

# CATEGORIAL COMPLEXITY OF RELATIVISATION AND CHILD ROMANCE PERFORMANCE

GLYN MORRILL AND ANNA GAVARRÓ

## 1 Introduction

Chomsky (1965) pointed out the contrast between non-centre embedding and centre embedding nesting of relativisation illustrated in (1):

- (1) a. The dog that chased the cat that saw the rat barked.  
b. ?The cheese that the rat that the cat saw ate stank.

While nesting of non-centre embedding subject relativisation as in (1a) is fully acceptable, the acceptability of nesting of centre embedding object relativisation as in (1b) deteriorates with the depth of embedding. Here we discuss relativisation in relation to Catalan and Italian first language acquisition. We report on an experiment on the child comprehension of Catalan relatives, replicating an earlier experiment on child Italian relatives (Adani in press) and we present an account of this child performance in terms of a categorial complexity metric (Johnson 1998, Morrill 2000).

## 2 Categorial Grammar and Incremental Processing

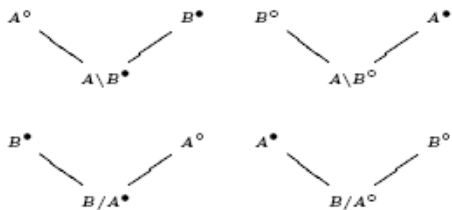
Categorial grammar (Lambek 1958) classifies words and expressions by means of fractional types built over basic types such as sentence (S) and nominal (which we parameterise here with number singular, N(sg), plural, N(pl), or unspecified, N(\_)). An expression of type  $A \setminus B$  is one which concatenates with any expression of type A to the left to form an expression of type B. An expression of type  $B / A$  is one which concatenates with any expression of type A to the right to form an expression of type B. Formally:

- (2)  $A \setminus B = \{s \mid \text{for all } s' \in A, s' + s \in B\}$   
 $B / A = \{s \mid \text{for all } s' \in A, s + s' \in B\}$

Morrill (2000) describes a complexity metric founded on incremental categorial processing in terms of proof nets (Girard 1987, Roorda 1991). In this view of

processing, types are marked with input polarity ( $\bullet$ ), meaning that a resource is given, or output polarity ( $\circ$ ), meaning that a resource is wanted. Polar types are unfolded upwards into polar type trees as follows:

(3)



We refer the reader to the references above for the details, which are quite involved, but we illustrate the basic idea here with the processing of this sentence:

(4) John loves Mary.

Initially, a sentence is sought and after hearing the first word its type is given:

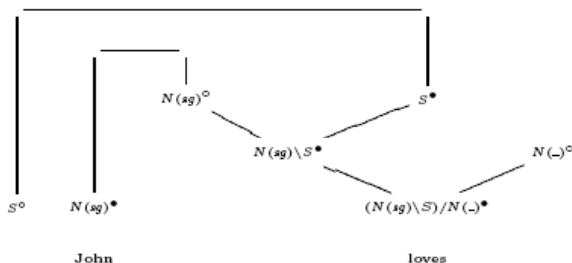
(5)

$S^{\circ}$        $N(\text{sg})^{\bullet}$

**John**

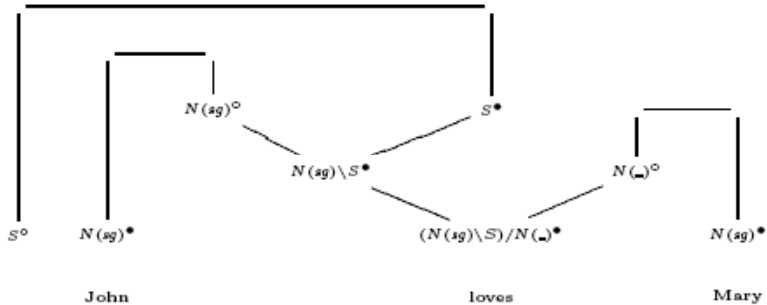
When the second word is heard, its unfolded type is connected by two dependencies: the subject sought is given by the first word *John* and the sentence projected is matched by the initial expectation of a sentence. We represent this as follows:

(6)



When the final word is heard, the parse is completed thus (the unspecified object number on the verb type becomes instantiated by unification with the type with which it is matched):

(7)



On this processing model, Morrill (2000) (cf. Gibson 1998, Johnson 1998) takes as a measure of complexity at each word boundary the number of unresolved valencies that must be kept on the working memory stack at that point, i.e. the number of overarching dependencies. Then the theoretical prediction is that the higher this complexity, the lower should be the acceptability to speakers. Here we invoke a refinement of that metric in which we count not the number of unresolved valencies, but the sum up to that point of the number of unmatched attributes: major category, and feature values (if there are any). E.g.  $N(sg)$  and  $N(pl)$  count for 2 and  $N(\_)$  and  $S$  count for 1. Hence for our example the complexity profile or curve is as follows:

(8)



Ignoring number just for this moment, the subject and object relatives in (1) are analysed, under the model outlined, as shown in Figures 1 and 2.

