

Wh-subextraction in child Catalan

Anna Gavarró & Jaume Solà
(Universitat Autònoma de Barcelona)

1. Introduction

The acquisition of wh-questions in child language has been relatively well studied, especially with respect to production. The common assumption is that comprehension of wh-questions does not present any deviation from the target adult one (Guasti 2002).

In this paper we present new evidence showing that the interpretation of wh-questions in child language deviates sometimes from the target, and in ways that parallel what has been shown in the literature to be the case for production. In particular, we show that wh-questions may be interpreted as involving wh-movement of the wh-word alone, without the pied-piping of the whole wh-complement, as found in the production of wh-sentences by Dutch children (van Kampen 1994). This possibility is known as subextraction or Left Branch Violation.

Child wh-subextraction in Dutch: production (Dutch: van Kampen 1994)

- (1) a. Welk_i wil jij [t_i boekje]? (age 3;7)
which want you book
'Which book do you want?'
- b. Hoe_i is het [t_i laat]? (age 6;5)
how is it late
'How late is it?'

Child wh-subextraction in Catalan: comprehension

- (2) a. Qui recull pedretes?
who picks-up pebbles
- b. Qui_i recull [t_i pedretes]? (age 2 & 3)
which picks-up pebbles
'Which stones does s/he pick up?'

Sentences such as (2a), which in adult language are unambiguous in involving a subject bare wh-word (*qui* 'who'), can be interpreted by children either in the adult way or also as involving a wh-determiner (meaning 'which') that has been extracted out of its noun phrase as in (2b).

Wh-subextraction is attested in adult language, where it varies cross-linguistically. Some languages with rich Case morphology, such as Slavic languages or Latin, allow it quite freely (Corver 1990). Other languages allow it in a

restricted way. For Romance languages, the presence of a preposition *de* between the *wh*-determiner and the rest of the phrase (as in *combien de livres*) seems to be a necessary (but not a sufficient) condition for subextraction.

Adult wh-subextraction:

- (3) a. Skolko Tania prochitala knig? (Russian, Gavrusseva & Thornton 2001)
 how-many T. she-read books-GEN
 'How many books did Tania read?'
 b. Jaki wykrecilés numer? (Polish, van Kampen 1994)
 which you-dialed number
 'Which number did you dial?'
 c. Combien as-tu lu de livres? (French, Obenauer 1984)
 how-many have-you read of books
 'How many books have you read?'
 d. Com és de gruixut? (Catalan)
 how is of thick
 How thick is it?

Regarding child language, subextraction has been attested in Dutch and English. Van Kampen (1994, 1997) found that Dutch children produced questions like those in (4a), which are not found in the input the child is exposed to. For English, subextraction cases have been reported in Hoekstra, Koster & Roeper (1992), Thornton & Gavrusseva (1996), and Chen, Yamane & Snyder (1998).

More specifically, Gavrusseva and Thornton (2001) investigate, through an elicited production experiment, *wh*-possessive questions, and find that, contrary to adult English, children allow *wh*-extraction of *whose*, without pied-piping of the entire *whose*-DP.

Child wh-subextraction: production

- (4) a. Welke wil jij liedje zingen?
 which want you song sing
 'Which song do you want to sing?'
 (Dutch: van Kampen 1994, age 3;7)
 b. How many do you think marbles are in there?
 (English: Chen, Yamane & Snyder, 1998)
 c. Who do you think's flower fell off?
 Whose do you think ball went in the cage?
 (English: Gavrusseva & Thornton 2001, mean age 5;4)

Subextraction is not, to our knowledge, attested in the literature on Romance. The case we report on bears a close resemblance to that in Dutch: the subextracted

quantifier is of the ‘which’ type (*welk* in Dutch/ *qui(n)* in Catalan) and is extracted from object position.

We will first introduce the experiment and its results. We will then proceed to discuss the theoretical options available to account for subextraction in general and our case in particular.

2. *The experiment*¹

In an experiment involving the interpretation of subject wh-questions with definite and bare objects (with Tom Roeper), we found a deviant interpretation of the question not envisaged in the experimental design. Children responded on occasion as exemplified in (6b) to question (5), for which the expected response was (6a).

