A SYNTACTIC APPROACH TOWARD THE INTERPRETATION OF SOME DISTRIBUTIONAL FREQUENCIES: COMPARING RELATIVE CLAUSES IN ITALIAN CORPORA AND IN ELICITED PRODUCTION

Adriana Belletti
Università degli Studi di Siena

Cristiano Chesi
Istituto Universitario di Studi Superiori di Pavia

Abstract: *A robust set of experimental results from previous studies on both production and comprehension of subject relatives (SRs) and object relatives (ORs) in Italian have confirmed the well known different status of SRs and ORs holding cross-linguistically in both children and adults, with ORs harder than SRs, in various dimensions. One crucial finding of these results concerns Italian-speaking adults who, in elicited production tasks, tend not to produce ORs in a systematic way and resort to the production of alternative structures: the privileged alternative is represented by use of passive in the relative, leading to the production of up to 90% of passive object relatives (PORs) in the studies considered.*

The contribution of this article is primarily comparative in nature, bringing into the picture a new dimension: a corpus study of (headed) SRs and ORs in standard Italian to be compared with the results from elicited production. The animacy feature is also manipulated in a new elicitation experiment adapting previous designs. Results indicate that, on the one hand simple frequency based considerations cannot be at the source of the ample resort to PORs in the elicited productions, as PORs are rather infrequent in the corpora of spontaneous production investigated; on the other hand, ORs with an inanimate relative head are relatively frequent in the same corpora, yet manipulation of the animacy feature does not play a role in favoring the elicited production of ORs headed by an inanimate noun phrase. We propose that the grammatical dimension in terms of intervention locality may offer a crucial key in interpreting the complex shape of the results and highlight that simple distributional frequency factors remain unreliable as for the expectations they can generate.

Keywords: *animacy, distributional frequency, featural Relativized Minimality, intervention locality, object relatives, passive object relatives, subject relatives*

1. Introduction

1. 1. Producing (and comprehending) types of relative clauses

A robust set of experimental results on both production and comprehension of subject relatives (SRs) and object relatives (ORs) in Italian (e.g. Adani et al. 2010; Arosio et al. 2008; Belletti & Contemori 2010; Contemori & Belletti 2013; Contemori & Garraffa 2010; Belletti & Guasti forthcoming for a review of the available acquisition data), have confirmed the different status of SRs and ORs, holding cross-linguistically 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 of these results concerns adults: in elicited production tasks (Belletti & Contemori 2010; Contemori & Belletti 2013), Italian adults tend not to produce ORs in a systematic way; specifically, there is a strong tendency to avoid ORs, in favor of the production of an alternative structure, typically a SR which is able to preserve the intended meaning. One privileged such alternative is offered by use of passive in the relative, that is exploited up to 90% in the different groups of adults investigated in the different experiments (see also Belletti 2009, 2014 for a first discussion; Contemori & Belletti 2013 for detailed presentation).

Following the references quoted, we will refer to subject relatives in the passive produced in response to the elicitation of an active object relative as Passive Object Relatives (PORs). PORs are thus the preferred option for adults and they also become the preferred option for children as well, as soon as passive becomes productively available to them, around age 5. PORs have also been tested in comprehension (Contemori & Belletti 2013), 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 speaking different languages (Friedmann et al. 2010)[1], and in self-paced reaction time experiments with adults (e.g. Lin & Bever 2006 on Mandarin Chinese).

Our contribution in this paper is primarily comparative and consists in bringing into the picture a different kind of empirical data: a corpus study of (headed) SRs and ORs in standard Italian. On the basis of the described experimental results, in this article we raise the following two main research questions:

i. In the overwhelming presence of PORs in the elicited production an effect of the frequency of these structures in the Italian naturalistic input?
ii. Should some type of (active) OR turn out to be relatively frequent in Italian corpora, are elicited ORs of the same type actually produced in experimental conditions?

As for question ii., in an elicitation task similar to one used in the previous quoted experimental studies, we will manipulate the animacy feature of the relative head and of the subject of the relative clause, since headed ORs with an inanimate head appear to be rather

[1] Resort to passive leading to PORs is a preferred option in several languages, but not all (e.g. Hebrew). This opens up an interesting comparative descriptive issue that we cannot address here.

frequently present in the naturalistic input (Kidd et al. 2007; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris 2005 for English).

We will mainly concentrate on results from adults, with occasional comparisons with results from children.

1.2. Interpreting frequency

Some considerations on frequency – i.e. what it may relate to and what its ultimate significance may be – are in order in these introductory remarks. Many different approaches to language processing share the intuition that what we hear more often is what we process with more ease. Distributional frequencies both of (sequences of) lexical items and syntactic "constructions"[2], can contribute to facilitating processing. This is a central assumption shared by many performance-oriented language processing models, like usage-based (Tomasello 2009), constraint-based (MacDonald et al. 1994), expectation-driven (Chambers et al. 2002) and, generally speaking, probabilistic models (Manning and Shutze 1999). Hence, our question i. will contribute to clarify whether there is indeed a direct relation between frequency in corpora and (elicited) production.

Documenting distributional asymmetries is crucial for building statistical models fitting empirical data (Roland et al. 2007). We are, however, legitimated in asking why an asymmetric distribution should be present at all: is it the case that there is an asymmetric distributional frequency in production because of an asymmetric distributional frequency in the input we receive when we learn a language? Notice the circularity in this argument: why in the first place should the input ever have such asymmetric distributional frequencies, if they exist at all? In this article we will touch upon this issue in our corpus analysis of relative clauses (RC) in Italian.