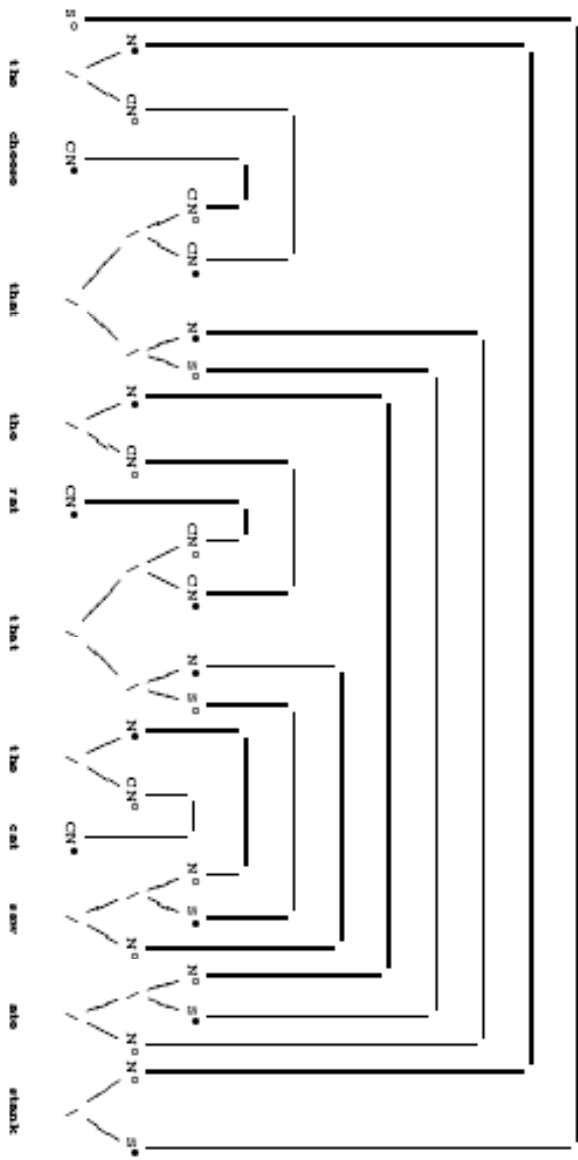
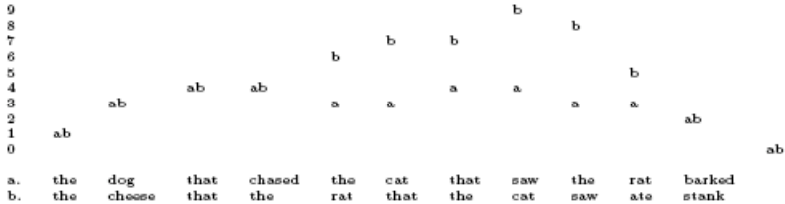


Figure 1: Centre-embedding relativisation



Let us compare the complexity profiles of the two sentences:

(9)



The complexity profile of (1b) is higher, thus predicting its lower acceptability.

### 3 Child Relativisation

Here we argue that this categorical metric of processing load can account for relativisation word order comprehension contrasts in child Catalan (and Italian).

It is widely known that object relativisation gives rise to more comprehension difficulties than subject relativisation. In particular, in relation to acquisition, Adani (in press) studied child comprehension in Italian with a picture identification task of subject relativisation, and object relativisation with either pre- or post-verbal subjects.<sup>1</sup> (The experiment was, with some variations, that of Arosio, Adani and Guasti (2009).) This experiment was replicated in Catalan for 33 children in the age range of 3;5–6;2 in Gavarró, Adani, Ramon, Rusiñol and Sánchez (2009) with very similar results. The three sentence types tested are exemplified in Catalan by the following: (10a) is a subject relative, and (10b,c) are object relatives with pre- and post-verbal subject respectively.

- (10) a. Assenyala el camell que segueix els elefants!  
 point to the.SG camel that follows the.PL elephants  
 ‘Point to the camel that follows the elephants!’  
 b. Assenyala el camell que els elefants segueixen!  
 point to the.SG camel that the.PL elephants follow  
 ‘Point to the camel that the elephants follow!’  
 c. Assenyala el camell que segueixen els elefants!  
 point to the.SG camel that follow the.PL elephants  
 ‘Point to the camel that the elephants follow!’

---

<sup>1</sup> Subjects in Italian and Catalan, as in all null subject languages, can be postposed.