- (5) Qui necessita sabates/les sabates?
who needs shoes/the shoes ‘Who needs shoes/the shoes?’
- (6) a. – La germana petita
the sister young ‘The younger sister’
b. – Les blaves.
the blue ‘The blue ones.’ (as in Qui_i necessita [t_i sabates]?)

The experiment was undergone by 27 monolingual Catalan speaking children of the Barcelona area, whose age ranged from 2;7,1 to 5;5 years of age; of these, 2 were 2 year-olds, 11 were 3-year-olds, 10 were 4 year-olds and 4 were 5. The results of the experiment were analysed by grouping the children in two age groups.

(7)	Subjects	#	age range
	2/3 year-olds	13	2;7,1–3;11,14
	4/5 year-olds	14	4;0,2–5;5

The general procedure was the following. The child was presented with a story at the end of which a question of the type in (5) was asked by the experimenter. Each child was presented with four stories, plus distractors. The results of this preliminary experiment, which appear in (8), indicate that 15% of the answers of the 2/3 year old group involved an answer of the type in (6b), while for the older group this kind of answer had reduced to 7%. All the non-target answers to the questions were of type (6b).

(8)	Results	target answer	left-branch interpretation
	2/3-year-olds	44/52 (84.6%)	8/52 (15.4%)

¹ We are grateful to the children, teachers and direction of the Escola Decroly de Barcelona and the Escola Bressol Gespa in Bellaterra, for taking part in the first and second experiment respectively.

4/5- year-olds 52/56 (92.8%) 4/56 (7. 1%)

Interestingly, non-target answers were found in the older group of children, including some 5 year-olds, which indicated that, whatever the reason for the production of these unexpected answers, it was quite persistent in some individuals. (Notice that this was also the case with the children considered in van Kampen 1994 and Gavruseva and Thornton 2001.)

(9) Number of children who produce some non-target answers

2/3-year-olds 5/13
4/5-year-olds 2/14

It was noted that all the experimental items involved the verb *necessitar* 'need', which accidentally could lead to confusion, as *qui necessita...* 'who needs' could have been reanalysed as *quines ...*'which'; this interpretation on the part of the child would be based on an ill-formed sentence, in several respects: it would involve a truncated verb and would have a deviant intonation pattern in Catalan. Nevertheless, to overcome these shortcomings, an experiment was designed to test specifically the interpretation of wh-questions.

We carried out the experiment with 21 monolingual Catalan speaking children in Bellaterra, near Barcelona. Their ages ranged between 2;5 and 3;2, so that older children were not tested. Ten controls also undertook the experiment.

(10) Subjects	#	age range	mean age
2-year-olds	12	2;5,27 – 2;10,25	2;7
3-year-olds	9	3;0,4 – 3;8,27	3;2
total	21		
controls	10		

Children were tested individually in a quiet room in the school; they were introduced to the characters depicted and the task took 10 to 15 minutes.

The task designed included a story told by the experimenter with the help of pictures, followed by a question. The story involved several characters carrying out an action with more than one object/animal. The distinctive features of the characters and of the objects/animals were pointed out, and the child was asked a subject wh-question with a transitive verb and an object. Each child was told four stories, plus distractors (which were also questions); the experimental items are listed in (11); the verbs which appear in them were chosen so that not only they belonged to the children's vocabulary, but also that they couldn't give rise to any phonological misanalysis. The expected answer consisted in identifying one of the several characters in the story, the one carrying out the action in the last picture shown to the child.

- (11) Experimental task: items
- a. Qui atrapa el cavall?
who chases the horse

- b Qui recull pedretes?
who picks-up pebbles
- c. Qui porta la cadira?
who carries the chair
- d. Qui mira el tòtem?
who looks-at the totem

The following exemplifies the stories told, and corresponds to question (11a).



The results in (12) show how non-adult answers were found both for 2 and 3 year-olds, although they decreased with age. Non-target answers were only marginally anything other than left-branch violations (6% for 2-year-olds). As a result, then, we can safely claim that the interpretation of wh-questions is rule-governed, and not the result of random behaviour. Left-branch violation answers were as many as 20.8% of answers by 2-year-olds, and dropped to 13.9% with 3 year-olds. Controls performed 100% as expected.

