Relative clauses from the input: syntactic considerations on a corpus-based analysis of Italian

ADRIANA BELLETTI & CRISTIANO CHESI
CISCL - University of Siena
belletti@unisi.it chesi@media.unisi.it

A well-known classical finding from both acquisition and processing is that (headed) Object Relative Clauses (ORs) are harder than (headed) Subject Relative Clauses (SRs) for children to acquire, and slower for adults to process (Adams 1990, Adani et al. 2010, Brown 1972, de Villiers et al. 1994, De Vincenzi 1991, Gordon et al. 2004, Tavakolian 1981, Warren & Gibson 2002, among many others). In this work, we aim at investigating which typologies of SRs and ORs are present in corpora of standard Italian and the way their frequency compares with some recent experimental findings from elicited production (Belletti 2009, Belletti & Contemori (2010), Contemori & Belletti (this volume)), and with the syntactic account that has been proposed in terms of a featural approach to locality (Friedmann et al. 2009). We also address the issue of the possible role of frequency in conditioning the linguistic performance in the domain of ORs and Passive Object Relatives (PORs).

1. Introduction

In this paper, we discuss the quantitative distribution of Relative Clauses (RCs) occurring in Italian corpora; in particular we evaluate:

1. what kind of RCs are present in child-directed speech and in Standard Italian: Subject Vs. Object Vs. Indirect Object headed RCs (henceforth SRs, ORs and IORs, respectively);
2. how many SRs are in the passive voice, hence they could have been realized as active ORs (following Belletti 2009, we will call such RCs Passive Object Relatives, PORs henceforth);
3. which distribution certain relevant syntactic properties/features have in ORs, specifically:
    a. the position of the subject when it is overtly realized;
    b. the nature of the Subject: Lexical Vs. Pronominal Vs. Null;
    c. the animacy feature associated with the head of the relative clause and with the Subject of the relative clause;
4. Whether there is any difference in the analyzed Italian registers (e.g. Standard Public-Broadcast Television Vs. Child-directed speech).

On the basis of such quantitative analysis, we want to verify whether or not the statistical distribution of the structural configurations considered is somehow predictive, and/or can be considered the cause, of the difficulties we know are related to ORs processing and acquisition. Moreover, we want to verify how the intervention account in terms of a featural approach to the locality principle Relativized Minimality (RM; Rizzi (1990, 2004), Starke (2001)), as developed in Friedmann et al. 2009 is coherent with the observed distribution in the naturalistic corpora investigated. One crucial issue that we address in this work is whether the (un-)frequency of the analyzed syntactic structures (SRs, ORs, PORs) could play a crucial role in determining the speakers' behavior in the (elicited) production of the complex OR structures. The complement of this question is also naturally raised, whether syntactic complexity may directly condition frequency in the input, such as the frequency of the complex OR structures.

2. Background

Recent experimental results on both production and comprehension of SRs and ORs in Italian (e.g. Adani 2010, Arosio et al. 2009, Belletti & Contemori 2010, Contemori & Garraffa 2010), have confirmed the different status of SRs and ORs in both children and adults, with ORs harder than SRs, in various respects (Adams 1990, Adani et al. 2010, Brown 1972, de Villiers et al. 1994, De Vincenzi 1991, Gordon et al. 2004, Tavakolian 1981, Warren & Gibson 2002, among many others over a long period of time). One crucial finding concerns adults: in an elicited production task (Belletti & Contemori (2010), Contemori & Belletti (this volume)), Italian adults tend not to produce ORs in a very systematic way; specifically, there appears to be an often strong tendency to avoid ORs, in favor of the production of an alternative structure, typically a SR which is able to preserve the same intended meaning. One privileged such alternative is offered by the use of passive, that is utilized up to almost 90% in the different groups of adults investigated in the experiments (see also Belletti 2009 for related findings).

The results have indicated that the production of what we refer to as Passive Object Relatives/PORs, is the preferred option for adults and becomes the preferred option for children as well, as soon as passive becomes productively available to them, around age 5. PORs have also been recently tested in comprehension (Contemori & Belletti this volume), and they have turned out to be significantly better comprehended by the children who master passive, than (active) ORs (with or without resumption; on child resumptive relatives, see Guasti & Cardinaletti 2003). Converging results have been found cross-linguistically in the same production experiment run with children of different languages (Friedmann et al. 2010), and in self-paced reaction time experiments with adults (e.g. Lin & Bever 2006 on Mandarin Chinese).

Our contribution in this paper is to bring into the picture a different kind of empirical data: a pilot corpus study of (headed) SRs and ORs in standard Italian. As a background, we first review the main recent experimental findings mentioned, and the syntactic account that has been proposed in terms of a featural approach to locality/RM (Friedmann et al. 2009). We then move to the novel corpus data and elaborate on their relevance for the assumed locality

approach as well as on their bearing on the issue of the respective role of syntactic complexity in grammar on the one side, and frequency in input on the other (Gennari & Mac Donald 2009, Tomasello 2003).

2.1. Experimental findings in production

In Belletti & Contemori (2010) and Contemori & Belletti (this volume), an adaption has been presented to Italian of a Preference task experiment from Novogrosky & Friedmann (2006) aiming at eliciting the production of SRs and ORs.[1] All relevant details of the design and the task are presented in the references quoted, to which the reader is referred. Here, we give the essential features of the design and of the results obtained. The task consisted in presenting the experimental subjects with a situation in which two children/persons were undergoing a certain event, bearing either the role associated with the subject or the one associated with the object. The experimental subjects were then asked to choose between the two situations, saying which person he/she would rather be. Depending on the different introductory story, the sentence elicited was either a SR or a OR. The experimental subjects were invited to begin each sentence with "I would rather be …". The eliciting story was built according to two conditions, a subject/object change condition, in which the subject/object present in the story changed, and a verb change condition, in which what changed was the verb presenting the event of the story. Two examples below give an illustration of the elicitation of a SR and of an OR in the object change and in the verb change condition, respectively (number mismatch conditions between the relative head and the subject of the relative clause were also tested in the references quoted, which are not relevant to the present discussion and will thus be ignored in the illustration, for which only examples from the match condition are presented in (1)):

(1) SR: *"There are two children. One child comb the neighbours one child comb the grandparents. Which child would you rather be? Start with ''I would rather be ...''*
Target answer:
"(Vorrei essere il bambino) che pettina i vicini/i nonni"
''(I would rather be) the child that combs the neighbours/grandparents"