Another important factor to take into account, is that distributional frequency appears to reflect contextual discourse properties and registers (i.e. written vs oral; child vs adult etc.). This might be at source of the asymmetries revealed across different kind of corpora in English, as documented in Roland et al. (2007), one of the most relevant studies on the frequency of structural configurations across English large-scale corpora.

Let us consider some of Roland et al. (2007)'s results in better detail. This study thoroughly discusses the intrinsic difficulties in gathering distributional frequencies from corpora: first of all, the extraction methods must be transparent, in order to make the count replicable; second, different corpora often show different distributional frequencies, especially when written and spoken collections are compared; for this reason, it is safer to inspect corpora closely related to the investigation area: e.g. if we aim at explaining the acquisition of certain phenomena, child-directed speech corpora should be much more adequate than newspapers corpora; third and last, we should clarify what fine-grained distinctions are relevant for counting, normalizing the raw data count (due to significant magnitude differences across corpora) and defining significant classes of structures to be compared.

[2] The term "syntactic construction" refers in fact to the set of computations that lead to a particular syntactic configuration. The term has a descriptive value. Theoretically, we know at least since Principle and Parameters (Chomsky 1981) that the derivational atoms of syntactic computations are shared by different "constructions". We will continue to use the term construction following the traditional and current use, keeping this proviso in mind.

As for relative clauses, Roland et al. (2007) note that different corpora show very large differences in frequency distribution: in particular, the English corpora analyzed appear to vary greatly with respect to the distribution of important features and properties that have been claimed to play a role in the processing of such structures such as the relative frequency of subject vs. object RC typology, the animacy of the head (see references above), whether the subject within the ORs is a full DP or a pronoun (Reali & Christiansen, 2007; Warren & Gibson, 2002; Belletti & Rizzi 2013). The main data on the distribution of RC macro typology are reported below, from Roland et al. (2007):

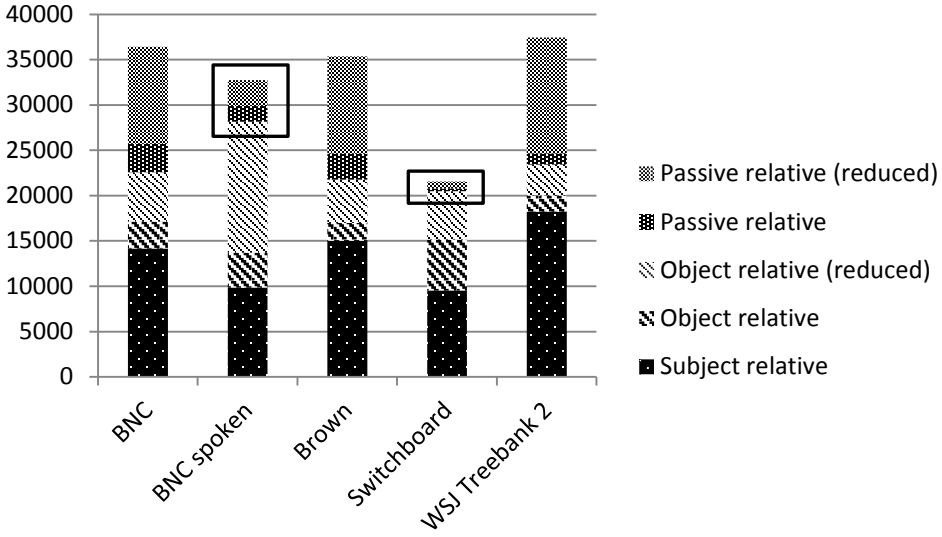


Fig. 1. RC macro typology in different corpora (data from Roland et al. 2007).

Type of relative clause	BNC	BNC spoken	Brown	Switchboard	WSJ
Subject relative	14182	9851	15024	9548	18229
Object relative	2943	3863	1976	5616	1802
Object relative (reduced)	5455	14423	4746	5314	3385
Passive relative	3118	1729	2867	302	1224
Passive relative (reduced)	10730	2886	10733	779	12788
Kind of production	*written*	*spoken*	*written*	*spoken*	*written*

Table 1. RC macro typology in different corpora (data from Roland et al. 2007, 355): normalize count w.r.t. 1 million of NPs. Corpora used: British National (BN) corpus, Brown corpus, Switchboard corpus, Wall Street Journal (WSJ) Treebank corpus. "Object relative (reduced)" refers to absence of complementizer in an active OR; "Passive relative (reduced)" only contain a (passive) past participle with both complementizer and auxiliary absent.

Comparing the distribution of macro typologies of relative clauses across corpora we should observe that ORs are much more frequent in spoken corpora (BNC spoken and Switchboard) than in written ones and in the very same spoken corpora, very few PORs (reduced or not) are present and this also contrasts with the written corpora where PORs seem to be quite represented. Roland et al. (2007) take this as evidence that discourse functions, register, and contextual considerations affect significantly the distribution of the different constructions considered. We can take this as clear evidence of the fact that plain distributional frequencies should not be taken naively as representative of the "standard" input received by native speakers.

We will speculate on the interpretation of some quantitative results obtained in our corpus study and endorse the view that a factor playing a crucial role is a grammatical formal factor expressed in terms of the locality principle featural Relativized Minimality (RM; Rizzi 1990, 2004; Starke 2001), as developed in Friedmann et al. (2009) to account for SRs vs ORs asymmetries in development. In the following section 2 we illustrate the essential aspects of the locality approach, which is mainly meant to explain why the hardest structure to compute are headed object relatives with an intervening lexical subject within the relative clause.