(12) Results by age group

	target answer	left-branch	other
2-year-olds	35/48 (72.9%)	10/48 (20.8%)	3/48 (6.2%)
3-year-olds	31/36 (86.%)	5/36 (13.9%)	
controls	40/40 (100%)		

Further, more than half of the 2 year-olds gave some left-branch violation answer, while this dropped to 1 in 3 of the 3 year-olds.²

(13) Number of children who produce some left-branch interpretations

2-year-olds	8/12	66.6%
3-year-olds	3/9	33%

3. Discussion

The results show that wh-subextraction is an available option for interpretation of wh-sentences in Child Catalan. Let us then proceed to consider the theoretical possibilities available to account for this and other cases of subextraction that have been reported. The kinds of possible approaches we will consider are summarised in (14):

Approaches to child wh -subextraction

- (14) a. Performance approaches
 b. Grammatical approaches
 – the structural approach
 – the “morphological” approach

Regarding performance explanations, we believe that the comprehension deviations provided by our data are unlikely to be attributed to performance. Performance explanations are generally based on the idea that there is some processing disturbance (overload) that causes a false-start, possibly followed by some repair strategy, which may consist in resorting to a (sort of) pronounced copy.

Performance approaches

- (15) a. Illicit movement: Wh_i ←—X— t_i
 b. Repair strategy: Wh_i ←—X— [pronounced copy]_i

This is the idea defended in Yamane, Chen and Snyder (1999) for child subextraction in English. According to them, in cases like:

- (16) a. How many do you think the witch has five magic wands in her coat?
 b. How many do you think how much animals like cookies?

² Other instances of subextraction are found on occasion in child Catalan, as in the spontaneous production in (i). Further research is needed to bring such cases to the light.

(i) Aquest no dada llibre? (Joan, 2;8,15)
 This not please book
 ‘Do you not like this book?’

In (16) there is an attempt at subextraction caused by the elicitation technique, which would give ungrammaticality as a result, and there is an attempt at repairing it by making the ungoverned trace overt.

In fact, it should be pointed out that quite often long wh-movement in child language involves copies of the sort in (16) independently of subextraction, as shown in Thornton (1990); in this case movement is not illicit and yet the pronounced copy is produced.

Medial wh-questions of Thornton (1990)

- (17) a. What do you think what Cookie Monster eats? (Katie, 5;5)

In our case, however, there is no room for performance factors to play a role. There is no apparent processing overload, since we are dealing with monoclausal 3- or 4-word sentences. The sentences, in addition, are clearly and smoothly pronounced in a highly contextualised setting, which does not favor the subextraction interpretation over the adult interpretation

Against a performance account there is also a continuity argument: left branch subextraction occurs in some adult languages and in child language it is attested as late as age 5 or 6, where performance errors likely decrease (see for English Gavrusseva and Thornton 2001 and for Dutch van Kampen 1994, 1997, 2000). Certainly, when the target language does not allow it, subextraction decreases with age, as should be expected. So we conclude that (at least in part of the cases), subextraction is a manifestation of (child or adult) grammar.

Among grammatical explanations, a classical one, developed by Corver (1990) for adult language and later adopted for child language by Jordens & Hoekstra (1991), Hoekstra, Koster & Roeper (1992) and Hoekstra (1994), is the idea that grammars may radically differ in the categorial and structural analysis of noun phrases. Let us call this approach the structural approach.

According to it, grammars with DPs do not allow left branch extraction out of DP, either because the DP is a barrier or because it is improper movement of some sort (head movement, non-constituent movement). There would be, however, languages in which noun phrases are bare NPs, while the apparent DP material simply constitutes an NP adjunct. Adjuncts to a maximal projection are free to move away from the maximal projection, since this projection can never constitute a barrier for it.

The structural approach

- (18) a. * Wh_i ... [DP t_i NP]
b. √ Wh-NP_i ... [NP t_i NP]

Even though we cannot discuss the issue at any length here, we discard this theoretical option on the grounds that it is both too permissive and too restrictive.

Problems with the structural approach

- (19) a. It is too permissive in predicting Conditions on Extraction Domains (CED) violations.
b. It is too restrictive in predicting absence of pied-piping in the presence of subextraction.
c. It is too loose in characterising possible language variation.

It is too permissive in predicting violations of Conditions on Extraction Domains: since adjuncts are not included in the category adjoined to, this category cannot constitute a barrier. As argued by van Kampen 1994, there are no CED violations attested in Dutch.