OR:*"There are two children. The grandpa looks for one child and the grandpa finds one child. Which child would you rather be? Start with ''I would rather be the child ...''*
Target answer:
"(Vorrei essere il bambino) che il nonno cerca/trova
''I would be the child that (the grandpa) looks for/finds"

The task has been adapted to be presented to either children or adults. Overall, 100 children aged 3:4-8:10 have been tested and 28 adults (see Contemori & Belletti, this volume, for the presentation of all details). We report in Table 1 below the results from the adults' productions (from a first tested group of 18

[1] First pilot adaptation to Italian, with similar results, in Utzeri (2007), with school age children 6-11.

adults; the subsequent 10 adults have confirmed the same pattern, Contemori & Belletti, this volume):

	# SR	*SR %*	*# OR*	*OR%*
Relatives produced	179/180	99.5%	25/234	10.6%
"si fa"/causative passive	-	-	-	-
Copular passive	-	-	89/234	38%
Reduced passive	-	-	117/234	50%
. . .	. . .	. . .	. . .	. . .
Change of character	-	-	2/234	0.4%
Change of verb	-	-	1/234	0.8%

Table 1. Summary of the relevant results from the adults' productions (Contemori & Belletti, this volume).

As Table 1 clearly indicates, the production of (active) ORs is extremely low in adults – around 10% –, highly significantly lower than the ceiling production of SRs. What adults do, overwhelmingly, is to produce PORs in place of (active) ORs – around 90% –. PORs can be either copular or reduced.[2] Adapting from the presentation in Belletti (2010), Figure 1 illustrates the tendency shown by young children in approaching the adults' type of production, so that as they grow older more PORs are produced in place of (active) ORs, in the same eliciting conditions:

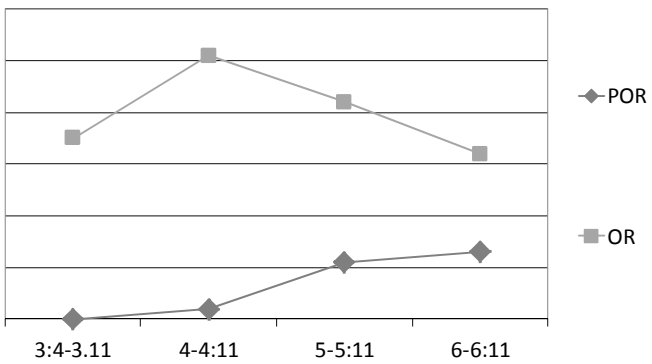


Figure 1. Summary of the results from the adults' productions (Contemori & Belletti, this volume).

Contemori & Belletti (this volume) present results from older children, up to age 8:10, showing that the tendency becomes stronger, and the older group, which masters passive well, produces a significantly higher number of PORs in place of ORs, similarly to adults. Furthermore, it is also shown that these results are not a task related effect due to some bias of the Preference task, as totally comparable results are obtained with a different elicitation design (a Picture description task, see the reference quoted for details). In conclusion, the

[2] No "si fa"/causative passive produced by adults, in contrast with the children's productions; for the interest of this difference in the kind of passives utilized by children and adults, see Contemori & Belletti, under submission)

experimental results on production in Italian have shown a clear and strong preference for the production of PORs when ORs were elicited, in both adults and children, depending on the developmental stage for the latter group.

2.2. Smuggling as a computation which eliminates intervention

These results open up the issue of a comparison of the complexity of different syntactic computations such as passive and (active) object relatives. A promising account has been proposed in terms of locality, specifically in terms of a featural approach to Relativized Minimality, as developed in Starke 2001, Rizzi 2004, which has been adapted to account for development in Friedmann et al. 2009, based on results from comprehension of SRs and ORs in Hebrew speaking children, aged 3:7-5 (see also Grillo 2008, for a related approach to agrammatism). According to the approach in Friedmann et al. (2009), in a structural situation meeting the locality/RM configuration

X … Z … Y …

where X = the target position – the position of the relative head in CP in the case of relative clauses –, Z = the intervener position – the subject position of the relative clause in the case of ORs – , Y = the origin position – the object position within the relative clause, where the relative head is merged in the case of the ORs

the dependency between the relative head in the target position X and its merge position Y within the relative clause, can be hard (sometimes even impossible) to establish for (young) children and may lead to slower processing for adults, if the target head X in CP and the intervener Z in the relative clause, share the feature labeled [NP]. The [NP] feature refers to presence of a "lexical restriction" in both the head of the relative clause and the intervening subject, so cases in which they both contain a full lexical noun phrase. Lexically headed ORs with an intervening lexical subject in the relative clause are thus singled out by this system as the hardest structures to compute. According to this system, the crucial property is not that much whether there is an intervener or the distance between X and Y, but rather whether the Target X and the Intervener Z share some computationally relevant feature on the attracting head. The hypothesis is that the feature [NP] is a crucially relevant attracting feature in lexically headed relative clauses. The schematic representation in (2) illustrates the intervention situation created in the OR, in which the [NP] feature of the intervening lexical subject Z is properly included in the feature set of the Target X (R in X corresponds to the attracting feature of relative heads):[3]

(2) il bambino che il nonno cerca/trova <il bambino>

+R +NP	+ NP	+R +NP
X	Z	Y

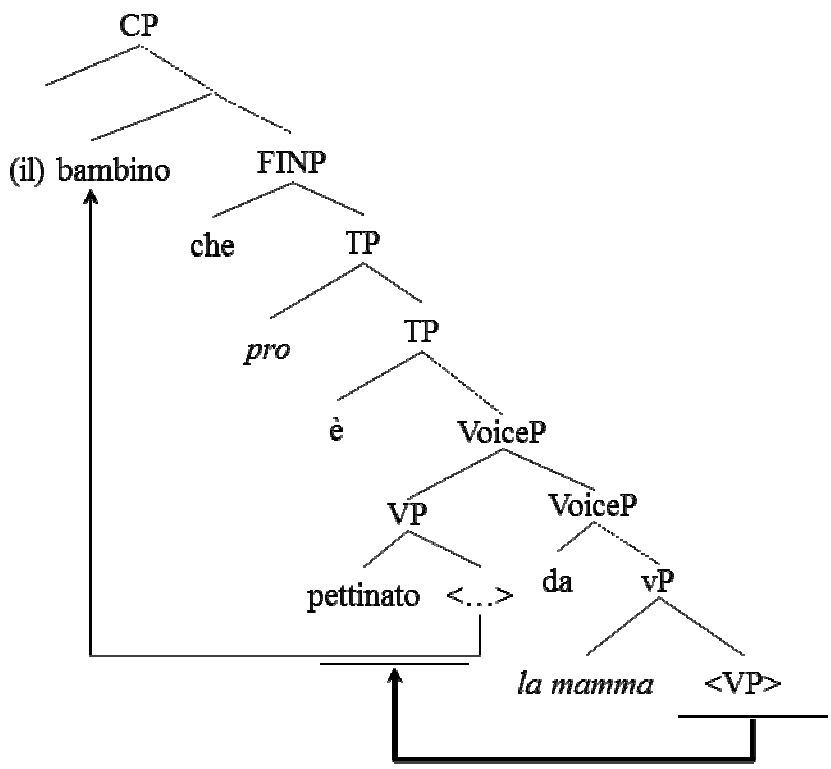
The intervention effect which arises in lexically headed ORs across an intervening lexical subject is the source of the difficulty in the processing of object relative clauses.