The article is organized as follows: section 2 presents the intervention locality account we will assume; we will then move to our corpus study in section 3, consisting of Italian child-directed speech taken from CHILDES database and two other standard Italian annotated corpora: the Siena Universtity Treebank (SUT, Chesi et al. 2008) and the Italian Television Corpus (CIT, Spina 2005). In this section we will present the counting methodology adopted and discuss the frequency distributions revealed across corpora. In section 4 a new elicitation experiment of ORs/PORs is presented, testing the role of animacy both on the RC head and on the subject of the relative clause. Section 5 concludes the article with a general discussion.

2. Lexically headed object relatives with a lexical subject in the relative clause: An intervention locality account

A featural approach to Relativized Minimality, as developed in Starke (2001), Rizzi (2004), has been adapted in Friedmann et al. (2009) to account for development, 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 by 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 the position Y where it is externally merged 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, such as 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 the intervention locality account in Friedmann et al. (2009), 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[3]. The schematic representation in (1) 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)[4]:

(1)	il bambino	che	il nonno cerca/trova	<il bambino>
	the child	*that*	*the grandpa seeks/finds*	*<the child>*
	+R +NP	+ NP		+R +NP
	X	Z		Y

The intervention effect that arises in lexically headed ORs across an intervening lexical subject is the source of the difficulty in the processing of object relative clauses. This has a reflex in development (later proper processing of lexically headed object relatives) as well as in (slower) adult parsing of these structures (Belletti & Rizzi 2013 for further discussion).

The robust results from production mentioned in the introductory section 1.1 can be amenable to the same type of locality account. Recall the major aspect of those results: when a lexically headed object relative containing a lexical subject in the relative clause is elicited, Italian-speaking adults overwhelmingly produce a POR instead of the expected active object relative. Italian-speaking children tend to approach the adult behavior as they grow older and master passive properly. In comprehension, (types of) PORs are the object relatives that are best comprehended by Italian speaking children at the age in which they have certainly no problem in mastering passive well (age range tested 6-8, Contemori & Belletti 2013).

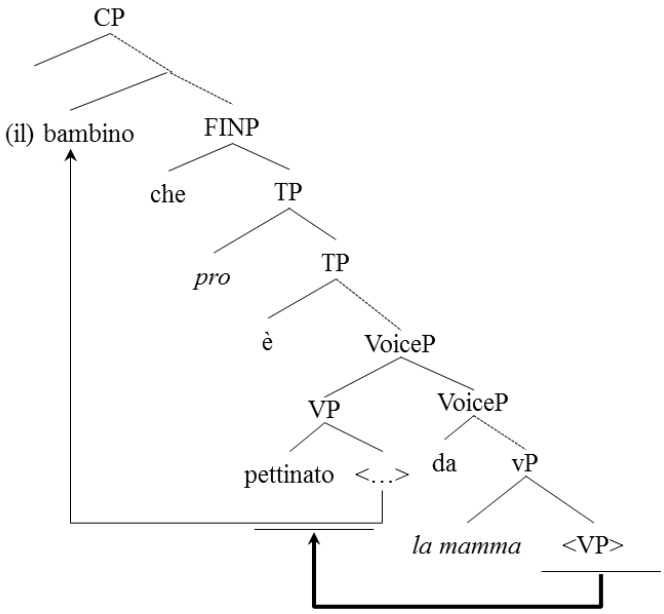
As discussed in previous work (Belletti 2009, 2014; Contemori & Belletti 2013), the intervention situation arising in lexically headed object relatives with a lexical subject in the relative clause can be overcome with the use of passive. Passive can be seen as an optimal way to overcome the described intervention effect, which inevitably arises in the relativization of a direct object across an intervening lexical subject. Assuming a derivation of what we call *passive* along the lines proposed in Collins (2005), movement of a verbal chunk containing (at least) the verb and the object and excluding the intervening lexical

[3] As also suggested by the facts discussed in Munaro (1999), quoted in Friedmann et al. (2009), according to which lexically restricted wh-phrases target different positions from those targeted by non-lexically restricted wh-phrases in the Northern Italian dialects discussed there.

[4] 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. (2012).

subject occurs; the chunk is attracted by a component of the passive voice, e.g. preposition *by* for convenience. Through this movement, referred to as *smuggling* in Collins (2005), intervention is eliminated. Thus, a principled reason is provided for the (overwhelming) appeal to passive in the syntactic computation of an OR in Italian (and also in other languages) that the experimental results have so clearly revealed. The assumed derivation is schematically illustrated in (2) for the Italian POR "il bambino che è pettinato dalla mamma" (the child that is combed by the mom):

(2)



In the following section 3 we now ask how the locality principle relates to distributional frequency in spontaneous productions and to what extent PORs are also found in naturalistic corpora. This will provide an answer to our first research question repeated here: Is the overwhelming presence of PORs in the elicited production an effect of the frequency of these structures in the Italian naturalistic input? Should this turn out not to be the case, the grammatical account in terms intervention locality will offer a suitable alternative line of explanation.

In the aim of further testing the intervention account we will also analyze the type of (active) ORs present in the corpora as for the nature and position of the subject in the relative clause. Finally, we will analyze whether animacy plays some role in conditioning the naturalistic production of (active) ORs.

3. The corpus research

The corpora chosen in the study to be presented here, reflect our research, theoretical questions and our approach to the study of frequency discussed in 1.2.

