Subjects, Topicalizations and Wh- Questions in Child German and Southern Romance

Patti Spinner and John Grinstead
University of Pittsburgh and The Ohio State University

1. Introduction

One of the functions of child language research, where possible, should be to tell us something about adult theories of language. In what follows, we continue an argument defending the position, based on child language evidence, that preverbal overt subjects are left-peripheral, clause-external constituents in the adult Southern Romance languages: Catalan, Spanish and Italian. We take this position within the context of a debate in adult syntactic theory, in which proponents of a traditional analysis argue that overt, and especially pre-verbal, subjects in Southern Romance are clause-internal Inflectional Phrase (IP) type constituents (Belletti, 1990; Cardinaletti & Starke, 1996; Goodall, 2002; Rizzi, 1990), while others propose analyses which take preverbal subjects to be topicalized or focused Complementizer Phrase (CP) type constituents (Ordóñez, 1997; Poletto, 2000; Zagona, 2000; Zubizarreta, 1998) and still others argue that there is more than one syntactic position for overt, preverbal subjects and that both "CP" and "IP" type subject positions exist (Camacho, 2005b; Costa & Galves, 2000; Suñer, 2003).

A second function of research into child language is to explain how children's grammars become adult grammars. In earlier work, Grinstead (2000, 2004) observed in child Spanish and Catalan that there is an early stage in which children do not seem to use overt subjects at all. This observation was explained as the result of two and three year-old children's failure to make syntactic use of their discourse-pragmatic understanding of new vs. old information. This failure implies that while these children may have access to the CP domain of their clause structure, they simply fail to use it because they assume that their interlocutors share their discourse knowledge. In this way, the observed difference in syntactic behavior of child and adult speakers of Southern Romance languages was explained as a function of a well-known, cross-linguistically documented phenomenon, namely, the problematic interface between syntax and discourse-pragmatics in child language (Avrutin, 1995; Schaeffer, 2000).

The basic observation that overt subjects failed to be used early on was combined with the observation that wh- questions and fronted objects – less controversially constituents of CP - begin to be used at close to the same point in longitudinally collected corpora as do overt subjects. Anecdotal evidence from child Dutch (van Kampen, 1997) was adduced to make the point that while these three constructions began to be used at the same time in Catalan, they did not begin to be used at the same time in all languages, and specifically not in a language in which overt subjects must be used as a function of finiteness, such as Dutch. This contrasts with Catalan, Spanish and Italian in which overt subjects are used as a function of discourse-pragmatics. The observation that the three constructions alluded to begin to be used at the same time in languages like Catalan would be much stronger were we able to demonstrate empirically that these three constructions do not emerge at the same time in the development of languages like Dutch, or German.

In what follows, we will attempt to show, using case studies and longitudinally-collected spontaneous child German data, that overt subjects in child German begin to be used significantly earlier than do fronted objects or wh- questions. We take this to be evidence of the distinct nature of the subject constituent in the two language families and further that children acquiring both languages converge on this fundamental property of the adult grammars even before the interface between syntax and discourse pragmatics is operational.
2. Interface Delay and Left Peripheral Delay

There is evidence that the integration of multiple domains of cognition may slow the development in children of behaviors which depend on that integration. For example, English-speaking children learn to count using language in an adult-like way at around three or four years of age; however, they are able to use non-verbal number from infancy (Cordes et al., 2001; Gallistel & Gelman, 1992; Gelman & Gallistel, 2004; Wynn, 1990, 1992). Is it the case that infants lack numerical cognition when they are still prelinguistic? Studies of infant numerical cognition would seem to cast doubt on such an idea (Xu & Spelke, 2000; Xu et al., 2005). Following this example, we propose that discourse-pragmatics and syntax are slow to integrate because the interface between the two domains of cognition itself is slow to develop. This hypothesis is distinct from other theories which posit that discourse-pragmatics is somehow underdeveloped in children. For example Schaeffer (2000) attributes child Dutch speakers' failure to scramble direct objects in an adult-like way to an "immature pragmatic system" while Avrutin (1994) attributes children's difficulties with pronoun interpretation to their "limited inferential resources". Another possibility, of course, is that the problem lies not in children's syntactic abilities, not in their pragmatic abilities and not in the interface between the two, but rather in their ability to take the perspective of their interlocutors (including their interlocutors' discourse-pragmatics representations) into account. This is commonly referred to as Theory of Mind (Baron-Cohen, 1995; Wimmer & Perner, 1983) and it is conceptually independent of the other variables proposed thus far to account for these phenomena. In any event, it would not appear to be due to problems in the discourse-pragmatic domain itself inasmuch as prelinguistic children appear sensitive to the notions of "old" and "new", necessary for making the grammatical distinctions referred to (Baker & Greenfield, 1988; Lempert & Kinsbourne, 1985; Muir & Field, 1979; Wachs & Chan, 1986).