As discussed in Belletti 2009, 2010, the use of passive can be seen as an optimal way to overcome the described intervention effect which inevitably arises in the relativization of a direct object across an intervening lexical subject. Assuming a

[3] On the difference between children and adults in the ability to compute the inclusion relation, see the discussion in Friedmann et al. (2009), and Belletti et al. (submitted).

derivation of what we call *passive* in the terms developed in Collins 2005, which involves movement of a verbal chunk containing the verb and the object across the intervening lexical subject, whereby intervention is eliminated - the process referred to as *smuggling* in Collins 2005 - a principled reason is provided for the (often overwhelming) appeal to passive in the syntactic computation of an OR in Italian (and also in other languages, as mentioned) that the experimental results have so clearly revealed. The assumed derivation is schematically illustrated in (3) for the Italian POR "il bambino che è pettinato dalla mamma" (the child that is combed by the mom):

(3)



A natural question to ask is: to what extent are PORs also found in naturalistic corpora? In the following section of this study, we address this question.

3. The analysis

3.1 Corpora used

The first kind of production we analyzed is the child-directed speech; to retrieve these productions, we inspected the Italian section of the CHILDES database (8 children, 113 files, plus 1 child, 19 files, whose data have been collected and transcribed at CISCL, Matteini 2011). Then, we compared the distribution of the RCs in these files with the distribution found in two other Italian corpora of adult speech: the Siena University Treebank (henceforth SUT, 29 television news taken from special editions of the national television news, shortened and simplified for on-line translation in Italian Sign Language, Chesi et al. (2008)) and the Italian Television Corpus (Corpus di Italiano Televisivo, henceforth CIT, 7 TV programs such as national editions of talk shows, standard news,

commercials etc., Spina (2005)). In the table below, we report the size of the corpora and their format.

Corpus Name	References	Size (in words)	Format
CHILDES	MacWhinney & Snow (1985)	132 files (390.511 words: 115.357 produced by children, 275.154 produced by adults)	chat format
SUT Siena University Treebank	Chesi et al. (2008)	29 TGs (17.981 words)	SUT (specific constituency/ dependency format, XML)
CIT Corpus di Italiano Televisivo	Spina (2005)	7 TV programs (42.668 words)	morphologically tagged text

Table 2. The corpora used for the analysis of RCs

3.2 Methods

Since the corpora were differently structured, we used different tools for retrieving relative clauses in a semi-automatic way: for simple-text encoded corpora (CHILDES) we used Regular Expressions through the GREP tool[4]. Regular Expressions are very flexible devices to define ordered sets of characters that correspond to specific morphological units: for instance, Italian SRs and ORs are (in almost all cases, but see the discussion on Reduced RCs in 3.3 and table 7) clearly marked with an invariable relative pronoun/complementizer (i.e. “che”); this can be productively encoded with a simple regular expression like the one in (4) that picks up all occurrences of “che” produced by a certain speaker (“TIER”) in a CHAT-encoded file (MacWhinney et al. 1985):

(4) Regular expressions using “grep”:

```
grep -i -n -E "TIER:([[:space:]]|[[:punct:]]|[[:alpha:]])*[[:space:]]che[[:space:]]"
```

Even though many occurrences of “che” introduce in fact declarative clauses and not RCs in Italian[5], this approach allows us to restrict the set of data to be manually inspected and it offers a precise way of counting linguistic phenomena. For instance, rather subtle regular expressions can be written for isolating past participles looking at the relevant morphological inflection; this allows one to restrict the set of data to be inspected for counting those past participles that can be Reduced RCs; the fact that such expressions isolate a certain number of verbs is a fact that can be precisely replicated.

[4] GREP is a Unix native Regular Expression interpreter that has been ported under many platforms; it is easy to use, free, reliable and fast; Given a Regular Expression it returns the line in the text where a matching occurs.

[5] The percentage of RCs with respect to all the occurrences of “che” ranges from a modest 12% in the adults section of CHILDES, to 83% in SUT.

On the other hand, with tagged corpora we can use a more precise counting system that relies on POS tags and on syntactic nodes annotation[6]: TGrep (Rohde 2004) is an extension of the Regular Expression Interpreter that allow us to search for specific syntactic patterns in a tagged corpus. For instance a non-reduced RC can be simply isolated using the pattern in (5).a, whereas an OR with the relative head and the subject of the relative both marked with the +animate feature can be retrieved with the expression in (5).b:

(5) a. tgrep 'NP.rel < C.rel'

b. tgrep 'NP.rel-obj.anim, NP-subj.anim'

3.3. Rough summary of the data collected

In this section, we present the main results of our quantitative analysis. In the tables below, we split the CHILDES corpus in the adult section (CHI A) and in the children section (CHI C).

Corpus	Tool used	# of analized words	# of "che" (%)		# of RCs (%)	
CHI A	Keyword [che]	275.154	5.580	(2,03)	677	(0,25)
CHI C	Keyword [che]	115.357	747	(0,65)	94	(0,08)
CIT	Tag [POS="pro:rela"]	42.668	1027	(2,4)	477	(1,1)
SUT	Tag [C.rel.pro]	17.981	210	(1,17)	174	(0,9)

Table 3. The frequency of the keyword "che" in all corpora compared to the frequency in which they correctly isolate RCs.

As mentioned, the table above shows that there is a substantial variability with respect to the "che" usage across corpora (as "che" can be either a declarative clause complementizer or a RC complementizer).
In table 4 the count of RCs with respect to their macro-typology is presented: SRs vs. ORs vs. IORs.