We checked for the occurrence of types of relative clauses in both children and adult Italian speakers in naturalistic speech. Our main concern was to verify the distribution of relevant properties and features discussed in the literature on relative clause processing and acquisition, with special attention to child-directed speech.

3.1. Corpora used

The corpora chosen in the study to be presented here, reflect our research, theoretical questions and our approach to the study of frequency discussed in 1.2.

We checked for the occurrence of types of relative clauses in both children and adult Italian speakers in naturalistic speech. Our main concern was to verify the distribution of relevant properties and features discussed in the literature on relative clause processing and acquisition, with special attention to child-directed speech.

Given the asymmetry revealed in other corpus studies (e.g. Roland et al. 2007) we decided to target first child-directed speech in our analysis; to retrieve these productions, we inspected the Italian section of the CHILDES database and we used Antelmi corpus (1 child, Antelmi 2004), the Calambrone Corpus (6 children, Cipriani and Cappelli 2004) and Matteini corpus (1 child, Matteini 2011). In total we considered 8 children, for a corpus consisting of 132 files (nearly 400.000 tokens).

We compared the distribution of 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

For this study we split the CHILDES corpus in the adult section (CHI A) and in the children section (CHI C).

3.1.1. Counting procedure

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[5]. 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 considerations on Reduced RCs in 3.1.2 and table 6) clearly marked with an invariable relative pronoun/complementizer *che*; this can be productively encoded with a simple regular expression like the one in (3) that picks up all occurrences of "che" produced by a certain speaker ("TIER") in a CHAT-encoded file (MacWhinney et al. 1985):

(3) 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[6], 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, precise 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.

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[7]: 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 (4a), 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 (4b):

(4) a. tgrep 'NP.rel < C.rel'
b. tgrep 'NP.rel-obj.anim, NP-subj.anim'

[5] GREP is a Unix native Regular Expression interpreter; it is easy to use, free, reliable and fast; given a Regular Expression pattern, it returns the line in the text where a matching occurs (options "-i -n –E" indicate a case insensitive search, with line number matching indication and the usage of extended regular expressions, e.g. "[[:alpha:]]" indicates any possible alphabetic char), or the exact count of occurrences (if "-c" option is used).

[6] 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.

[7] 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).

3.1.2. Results

In this section, we present the main results of our quantitative analysis. Consider first Table 3.

Corpus	Tool used	# of analiyed 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.

Table 3 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 (SRs = Subject Relatives, ORs = Object Relatives, IORs = Indirect Object Relatives; the counting includes only those colloquial IORs introduced by "che"[8].

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

[8] Such as:

(i) Un bimbo che ci va insieme all'asilo
A child that he goes with-him CL to kindergarden

with a resumptive clitic. Or:

(ii) Quel giorno che sei stato così bravo
That day that you have been so good

with a temporal interpretation.

(SRs are roughly twice more frequent than ORs[9]), 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 A1 in the Appendix), 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.

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:

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

Table 5 shows that the presence of full PORs is almost unattested across all corpora. This is in striking contrast with the experimental results from elicited production described in sections 1.1. and 2.

Including in the counting also all possible reduced PORs (e.g. "the boy chased (by the policemen)"[10]) 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 6. 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

[9] 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).

[10] Both long, with the by-phrase, and short, without by-phrase reduced relatives are included.

ones are included[11]. 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.

We also observed that, generally, the passive voice is very poorly represented in child directed speech: inspecting about 10% of a random sample of the child directed speech productions we revealed that less than 5% of the verbs used were in the passive form.

In the same vein, we also checked more generally how frequent the passive voice is throughout other corpora and found that, in fact, generally speaking, it is not so infrequent as in child directed speech to justify the low rate of PORs in spontaneous productions.

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

Table 7. Passive voice (pass) compared to active verbs (transitive and di-transitive) in SUT.

3.1.2.1. (Active) ORs and the position and nature of the subject in the relative clause

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:

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 preference for having an null subject is clearly present in the CIT corpus, in the CHI A, and, marginally, also in the SUT, a less straightforward tendency can be drawn from the pre-/post-verbal opposition: both children (CHI C) and adults in their child directed speech preferably locate the subject (often pronominal) in the post-verbal position, while the CIT shows a slight tendency in favoring the preverbal lexical alternative.

[11] 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, found in the elicitation experiments referred to in section 1.1, 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 section 5 on this), choice of the optimal solution in SUT may not be surprising.

3.1.2.2. (Active) ORs and the animacy of the relative head

Our final investigation targeted animacy distribution. Recall that according to Kidd et al. (2007), among others, this represents a major bias for ORs that are mainly formed on inanimate heads; the table below seems to support this claim, as ORs with an animate head are rather rare:

Corpus	# of ORs	# animate heads (%)	
CIT	117	6	(5%)
CHI A	228	33	(14%)
SUT	12	0	(0%)
CHI C	11	0	(0%)

Table 9. Animacy distribution in ORs.

Going deeply into CHI A occurrences (this is the corpus where the biggest number of animate heads is attested), it might be interesting to notice that most of the occurrences of ORs with animate heads present either a null (animate) subject (26 occurrences) or an overt (animate) pronominal subject (5 occurrences); the only two occurrences with a lexical subject are realized using a post-verbal subject.

3.1.2.3. SRs vs (active) ORs in the corpora

Looking closer at the SRs vs ORs asymmetry in the naturalistic corpora summarized in Table 2, we note that the significance of what is or is not (in the domain of relative clauses) frequently present in the analyzed corpora, must be treated with caution.