In adult language the postposed type (10c) sentences would be ambiguous, having also a subject relativisation interpretation, were it not for the disambiguating number features. But the experiment showed a surprising effect for this sentence type in child language: there is a preference for the subject relativisation reading even when this involves number disagreement.

### 3.1 The Catalan Experiment

The experiment was an agent selection task in which the child heard sentence types in the quantities given in (11).

- (11) a. Subject relatives #6
- b. Object relatives, preverbal subject #6
- c. Object relatives, postverbal subject #5
- d. Distractors #3/4

The experimental items consisted of sentences with the transitive verbs listed in (12) (the same used for Italian by Adani in press), presented in pseudo-random order.

- (12) perseguir ‘to chase’
- estirar ‘to pull’
- picar ‘to peck’
- seguir ‘to follow’
- rentar ‘to wash’
- mirar ‘to look at’
- mossegar ‘to bite’
- empènyer ‘to push’

Subjects and objects were animate and pragmatically reversible, so that the child could only rely on the linguistic input to carry out the task. Fillers involved intransitive verbs or transitive verbs with inanimate objects. The materials complied with the constraints established by Hamburger and Crain (1982) to satisfy felicity conditions for testing. Samples of experimental materials can be seen in Figure 3.

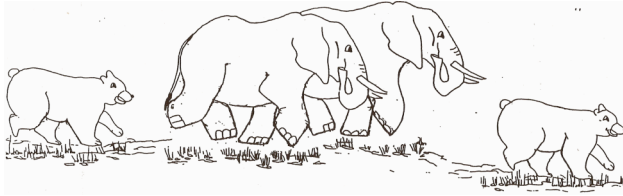


Figure 3: Sample of experimental materials

The subjects were all native speakers of (Central) Catalan from Barcelona, Prats de Lluçanès, and L'Hospitalet de Llobregat. The ages and sizes of the subject groups were as shown in Table 1.

group	#	age range	mean age
<4;6	12	3;5,9–4;4,11	3;11,26
4;6–5;6	11	4;6,10–5;3,25	4;11,6
5;6<	10	5;10,5–6;2,30	6;0,12
total	33	3;5,9–6;2,30	4;11,4
controls	22		

Table 1: Subjects

The results of comprehension for adult controls appear in Tables 2a and 2b, and those for the children in Tables 3a and 3b. Fillers were correctly interpreted by all subjects in all age groups.

a. Results, adults, subject relatives and fillers

group	SR			fillers			
	correct		object		other		
total	131/132	99%	1/132	0.7%	0	71/71	100%

b. Results, adults, object relatives

group	OR			ORp					
	correct		subject	other	correct	subject	oth		
total	132/132	100%	0	0	111/120	92%	9/120	7%	0

Table 2: Catalan adult performance



a. Results, children, subject relatives and fillers

group	SR			fillers				
	correct		object	other	correct			
<4;6	57/72	79%	12/72	16,5%	3/72	4%	44/44	100%
4;6–5;6	64/66	97%	1/66	1,5%	1/66	1,5%	37/37	100%
5;6<	60/60	100%					45/45	100%
total	181/198	91%	13/198	7%	4/198	2%	126/126	100%

b. Results, children, object relatives

group	OR					
	correct		subject	other		
<4;6	42/72	58%	21/72	29%	9/72	12.5%
4;6–5;6	42/66	64%	20/66	30%	4/66	6%
5;6<	55/60	92%	3/60	5%	2/60	3%
total	139/198	70%	40/198	22%	15/198	7.5%

group	ORp					
	correct		subject	other		
<4;6	5/60	8%	51/60	85%	4/60	6.5%
4;6–5;6	11/55	20%	44/55	80%		
5;6<	8/50	16%	42/50	84%		
total	24/165	14.5%	137/165	83%	4/165	2.5%

Table 3: Catalan child performance

The subject relatives (type a) were understood at least 79% of the time by even the youngest children. The object relatives with preverbal subject (type b) progressed across the age groups from about 58% to almost ceiling comprehension. But with object relatives with postverbal subjects there was no progression above 20% comprehension: more than 80% of the time it was the subject relativisation interpretation which was made, despite failure of number agreement.

The percentage of target responses for children is represented in Figure 4. (We refer the reader to Gavarró, Adani, Ramon, Rusiñol and Sánchez 2009 for detailed statistical analysis of the results.)