Corpus	# of Rs	# SRs (%)		# ORs (%)		# IORs (%)	
CIT	477	314	(66%)	117	(25%)	46	(9%)
CHI A	677	441	(65%)	228	(34%)	8	(1%)
SUT	174	162	(93%)	12	(7%)	-	
CHI C	94	83	(88%)	11	(11%)	-	

Table 4. RC macro-classes.
As expected, the number of SRs is significantly higher than the number of ORs. IORs are the less frequent type of RCs. While CIT and CHI A show comparable ratios SRs/ORs (SRs are

[6] Part-Of-Speech (POS) tags are morphosyntactic classes associated to the words in an annotated corpus (e.g. "(D-MS il)" indicates that "il" is a Determiner, Masculine, Singular); the syntactic annotation includes features related to the thematic dependency (e.g. "(VP (NP-subj (D-MS il) (NN-MS cane)) (V-IP3S abbaia))". The standard annotation (PENN-TREEBANK-II) has been expanded in order to include the relevant features under analysis (e.g. animacy: "(NP-subj-anim …)"; on animacy see below).

roughly twice more frequent than ORs[7]), this is highly contrasting with respect to the ratio we found in SUT and CHI C. While the CHI C count is expected, as in the CHILDES database children are registered up to age 3;4 (table 5), and the production of ORs (and relatives in general) is poorly attested at this young age, the SUT frequency seems to interestingly reveal that the "naïve" intuition behind the notion of "simplified Italian suitable for on-line translation" toward LIS leads to avoid ORs.

Corpus	Camilla	Diana	Guglielmo	Marco	Martina	Raffaella	Rosa	Sabrina	Viola
1;5									
1;6									
1;7									
1;8									
1;9									
1;10									
1;11									
2;0				2 - 0					
2;1		1 - 0		1 - 0					
2;2	3 – 0		5 - 0	1 - 0					
2;3				3 - 0				0 - 1	
2;4	5 – 0			2 - 0				1 - 0	
2;5				2 - 1					1 - 0
2;6		2 - 0						0 - 1	
2;7			1 - 0		2 - 0			9 - 0	
2;8						1 – 0			
2;9	1 – 4		2 - 0			1 – 1		1 - 0	
2;10							1 – 0	3 - 0	
2;11			3 - 2			5 – 0	10 – 0		
3;0									
3;1	1 – 0						1 – 0		
3;2									
3;3							1 – 1		
3;4	6 – 2								
3;5									

Table 5. RC macro-classes in CHI C: gray cells corresponds to the files present in CHILDES; the two numbers in the cells (*n* - *m*) represent the number of SRs - ORs.

To answer the main question of this study, whether and to what extent PORs are present in spontaneous production, we split the SR typology in active (labeled SRs) and passive voiced SRs (i.e. PORs). The result of this is reported in table 6:

[7] The general ratio between SRs and ORs seem to be steady cross-linguistically (see the values presented for very diverse languages such as e.g. Hamann & Tuller 2010 on French, Carreiras et al. 2010 on Basque).

Corpus	# of Rs	# SRs (%)		# ORs (%)		# PORs (%)	
CIT	477	295	(62%)	117	(25%)	19	(4%)
CHI A	677	440	(65%)	228	(34%)	1	(0,1%)
SUT	174	159	(91%)	12	(7%)	3	(2%)
CHI C	94	83	(88%)	11	(11%)	-	

Table 6. RC macro-classes with SRs split in active (SRs) and passive (PORs) SRs.

This table shows that the presence of full PORs is almost unattested across all corpora. This is in striking contrasts with the experimental results of elicited production described in section 2.1 (Belletti & Contemori 2010, Contemori & Belletti this volume).
Including in the counting also all possible reduced PORs (e.g. "the boy chased (by the policemen)"[8]) the situation does not change significantly, (with the exception of the SUT data):

Corpus	# of Rs	# SRs (%)		# ORs (%)		# PORs (%)	
CIT	477+48	295	(56%)	117	(22%)	19+48	(13%)
CHI A	677+78	440	(58%)	228	(30%)	1+78	(10%)
SUT	174+22	159	(81%)	12	(6%)	3+22	(13%)
CHI C	94	83	(88%)	11	(11%)	0+15	(?)

Table 7. RC macro-classes with SRs split in active (SRs) and passive (PORs, full + reduced) SRs. (PORs in CHI C cannot be safely quantified since the reduced forms used are probably simple adjectival modifications, whence the question mark).

PORs are mostly realized in a reduced format in all corpora; in CIT and in CHI A they are less frequent than ORs; in SUT, PORs turn out to be more frequent than ORs if reduced ones are included.[9] Children do produce some pseudo-reduced PORs (e.g. "mamma io ho le mani occupate"/lit: I have the hands occupied, Camilla 3;4.9), but since passive is unattested in simple declaratives at this stage in the same corpora, we concluded that these utterances are instances of adjectival modifications.
In the end, we looked closer at the typology and position of the subject in the attested ORs: in particular we considered in how many ORs the subject was lexical or null and, in the first case, with which frequency it appeared pre- or post-verbally:

[8] Both long, with the by-phrase, and short, without by-phrase reduced relatives are included.
[9] We do not have any precise hypothesis to offer as to why PORs including reduced ones should more numerous than ORs in SUT; we speculate that this fact may correlate with the high presence of reduced PORs in the elicited production by adults (Tables 11-12 and the surrounding discussion), which may be considered the optimal solution to the production of an ORs, under the eliciting conditions. Since the simplified Italian of SUT involves a "planned" simplification (see § 5), choice of the optimal solution in SUT may not be surprising.

Corpus	# of ORs	# pro V (%)		# S V (%)		# V S (%)	
CIT	117	72	(61%)	19	(25%)	10	(13%)
CHI A	228	139	(61%)	10	(4%)	80	(35%)
SUT	12	5	(42%)	3	(25%)	4	(33%)
CHI C	11	2	(8%)	-		9	(82%)

Table 8. Subject typology and distribution in ORs ("pro V" = null subject; "S V" = pre-verbal lexical subject; "V S" post-verbal lexical subject).

Whereas the preference for having an empty subject is clearly present in the CIT, in the CHI A, and, marginally, also in the SUT, a less straightforward tendency can be drawn from the pre-/post-verbal opposition: in this sense, both children (CHI C) and child directed speech seem to prefer the post-verbal (often pronominal) solutions, while the CIT shows a slight tendency in favoring the preverbal lexical alternative.

