Abstract

“It can exist a better world”: L2 syntax acquisition at the interfaces – new corpus evidence

In this presentation I will discuss a frequent phenomenon in second language acquisition: postverbal subjects (VS order) are produced by learners of L2 English with different L1 backgrounds (e.g., Spanish, Italian, Japanese, Arabic): *It can exist a better world*; *It has appeared some cases of women that have killed their husbands*; *It will not exist a machine or something able to imitate the human imagination*.

The data in this talk were drawn from learner and native corpora: a corpus of English native speakers and two comparable corpora of Spanish and Italian learners of English.

Based on our review of the theoretical literature and previous research findings, it was hypothesized (Hypothesis 1, H1) that for VS order to occur in L2 English, the verb must be a type of intransitive known as unaccusative (e.g., exist, occur, appear, arrive), as the Unaccusative Hypothesis predicts. Additionally, (H2) predicts that the postverbal subject must be focus (new information), according to the End-focus Principle. Finally, (H3) states that there is a tendency for the postverbal subject to be phonologically heavy (long) as part of a general processing mechanism by which long and complex constituents tend to be placed towards the end of the clause (the End-weight Principle).

Importantly, while H1 has found confirmation in the L2 literature, H2 and H3 have, to our knowledge, been untested and ... as English natives, though learners ‘overuse’ the construction and show persistent errors in their syntactic encoding.

Thus, a full account of the production of inverted subjects in L2 English must look at properties which operate at three interfaces:

- the lexicon-syntax interface (Unaccusative Hypothesis).
- the syntax-discourse interface (End-focus Principle).
- the syntax-phonology interface (End-weight Principle).

The phenomenon in SLA

Production of postverbal subjects (VS order) in L2 English


- L1 Spanish/Italian/Arabic/Japanese – L2 English:
  1. *...it arrived the day of his departure* (L1 Spanish)
  2. *And then at last comes the great day.* (L1 Spanish)
  3. *In every country exist criminals* (L1 Spanish)
  4. *...after a few minutes arrive the girlfriend with his family too.* (L1 Arabic)
  5. *Sometimes comes a good regular wave.* (L1 Japanese)
  6. *...it happened a tragic event.* (L1 Italian)
AIM OF THE PRESENTATION (1)

SPECIFIC AIMS:
- To characterize the interlanguage of advanced learners of L2 English (L1 Spanish/Italian) by examining their production of VS structures.
- Source of data: large learner corpora.
- To confirm previous research:
  - Postverbal subjects appear only with a type of intransitives (unaccusatives).
- But previous research has ignored that unaccusativity is a necessary but not sufficient condition for postverbal subjects to be produced.
- We argue that the production of postverbal subjects is constrained at 3 interfaces:
  - Lexicon-syntax interface: unaccusative hypothesis
  - Syntax-phonology interface: end-weight principle
  - Syntax-discourse interface: end-focus principle
- So, a full account of L2 word order acquisition needs to take into account properties at these three interfaces.

AIMS OF PRESENTATION (2)

GENERAL AIMS:
- Current SLA literature: role of interfaces in acquisition
  - Syntax-semantics
  - Syntax-discourse
- General findings:
  - Formal (syntactic) features are acquired easily and from early stages in SLA.
  - Interface features are acquired late and pose persistent problems even at very advanced stages.

THEORETICAL BACKGROUND:

Word order

Postverbal subjects in SLA

- L1 Spanish/Italian/Arabic – L2 English:
  1. *...it arrived the day of his departure...
  2. *And then at last comes the great day.
  3. *In every country exist criminals
  4. *...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)
There arrived three girls.

Unaccusative Hypothesis
Lexicon-syntax interface
Burzio (1986), Levin & Rappaport-Hovav (1995), etc...

Well known in previous studies:
- L2 learners discriminate argument structure of unaccusative vs unergative verbs:
  - With different manifestations of unaccusativity: word order, interpretation of quantifiers, clitic climbing, auxiliary selection, etc.
  - With different L1 – L2 backgrounds (Japanese, Chinese, English, Spanish, Italian, etc)
- Learners use this knowledge as a guiding principle to construct L2 mental grammars.