If we reconsider the frequency of SRs and ORs with respect to verb classes, we observe that the SRs/ORs asymmetry in fact disappears:

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

Table 10. 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). We conclude that 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 trivially reflect the complexity of a given structure.

3.2. Interim discussion 1

Given the frequency distributions presented in the previous section we can answer the question raised on the distribution of PORs, by concluding that PORs are not frequent structures at all in the naturalistic input. Since PORs have turned out to be the most frequently produced structure in the elicited production experiments reported in section 1.1., for adults and also for (older) children, the conclusion must then be drawn that PORs are nevertheless resorted to in the production experiments, despite their poor frequency in spontaneous speech[12]. Hence, the linguistic performances revealed by the experimental results do not simply reflect the shape of the linguistic naturalistic input.

We can conclude that PORs, which are the preferred structures in the elicited productions, must be preferred on different grounds rather than being a simple and straightforward consequence of a frequency effect. We submit the proposal that PORs count as the optimal structures in the elicited productions; preference for PORs in the elicited productions may be a consequence of the optimal way to eliminate intervention that use of passive in ORs offers, as illustrated in section 2. We delay until section 5 a possible hypothesis on the origin of the tension that 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.

What frequency in corpora may reveal is thus not a trivial matter. This is so in at least two complementary directions:

i. it is not the case that speakers always tend to produce those structures which are more frequent in the input 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 the case that (adult) speakers always tend to produce those structures which are computationally less complex, as revealed by the balanced presence in the corpora of SRs and ORs with transitive verbs.

This latter point is also consistent 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 want to give a word of caution. We note that our corpus study also suggests that bare frequency does not directly reflect the complexity of potentially alternative structures in a trivial way. Looking at the distribution of the subject within the ORs present in the corpora illustrated in table 8, we observed that in all corpora the empty subject is the most attested option (61% in the SUT and CHI A). From the point of view of a feature-based intervention approach, along the lines of Friedmann et al. (2009), intervention is weaker/absent in these cases as no [+NP] feature is shared by the target and the intervener, in the sense illustrated in section 2, under the assumption that a (null) pronominal has no such feature in its composition as it has no lexical restriction. Although this is likely to be a crucial factor in determining preference for these structures, yet it cannot be the only relevant factor. Discourse considerations also play a role, as it must be

[12] Note that PORs have been resorted to in different production tasks, a Preference task and a Picture description task (Contemori & Belletti 2013), both adapted from Novogrodsky & Friedmann (2006). Hence, resort to PORs cannot be just considered a simple task related effect.

the case that the relevant sentence allows for the occurrence of a null pronoun as its subject. In this respect we note that the percentage of null subjects found in the ORs of the analyzed corpora is in fact lower than the one found in simple declaratives, as reported by Lorusso (2003) who calculated that null subjects appear in 79% of the adults' verbal utterances, 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. This result must be due to contextual reasons as clearly simple declaratives are less complex than relative clauses, in any sense of complexity.

The tension which has emerged between the corpus analysis and the results from elicited production opens up new questions that we want to better investigate, in particular: why should PORs be so pervasively present in the elicited production given that they are rather rare in spontaneous production?[13]

In the aim to look for an answer to this question, we now move to our production study in which 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: whereas the experimental design of the previous quoted studies elicited productions in which the relative head (of the ORs) was mainly animate, in the corpora only 14% of the relative heads were animate (data from CHI A, table 9). Hence, the natural question arises whether animacy was responsible for the lower production of ORs and the consequent resort to PORs in the elicitation experiments, thus suggesting an (at least partial) answer to the question raised above. We then decided to test the elicited production of ORs through a preference task of the type utilized in the previous studies, in which the animacy feature was manipulated.

4. The Elicitation study

The goal of this study was to see if a [– animate] head favors the production of ORs better than a [+ animate] head. In order to do so, we run two experiments (an adaptation of Belletti & Contemori 2010 design): the experimental subjects were asked to listen to a certain number of minimal pairs of introductory 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. Method

4.1.1. Participants

For our experiment we enrolled 52 adult subjects (master students, age range = 22-25); we tested 24 subjects in one condition (verb change/Experiment 1) and 28 subjects in the other condition (subject change/Experiment 2) as described below in the materials section.

[13] The converse question, why PORs are rare in spontaneous production since they are in fact so pervasively present in elicited production, is also raised by our results. On this we will only offer some preliminary speculative considerations in the discussion section.

4.1.2. Materials

We used the same lexical materials (with minimal variations related to the condition) in two experiments and we implemented a Latin square design exhausting any logical possible dependent variable combination to be tested:

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

In the first experiment, the verb change condition, the cue sentence was modified at the verb segment: "the policemen chase a child" vs "the policemen greet a child".

In the second experiment, the subject change condition, the cue sentence was modified at the subject segment: "the policemen chase a child" vs "the shopkeepers chase a child".

All grammatical subjects in the cue sentences were definite, masculine and plurals[14] 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.

Below, one sample for each experimental animacy condition (cue sentences and elicitation sentences) in both verb change and subject change experiments:

Cond.	RC head	Subj	cue sentence	elicitation sentence
1	+anim	+anim	I poliziotti salutano un ragazzo	tu quale ragazzo vorresti incontrare?
			the policemen greet a child	*Which child would you rather meet?*
			I poliziotti rincorrono un ragazzo	"vorrei incontrare il ragazzo..."
			the policemen chase a child	*I would rather meet the child…*
2	+anim	–anim	I secchi sbilanciano un imbianchino	Tu quale imbianchino vorresti aiutare?
			The buckets unbalance a decorator	*Which decorator would you rather help?*
			I secchi sporcano un imbianchino	"vorrei aiutare l'imbianchino..."
			The buckets dirty a decorator	*I would rather help the decorator…*

[14] 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.