3.4. Discussion 1

Given the frequency distributions presented in the previous section, the question raised in 2.2 has the following answer: PORs are not a frequent structure in the naturalistic input. Since PORs have turned out to be the most frequently produced structure in the elicited productions summarized in section 2.1, for both children and adults, the conclusion must then be drawn that, despite their poor frequency in spontaneous speech, the linguistic performances revealed by the experimental results do not simply reflect the shape of the linguistic input. Hence, we conclude that PORs, which are the preferred structures in the elicited productions, must be preferred on different grounds than as a simple and straightforward consequence of a frequency effect. We submit the proposal that the preference for PORs in elicited production is a consequence of the optimal way to eliminate intervention that use of passive in ORs offers, as illustrated in 2.2. We delay until section 5 a possible hypothesis on the origin of the tension which has emerged between the results from elicited production on the one side and the new results from the naturalistic performance on the other, revealed by the corpus analysis. We now make some considerations on a related aspect of the issue concerning frequency in the input, and point out that the significance of what is or is not (in the domain of relative clauses) most frequently present in the analyzed corpora, must be treated with caution.

Looking at the distribution of relatives in the corpora, the SRs vs. ORs asymmetry could directly fit with the hypothesis that SRs are the most frequent type of relative clause since they involve a less complex syntactic derivation, than the one of ORs, which, in the case of headed ORs with a preverbal lexical subject in particular, typically gives rise to the intervention effect discussed in 2.2. Hence, one could interpret the more frequent presence of SRs in corpora as a consequence of their less complex derivation compared to ORs. In fact, the picture is less straightforward and more articulated; if we reconsider the frequency of SRs and ORs with respect to the verb classes and their subcategorization frame, we observe that the SRs/ORs asymmetry is not there:

Verb class	# SR	# OR
Unacc.+Unerg.+be	231	0
Transitive	161	193
Di-transitive	22	35

Table 9. SRs and ORs distribution across verb subcategorization classes (CHI A corpus).

In the relevant cases, i.e. with transitive verbs (and di-transitives), the difference between the number of SRs and ORs is not significant (t = 1.5934, df = 41.355, p-value = 0.1187). Adult speakers who have the computational capacity to process the complex OR structure, do so in spontaneous production to an extent which is comparable to the production of SRs with transitive verbs; in the analyzed corpora they have produced even more ORs than SRs in absolute numbers. Hence, bare frequency does not directly reflect the complexity of a given structure.

In conclusion, what frequency in corpora may reveal is not a trivial matter in both directions: i. it is not the case that speakers always tend to produce those structures which are more frequent in corpora, as revealed by the ample presence of PORs in elicited production and their very limited presence in the Italian corpora analyzed; ii. nor is it true that speakers always tend to produce those structures which are computationally less complex, as revealed by the balanced presence in the input of SRs and ORs with transitive verbs. This latter point is also coherent with the experimental results on adults' elicited production, in which the ample production of PORs witnesses the preferred use of a relatively complex computation (e.g. a computation which needs some time to fully develop in children).

As a last point, we note that the conclusion that bare frequency does not immediately reflect the complexity of certain potentially alternative structures (e.g. SR as POR instead of OR), is also supported by the distribution of the subject within the ORs present in the corpora: as illustrated in table 8, in all corpora the empty subject is the most attested option (61% in the SUT and CHI A). This could be interpreted as a tendency in favoring above chance a null pronominal subject. A null subject allows for a computation in which intervention is less strong, given a feature-based intervention approach, along the lines of Friedmann et al. 2009, as no NP feature is shared by the target and the intervener, in the sense illustrated in 2.2. However, if we look at the null subject rate in declarative sentences, we notice that the percentage of null subjects found in the ORs of the analyzed corpora is lower than the one found in simple declaratives: Lorusso 2003, reported that null subjects appear in 79% of the verbal utterances of adults, in the CHILDES files he analyzed; removing occurrences of null subjects in (I)ORs and (Indirect) Object wh-questions from his count, null subjects occur up to 72% of cases in declarative sentences. Then, again, the preference to use a null subject in ORs cannot be taken to be an indicator of the complexity of the involved syntactic computations under discussion.

The tension which has emerged between the corpus analysis and the results from elicited production opens up a new question: we now want to better investigate why PORs should be rare in spontaneous production and, conversely, why they should be so pervasively present in the elicited production.

Looking for an answer to this question(s), we first checked how frequent the passive voice is throughout the corpora and found that, in fact, it is not so infrequent to justify the low rate of PORs in spontaneous productions. As a first preliminary sample we checked the SUT corpus:

Corpus	# of verbs	# trans (%)		# ditrans (%)	# pass (%)	
SUT	872	645	(74%)	50 (6%)	177	(20%)

Table 10. Passive voice (pass) compared to active verbs (transitive and di-transitive) in SUT.

Then, we controlled for the animacy feature on both the relative head and the subject of the object relative clauses. Here we found an important asymmetry that asked for a deeper investigation: while the experimental design elicited productions in which the relative head (of the ORs) was always animate, in the corpora only 43% of the relative heads were animate (data from CHI A). We then decided to test the elicited production of ORs, manipulating the animacy feature.

4. Testing head animacy

To see if a [– animate] head favors the production of ORs better than a [+ animate] head, we run two experiments that are an adaptation of Belletti & Contemori 2010 design: the subjects were asked to listen to a certain number of minimal pairs of cue sentences and to answer in the most natural and complete way, choosing one of the two situations described. The answer, in most of the cases, resulted to be a RC, as expected.

4.1. Methods

In both experiments we used four conditions that exhausted the logical possibilities to be tested:

1. [+ animate] Head, [+ animate] Subject
2. [+ animate] Head, [– animate] Subject
3. [– animate] Head, [+ animate] Subject
4. [– animate] Head, [– animate] Subject

We first provided the experimental subject with a short context (e.g. “in a park, there are children playing with an apple…”), then we made the subject listening to a minimal pair of cue sentences (e.g. “the children wash the apple”, “the children throw the apple”) and we finally asked to answer a question in the most natural and complete possible way (e.g. “which apple would you eat?”… Target sentence: “I would eat the apple that the children wash/throw”).
All grammatical subjects in the cue sentences were definite, masculine and plurals (this is because we wanted to eliminate a potential ambiguity and discriminate between non target productions of SRs with post-verbal object, and true ORs with a post verbal subject; both options are realized with the very same word order in Italian, but in the latter case we could rely on the verb-subject agreement), all objects were masculine and singular, all the verbs were inflected at present tense.

We used three items per condition (then, in the end, we had 12 experimental items), we balanced the lexical material in terms of frequency and imaginability and we took 28 fillers to separate the experimental items. We semi-automatically created four randomizations such that: every randomization started with an item taken from a different condition, at least two fillers separated two experimental items, no experimental items of the same condition appeared in sequence, the first 4 experimental items in all 4 randomizations exhausted all 4 possible conditions.
We digitally recorded the audio materials (contexts, cues and elicitation sentences) and we created a PowerPoint presentation where, for every slide, the context was first played, then the cues and at the same time the discriminating words were briefly displayed (in case of verbs, the infinitive forms was chosen for not priming a finite RC) on the screen to help the experimental subjects to memorize the two proposed situations; in the end, the question was played and the beginning of the answer was displayed on the bottom of the screen.