However:
- Unaccusativity is a necessary but not sufficient condition for the production of postverbal subjects in L2 English.

Word Order in native English

Fixed SV(O) order

VS: Restricted use of postverbal subjects:

XP V S
(Inversion structures with an opening adverbial)

(6) a. [On one long wall] hung a row of Van Goghs.
   b. [Then] came the turning point of the match.
   c. [Within the general waste type shown in these figures] exists a wide variation.  
     [Biber et al. 1999: 912-3]

There-constructions

(7) a. Somewhere deep inside [there] arose a desperate hope that he would embrace her
   b. In all such relations [there] exists a set of mutual obligations in the instrumental and economic fields
   c. [There] came a roar of pure delight .... 
     [Biber et al. 1999: 945]

The psychological reality of the Unaccusative Hypothesis

Word Order in native English (VS)
Word Order Spanish/Italian

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

(9) a. Ha telefoneado María al presidente.
    has phoned Mary the president

   (transitive).

b. Ha hablado Juan.
    has spoken Juan

   (unergative).

c. Ha llegado Juan
    has arrived Juan

   (unaccusative).

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

(10) ¿Quién ha llegado/hablado? (Spa)
Who has arrived/spoken?

i. Ha llegado/hablado Juan
   has arrived/spoken Juan

ii. #Juan ha llegado/hablado
   #Juan has arrived/spoken

(11) Chi è arrivato/parlato? (Ital)
Who has arrived/spoken?

i. È arrivato/ Ha parlato Gianni
   has arrived/ has spoken Gianni

ii. #Gianni è arrivato/ha parlato
   #Gianni has arrived/has spoken

---

RESEARCH QUESTIONS

Main questions:

1) What are the conditions governing the production of VS structures in English?
   - We have just seen them.

2) Do learners of English produce inverted subjects (VS) under the same conditions as English natives do, regardless of problems to do with syntactic encoding (grammaticality)?
   - We are about to see.

3) Can these findings inform us about the interfaces in L2 acquisition?
   - We will see at the end.

   - COMPARATIVE FRAMEWORK: to answer these questions, we compared:
     - Study #1: Learner corpora vs. native corpora
     - Study #2: Learner corpora vs. learner corpora

---

Word Order in Spa / Ital (VS order)

1 Lexicon-syntax interface
   - No restrictions: postverbal subjects occur with all verb classes

2 Syntax-discourse interface
   - Postverbal subjects in Spanish and Italian are focus

3 Syntax-Phonological Form (PF) interface
   - Heavy subjects show a tendency to be postposed – a universal language processing mechanism: placing complex elements at the end of a sentence reduces the processing burden (J. Hawkins 1994).

Subjects which are focus, long and complex tend to occur postverbally, with no restrictions at the lexicon-syntax interface.

---

Corpora used in the 2 studies

- **ICLE**: International Corpus of Learner English
  - L1 Spa – L2 Eng
  - L1 Ital – L2 Eng

- **WriCLE**: Written Corpus of Learner English; Universidad Autónoma de Madrid (Rollinson, O'Donnell, Mendikoetxea, in progress)
  - http://www.uam.es/woslac
  - L1 Spa – L2 Eng

- **LOCNESS**: Louvain Corpus of native English Essays, UCL/CECL, Louvain-la Neuve
  - English native speakers
Hypotheses

GENERAL HYPOTHESIS: Conditions licensing VS in L2 Eng are the same as those in native Eng, DESPITE differences in syntactic encoding (ungrammatical sentences).

