ON THE ACQUISITION OF OVERT SUBJECTS, TOPICS AND \textit{WH}-QUESTIONS IN SPANISH

JULIO VILLA-GARCÍÁ AND WILLIAM SNYDER

1. Introduction

The principal objective of this investigation is to shed new light on the much-debated analysis of overt, lexical subjects in Mainland Spanish (henceforth, Spanish) from the standpoint of First Language Acquisition. This will be achieved by providing acquisitional evidence gathered from four longitudinal case studies of children acquiring Spanish in Spain, which will be brought to bear on the choice of analysis of the adult grammar regarding overt subjects. More specifically, this acquisitional study aims to illuminate the longstanding lack of consensus over whether subjects in languages like Spanish are located in the same structural position as subjects in languages like English, i.e., in [Spec, IP/TP]/[Spec, AgrSP], or whether they are CP-related phenomena in the left periphery/CP. For this purpose, the paper investigates the time-course of acquisition of overt subjects and unambiguously CP-related phenomena such as \textit{wh}-questions and topicalizations.

Our results suggest that lexical subjects emerge on a par with topicalization in Spanish, providing novel acquisitional support for the prominent view championed by Ordóñez (1997) that subjects in \textit{pro}-drop languages like Spanish are left-peripheral phenomena located in the CP field. In contrast to the findings of Grinstead (1998, 2004), Spinner & Grinstead (2006) and Grinstead & Spinner (2009), however, our results indicate that overt subjects begin to appear significantly earlier than \textit{wh}-questions in child Spanish. Be that as it may, we submit that a weaker version of Grinstead’s “interface delay” hypothesis can be maintained.

The paper is organized as follows: Section 2 offers a cursory look at different analyses of Spanish overt subjects proposed in the literature; Section 3 summarizes a number of existing L1 acquisition studies looking at the acquisition of subjecthood in Spanish; Section 4 is devoted to the methods employed in this study; Section 5 is concerned with the results and discussion; Section 6 offers some concluding remarks.
2. The view from syntax

Mainland Spanish is a paradigmatic pro-drop language in which subjects can be left unpronounced and ‘free subject inversion’ is readily available, as shown in (1):

(1) a. María llamó [SV]
    Mary  called

b. Llamó María [VS]
    called Mary

Both: ‘Mary called.’

The standard analysis of subjects in pro-drop Romance languages put forward by Rizzi (1982) assumes that preverbal subjects, whether overt or null, sit in [Spec, IP/TP], in much the same way as in English. On this view, postverbal subjects (cf. (1b)) result from a rule moving the subject to the right, adjoining it to the VP, [Spec, TP] being filled by pro. Authors who currently pursue the hypothesis that preverbal subjects in Spanish sit in [Spec, TP] include Ortega-Santos (2006), inter alia.

However, research spanning more than a quarter of a century has shown that the [Spec, TP]/EPP analysis of subjects may be empirically inadequate. It has been demonstrated that the word-order possibilities in (1) cannot be accounted for by means of a quintessentially syntactic parameter; other factors such as the type of predicate involved, the categorical-thetic distinction, and discourse pragmatic considerations such as topic and focus seem to play a key role in determining word order configurations in Spanish. Thus, an influential venue of research hypothesizes that overt subjects in languages like Spanish are not grammatical subjects per se, but rather discourse-sensitive constituents in a left-peripheral/CP specifier. Proponents of this view include Contreras (1991), Olarrea (1996), Ordóñez (1997), Grinstead (1998) and Barbosa (2009), among many others. The technical implementations and theoretical assumptions of each of the existing analyses naturally differ: for certain authors, preverbal subjects are base-generated clitic-left-dislocated elements, and for others they are moved topics, whereas under certain accounts postverbal subjects are in situ or else they are located in a (low or high) Focus position, in line with the focus-last requirement in Spanish.

An additional issue contributing to the above-mentioned debate concerns that analysis of postverbal subjects (e.g., VS/VSO/VOS). Aside from Rizzi’s classical account, the VOS order can be derived by assuming that it results from scrambling of the object, remnant movement of the TP over the subject in FocusP, remnant movement of the VP over the subject in the VP periphery, p(rosodic)-movement out of focus position past the subject, or LF movement of the subject to [Spec, TP] (assuming a weak-strong features/early minimalism type of framework), to
mention just a few. The reader is referred to Ortega-Santos (2006) and references therein for discussion of existing proposals.