Figure 2. Experimental screenshot with all components displayed.

The experimental session was preceded by a short warm-up with three items.
The only difference between the two experiments was that in the first one we used the "verb change" elicitation condition (i.e. the only thing that distinguished the minimal pair in the cue sentences was the verb used: "the children wash the apple" vs. "the children throw the apple"), whereas in the second experiment we implemented the "subject change" elicitation condition (i.e. the only thing distinguishing the cue sentences was the subject used: "the children wash the apple" vs. "the parents wash the apple"). The lexical material and the randomization were the same (except for the extra verbs/subject added to comply with the different design).
Below, one sample for each experimental animacy condition (cue sentences and elicitation sentences) in both designs:

Cond.	RC head	Subj	cue sentence	elicitation sentence
1	+anim	+anim	I poliziotti salutano un ragazzo *the policemen greet a child* I poliziotti rincorrono un ragazzo *the policemen chase a child*	tu quale ragazzo vorresti incontrare? *Which child would you rather meet?* "vorrei incontrare il ragazzo..." *I would rather meet the child…*
2	+anim	–anim	I secchi sbilanciano un imbianchino *The buckets unbalance a decorator* I secchi sporcano un imbianchino *The buckets dirty a decorator*	Tu quale imbianchino vorresti aiutare? *Which decorator would you rather help?* "vorrei aiutare l'imbianchino..." *I would rather help the decorator…*
3	–anim	+anim	I giornalisti scrivono un articolo *The journalists write an article* I giornalisti copiano un articolo *The journalists copy an article*	Tu quale articolo vorresti leggere? *Which article would you rather read?* "vorrei leggere l'articolo..." *I would rather read the article…*
4	–anim	–anim	I camini riscaldano un appartamento *The fireplaces warm an apartment* I camini affumicano un appartamento *The fireplaces smoke an apartment*	Tu quale app i riscaldamenti è acceso artamento vorresti scegliere? *Which apartment would you rather choose?* "vorrei scegliere l'appartamento…" *I would rather choose the apartment…*

Table 11. Experiment 1, verb change. 4 conditions.

Cond.	RC head	Subj	cue sentence	elicitation sentence
1	+anim	+anim	I poliziotti rincorrono un ragazzo *the policemen chase a child* I commercianti rincorrono un ragazzo *the shopkeepers chase a child*	tu quale ragazzo vorresti incontrare? *Which child would you rather meet?* "vorrei incontrare il ragazzo..." *I would rather meet the child…*
2	+anim	–anim	I secchi sporcano un imbianchino *The buckets dirty a decorator* I pennelli sporcano un imbianchino *The paintbrushes dirty a decorator*	Tu quale imbianchino vorresti aiutare? *Which decorator would you rather help?* "vorrei aiutare l'imbianchino..." *I would rather help the decorator…*
3	–anim	+anim	I giornalisti scrivono un articolo *The journalists write an article* I pubblicisti scrivono un articolo *The publicists write an article*	Tu quale articolo vorresti leggere? *Which article would you rather read?* "vorrei leggere l'articolo..." *I would rather read the article…*
4	–anim	–anim	I camini riscaldano un appartamento *The fireplaces warm an apartment* I termosifoni affumicano un appartamento *The heaters warm an apartment*	Tu quale appartamento vorresti scegliere? *Which apartment would you rather choose?* "vorrei scegliere l'appartamento…" *I would rather choose the apartment…*

Table 12. Experiment 2, subject change. 4 conditions.

4.2. Results

We tested 24 subjects with the verb change elicitation condition and 28 subjects with the subject change elicitation condition.

Here we only report the rough results (see Belletti, Chesi, Contemori and Laudanna, in progress, for a detailed analysis) since this is sufficient to answer the relevant question we posed, that is: do [– animate] heads favor the production of a certain amount of ORs?

	H+anim S+anim	H+anim S-anim	H-anim S+anim	H-anim S-anim
POR all	57 (79%)	60 (83%)	65 (90%)	63 (87%)
POR	11	20	5	5
POR r.	37	37	50	55
POR r. by	6	1	9	3
POR by	3	2	1	0
OR all	14 (20%)	4 (6%)	7 (10%)	8 (11%)
OR	2	0	2	2
OR VS	4	1	1	2
OR pro	8	3	4	4
ALT	1 (1%)	8 (11%)	0	1 (1%)
ALT SR	1	7	0	0
ALT PP	0	1	0	1

Table 13. Experiment 1 (verb change) results (24 subjects); r. = reduced, by = by-phrase present, VS = post-verbal subject, pro = null subject, ALT SR = SR produced instead of OR, ALT PP = Prepositional Phrase produced instead of OR.

	H+anim S+anim	H+anim S-anim	H-anim S+anim	H-anim S-anim
POR all	64 (76%)	64 (76%)	50 (60%)	59 (70%)
POR	0	0	0	0
POR r.	0	0	0	0
POR r. by	52	52	45	52
POR by	9	12	5	7
OR all	9 (11%)	3 (4%)	5 (6%)	3 (4%)
OR	1	2	2	0
OR VS	8	1	3	3
OR pro	0	0	0	0
ALT	11 (13%)	17 (20%)	29 (34%)	22 (26%)
ALT SR	0	6	0	59
ALT PP	11	11	29	0

Table 14. Experiment 2 (subject change) results (28 subjects); r. = reduced, by = by-phrase present, VS = post-verbal subject, pro = null subject, ALT SR = SR produced instead of OR, ALT PP = Prepositional Phrase produced instead of OR.

Despite a non negligible tendency to avoid the production of ORs if favor of a (genitive) PP when the subject is animate and the head inanimate (e.g. “the paper of the journalists” instead of “the paper that the journalist write”) in the subject-change experiment, we can easily see that the great majority of experimental subjects clearly keep preferring the POR solution also in the new experiment manipulating the animacy feature in the described conditions (in the great majority of cases, reduced PORs were produced, e.g. “the child chased” in the verb-change design and “the child chased by the policemen” in the subject-change design). The by-phrase is often unrealized in the verb-change experiment, whereas the use of PORs with the by-phrase is the preferred solution in the subject-change experiment (it is significantly more used than the possible equivalent alternative of OR with post-verbal subject).