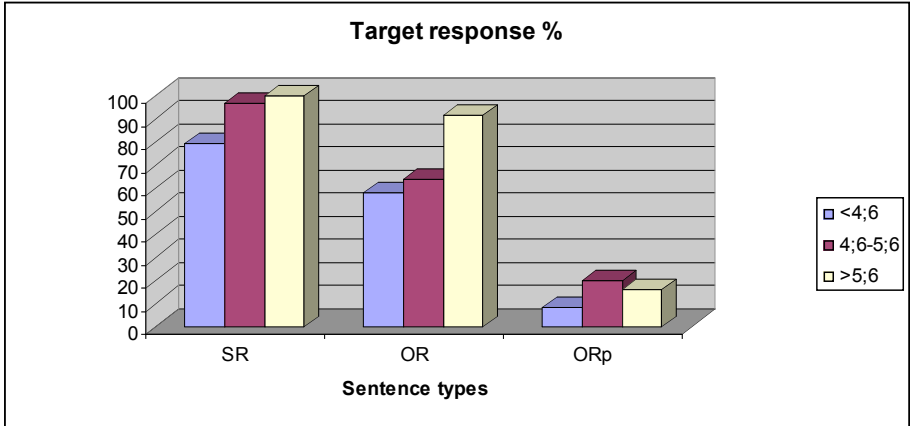
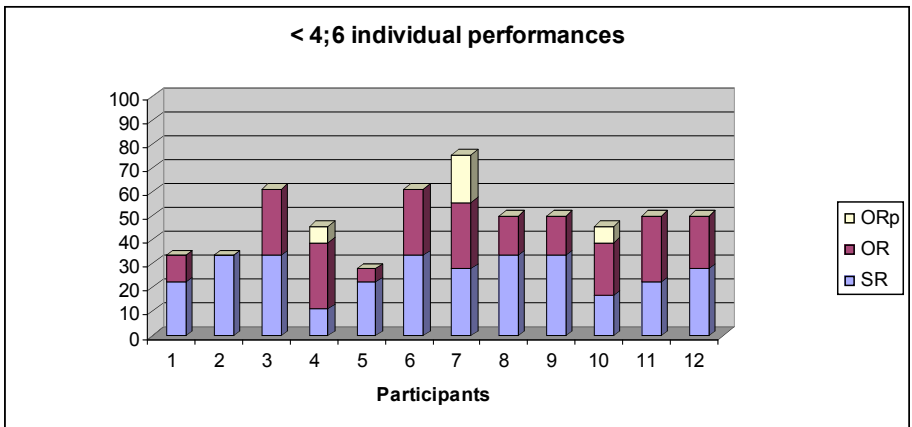


Figure 4: Child performance by groups

There was considerable individual variation, and some children achieved almost adult-like performance at a much earlier age than others. Individual performance for each age group is represented in Figure 5.



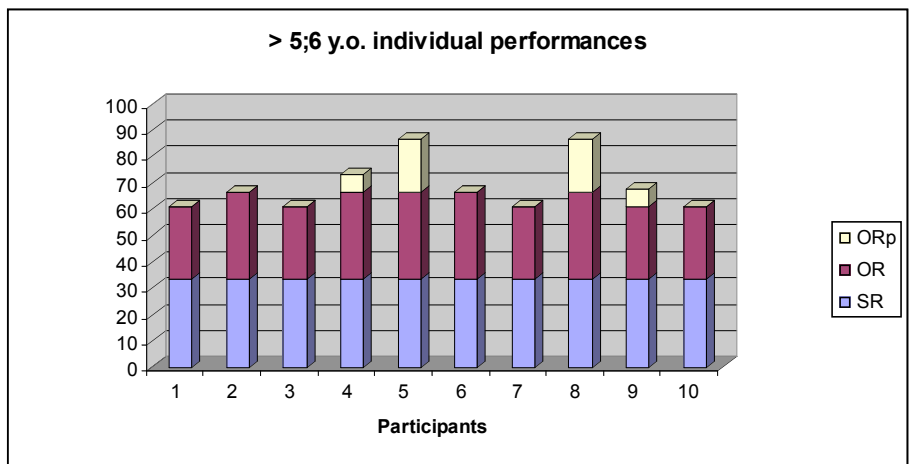
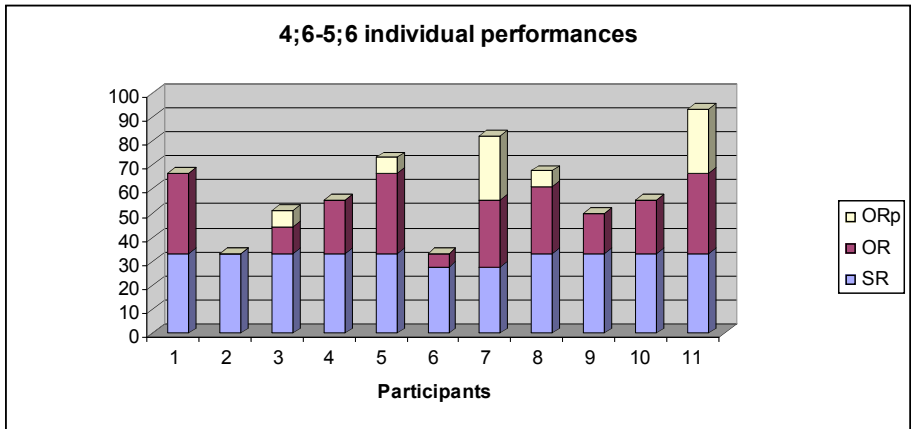