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 appartamento 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.

4.1.3. Procedure

In both experiments we first provided all subjects 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”).

We recorded digitally 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 cue sentences 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 (see Figure A1, in the Appendix).

The experimental session was preceded by a short warm-up with three items.

4.1.4. Coding

Answers have been transcribed and the results have been coded using the following categories:

Abbreviation	Category
POR all	Overall number of Passive Object Relatives
POR	Full Passive Object Relatives
POR r.	Reduced Passive Object Relatives
POR r. by	Reduced Passive Object Relatives with by phrase
POR by	Full Passive Object Relatives with by phrase
OR all	Overall number of Object Relatives
OR	Object Relatives
OR VS	Object Relatives with post-verbal Subject
OR pro	Object Relatives with null/pronominal Subject
ALT	Overall number of Alternative structure produced
ALT SR	Subject Relatives produced instead of (P)OR
ALT PP	Prepositional Phrase produced instead of (P)OR

Table 13. Coding

4.2. Results

Here we only report the rough results 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 (15%)	20 (28%)	5 (7%)	5 (7%)
POR r.	37 (51%)	37 (51%)	50 (69%)	55 (76%)
POR r. by	6 (9%)	1 (1%)	9 (12%)	3 (3%)
POR by	3 (4%)	2 (3%)	1 (1%)	0
OR all	14 **(20%)**	4 **(6%)**	7 **(10%)**	8 **(11%)**
OR	2 (3%)	0	2 (3%)	2 (3%)
OR VS	4 (6%)	1 (1%)	1 (1%)	2 (3%)
OR pro	8 (11%)	3 (4%)	4 (6%)	4 (6%)
ALT	1 **(1%)**	8 **(11%)**	0	1 **(1%)**
ALT SR	1 (1%)	7 (10%)	0	0
ALT PP	0	1 (1%)	0	1 (1%)

Table 14. 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 (62%)	52 (62%)	45 (54%)	52 (62%)
POR by	12 (14%)	12 (14%)	5 (6%)	7 (8%)
OR all	9 **(11%)**	3 **(4%)**	5 **(6%)**	3 **(4%)**
OR	1 (1%)	2 (3%)	2 (2%)	0
OR VS	8 (10%)	1 (1%)	3 (4%)	3 (4%)
OR pro	0	0	0	0
ALT	11 **(13%)**	17 **(20%)**	29 **(34%)**	22 **(26%)**
ALT SR	0	6 (7%)	0	22 (26%)
ALT PP	11 (13%)	11 (13%)	29 (34%)	0

Table 15. 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 preferred the production of a POR also in the new experiment manipulating the animacy feature (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).

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):

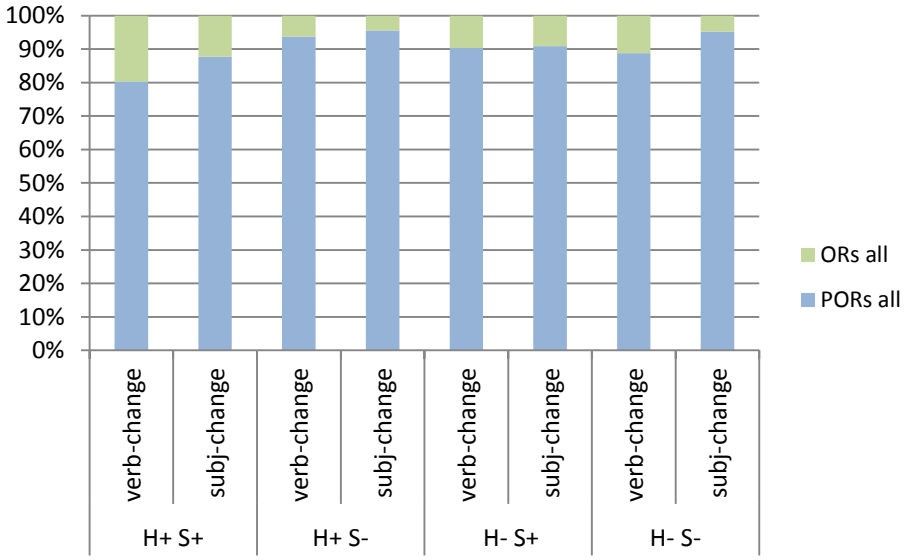


Fig. 2. Aggregated results of the elicitation task (H+/– = [+/– animate] relative head, S+/– = [+/– animate] relative subject).

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.[15]

4.3. Interim discussion 2

First of all, from the very neat results of the elicitation study we observe once again 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, despite the higher frequency of ORs with an inanimate head in naturalistic productions.

We now further 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

[15] 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.