4.3. Discussion 2

To better visualize the results, we report a histogram with the relative distribution of RCs produced both in the verb-change and in the subject-change experiments (we collapsed together all three items per condition and we removed non-RCs productions):

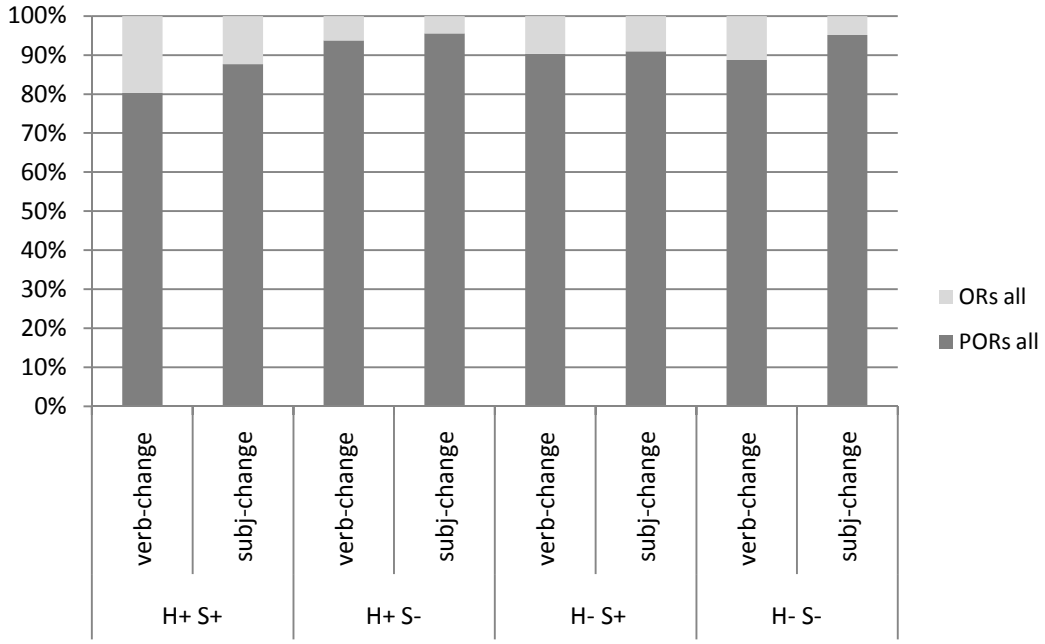


Table 15. Aggregated results of the elicitation task (H+/– = [+/– animate] relative head, S+/– = [+/– animate] relative subject)

Here it is clear that the animacy (mis)match does not play any role in favoring or disfavoring the production of (active) ORs, in the adopted experimental conditions.[10] Again, we observe lack of a direct correlation between frequency in the input and the behavior in the elicited production. PORs remain the preferred structure produced also in the new experiments manipulating animacy.

Here we observe that the intervention account proposed in Friedmann et al. 2009, correctly predicts the ranking of the produced relatives in the new experiments: ORs with a preverbal lexical subject, are the least produced ORs in the overall results (only 11 out of 535 relatives produced, Tables 11, 12): these are indeed the structures singled out as those in which intervention is stronger hence the structure harder to compute, as the NP feature of the intervening lexical subject is properly included within the feature set of the target relative lexical head. ORs with a post-verbal subject and ORs with a null pronominal subject are more often produced (Tables 11, 12). Assuming a derivation through smuggling for (active) ORs with a postverbal subject (Belletti & Contemori (2010)), this solution eliminates intervention in a way parallel to PORs; a further complicating factor is however involved in (active) ORs with a postverbal subject, which displays crossing between the dependency of the (expletive) null

[10] In fact, ORs are slightly more often produced in the [+ animate] head, [+ animate] subject condition, where, if anything, one would have expected a higher intervention effect due to animacy matching, if animacy was a relevant feature in the computation.

subject in the EPP position and the lexical subject in the postverbal position, with the chain relating the relative head and the gap in the object position of the smuggled VP chunk (structure 3 in §5 below). No such crossing is involved in PORs (structure 1 in §5 below). In ORs with a null (pronominal) subject, intervention should be less strong in principle, as no NP feature is contained in the intervening subject; hence, a null (pronominal) subject does not constitute as a strong intervener as a lexical subject (see also Gordon et al. 2004). PORs are by far the best solution: they are the only case in which intervention is totally eliminated, and no further complicating crossing is involved in the computation, as noted. In conclusion, the assumed intervention approach expressed in featural terms, accounts for the preferences revealed by the elicited productions of the new experiments.

5. Final considerations

Our corpus analysis has revealed that adults can process ORs and, in their spontaneous production, they do produce some ORs. This happens to a significantly smaller extent than SRs. These data are coherent with the assumed intervention account, which constitutes the key factor for interpreting the crucial fact that ORs are generally harder to process, also for adults, in various respects. However, we have also pointed out that the higher frequency of SRs in the corpora cannot be linked in a simple minded way to the complexity of the syntactic computation, as SRs and ORs are evenly distributed when the verb of the relative is a transitive verb, thus confirming that ORs can be properly processed by adult speakers and productively used in real communicative situations; hence, they are not just avoided on the basis of a complexity measure. The rareness of PORs in spontaneous productions in turn, may suggest a residual disfavoring of passive over active in naturalistic productions; presumably more so in contexts in which an already articulated computation is processed, such as a relative clause. A conclusion in need of further investigation, which we leave at this speculative stage here.

In contrast, in the elicited production, speakers tend to select the best computation, which is the one where no intervention arises. This explains the clear preference for the passive derivation through *smuggling* in computing the relativization of a direct object (§2.1), yielding the production of PORs.

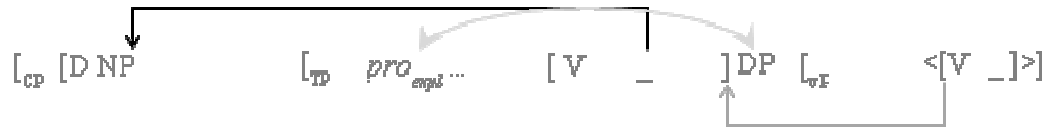
We suggest that the asymmetry in spontaneous production and in the elicited context plausibly derives from the fact that in the elicited production, but not in the spontaneous production, a (semi-conscious) “planning” of the sentence structure is made possible by the fact that all lexical material (the relative head, the subject and the verb) is provided to the experimental subjects in the introductory story. This allows the speakers to compute the best possible computation which, according to the analysis adopted here, is the one which, eventually, totally eliminates intervention, as is the case in PORs.