Figure 5: Individual child performance

### 3.2 The Italian Experiment

Italian is, in all respects relevant here, identical to Catalan, and our results can be validly compared with those of Adani (in press). From her results in Tables 4 and 5 we see that the pattern in Italian is similar to that in Catalan: subject relatives are understood at the earliest age recorded, while object relatives are not, but also

object relative comprehension varies as a function of word order. There is a clear progression with object relatives with preverbal subjects for the period investigated and 89% comprehension is reached at age 7; with object relatives with a postverbal subject, there is also a progression, but these object relatives are still interpreted as subject relatives in 30% of cases at age 7.

group	SR		
	correct	object	other
3	91%	9%	0%
4	93%	7%	0%
5	90%	9%	1%
6	96%	4%	0%
7	93%	7%	0%

Table 4: Italian results, children, subject relatives

group	OR			ORp		
	correct	subject	other	correct	subject	other
3	53%	34%	13%	36%	60%	3%
4	83%	10%	7%	59%	41%	0%
5	74%	15%	11%	54%	44%	2%
6	85%	14%	1%	55%	45%	0%
7	89%	11%	0%	70%	30%	0%

Table 5: Italian results, children, object relatives

The results of Arosio, Adani, and Guasti (2009), also for Italian, are represented in Figure 6. Comprehension was poorer than in Adani’s results although the pattern was the same. As suggested by Adani, it is quite possible that the small differences between one set of results and the other are a consequence of differences in the experimental design; we leave this issue open.

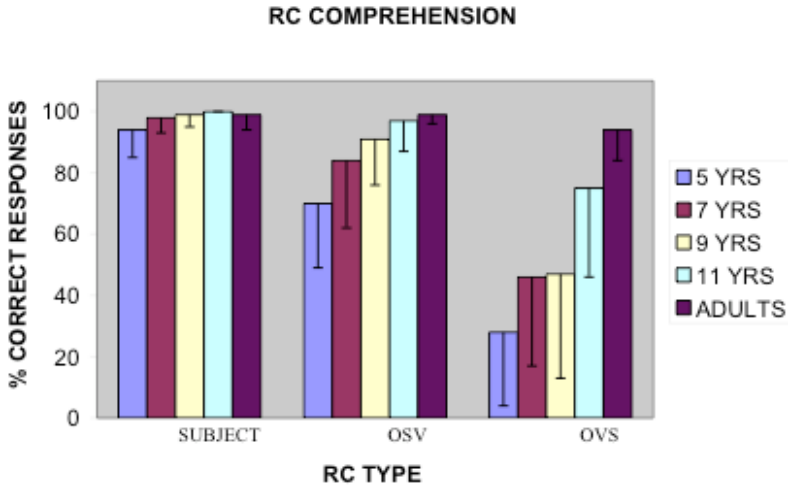


Figure 6: Italian child performance

## 4 Our account

Variation such as that observed in the experiments considered here invites the application of some kind of processing metric, but these are rarely non-arbitrary. Here we apply our simple measure of successive load on memory of unmatched attributes of categorial grammar in sentence processing based on the promptness of incremental resolution of categorial dependencies.

For our categorial analyses of the experimental materials we invoke two enrichments of basic categorial grammar. Firstly, to allow both preverbal and postverbal subjects, we invoke an ambidextrous categorial slash | such that an expression of type B|A is one which concatenates with an A both on the left or on the right to form a B (Morrill 1994, p.101):

$$(13) \quad B|A = \{s \mid \text{for all } s' \in A, s'+s \in B \ \& \ s+s' \in B\}$$