11 out of 535 relatives produced, Tables 14, 15): 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 14, 15). Let us assume a derivation through movement of a chunk of the verb phrase/smuggling for (active) ORs with a postverbal subject, along the lines proposed in Belletti & Contemori (2010). Under this analysis, intervention is eliminated in ORs with a postverbal subject in a way parallel to PORs. A further complicating factor is however involved in (active) ORs with a postverbal subject: beside the chain relating the relative head and the gap in the object position of the smuggled VP chunk, a further relation is also established between a (expletive) null pronominal in the preverbal subject position and the lexical subject in the postverbal position (De Vincenzi 1991). Furthermore, the two long distance dependencies cross each other (as illustrated in structure 3 in the following section 5). No such further relation nor crossing is involved in PORs (as evidenced in structure 1 in the following section 5). 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 compared to a full lexical subject (see also Gordon et al. 2004; Belletti & Rizzi 2013). PORs are by far the best computation: 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, provides a line of account for the preferences revealed by the elicited productions of the new experiments.

5. General discussion

As revealed by other corpus studies (e.g. Roland et al. 2007), our study confirmed that different corpora present some differences in the distribution of relevant syntactic configurations. Even though, generally, distributions are coherent with specific featural patterns, e.g. ORs are usually headed by inanimate heads and have animate subjects in naturalistic corpora, this does not produce a frequency effect. An intervention-based approach, such as, specifically, the one adopted here from Friedmann et al. (2009) can interpret the naturalistic distribution as the effect to disfavor intervention configurations, if possible (see the results on the distribution and nature of the subject in active ORs).The approach is also better equipped to interpret why in the elicited production, despite their infrequency in naturalistic corpora, PORs are nevertheless the overwhelmingly produced structures by Italian speaking adults.

Our corpus analysis has revealed that adults can process ORs and, in their spontaneous production, they do produce active ORs. Overall, this happens to a significantly smaller extent than SRs. These data are consistent with the assumed intervention account, which constitutes the key factor for interpreting the robust fact that ORs are generally harder to process than SRs, 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,

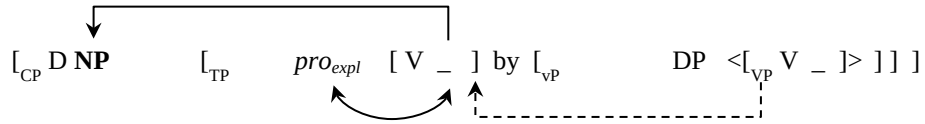
they are not just "avoided" on the basis of a complexity measure. As for PORs, we speculate that their rareness in spontaneous productions in turn, may result from a residual disfavoring of passive over active in naturalistic productions; presumably more so in contexts in which an already fairly articulated computation is being processed, such as a relative clause. This is, however, a conclusion in need of further investigation, which should bring into the picture also precise quantitative data on the occurrence of passive in naturalistic corpora compared to active. Thus, we leave this possible interpretation at this speculative stage here.

In contrast to naturalistic data, results from our elicited production experiments have confirmed previous results from other studies indicating that, in those experimental conditions, speakers tend to select the best/optimal computation; namely, the one where no intervention arises. This explains the clear preference for PORs, assuming a derivation of passive within the relative clause in which a chunk of the verb phrase is moved/smuggled, thus eliminating intervention of the lexical subject.

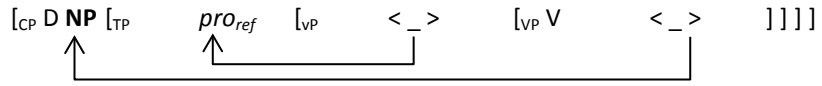
We suggest that the asymmetry between spontaneous and elicited production plausibly derives from the fact that in the latter, but not in the former, a (semi-conscious) "planning" of the sentence structure to resort to is made possible by the fact that all lexical material (the relative head, the subject 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 discussed, is the one that, eventually, totally eliminates intervention, as is the case in PORs.

We conclude by presenting the schematic derivations assumed, illustrating the predictions and the associated rankings of complexity that are immediately derived by the assumed intervention locality account in terms of featural Relativized Minimality:

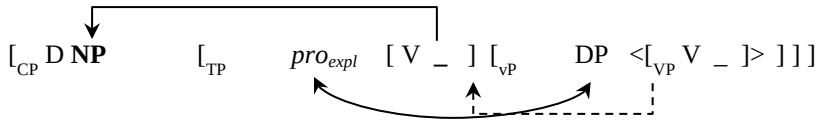
1. PORs:



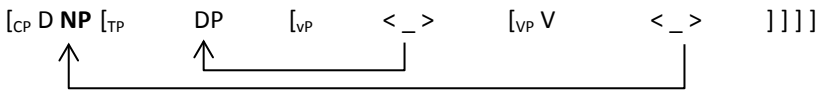
2. OR with null subject:



3. OR with post-verbal subject through smuggling:



4. OR with pre-verbal subject:



On one extreme -1-, PORs are the best structure in terms of intervention, given the *smuggling* analysis assumed (the dotted arrow shows the VP chunk movement), since there is no intervention at all in these configurations. On the other extreme -4-, ORs with a preverbal lexical subject are the worst structure in terms of intervention since there is intervention in the strongest form, due to the presence of the relevant [+NP] feature in both target and intervener. Intermediate configurations are ORs with a null (pronominal) subject -2-, and ORs with a post-verbal subject -3-. In the former, no [+NP] feature is present on the subject, which has no lexical restriction (only present on the relative head). Arguably in the latter structures the postverbal lexical subject intervenes to a lesser extent than a preverbal subject as proposed in Guasti et al. (2012) for similar structures in wh interrogatives, following Franck, Lassi, Frauenfelder, & Rizzi (2006). However, although as proposed in section 4.3, intervention is eliminated through *smuggling* of a verbal chunk in a derivation like 3, the further relation between the (expletive) null subject in the relevant preverbal subject position (Cardinaletti 2004; Rizzi & Shlonsky 2007) and the low postverbal lexical subject implies crossing of dependencies (De Vincenzi 1991) that likely makes the structure less optimal than a POR structure.