3. Previous Findings

Examining the spontaneous speech of 4 monolingual Catalan children from the CHILDES Database (MacWhinney, 2000; Serra & Solé, 1986), Grinstead (2004) showed, as illustrated in Table 1, that overt subjects begin to be used very close to the same point in development as do wh-questions and fronted objects.

<table>
<thead>
<tr>
<th></th>
<th>Overt Subjects</th>
<th>Wh- Questions</th>
<th>Fronted Objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gisela</td>
<td>2;1.23</td>
<td>2;8.0</td>
<td>2;1.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p &lt; .0001</td>
<td>Same session</td>
</tr>
<tr>
<td>Guillem</td>
<td>1;11.13</td>
<td>2;3.28</td>
<td>1;11.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p = .051</td>
<td>Same session</td>
</tr>
<tr>
<td>Laura</td>
<td>2;4.11</td>
<td>2;4.11</td>
<td>2;8.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Same session</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>p = .248</td>
<td></td>
</tr>
<tr>
<td>Pep</td>
<td>1;10.6</td>
<td>1;11.6</td>
<td>1;10.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p = .248</td>
<td>Same session</td>
</tr>
</tbody>
</table>

Table 1 – Age of Acquisition of Subjects, Wh-Questions and Fronted Objects in the 4 Child Catalan Speaker of Serra & Solé (1986) with Probability of Constructions Arising At the Same Time

In Table 1, we give the ages (in years, months and days) of the children for the recording session in which the construction listed at the top of the column began to be used. Below the ages listed for wh-questions and fronted objects, there appear the results of a binomial statistical test to determine whether there is a significant difference between the point that overt subjects began to be used and the point at which the construction corresponding to that column began to be used. Notice that of the eight possible comparisons between construction onsets, four of them began to be used in the very same recording session as overt subjects. Three of the compared constructions did not begin to be used in the same session, but nevertheless began to be used sufficiently soon after overt subjects, such that the difference was not statistically significant. One comparison, wh-questions and overt subjects for
Gisela, showed that wh- questions began to be used significantly later than overt subjects. We take these results to be strong confirmation of the hypothesis that overt subjects are CP constituents.

The binomial test, suggested by William Snyder, was done in the following way: we examined the recording sessions following what was determined to be first use of a construction, wh- questions for example. We then counted the number of wh- questions (WH) and overt subjects (OS) and calculated the ratio of WH/WH + OS in the following files. Let us say that a child has 34 recording sessions after her first use of an overt subject, and that there were 302 overt subjects and 14 wh- questions in those files. The calculation would then be: 302/302 + 14 = 0.96. Having calculated this ratio, we look between the first overt subject the child used and the first wh- question she used and find that there were 15 overt subjects (across 3 recording sessions which covered about 3 weeks). To determine the probability that a wh- question should have been produced somewhere among those 15 overt subjects, we calculate 0.96^15 = 0.542, which is not significant. Again, all but the wh- question comparison with overt subjects showed no significant difference between the onsets of these constructions. Below we use this same test with the child German data.

4. Child German

But, does everything syntactic develop at the same time in all languages? Could the observed simultaneity be language-particular? What happens in a language in which overt subjects are produced fundamentally as a function of finiteness and not as a function of discourse-pragmatics? Our argument that these constructions arise at the same time because of an underlying commonality of syntactic and pragmatic features would be much stronger if it were shown that in a language such as German, overt subjects, fronted objects and wh- questions do not emerge at the same time.