As for Slavic languages, we know that in Russian extraction is only possible out of a subject if the subject is postverbal and does not trigger agreement on the verb (E. Gavrusheva, p.c.). We assume that these are cases of internal subjects which are somehow “governed”.

- (20) a. Skolko na vstrechu prishlo studentov?
(unaccusatives)
how-many to meeting came-PastNeutre students-GenMascPl
'How many students came to the meeting?'
- b. *A eti knigi skolko chitali studentov?
(transitives)
Prt these books, how-many read-PastPlural students GenPLMasc?
- b'. A eti knigi skolko chitalo studentov?
Prt these books how-many read-Past.Neutre students
'How many students read these books?'
- c. *A na vstreche skolko smejalis' studentov?
(unergatives)
Prt at meeting how-many laughed-PastPlural students-Gen.PLMasc?
- c'. A na vstreche skolko smejalos' studentov?
Prt at meeting how-many laughed-PastNeuter students-Gen.PLMasc?
'How many students laughed at the meeting?'

The structural analysis is too restrictive, we believe, in that it would rather predict obligatoriness of subextraction: if *wh*-XPs are adjuncts, there seems to be no reason why the whole constituent should be pied-piped. [There may be technical ways to make percolation of the *wh*-feature (hence pied-piping) possible, but this runs against any minimalist view on possible derivations]. In addition, pied-piping is never a marginal option when subextraction occurs. This argument is advanced in van Kampen (1994).

On the theoretical side, it is questionable that UG should allow for such an essential part of syntactic structure as is DP to be reshuffled as something as different as an adjunction structure. Anything departing from universal functional structure poses the question of how to define the limits of syntactic variation in a non-stipulative way. We adhere to the view, defended by Cinque (1999) and

others, that there is no variation in the hierarchical structure of FCs. Also, as pointed out in van Kampen (1994), the child learning Dutch seems to fully master the Determiner system of Dutch when she keeps producing subextractions.

Let us, then, consider the hypothesis that subextraction involves no structural variation (or minimal variation) with respect to pied-piping. One version of this theory is the minimalist contention that it is not syntax itself, but the phonological component, that imposes pied-piping. As for syntax, the only requirement is that a feature be moved or somehow checked. Moving the whole word containing the feature is already a minimal case of pied-piping. Moving larger constituents is a stronger case of pied-piping that must be due to further phonological requirements. In other words, to move a feature F, F “must carry enough material for convergence”.

Let us call this approach the “morphological approach”, a loose term intended to cover any phenomena whose licensing is sensitive to the PF component. Some ideas in this connection are summarised in (21):

The “morphological” approach: possible implementations

- (21) a. Keep together enough material for PF convergence (Minimalist Program)
- b. Left branches may be morphologically marked as <+attributive>, which requires adjacency to the lexical head. (van Kampen 1994)
- c. Rich Case and agreement morphology on both determiners and nouns (in languages featuring subextraction such as Slavic languages or Latin) play a crucial role (Ross 1967) (probably in making it possible that Determiner and N are licensed as independent morphological units).
- d. Genitive case blocks *whose*-extraction in English-like languages, while the absence thereof frees subextraction of possessives in Hungarian-like languages (Gavruseva & Thornton 2001).
- e. Case licensing is crucial for (some cases of) determiner subextraction (Kayne 2002).

An interesting prediction of the “morphological” approach is that mistakes in child language should occur to the extent that the morphological intricacies of the adult language are hard to acquire. Also, it predicts that there is no macro-parameter of the type [+/-subextraction language]: morphological intricacies are expected to be tied to lexical items or lexical classes of items. This seems to be the case, as exemplified in (22):

(22) *Subextraction is not uniform within languages*

	<i>Subextraction of</i>		
<i>Language</i>	‘which’	‘how-many’	‘whose’
Russian	+	+	+
Hungarian	-	(?)+	+
French	-	+	?+

- d. [V [t_i combien]] [de [llibres_i t_k]]
- e. [t_i combien]_j ... [V t_j] [de [llibres_i t_k]]

where you start with the phrase [*llibres combien*] in object position, where the NP *llibres* is a specifier.

