr>
<tr>
<td>+ recent</td>
<td>+ recent</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Evoked (text +lam)</td>
<td>2. Evoked (text +lam)</td>
</tr>
<tr>
<td>(E.g)</td>
<td>(E.g)</td>
</tr>
<tr>
<td>h-old</td>
<td>d-new</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Unused</td>
<td></td>
</tr>
</tbody>
</table>

---

Result: VS and (in)definiteness

![Figure 1: Production of postverbal subjects according to their definiteness.](image_url)

---

...some decades ago, it appeared a new invention: the television.

The play was very well performed and also appeared new elements in the stage.

...it has appeared some cases of women that have killed their husbands...

---

DEFINITE

...because later could appear the real evidence and the real guilty.

...and usually appears the noble young man that either waste or has wasted his fortune.

In the main plot appear the main characters: Volpone and Mosca.
The psychological reality of the Unaccusative Hypothesis

- **Previous studies:**
  - L2 learners discriminate argument structure of unaccusative vs unergative Vs
  - they use this as a guiding principle to construct L2 mental grammars.

- **However:**
  - they have difficulty in determining the range of appropriate syntactic realizations of the distinction
  - this difficulty can persist into near-native levels of proficiency (see R. Hawkins 2001: 5.4).

CRUCIAL DIFFERENCE: these previous studies focused on ERRORS, thus emphasising the differences between native and non-native structures. By contrast, our study emphasises the similarities between native and non-native structures.