In the middle of a spectrum polarized by the two dissenting views presented above, there is a third type of analysis which assumes that overt preverbal subjects can occupy different positions (e.g., either a CP position or the traditional [Spec, TP] position), as argued by Casielles (2001) and Camacho (2006), among others.

In sum, there is disagreement in the literature as to whether subjects in Spanish are located in [Spec, TP] (cf. (2a)), or whether they are CP-related, left-peripheral elements in the CP layer (cf. (2b)). Thus, a Spanish declarative sentence like *Mikel visitó Lisboa* ‘Mikel visited Lisbon’ can be assigned (at least) two derivations, as shown by the simplified labelled bracketings in (2):

\[
\begin{align*}
\text{(2) a. } & \text{[TP Mikel [T visitó] [vP ... [VP Lisboa]]]} \\
\text{b. } & \text{[CP/TopicP Mikel [C/Topic Ø] [TP ... [T visitó] [vP ... [VP Lisboa]]]]}
\end{align*}
\]

### 3. The view from child language acquisition

As far as the emergence of overt subjects in child Spanish is concerned, Grinstead (1998, 2004), drawing on the seminal work of Ordóñez (1997), has advanced the “interface hypothesis,” according to which children acquiring null-subject languages like Spanish do not produce overt subjects at an earlier stage, because overt subjects are located in a topic/focus position in the CP layer, to which the child has no access in production from the beginning. This may be due to a delay in the development of the grammar-pragmatics interface. In this connection, the acquisitional prediction derived from the theoretical claim that overt subjects are topics or *foci* in the CP domain is that overt subjects should start to appear concurrently with less controversially CP-related constituents such as fronted objects and *wh*-questions in child Spanish-style null-subject languages, once the syntax-pragmatics interface is accessible to the child.

In order to test this prediction, Grinstead (1998, 2004) analyzed data from children learning Catalan and Mexican-Spanish. It is important to note that the results reported in Grinstead (1998) are solely based on chronological proximity of onset of the relevant constructions. The Grinstead (2004) study, for its part, provides supporting statistics, though no *wh*-questions were found in the corpora of the Spanish-speaking children analyzed, so the conclusion of the investigation is essentially based on the behavior of the Catalan-acquiring children of the study. Yet, Grinstead & Spinner (2009) tackle this issue and provide data about the acquisition of *wh*-questions in (Mexican) Spanish. The results of the authors’ statistical analysis reveal that whereas for one child the *p*-value was not significant, it was so for another child; the same occurred for one of the Catalan-acquiring
children studied by Grinstead (2004). In the case of one of the three Spanish-acquiring children studied by Grinstead & Spinner (2009), the first uses of wh-questions appeared in the same transcript as the first occurrence of overt subjects, suggesting that both constructions emerged together. In checking for the concurrent acquisition of subjects and fronted objects, Grinstead & Spinner (2009) found two null results and one significant $p$-value.

Despite these arguably inconclusive results, Grinstead (1998, 2004) and Grinstead & Spinner (2009) maintain that in null-subject Southern Romance languages, overt subjects and unambiguously CP-related constituents such as wh-questions and fronted objects emerge concurrently in development. However, Grinstead’s (1998, 2004) strong prediction implies that no single child should acquire overt subjects significantly earlier than unambiguously left-peripheral phenomena like wh-questions, which casts doubt on the interpretation of the results yielded by the investigations at issue. Consequently, the current study presents novel Spanish acquisitional data with an eye on further testing the “interface delay” hypothesis advocated by John Grinstead.

### 4. Research methods and data collection

As shown in Table 1, the corpus data utilized were retrieved from the CHILDES database (Child Language Data Exchange System (MacWhinney 2000)) and from the CLESS database (the University of Connecticut’s Cross-Linguistic Early Syntax Study project).

<table>
<thead>
<tr>
<th>Child</th>
<th>Corpus</th>
<th>Database</th>
<th>Date of Retrieval</th>
<th>Age Span</th>
<th># of Trans. and Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emilio (♂)</td>
<td>Vila (Vila 1984)</td>
<td>CHILDES</td>
<td>March 7, 2009</td>
<td>0:11.09-2:11.24</td>
<td># = 27; monthly/biweekly</td>
</tr>
<tr>
<td>Inés (♀)</td>
<td>Inés</td>
<td>CLESS</td>
<td>April 1, 2009</td>
<td>1:02.00-2:02.11</td>
<td># = 36; (bi)weekly</td>
</tr>
<tr>
<td>Irene (♀)</td>
<td>Llinàs-Grau/Ojea</td>
<td>CHILDES</td>
<td>October 16, 2008</td>
<td>1:05.27-2:04.13</td>
<td># = 31; biweekly</td>
</tr>
<tr>
<td>Juan (♂)</td>
<td>Linaza</td>
<td>CHILDES</td>
<td>March 10, 2009</td>
<td>1:07.02-3:09</td>
<td># = 18; mostly monthly</td>
</tr>
</tbody>
</table>