- In (23b) you extract the NP [*llibres*] to the specifier of some FC responsible for Case licensing.
- Step (23c) merges the functional head *de*, which does not form a constituent with *llibres*.
- Step (23d) is VP movement to Spec,*deP*.
- Finally, and essentially to our proposal, in step in (23e) remnant movement of [*t combien*] takes place.

Essential to Kayne's proposal is that step (23b,c) is determined by Case requirements: *llibres* moves to the specifier of a FC that licences Case. While step (23e) (wh-movement) is universal (or uniform across many languages), the possibility of step (23b) may vary from language to language and, apparently, also within a language. For an NP like [*llibres*], in some languages or in some constructions Case would be licensed via movement to some specific Case FC; others would licence Case in a more local way, not involving movement. Let us call these possibilities split Case-licensing and local Case-licensing. So subextraction is predicted to occur only with split Case-licensing.

Admittedly, this proposal looks like a step back from the minimalist ideal of a "uniform syntax except for PF requirements". The non uniform analysis, however, can be defended if it can be shown that the two derivations find independent motivation.

Let us see whether these factors can play a role in predicting the availability of subextraction in Catalan child language. In adult Catalan, subextraction is only available in one case:

- (24) a. *Quants has comprat t (de) llibres?
how-many have-you bought (of) books
- b. *Quins has comprat t (de) llibres?
which have-you bought (of) books
- c. Com és de llarg?
how is-it of long

Specifically, (24b) is not allowed in contrast with our reported cases in child language. What evidence could lead the child to allow subextraction in this case? We propose that independent evidence in the adult language provides a clue. Consider (25):

- (25) a. Quants n'has comprat, de llibres?
how-many NE-have-you bought (of) books
- b. Quins has comprat, de llibres?
which have-you bought (of) books
- c. Com n'és, de llarg!
how NE-is-it of long

(24) and (25) minimally differ in that in the latter the inner NP (*de llibres*) appears as right-dislocated. It could also appear as left dislocated. The point is: why are the dislocated NPs marked with the preposition *de*? The uninteresting answer is that this is an idiosyncratic requirement on dislocated NPs (and APs). A more interesting answer is that these dislocated constituents have been Case-licensed through split Case-licensing. (25b) would have the derivation in (26):

- (26)
- | | | | |
|----|--|---------------------------------|--------------------------------|
| a. | | comprat | [<i>llibres quins</i>] |
| b. | <i>llibres_i ...</i> | [comprat [| <i>t_i quins</i>]] |
| c. | <i>de [llibres_i ...</i> | [comprat [| <i>t_i quins</i>]]] |
| d. | [comprat [<i>t_i quins</i>]] [<i>de</i> | <i>llibres_i</i> | <i>t_k</i>]] |
| e. | <i>[t_i quins]_j ...</i> | [[comprat | <i>t_i]]</i> |
| | | <i>[de [llibres_i</i> | <i>t_k</i>]] |
| | | |]]] |

Now, suppose that the constituent [*de llibres t_k*] in (26e) is obligatorily assigned a Topic feature in Catalan, and that this feature forces it to move to the appropriate specifier, the specifier of a Topic phrase that licences dislocated elements, according to Villalba (2000). It is well known that languages differ in the obligatoriness of overt syntactic focus/topic marking. Now let us try to put things together:

- (27)
- a. Languages like French or Catalan share the possibility of split Case-licensing, as in derivations (23) and (26).
 - b. All languages share the possibility of local Case-licensing.
 - c. French and Catalan differ in the obligatoriness of marking the stranded part of step (d) as Topic, which must end up as a dislocated constituent.
 - d. For Catalan, examples like those in (25), which are frequent in adult speech, constitute robust evidence for split Case-licensing.
 - e. Let us assume that, for the child, they are not robust enough evidence for setting the obligatoriness of Topic marking, so that the child is led to admit the possibility of split Case-licensing without Topic marking.

The analysis we propose to account for the deviant interpretations found in Catalan child language is that in (28) (for question (2) above).

- (28)
- | | | | |
|---|-----------------------------------|--------------------------|-------------------------|
| [<i>t_i qui</i>] _j ... | [[<i>recull t_j</i>]] | [[<i>(de) [pedretes</i> | <i>t_k</i>]] |
| which | picks-up | stones | |
| ‘Which stones does s/he pick up?’ | | | |

This child interpretation implies to adjustments with respect to the adult grammar: the *wh*-word *qui* 'who' is interpreted as *quin* 'which'³, and the Case-related element *de* is phonetically null.