In order to pursue this analysis we make the following assumptions with respect to German syntax. Overt subjects are produced as a function of finiteness, with the exception of the narrow pragmatic register which allows subject-drop in the adult language. Fronted objects (topicalizations hereafter) are part of the general V2 phenomenon which moves a wide variety of constituents (XPs, heads, nouns, adverbs, etc.) to first position. Note that fronting does not always take place for discourse-pragmatic reasons, as it is the norm for adverbs of time to be fronted regardless of the context. The fronting of objects, however, is generally motivated by discourse needs. In wh-questions, we assume that XPs occur at the left edge of the clause in a focus position, as in Spanish, English and other languages. Wh- questions in German and other languages involve an interface between syntax and discourse inasmuch as wh- questions always include a presupposition and one focus (e.g. “What did you eat?” presupposes that you ate something and asks what it was).

Fortunately, there was an extensive debate in the eighties and nineties regarding the nature and existence of CP in child German, growing out of the observation that German-speaking children seem to have some problems early on with the left edge of the clause. In this literature (e.g. Clahsen et al., 1995a; Clahsen et al., 1995b; Penner, 1994; Penner et al., 2000; Penner & Weissenborn, 1996; Penner et al., 2001; Poeppel & Wexler, 1993; Weissenborn, 1990, 1992, 1994; Weissenborn et al., 1995), the finding of relevance to our study is that there is an early absence of adult-like wh- questions and fronted objects, while overt subjects are used from the very earliest files of all children studied. Thus, the existing literature supports our hypothesis that overt subjects do not pattern with CP constructions in the development of child German. In spite of this significant literature, there did not appear to be a precise answer to our question of whether the points of onset of these three constructions were close or disparate. Consequently, we studied the available data ourselves to determine the points of onset of these constructions in child German.

5. Methods

Our study looked at both the longitudinally collected data of Simone and Caroline (Powers et al., 2001) and the case study data of Andreas and Katrin (Wagner, 1985). All were obtained from the CHILDES Data Base (MacWhinney, 2000). With respect to the longitudinal data, Simone was recorded over the course of two years, three months (from 1;9 to 4;0), which included 71 recording sessions. Caroline was recorded over the course of three years, 2 months (from 0;10.0 to 4;0), which
included 239 recording sessions. Regarding the case study data, Katrin was recorded for 3 hours and 22 minutes on a single day, when she was 1;5.15, providing an in-depth snapshot of the state of her grammar, while Andreas, similarly, was recorded for 3 hours and 33 minutes on a single day when he was 2;01.0.

Our criteria for inclusion of these children's utterances in our analysis were:

- spontaneous utterances only (no repetitions, song lyrics, etc.)
- reasonably interpretable utterances only
- utterances with verbs only
- imperatives excluded
- clear cases of "chunking" or lexicalized, unproductive expressions excluded

Our criteria for productive use for each of the three constructions were the following:

- **Subjects:**
  - Several different verbs used
  - Several different subjects used
  - Not an isolated case

- **Wh-questions:**
  - Several different wh-words used
  - Several different verbs used
  - Not an isolated case

- **Topicalizations**
  - Several different topicalized elements used
  - Several different verbs used

  NB: It is difficult to recognize a topicalized element without a subject being present; therefore, these data are to be interpreted with some caution.

### 6. Results and Discussion

Before presenting the actual ages of onset of the constructions, it is worthwhile to note that there are substantial differences in the Mean Length of Utterance values, Measured in Words (MLUw), – a measure of linguistic and not simply chronological development – at which the longitudinally studied children began to use each of the three constructions, as illustrated in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Subjects</th>
<th>Wh- Questions</th>
<th>Topicalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caroline</td>
<td>1.36</td>
<td>1.66</td>
<td>1.69</td>
</tr>
<tr>
<td>Simone</td>
<td>1.45</td>
<td>2.65</td>
<td>1.86</td>
</tr>
</tbody>
</table>

Table 2 – The Mean Length of Utterance, Measured in Words (MLUw), at Onset of Subjects, Wh-Questions and Topicalizations for Caroline and Simone.

In Table 2 we see that in both cases the MLUw value for the children's speech was much lower when they began using subjects than it was when they began using either of the other constructions. This basic pattern is consistent with our hypothesis regarding the early onset of overt subjects in German and the later onset of CP-related constructions such as wh-questions and topicalizations.