Table 1: Longitudinal data
The data at hand were counted manually, since currently there is no reliable computer-assisted method to find occurrences of the constructions of interest. In so doing, imitations, repetitions, and lexicalized expressions were not included. All sentences containing overt subjects along with any instances of fronting phenomena including topics and wh-questions were tabulated. Note that the notion of “topic” adopted here is that of a fronted element entering the topic-comment articulation (old vs. new information). This category includes uncontroversial cases of topicalization in Romance, expressed by Cinque’s (1990) Clitic Left Dislocation (CLLD) construction—which co-occur with a coreferent clitic—, clearly topicalized subjects appearing before a wh-item, and subjects in front of bona fide examples of topics or overt complementizers.

The measure of acquisition employed in this study was first clear use, followed soon after by additional, distinct uses (cf. Stromswold 1996).

The statistical method used to check for concurrent acquisition was the Binomial Test, as outlined in Snyder (2007, Ch. 5). This method is an exact test of the statistical significance of deviations from a theoretically expected distribution of observations into two categories. The question which this statistical method aims to answer concerns the likelihood of a given outcome (e.g., several uses of overt subjects (construction A), before the first occurrence of a bona fide topic (construction B)), under the null hypothesis that construction B was available to the child as early as A, and had the same relative frequency of use as in later transcripts (e.g., the next ten transcripts after the second construction appeared). In other words, the Binomial Test addresses the question of whether the apparent gap between two given constructions is plausibly due to a lower frequency of use for the construction emerging later, or whether there is in fact a statistically significant difference between the two, as expected when the two constructions demand different prerequisites that the child needs to have prior to using the particular constructions successfully.

5. Results and discussion

This section presents and discusses the results of the investigation, organized according to the two major predictions of this study, namely the concurrent acquisition of overt subjects and topicalizations, and the simultaneous emergence of lexical subjects and wh-questions.

5.1. The timing of acquisition of subjects and topics
Regarding overt subjects and topicalizations, the findings reported in Table 2 show that, attending to chronological age, the former arise in the spontaneous speech of all the children under consideration before the first clear uses of topicalization. It is well known that language acquisition proceeds very quickly at the ages with which this study is concerned. In any single week it is entirely possible—indeed probable—that the child's grammar will change in some respect. Nonetheless, it is not the case that everything changes all at once. The problem, then, is that chronological age, and differences in age, tell us extremely little about the actual process of language acquisition. The Binomial Test provides a solution: The unit of measurement now becomes, not days or weeks, but rather occurrences of a particular grammatical construction (Construction A). A significant difference by Binomial Test thus tells us that the age difference was “meaningful.”

<table>
<thead>
<tr>
<th>Child</th>
<th>Onset of Overt Subj.</th>
<th>Onset of Topics</th>
<th># of Earlier Constr.</th>
<th>Relative Frequency Ov. S. Top.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emilio</td>
<td>1;09.19</td>
<td>2;03.01</td>
<td>15 (Ov. S.)</td>
<td>144</td>
<td>21</td>
</tr>
<tr>
<td>Inés</td>
<td>1;06.05</td>
<td>1;09.03</td>
<td>26 (Ov. S.)</td>
<td>117</td>
<td>8</td>
</tr>
<tr>
<td>Irene</td>
<td>1;07.05</td>
<td>1;11.13</td>
<td>47 (Ov. S.)</td>
<td>222</td>
<td>6</td>
</tr>
<tr>
<td>Juan</td>
<td>1;09.02</td>
<td>2;03</td>
<td>14 (Ov. S.)</td>
<td>90</td>
<td>12</td>
</tr>
</tbody>
</table>

Table 2: Results of the statistical analysis. Overt subjects and topics

The null results of the statistical analysis in Table 2 suggest that it is possible to retain the claim made by Grinstead and colleagues that overt subjects and topics emerge together in child Spanish. The fact that subjects appear chronologically earlier than topics may well be due to the lower frequency of topics versus overt subjects. Such a difference in frequency might arise because the child is copying the relative frequencies in the adult input, for example, or because the child’s immature processing system finds the second construction more difficult. The findings in Table 2 thus provide weak evidence against the null hypothesis that the two structures emerge together. Therefore, the hypothesis advanced by Grinstead and colleagues that overt subjects and topics appear concurrently in development in the acquisition of Spanish is substantiated by the data available. If correct, these results constitute novel acquisitional support for the preponderant view that subjects in languages in Spanish have topic-like properties, and that they are
located in an A’-position in the CP (cf. Ordóñez 1997, *inter alia*). The following is an example of early topics produced by one of the children at issue.