To conclude, we argue that the deviations in the interpretation of *wh*-questions found in child Catalan are grammatical in nature, and stem from the fact that UG makes subextraction available if there is split Case assignment to internal arguments. Adult Catalan provides the child with evidence for split Case assignment in dislocation, and the child generalises this option to cases of non-dislocation. This analysis would make us expect similar cases of *wh*-interpretation in child French, for instance, which presents split Case assignment, and not in other cases (e.g. Italian?). How the analysis extends to what has been found in other child languages remains a topic for future research.

REFERENCES

- Chen, D., M. Yamane & W. Snyder. (1998). 'On the nature of Children's Left Branch Extractions' *Proceedings of the 22nd Annual Boston University Conference on Language Development*.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge, Mass.: The MIT Press.
- Cinque, G. (1999). *Adverbs and Functional Heads: A Cross-linguistic Perspective*. Oxford: Oxford University Press.
- Corver, N. (1990). *The Syntax of Left-Branch Extractions*. PhD dissertation, University of Tilburg.
- Gavruseva, E. & R. Thornton. (2001) 'Getting it right: acquisition of whose-questions in child English'. *Language Acquisition* 9(3): 229–267.
- Guasti, M. T. (2002). *Language Acquisition. The Growth of Grammar*. Cambridge, Mass.: The MIT Press.
- Hoekstra, T. (1994). 'The acquisition of functional structure'. In. Hoekstra, T & B. Schwartz (eds.). *Language Acquisition Studies in Generative Grammar*. Amsterdam/Philadelphia: John Benjamins.
- Hoekstra, T., J. Koster & T. Roeper. (1992) 'Left Branch violations in acquisition'. Paper presented at the *Boston University Conference on Language Development*. October 24, Boston.
- Jordens, P. & T. Hoekstra. (1991). 'The acquisition of negation'. Paper presented at the GLOW Workshop 'The development of bound variables and operators'. March 28, Leiden.
- Kampen, J. van. (1994) 'The learnability of the left-branch condition'. In R. Bok-Bennema & C. Cremers (eds.) *Linguistics in the Netherlands 1994*, John Benjamins, Amsterdam, 83–94.
- Kampen, J. van. (1997) *First Steps in Wh-Movement*. Doctoral dissertation, Utrecht University.

³ Were it not the case that *qui* and *quin* are so phonetically similar, the subextraction interpretation may not have occurred. We have found no way to get around this problem experimentally.

- Kampen, J. van. (2000) 'Left-branch extractions as operator movement: evidence from child Dutch'. In S. Powers & C. Hamann (eds.) *The Acquisition of Scrambling and Cliticization*. Kluwer, Dordrecht.
- Kayne, R. (2002). 'On some prepositions that look DP-internal: English *of* and French *de*'. *Catalan Journal of Linguistics* 1: 71-115.
- Obenauer, H. (1984). 'On the identification of empty categories'. *The Linguistic Review* 4-2: 153-202.
- Ross, J. R. (1967). *Constraints on Variables in Syntax*. Doctoral dissertation, MIT.
- Thornton, R. (1990) *Adventures in long-distance moving: The acquisition of complex wh -questions*. Doctoral dissertation, University of Connecticut.
- Thornton, R. & E. Gavrusseva. (1996). 'Children's split "Whose-questions" and the structure of possessive Nps'. Paper presented at the *21st Annual Boston University Conference on Language Development*.
- Villalba, X. (2000). *The syntax of sentece periphery*. Doctoral dissertation, Universitat Autònoma de Barcelona.
- Yamane, M., D. Chen & W. Snyder (1999). 'Subject-object asymmetries and Children's Left Branch Extractions' *Proceedings of the 23rd Annual Boston University Conference on Language Development*, vol. 2: 732-740.

anna.gavarro@uab.es

jaume.sola@uab.es

<http://seneca.uab.es/ggt/Gavarro/gavarro.html>

<http://seneca.uab.es/ggt/Sola/sola.html>

Departament de Filologia Catalana
Universitat Autònoma de Barcelona
08193 Bellaterra