In Figures 1 and 2, we see the overall trend of development of the three constructions in the two longitudinally studied children, Caroline and Simone, and again we see that overt subjects begin to be used long before either topicalizations or wh-questions.
In Figure 1, there is no representation of the onset of wh-questions because it happens so much later (2;10.04) than do the onsets of subjects and wh-questions. In Figure 2, we see the early onset of overt subjects, with hundreds of them being produced before the first topicalization or wh-question, both of which begin to be used at approximately the same time in the speech of Caroline.

Finally, we see in Table 3 that using the binomial test, described above, there is a statistically significant difference between the onset of subjects and the onsets of wh-questions and topicalizations, for both of the longitudinally studied children.
<table>
<thead>
<tr>
<th>Subjects</th>
<th>Wh-questions</th>
<th>Topicalizations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caroline</td>
<td>1;10.17</td>
<td>2;2.10* ( .895^{125} &lt; 0.0001 )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2;3.0*</td>
</tr>
<tr>
<td>Simone</td>
<td>1;10.20</td>
<td>2;10.04* ( .857^{350} &lt; 0.0001 )</td>
</tr>
</tbody>
</table>

Table 3 – Ages of Onset of the Three Constructions for Simone and Caroline, with the Results of the Binomial Tests

The results in Table 3 show that for the longitudinally studied children, there was a difference between when they started to use subjects and when they started to use the other constructions. In Table 3, we also see the calculations for each binomial test. Notice that in Simone's data, for example, that though there were only 3 chronological months between the onset of overt subjects and the onset of topicalizations, there were 112 subjects produced before the first topicalization, which in combination with Simone's subject-to-topicalization ratio, demonstrates the improbability of their being any topicalizations produced in the intervening time. For the other comparisons, the difference is even more dramatic both in terms of the absolute time intervening between constructions and in terms of the binomial test.

For the case study data of Andreas and Katrin, it was of course impossible to show differences in onset of the different constructions because each of their very large files was from a single day. Nonetheless, the fact that Andreas was 8 months older and had a much higher MLU\(_w\) than Katrin, as illustrated in Table 4, when the data was collected and the fact that so much data was collected for each child appeared to offer a fine-grained picture of their respective moments in development, particularly with respect to their use of the three constructions of interest.

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>MLU(_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katrin</td>
<td>1;05.15</td>
<td>1.6</td>
</tr>
<tr>
<td>Andreas</td>
<td>2;01.0</td>
<td>2.35</td>
</tr>
</tbody>
</table>

Table 4 – The Ages and MLU\(_w\) Values for Katrin and Andreas for Their One Recording Session

Indeed, our examination of their data suggests that subjects were well-established in both of their grammars, including the younger child, Katrin's data. Similarly, both children appeared to use topicalizations in an adult-like way, as illustrated in the following examples.

*Topicalizations*

Katrin uses topicalizations with a variety of verbs and subjects.

1. Page 1, line 8
   "Weg ist sie"
   "gone is she"
   “she’s gone.”

2. Page 2, line 5
   "Ei isst Papa"
   "egg eats Papa"
   “Papa is eating an egg.”
3. Page 33, line 22
Eine Puppe darf ich haben
a puppet may I have
“I may have a puppet.”

4. Page 39, line 34
Kerne frisst es
uts eats it
“It eats nuts”

5. Page 55, line 44
Katze sehe ich
cat see I
“I see a cat.”

Andreas also has fairly adult-like use of topicalizations and also uses them with a variety of subjects and verbs.

6. File 1, line 228
Zwei Teddys habe ich
two teddys have I
“I have two teddys”

7. File 2, line 393
Carrerabahn fahre ich
Toy-car drive I
“I’m driving a (toy) car.”

8. File 1, line 125
Da guckt er raus
there peeks he out
“He’s peeking out there.”

9. File 1, line 480
So macht er dann
so makes he then
“That’s how he does it, then.”

10. File 1, Line 841
So mein Hubschrauber mache ich fertig
so my helicopter make I finished
“So I’m finishing my helicopter.”