- **H1 [LEXICON]: Lexicon-syntax interface**: 
  - Postverbal subjects with unaccusatives (never with unergatives)  
  Known from previous research

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

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

Corpora

- **Corpora**:  
  - L1 Spa – L2 Eng  
  - Eng natives

<table>
<thead>
<tr>
<th>Learner corpora</th>
<th>Native corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLE-Spanish</td>
<td>200,376</td>
</tr>
<tr>
<td>WriCLE</td>
<td>63,836</td>
</tr>
<tr>
<td>LOCNESS Alevels</td>
<td>60,209</td>
</tr>
</tbody>
</table>

Total no. of words: 264,212

**Query software**: WordSmith v. 4.0 (Scott 2004)

<table>
<thead>
<tr>
<th>Learner</th>
<th>Verb type</th>
<th>Usable concordances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unerg</td>
<td>181</td>
<td></td>
</tr>
<tr>
<td>Unac</td>
<td>820</td>
<td></td>
</tr>
<tr>
<td>Native</td>
<td>185</td>
<td></td>
</tr>
<tr>
<td>Unac</td>
<td>219</td>
<td></td>
</tr>
</tbody>
</table>

**TOTAL**: 1905
H1 results: syntax-lexicon

Table 1: Frequency of postverbal subjects produced

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Verb type</th>
<th>Postverbal subjects</th>
<th>Usable concordances</th>
<th>% Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner</td>
<td>Unac</td>
<td>0</td>
<td>181</td>
<td>0%</td>
</tr>
<tr>
<td>Native</td>
<td>Unac</td>
<td>58</td>
<td>820</td>
<td>7.1%</td>
</tr>
<tr>
<td>Learner</td>
<td>Unac</td>
<td>0</td>
<td>185</td>
<td>0%</td>
</tr>
<tr>
<td>Native</td>
<td>Unac</td>
<td>16</td>
<td>719</td>
<td>2.3%</td>
</tr>
</tbody>
</table>

H1: Unaccusative: grammatical vs ungrammatical VS

- According to structure type

H1: Syntax-lexicon
Ns vs NNs: Verbs in VS structures
**H2 results: syntax-phonology**

![Diagram showing the weight of words for SV and VS speakers, with learners and natives compared.]

**H2: measuring weight**

Table 1: A syntactic scale for measuring syntactic weight

<table>
<thead>
<tr>
<th>SYNTACTIC WEIGHT</th>
<th>SYNTACTIC STRUCTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOMINAL SCALE</td>
<td>ORDINAL SCALE</td>
</tr>
<tr>
<td>LIGHT</td>
<td>HEAVY</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(D) N</td>
<td>(D) ADJ N</td>
</tr>
<tr>
<td>PRN</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>(D) ADJ N*</td>
<td>(D) ADJ N</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(D) ADJ N PP*</td>
<td>(D) ADJ N</td>
</tr>
<tr>
<td>(D) ADJ N PP*</td>
<td>(D) ADJ N PP*</td>
</tr>
<tr>
<td>HEAVY</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>(D) ADJ N* PP</td>
<td></td>
</tr>
<tr>
<td>Notes: (i) The asterisk (*) represents a complex (i.e., recursive) categorical or phrasal structure. (ii) Parentheses indicate the optional realization of the bracketed category or phrase.</td>
<td></td>
</tr>
</tbody>
</table>

**Examples H2: syntax-phonology**

**SV: typically LIGHT (Pronoun, D + N)**

Learners: ...but they may appear everywhere. 
...since the day eventually came...

Natives: These debates began over two decades ago. 
...a great controversy exists over the topic.

**VS: typically HEAVY (postmodification)**

Learners: Against this society drama emerged an opposition headed by Oscar Wilde and Bernard Shaw.
...exists yet in Spain a group of people who are supposed to be professional soldiers.