(3) Mami, eso, ¿qué es?  
‘Mum, as for that, what is it?’ [Irene, 1;11.30]

Lastly, as Grinstead notes, these results would be strengthened and their interpretation would become more clear-cut if a significant discrepancy were found in languages/dialects where concurrent appearance of lexical subjects and topics is not predicted. This prediction is in fact borne out for German and Caribbean Spanish, as argued by Grinstead & Spinner (2009) and Villa-García et al. (2010), respectively.

### 5.2. The timing of acquisition of subjects and *wh*-questions

The second uncontroversially CP-related phenomenon considered in this study is *wh*-questions, and what follows addresses the question of whether there is a contingency between lexical subjects and *wh*-questions in acquisition. The following child utterance illustrates *wh*-questions in child Spanish:

(4) ¿Cómo se llama esto?  
‘What is this called?’ [Emilio, 2;03.01]

The results summarized in Table 3 falsify Grinstead’s (1998, 2004), Grinstead & Spinner’s (2006) and Grinstead & Spinner’s (2009) claim that overt subjects emerge at the same time as *wh*-questions in development in child Spanish, since not only do overt subjects appear chronologically earlier than *wh*-questions, but also the *p*-values yielded by the Binomial Test are significant in three cases. A significant result by Binomial Test directly contradicts an explanation based on the lower frequency of the construction appearing later (cf. Section 5.1.), and instead supports a grammar-based hypothesis: the child had to learn some new information, beyond what she had to know for the first construction, before her grammar allowed her to produce the second construction (i.e., *wh*-questions).
On the Acquisition of Overt Subjects, Topics and Wh-Questions in Spanish

<table>
<thead>
<tr>
<th>Child</th>
<th>Onset of Overt Subjects</th>
<th>Onset of Wh-Qs</th>
<th># of Earlier Constr.</th>
<th>Relative Frequency Ov. S.</th>
<th>Wh-Qs</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emilio</td>
<td>1;09.19</td>
<td>2;03.01</td>
<td>14 (Ov. S.)</td>
<td>144</td>
<td>48</td>
<td>p = .018 significant</td>
</tr>
<tr>
<td>Inés</td>
<td>1;06.05</td>
<td>1;09.03</td>
<td>36 (Ov. S.)</td>
<td>141</td>
<td>6</td>
<td>p = .223 not significant</td>
</tr>
<tr>
<td>Irene</td>
<td>1;07.05</td>
<td>1;08.26</td>
<td>33 (Ov. S.)</td>
<td>143</td>
<td>32</td>
<td>p = .001 significant</td>
</tr>
<tr>
<td>Juan</td>
<td>1;09.02</td>
<td>2;03</td>
<td>23 (Ov. S.)</td>
<td>90</td>
<td>13</td>
<td>p = .045 significant</td>
</tr>
</tbody>
</table>

Table 3: Results of the statistical analysis. Overt subjects and wh-questions

The findings above provide grounds to refute the hypothesis that once the child has access to the syntax-pragmatics interface, she starts to produce overt subjects, wh-questions, and topics at the same time. In other words, our results weaken Grinstead’s contention that all CP-related phenomena arise simultaneously at a certain point in development, prior to which the child does not have access to the interface between syntax and information structure. As has been pointed out, Grinstead’s (1998, 2004) predictions imply that no single child should acquire overt subjects prior to unambiguously left-peripheral phenomena such as wh-questions. An important question posed by the state of affairs in Table 3 is the manner in which the Spanish-acquiring child’s grammar develops with regard to the constructions of interest. Three possibilities will be considered, namely:

- Possibility 1: the prerequisites for overt subjects and for wh-questions constitute non-overlapping sets, that is, the two structures have completely different prerequisites.
- Possibility 2: the linguistic prerequisites for overt subjects are a proper subset of those for wh-questions, which strongly predicts an ordering effect in acquisition (Snyder 2007).
- Possibility 3: overt subjects and wh-questions share certain prerequisites (i.e., the prerequisites are contained in the intersection of the two sets), wh-questions crucially necessitating additional prerequisites that the child typically acquires later on.