Wh-Questions

Katrin uses a few questions, but in a formulaic manner. She uses “what” and “where” only with the verb sein (to be), although she is able to inflect the verb appropriately for the subject. The other question - “What do you want?” - she addresses to all manner of objects (including her own feet, apparently), without altering the structure at all. The following is a near-complete list of all her question structures.
11. Page 11, line 19
Was das denn?
what that then
“What’s that, then?”

12. Page 12, line 42
Wo   ist?
Where is
“Where is (it)?”

13. Page 15, line 38
Wo sind die Ringe?
Where are the rings?

The following example was repeated often and appeared to be a lexicalized expression.

14. Page 27, line 17 & Page 34, line 26 & Page 47 line 19
Was  will du, Auto? / Eisenbahn / Fuss
What want you, car / train / foot
“What do you want, car / train / foot?”

Andreas uses very few wh-questions, all with was (what). He does not use multiple question words or multiple subjects/verbs.

15. File 3, line 15
Was ma(chst) du?
What do you?
“What are you doing?”

16. File 7, line 87
was is(t) da dann .
what is that then
“What’s that, then?”

Andreas also produces a number of unclear forms, as in 17. and 18.

17. File 6, Line 677
Schenkst du mir zu Weihnachten ?
give you me for Christmas?

The utterance in 17. could be either "Are you giving me something/this for Christmas?" or "What are you giving me for Christmas?" with a dropped wh-word. 18. is another example of the same phenomenon, which has been noted elsewhere (e.g. van Kampen, 1997).

18. File 8, line 357
Papa was    mitgebracht?
papa what/something with-brought
“What did Papa bring” OR “did you bring something, Papa?”

In summary, the case study data shows subjects and topicalizations being used in both the younger and older child, but wh- questions do not seem to be used productively in either of their speech samples. We take this to be supportive of our contention that subjects develop earlier than wh-questions in child German. With respect to topicalizations, it is harder to say, since there could have been an earlier period, not captured in this data sample, during which subjects were used and
topicalizations were not yet productive. The longitudinal data from Simone and Caroline, which both show precisely this pattern, suggest that this is likely to be the case for Andreas and Katrin.

6. Conclusion

In this study we have shown child learners of German, a language in which overt subjects are not produced as a function of discourse pragmatics but rather as a function of finiteness, acquire them much earlier than wh-questions, which may be considered a discourse-sensitive, left-peripheral constituent. Similarly, subjects began to be produced significantly earlier than topicalizations, another left-peripheral construction which also involves an interface with discourse-pragmatics. These observations contrast sharply with the facts of child Catalan and Spanish in which overt subjects, fronted objects and wh-questions appear to emerge in spontaneous production at close to the same point in development.

In this way, child German confirms the observed pattern that constructions which involve an interface between syntax and discourse-pragmatics (fronted objects and wh-questions) begin to be used late relative to those which are not (overt subjects in German). This data is confirmatory of the pattern in Southern Romance in the sense that the same pattern obtains in child Spanish and Catalan, but with different constructions. Namely, child Spanish and Catalan speakers are already aware that overt subjects are discourse-sensitive in their languages, in the same way that wh-questions and fronted objects are, which is why they emerge at the same point.

The significance of these findings is that even at MLUs below 2, children are aware of whether their language is a null subject or an overt subject language. Further, our results may be taken as evidence for adult syntactic theories which place at least most overt subjects in discourse-pragmatically sensitive positions in Southern Romance. With respect to Interface Delay and Left-Peripheral Delay, little is currently understood about exactly why it occurs, but it will hopefully be an active area for future research. Finally, inasmuch as we have shown that child Catalan and child German speaking children appear to produce Complementizer Phrase constituents before the age of 3, our results should be taken as evidence against development theories which posit that children younger than 3 years old have no “syntagramatic competence” ( Tomasello, 2003), in that the complementizer phrase is such a syntagramatic constituent and appears well developed in both languages before 3 in all children examined.

Acknowledgments

We gratefully acknowledge the helpful discussion of the issues raised here with Teun Hoekstra, Nina Hyams, Detmar Meurers, Francisco Ordóñez, William Snyder, Jürgen Weissenborn and two HLS reviewers.

References

Camacho, José. (2005b). Do subjects have a place in Spanish: Rutgers University.
In H. Clahsen (Ed.), Generative perspectives on language acquisition (pp. 129-160). Amsterdam: Benjamins.