The following schematic derivations illustrate the predictions/rankings of the assumed intervention account:

1. PORs:

2. OR with null subject:

3. OR with post-verbal subject through smuggling:



4. OR with pre-verbal subject:

On one extreme -1- , PORs are the best possible solution, given the smuggling analysis *à la* Collins, since there is no intervention, on the other extreme -4-, ORs with preverbal lexical subject are the worst possible solution, since there is intervention in the strongest form. Intermediate solutions are ORs with a null (pronominal) subject -2-, where no NP/lexical restriction feature is present on the subject (but only on the relative head), and ORs with a post-verbal subject -3-: arguably in the latter structures a lexical subject intervenes to a lesser extent than a preverbal subject as proposed in Guasti et al. (2010) for similar structures in wh interrogatives, following Franck, Lassi, Frauenfelder, & Rizzi (2006). However, as noted in 4.3, although intervention is eliminated through *smuggling* in 3, the crossing of dependencies that the structure implies makes it less optimal than a POR structure.

References

Adams, C. (1990). Syntactic comprehension in children with expressive language impairment. *British Journal of Disorders of Communication* 25, 149–71.

Adani, F., van der Lely, H.K.J., Forgiarini, M., Guasti, M.T. (2010). Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children. *Lingua*.

Belletti, A. (2009). Notes on Passive Object Relatives. To appear in P. Svenonius ed. *Functional Structure from Top to Toe*.

Belletti, A. (2010) Considering the complexity of relative clauses and passive from the Italian perspective, Romance Turn 4, to appear in the *Proceedings.*

Belletti, A., & Contemori C. (2010). Intervention and Attraction. On the production of Subject and Object Relatives by Italian (young) children and adults. In, J.Costa, A,. Castro, M. Lobo , F. Pratas eds., Language Acquisition and Development, 3. Proceedings of Gala 2009 , CSP, Cambridge, UK, pp.39-52.

Belletti, A., C. Chesi, C. Contemori & A. Laudanna (in progress) Animacy in the elicitation of Relative Clauses.

Brown, H., 1972. Children's comprehension of relativized English sentences. *Child Development* 42, 1923-1936.

Carreiras, M., J.A.Dunabeita, M.Vergara, I. De la Cruz-Pavia, I.Laka (2010) "Subject relatives are not universally easier to process. Evidence from Basque", *Cognition* 115, 79-92.

Chesi C., G. Lebani e M. Pallottino (2008). A Bilingual Treebank (ITA-LIS) suitable for Machine Translation: what Cartography and Minimalism teach us. Studies in Linguistics, *StiL*, 2:165-185

Collins, C. (2005) A Smuggling approach to the passive in English. Syntax, 8(2), 81-120.

Contemori, C. & Belletti, A. (this volume) Disentangling the mastery of object relatives in children and adults. Evidence from Italian.

Contemori, C. & Belletti, A. (under submission) "Relatives and Passive Object Relatives in Italian speaking children and adults: Intervention in production and comprehension", ms. Ciscl, University of Siena

Contemori, C., Garraffa M. (2010) Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences. doi:10.1016/j.lingua.2010.02.011.

de Villiers, J.G., de Villiers, P.A., Hoban, E (1994) The central problem of functional categories in the English syntax of oral deaf children. In: H. Tager-Flusberg (Ed.), *Constraints on Language Acquisition: Studies of Atypical Children*, (pp. 9–47). NJ: Erlbaum, Hillsdale.

De Vincenzi, M. (1991) *Syntactic Parsing Strategies in Italian*. The Netherlands: Kluwer, Dordrecht.

Franck, J., Lassi, G., Fraunefelder, U. H., Rizzi, L. (2005) Agreement and movement: A syntactic analysis of attraction, *Cognition* 30, 1-44.

Friedmann, N., Belletti, A., & Rizzi, L. (2009) Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119(1), 67-88.

Gennari, S. P., & MacDonald, M. C. (2009) Linking production and comprehension processes: The case of relative clauses, Cognition, doi:10.1016/j.cognition. 2008.12.006.

Gordon, P., Hendrick, R., Johnson, M., (2004) "Effects of noun phrase type on sentence complexity", *Journal of Memory and Language* 51, 97-114.

Grillo, N. (2008) *Generalized Minimality*. Utrecht Institute of Linguistics, OTS.

Guasti, M.T. & A., Cardinaletti (2003) Relative clause formation in Romance child production. *Probus* 15, 47-8.

Hamann, C. and L.Tuller (2010) *Relative Clause Production in French Children and Adolescents*, University of Oldenburg-Unversité de Tours (submitted)

Lin C. and T. G. Bever (2006) Subject Preference in the Processing of Relative Clauses in Chinese. *Proceedings of the 25th WCCFL*, ed. Donald Baumer, David Montero, and Michael Scanlon, 254-260. Somerville, MA: Cascadilla

MacWhinney, B., & Snow, C. (1985) The child language data exchange system. Journal of child language, 12(2), 271.

Matteini S. (2011) Ms.

Novogrodsky, R., Friedmann, N. (2006) The production of relative clauses in SLI: A window to the nature of the impairment. *Advances in Speech-Language pathology* 8(4), 364-375.

Rizzi, L. (1990) *Relativized Minimality*, MA: MIT Press.

Rizzi, L. (2004) Locality and the left periphery. In A. Belletti (Ed*.), Structures and Beyond: The cartography of syntactic structures*, Vol. 3, Oxford University Press, 223-251.

Rohde, D. L. T. (2004). Tgrep2 user manual. URL: http://tedlab. mit. edu/ dr/Tgrep2.

Spina, S. (2005). Il Corpus di Italiano Televisivo (Cit): struttura e annotazione. In E. Burr (Ed.), Tradizione & Innovazione. Il parlato: teoria - corpora - linguistica dei corpora. Proceedings of the VI SILFI Conference. Firenze: Franco Cesati: 413-426.

Starke, M. (2001) *Move Dissolves into Merge: A Theory of Locality*. Doctoral Diss., University of Geneva.

Tavakolian, S.L. (1981) The conjoined-clause analysis of relative clauses. In S.L. Tavakolian (Ed.), *Language Acquisition and Linguistic Theory* (pp. 167–187) Cambridge, MA: MIT Press.

Tomasello, M. (2003) Constructing a Language: A Usage-Based Theory of Language Acquisition. Harvard University Press.

Utzeri, I.(2007) The production and acquisition of subject and object relative clauses in Italian. Nanzan Linguistics Special Issue 3, 283-314.

Warren, T., Gibson, E. (2002) The influence of referential processing on sentence complexity. Cognition 85, 79–112.