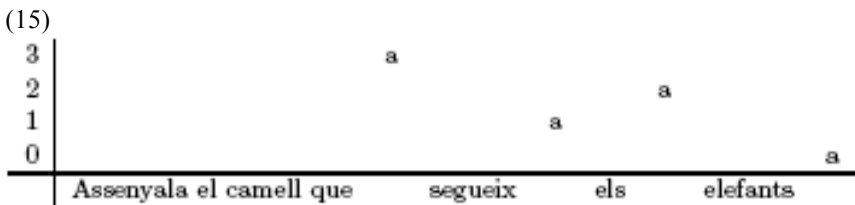
Thus a transitive verb such as *segueix* ('follows') will have type  $(S|N(\text{sg}))/N(\_)$ . Secondly, to allow gaps to be clause-initial (subject relativisation) or clause-medial or clause-final (object relativisation) in the bodies of relative clauses, we invoke residuated unary modalities with structural properties  $(p(s)+s=s+p(s))$ :

$$(14) \quad \langle \rangle A = \{p(s) \mid s \in A\}$$

$$[\ ]^{-1} A = \{s \mid p(s) \in A\}$$

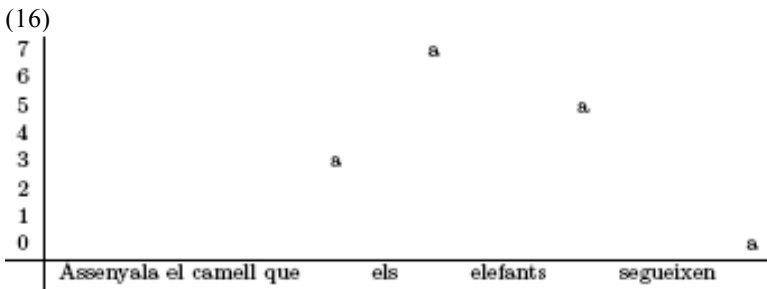
See e.g. Moortgat (1999). Thus the relative pronoun *que* will have type  $(CN(\alpha) \setminus CN(\alpha)) / (S \langle \rangle [\ ]^{-1} N(\alpha))$  where the variable  $\alpha$  indicates the sharing of agreement features (here we have just number).

Consider the analysis of the relevant part of sentence type (10a) above, as shown in Figure 7. The number of features which are unmatched and which must therefore be held in short term working memory at the boundaries are as follows:



The complexity never rises above 3 and there are no processing difficulties: children perform in an adult-like manner.

For type (10b), our analysis is as in Figure 8. The complexity profile is as follows:



This word order pattern can only correspond to an object relative and we expect the object relative reading. Children provide the object relative reading over half of the time even at the earliest age considered in our experiment, and give the adult interpretation to the sentence in 92% of cases in the oldest group. We attribute the lower performance relative to sentence type (10a) to the greater load on memory.

*Assenyola el canell*

*que*

*seguir*

*els*

*elefants*

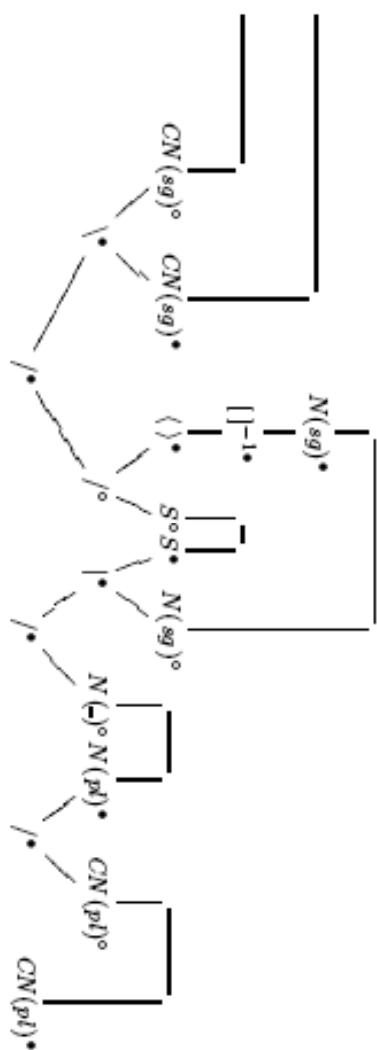


Figure 7: Net for sentence type (a)

Assenyala el canell

que

els

elefants

sequitzen

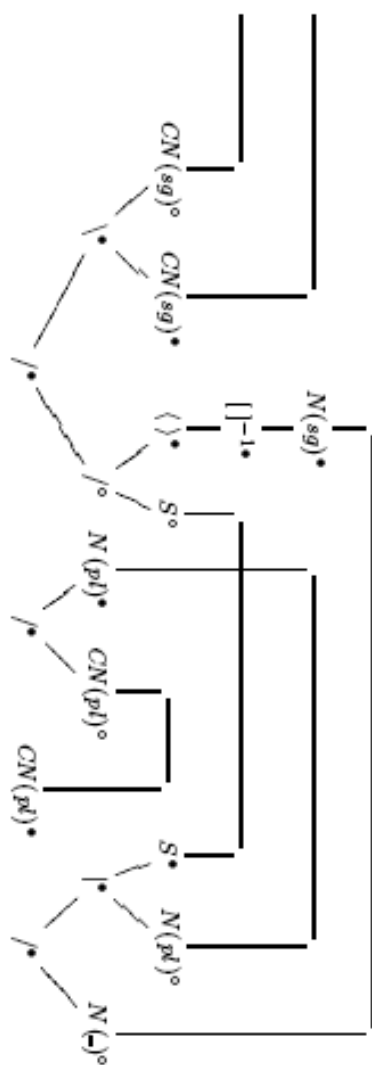


Figure 8: Net for sentence type (b)