Natives: With this theory also came the area of quantum mechanics. Thus began the campaign to educate the public on how one contracts aids.
H3: syntax-discourse

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Weight</th>
<th>SV order</th>
<th>VS order</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learner</td>
<td>Light</td>
<td>65/96 (67.7%)</td>
<td>11/58 (19.0%)</td>
</tr>
<tr>
<td></td>
<td>Heavy</td>
<td>31/96 (32.3%)</td>
<td>47/58 (81.0%)</td>
</tr>
<tr>
<td>Native</td>
<td>Light</td>
<td>62/91 (68.1%)</td>
<td>3/16 (18.8%)</td>
</tr>
<tr>
<td></td>
<td>Heavy</td>
<td>29/91 (31.9%)</td>
<td>13/16 (81.3%)</td>
</tr>
</tbody>
</table>

Examples H3: syntax-discourse

**VS: FOCUS**

**Learners:** The existence of many, and let’s say, misleuqns programmes en[|]ourages people to keep watching TV. In Spain we h[a]ve four different channels, and some provinces have their own channels. Furthermore there also exists a wide variety of optional channels which have to be paid.

**Natives:** Humanity witnessed one of histories (recorded history) most incredible minds at work when Albert Einstein came onto the scene. Although his theory (his and his wife’s) was basically scientific in nature, it can and has been applied to all areas of human existence. The theory i’m speaking of is relativity. With this theory also came the area of quantum mechanics.

**SV: typically TOPIC**

**Learners:** 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...

**Natives:** However, Hugo is not prepared ... Louis took such exception to Hugo ... Hugo came from a bourgeoisie family.

Conclusion

<table>
<thead>
<tr>
<th>Interfaces:</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Lexicon-syntax</td>
</tr>
<tr>
<td>● Syntax-discourse</td>
</tr>
<tr>
<td>● Syntax-phonology</td>
</tr>
</tbody>
</table>

**Contingency Table**

<table>
<thead>
<tr>
<th>Information status</th>
<th>Light</th>
<th>Preverbal</th>
<th>---</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>Heavy</td>
<td>---</td>
<td>Postverbal</td>
</tr>
</tbody>
</table>
Conclusion (2)

- These results confirm that Spanish L2 learners of English produce postverbal subjects under exactly the same 3 interface conditions as in L1 English (unaccusativity being a necessary but not a sufficient condition).
  - Unaccusativity Hypothesis: postverbal subjects appear with unaccusatives.
  - End-weight principle: postverbal subjects tend to be long and complex.
  - End-focus principle: postverbal subjects tend to be focus.

- So, learners do not show a pragmatic deficit at the syntax-discourse interface.

- Learners show rather a persistent problems in the syntactic encoding of the construction → syntactic deficit
  - High production of ungrammatical examples (it-insertion, Ø-insertion).
  - Spanish learners overuse the construction and show a lexical bias for the V exist.
  - Example: "... it will not exist a machine or something able to imitate the human imagination.

Next question

- Our results: Language specific or universal pattern or just random behaviour?
- Are other learners guided by the same 3 principles?
- Second corpus study
  - L1 Italian – L2 Eng vs. L1 Spa – L2 Eng (Lozano & Mendikoetxea 2008): same results
    - Unaccusativity Hypothesis: postverbal subjects appear only with unaccusatives.
    - End-weight principle: postverbal subjects tend to be long and complex.
    - End-focus principle: postverbal subjects tend to be focus.

- Also: evidence from L1 French – L2 Eng (unpublished results yet).

CORPUS STUDY #2

V-S structures in:
- L1 Spa – L2 Eng
- L1 Ital – L2 Eng


Corpus Study #2

Italian/Spanish L1 - English L2

- Main question:
  Do learners (with different L1s) produce postverbal subjects under the same conditions as Eng natives do, irrespective of problems with their syntactic encoding (grammaticality)?