The third option allows us to retain Grinstead’s original analysis, mutatis mutandis.
6. Conclusions

The findings of this paper are compatible with the claim that children learning Mainland Spanish begin to use overt subjects on a par with topicalization, which supports the theoretical claim that overt subjects in null-subject languages like Spanish are CP-related phenomena in a left-peripheral position above TP. However, the results of the present paper point to the conclusion that overt subjects emerge significantly earlier than wh-questions, which refutes the claim made by Grinstead (1998 et seq.) that the acquisition of subjects and wh-questions proceeds in parallel. In this sense, we have put forward the suggestion that overt subjects and wh-questions may have certain prerequisites in common, wh-questions critically having extra prerequisites that the child usually acquires later on. This suggestion can be extended to topicalizations if future studies eventually encounter significant results when checking for the concurrent acquisition of overt subjects and topics, as was the case in some of the children analyzed by Grinstead and colleagues. Hence, this paper ultimately argues for a weaker version of Grinstead’s “interface delay” hypothesis in the acquisition of pro-drop languages like Spanish.

Acknowledgments

This paper has benefited enormously from the comments of Eva Bar-Shalom, Jonathan Bobaljik, Željko Bošković, Eugenia Casielles, John Grinstead, Diane Lillo-Martin, Mireia Llinàs-Grau, Ana de Prada Pérez, Koichi Ohtaki, Antxon Olarrea, Iván Ortega-Santos, Yael Sharvit, and Karen Zagona, as well as the observations made by the audiences at UUSLAW Smith College 2008 and GALA Lisbon 2009, especially Nina Hyams, Letitia Naigles and Tom Roeper. This investigation owes much to the CLESS project at the University of Connecticut. This research has been supported in part by National Institute of Health grant #NIDCD 00183 to Diane Lillo-Martin and the second author. Moreover, this investigation has also been partially funded by the Full Graduate Fellowship “la Caixa,” awarded to the first author by the savings bank “la Caixa” (Barcelona, Spain). The usual disclaimers apply.

References


Ordóñez, Francisco (1997) Word Order and Clausal Structure of Spanish and Other Romance Languages, Ph.D. Dissertation, CUNY.


Villa-García, Julio and Neda Todorović (in preparation) “What will the Serbo-
Croatian-acquiring child use first, SV or VS?” Ms., University of Connecticut, 
Storrs.
Bearing on the Controversial Account of Preverbal and Postverbal Subjects in 
Spanish,” Ms., University of Connecticut, Storrs.
Villa-García, Julio, William Snyder, and José Riqueros-Morante (2010) “On the 
analysis of lexical subjects in Caribbean and Mainland Spanish: Evidence from 
Somerville, Mass.: Cascadilla Press.

Notes

1 Binomial Test \( p = \frac{X}{X + Y} \)
\( p \) stands for ‘p-value;’ \( X \) corresponds to the number of times construction A appears in the 
ten transcripts following the first clear use of B; \( Y \) stands for the times B occurs in the ten 
transcripts after the first use of B; \( Z \) refers to the uses of A before the first clear use of B.
2 Note that preverbal and postverbal subjects emerge simultaneously in child Spanish, as 
shown by Grinstead (1998, 2000) for Mainland Spanish, Villa-Garcia & Snyder (in 
preparation) for Mainland and Caribbean Spanish, and Villa-García & Todorović (in 
preparation) for Serbo-Croatian, a null-subject language of the Spanish/Italian type. For an 
opposing view, see Casielles et al. (2006).
3 In this study, we followed John Grinstead and colleagues in considering only those 
elements which are unambiguous cases of topicalization, disregarding the rather scarce 
instances of foci. It would be reasonable to compare the appearance of overt subjects and 
foci, since postverbal subjects usually bear a focal interpretation (cf. Section 2). More 
generally, it would be far from implausible to explore the acquisition of overt subjects and 
non-wh-fronting phenomena like topicalization and focalization. By way of illustration, if 
we run the Binomial Test for Irene’s data regarding overt subjects and non-wh-fronting 
phenomena, the result is still not significant \( (p = .732) \), consistent with the results for topics 
presented in the main text. Future research will determine whether this preliminary result is 
on the right track.