For sentence type (10c) the results are surprising. The word order here is the same as for subject relatives (the sequence of the embedded clause consists in a transitive verb followed by a nominal). The number agreement used in the experimental materials removes this ambiguity, being consistent with only the postposed subject object relativisation reading. But the interpretation that children give to sentences of type (10c) is overwhelmingly the subject relative interpretation (and other interpretation errors are marginal). Why should this be?

The explanation we offer comes in three parts. Firstly, we believe that the children are not registering the plural morphology of the verb, and invoke both the singular and the plural verb grammatical types in their processing. Secondly, because of grammatical agreement in the derivations, the singular verb type gives rise to only the subject relative analysis and the plural verb type gives rise to only the postposed subject object relative analysis, and to the child the sentences are in principle ambiguous accordingly. But the subject relativisation derivation is of lower complexity than the object relativisation derivation. This, we would like to suggest, is part of the reason why children obtain the subject relative reading. But additionally, a noun phrase occurring after a transitive verb is canonically interpreted as the object of this verb (underlying SVO word order) and only non-canonically as the subject (underlying VOS word order). Hence by default a sequence comprising relative pronoun, transitive verb and noun phrase is interpreted as subject relativisation. Thus we attribute the overwhelming tendency of the children to obtain the subject relative reading of the type (10c) sentences to a double cause of lower complexity and default word order.

In more detail, Johnson, de Villiers and Seymour (2005) reported on the inability of four- and five-year-olds to disambiguate sentences in English on the basis of verb inflection alone: when exposed to sentences of the type *The cat sleeps on the bed* or *The cats sleep on the bed*, in which the plural marker on the noun is shadowed by the first consonant of the verb, children were unable to interpret the sentence on the basis of the number marking on the verb, even though at the same age these children were able to produce adult-like subject-verb agreement. Pérez-Leroux (2005) carried out a similar experiment with Dominican Spanish-speaking children. She compared the interpretation of sentences with null subjects such as *Duerme en la cama* '(He) sleeps on the bed' vs. *Duermen en la cama* '(They) sleep on the bed'; given the absence of overt subjects, the interpretation rests solely on the subject agreement inflection on the verb. But three and four-year-olds were not able to differentiate between the two different verb forms, while five and six-year-olds were in the 60–70% range in accuracy. Again, the results indicate that the detection of verb inflection in comprehension lags behind.

We do not know why there should be this deficit in comprehension as opposed

to production, but, whatever the reason,<sup>2</sup> we take this as sufficient independent evidence to propose that, in the type (10c) sentences of the Catalan and Italian experiments, the children do not pick up on the plural morphology of the transitive verb and allow into derivations both the singular lexical category (S|N(sg))/N( ) and the plural lexical category (S|N(pl))/N( ). Then the former allows (only) the (subject relative) analysis of Figure 9 and the latter allows (only) the (postposed subject object relative) analysis of Figure 10.

The complexity profiles are as follows:

(17)

3			ab			
2				b	ab	
1				a		
0						ab
a.	Assenyala el camell que	segueixen.	SG	els	elefants	
b.	Assenyala el camell que	segueixen.	PL	els	elefants	

The complexity of the former is lower (just) than the latter because the subject relative analysis allows checking off of the (mis)assigned singular number feature attributed to the verb as soon as the verb is heard, whereas the latter does not allow the checking off of the (correctly) assigned plural number feature attributed to the verb until after hearing the following word. The difference is minimal and we believe that the children's strong preference is due in addition to an expectation of canonical word order.

In relation to the default word order in Catalan, in production, Gavarró and Cabré-Sans (in press) show that the distribution of subjects with transitive verbs is preferably preverbal both in children (after MLU 2.5) and in adults; in fact, there is a statistically significant difference between children and adults in that children present even more preverbal subjects with transitives than adults (80% for children vs. 56% for adults on average).

We have offered an account of child Romance relativisation performance largely on the basis of a categorial complexity measure, but including other factors. We believe that such an approach is theoretically well-founded and may offer a principled basis to the explanation of other performance phenomena in the linguistic behaviour of children and other populations.

---

<sup>2</sup> An anonymous reviewer suggests that the low phonetic salience of coronal nasals may result in poor perception. This could be tested experimentally with a phonetically more salient plural marker.