- 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.
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>200,376</td>
</tr>
<tr>
<td>ICLE Italian</td>
<td>392</td>
<td>227,085</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>643</strong></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></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><strong>TOTAL</strong></td>
<td><strong>1510</strong></td>
<td></td>
</tr>
</tbody>
</table>

H1 results: syntax-lexicon

<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></td>
<td>Unaccusative</td>
<td>52</td>
<td>640</td>
<td>52/640 (8.1%)</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>15</td>
<td>574</td>
<td>15/574 (2.6%)</td>
</tr>
</tbody>
</table>

H1: Unaccusative VS:
grammatical vs ungrammatical

- **Locative inversion:**
  - In some places still exist popularly supported death penalty. (L1 Spa)
  - …on the earth lived people which were born-criminal. (L1 Ital)
- **There-insertion:**
  - …there also exists a wide variety of optional channels which have to be paid. (L1 Spa)
  - …there still remains a predominance of men over women. (L1 Ital)
- **AdvP-insertion:**
  - Then come the necessity to earn more… (L1 Spa)
  - Later came a world of disorder… (L1 Ital)
  - **’it’-insertion:**
    - “In the name of religion it had occurred many important events …” (L1 Spa)
    - “…g still live some farmers who have field and farmhouses. (L1 Ital)
  - **’cf’-insertion:**
    - “…exist volunteers with such a feeling against it” (L1 Spa)
    - “…exists factors which, on long term, can predispose human mind to that crime…” (L1 Ital)
  - **’XP’-insertion:**
    - “…and from this moment begins the avarice” (L1 Spa)
    - [no instances found in Italian corpus]

Result: Unaccusative:
Type of VS structures
Result H1: VS and specific unaccusative verbs L1 Spa vs L1 Ital

H2 results: syntax-phonology

Examples H2: syntax-phonology

SV: typically LIGHT
...these people should exist, ...
Violence does exist...
(L1 Spa)

VS: typically HEAVY
...it will not exist a machine or something able to imitate the human imagination (L1 Spa)
...emerges the people’s ignorance in having prejudices
(L1 Ital)

H3: syntax-discourse

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**

In the world, dominated by science, technology and industrialisation, there is no place for dreaming and imagination. Thanks to science and its consequences, technology and industrialisation, appeared the big factories and the capitalism system. (L1 Spa)

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. (L1 Ita)

**SV: typically topic**

The approval of acting of women were something essential. Women started to perform female characters and this contribute to give a sexual and realistic atmosphere. [...] Female characters appear with a stronger personality they really love these men. (L1 Spa)

The idea of Europe doesn’t ignore these differences, but inglobes them, accept them and upon them construct its identity. [...] If I think of the concept of Europe I cannot think of anything else that of a whole of different countries, but that all together produce the European identity. The differences have always existed in the Europe and for ages its peoples fought one against the other. (L1 Ita)

---

**OVERALL PICTURE CORPUS DATA**

V-S structures in:
- L1 Spa – L2 Eng
- L1 Ital – L2 Eng
- L1 Fre – L2 Eng

---

**Conclusion**

**Interfaces:**
- Lexicon-syntax
- Syntax-discourse
- Syntax-phonology

**NNS vs. NS: comparisons across different NNS**

- Overall picture corpus data
- Frequency (%) of VS production
CONTRIBUTIONS TO THE FIELD

- **Linguistic Theory**: better understanding of interfaces
  - lexicon-syntax
  - syntax-discourse
  - syntax-phonology

- **L2 acquisition**: better understanding developmental/universal phenomena and L1 transfer.

- **Corpus studies**: use of corpora for the study of formal features. Creation of *Spanish* learner corpus (CEDEL2).

- **Converging evidence**: use of naturalistic (corpus) data and experimental (AGT) data.

CURRENT RESEARCH PROJECTS

- **Universidad de Granada**
  - Psycholinguistics: ongoing study
  - Processing of *information structure* in Spa-Eng bilinguals
  - Cleft sentences: “It was John who broke the glass (not Mary)”
  - ERP technique

- **Universidad Autónoma de Madrid**
  - New research project on *optionality* in L2 acquisition
  - CEDEL2 corpus: L1 Eng – L2 Spa
  - WriCLE corpus: L1 Spa – L2 Eng
  - Interfaces: syntax-discourse
  - Corpus data and quasi-experimental data
  - Our own tagging and concordance software: UAM CorpusTool
• REFERENCES (2)