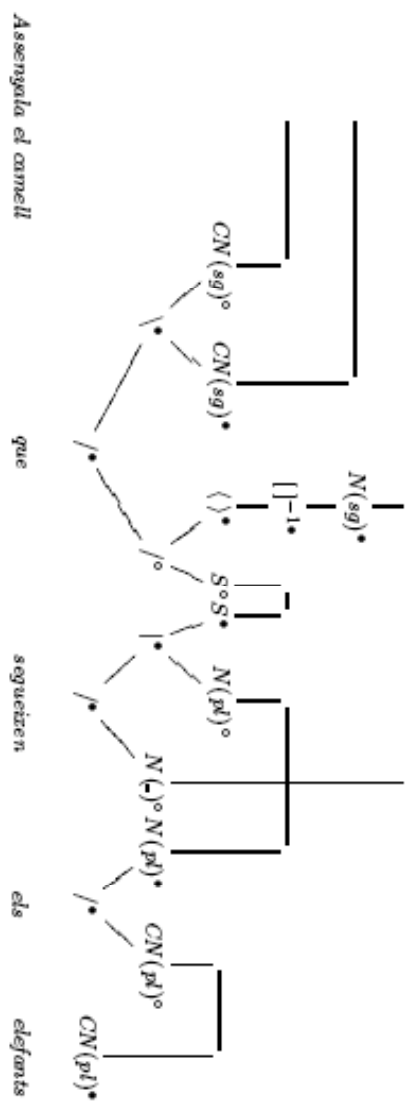


Figure 10: Net for sentence type (c) with assignment of verb as plural

## Acknowledgments

We thank two anonymous reviewers for their very helpful comments. The authors acknowledge the support of DGICYT projects TIN2008-06582-C03-01 (SESAAME-BAR), HUM2006-13295-02-01 and 2009SGR1079.

## References

- Adani, Flavia (in press) Re-thinking the acquisition of relative clauses in Italian: Towards a grammaticality-based account. *Journal of Child Language*.
- Arosio, Fabrizio, Flavia Adani, and Maria Teresa Guasti (2009) Grammatical features in the comprehension of Italian relative clauses by children. In José M. Brucart, Anna Gavarró, and Jaume Solà, editors, *Merging Features: Computation, Interpretation, and Acquisition*, Oxford University Press, Oxford, 138–155.
- Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.
- Gavarró, Anna, Flavia Adani, Marta Ramon, Sandra Rusiñol, and Raquel Sánchez (2009) On the comprehension of relative clauses in child Catalan. GGT research report GGT-09-11, UAB.
- Gavarró, Anna, and Yolanda Cabré-Sans (in press) Subjects, verb classes and word order in child Catalan. In John Grinstead, editor, *Perspectives on Typical and Atypical Hispanic Language Development*, John Benjamins, Amsterdam/Philadelphia.
- Gibson, E. (1998) Linguistic complexity: locality of syntactic dependencies. *Cognition*, 68:1–76.
- Girard, Jean-Yves (1987) Linear logic. *Theoretical Computer Science*, 50:1–102.
- Hamburger, H. and S. Crain (1982) Relative acquisition. In S. Kuczaj, editor, *Language Development: Syntax and Semantics*. Erlbaum, Hillsdale, NJ.
- Johnson, M. E. (1998). Proof nets and the complexity of processing center-embedded constructions. *Journal of Logic, Language, and Information*, 4(7).
- Johnson, V., J. de Villiers and H. Seymour. (2005) Agreement without understanding? The case of third person singular /s/. *First Language*, 25.
- Lambek, Joachim (1958) The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170. Reprinted in Buszkowski, Wojciech, Wojciech Marciszewski, and Johan van Benthem, editors, 1988, *Categorial Grammar, Linguistic & Literary Studies in Eastern Europe* volume 25, John Benjamins, Amsterdam, 153–172.
- Moortgat, Michael (1999) Constants of grammatical reasoning. In G. Bouma, E.

- Hinrichs, G.-J. Kruijff, and R.T. Oehrle, editors, *Constraints and Resources in Natural Language Syntax and Semantics*, CSLI Publications, Stanford, 195–219.
- Morrill, Glyn (2000) Incremental Processing and Acceptability. *Computational Linguistics*, 26(3):319–338.
- Morrill, Glyn V. (1994) *Type Logical Grammar: Categorical Logic of Signs*. Kluwer Academic Press, Dordrecht.
- Pérez-Leroux, Ana-Teresa (2005) Number problems in children. In C. Gurski, editor, *Proceedings of the 2005 Canadian Linguistics Association Annual Conference*, 1–12. <http://ling.uwo.ca/publications/CLA-ACL/CLA-ACL2005.htm>.
- Roorda, Dirk (1991) *Resource Logics: Proof-Theoretical Investigations*. PhD thesis, Universiteit van Amsterdam.