To conclude, from our comparative studies over the same syntactic domain of relative clauses in both naturalistic corpora and elicited production experiments we can claim that frequency is indeed a complex notion; this should be expected since, as noted, it depends upon various different factors and dimensions – purely grammatical factors, discourse-contextual factors, registers, developmental factors. Overall, we think that the grammatical dimension plays a central role, which is consistent with the intervention locality approach assumed, one especially relevant case in point in this respect being the nature and distribution of subjects in the naturalistic ORs analyzed. Bare frequency across corpora, however, cannot plainly map the ranking of complexity in 1-5 above, precisely because of its articulated nature. Thus, distributional frequencies remain unreliable as for the expectations they can generate on speakers' linguistic performances, since speakers do not always opt for the grammatically less complex and optimal computation in their natural conversations.

Our results from elicited production have revealed that the animacy feature does not appear to play any role in modulating intervention in production, as PORs remain the preferred structures resorted to by Italian speaking adults in their elicited responses. In the dialogue with frequency considerations inspired by naturalistic corpora, we take this to be a very relevant conclusion, as ORs with inanimate head (and animate subject within the relative clause) are the most frequent type of OR found in the naturalistic corpora considered. Hence not only it is the case that (adult) speakers resort to infrequent structures such as PORs in their (elicited) productions and do so overwhelmingly, it also the case that they do not resort to frequent structures such as ORs with an inanimate head. These at first sight unexpected conclusions constitute a warning on any simple and quick implication one may want to draw on frequency effects and their role.

Appendix

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	2 – 1		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 A1. 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.

Figure A1. Experimental screenshot with all components displayed.

References

Adams, Catherine. 1990. "Syntactic comprehension in children with expressive language impairment". *British Journal of Disorders of Communication* 25 (2): 149-71.

Adani, Flavia, van der Lely, Heather K. J., Forgiarini, Matteo, and Maria Teresa Guasti. 2010. "Grammatical feature dissimilarities make relative clauses easier: a comprehension study with Italian children". *Lingua* 120 (9): 2148-2166.

Antelmi, Donella. 2004. *The Antelmi corpus*. CHILDES Database.

Arosio, Fabrizio, Flavia Adani, and Maria Teresa Guasti. 2008. "Children's processing of subject and object relatives in Italian". In *Language Acquisition and Development: Proceedings of Gala 2005*, ed. by Adriana Belletti, Elisa Bennati, Cristiano Chesi, Elisa Di Domenico and Ida Ferrari, 15-27. Newcastle: Cambridge Scholars Press.

Belletti, Adriana. 2009. "Notes on Passive Object Relatives". In *Functional Structure from Top to Toe*, ed. by Peter Svenonius, 97-114. Oxford: Oxford University Press.

Belletti, Adriana, and Carla Contemori. 2010. "Intervention and Attraction. On the production of Subject and Object Relatives by Italian (young) children and adults". In *Language Acquisition and Development: Proceedings of Gala 2009*, ed. by João Costa, Ana Castro, Maria Lobo and Fernanda Pratas, 39-52. Newcastle: Cambridge Scholars Press.

Belletti, Adriana, Naama Friedmann, Dominique Brunato, and Luigi Rizzi. 2012. "Does gender make a difference? Comparing the effect of gender on children's comprehension of relative clauses in Hebrew and Italian". *Lingua* 122 (10): 1053-1069.

Belletti, Adriana, and Maria Teresa Guasti. 2015. *The Acquisition of Italian*. Amsterdam-Philadelphia: John Benjamins.

Belletti, Adriana, and Luigi Rizzi. 2013. "Intervention in grammar and processing". In *From Grammar to Meaning*, ed. by Ivano Caponigro and Carlo Cecchetto, 294-311. Cambridge: Cambridge Scholar Press.

Brown, H. Douglas. 1972. “Children's comprehension of relativized English sentences”. *Child Development* 42 (6): 1923-1936.

Cardinaletti, Anna. 2004. “Toward a cartography of subject positions”. In *The Structure of CP and IP*, ed. by Luigi Rizzi, 115-165. Oxford: Oxford University Press.

Carreiras, Manuel, Jon Andoni Duñabeitia, Marta Vergara, Irene de la Cruz Pavía and Itziar Laka. 2010. “Subject relatives are not universally easier to process. Evidence from Basque”. *Cognition* 115: 79-92.

Chambers, Craig G., Michael K. Tanenhausb, Kathleen M. Eberhard, Hana Filip and Greg N. Carlson. 2002. “Circumscribing referential domains during real-time language comprehension”. *Journal of memory and language* 47 (1): 30-49.

Chesi, Cristiano, Gianluca Lebani, and Margherita Pallottino. 2008. “A Bilingual Treebank (ITA-LIS) suitable for Machine Translation: what Cartography and Minimalism teach us”. *STiL - Studies in Linguistics 2:* 165-185.

Chomsky, Noam. 1981. *Lectures on government and binding: The Pisa lectures*. Berlin: Walter de Gruyter.

Cipriani, Paola, and Giuseppe Cappelli. 2004. *Italian Calambrone Corpus.* TalkBank. oai:childes.psy.cmu.edu:Romance-Italian-Calambrone.

Collins, Chris. 2005. “A Smuggling approach to the passive in English”. *Syntax* 8 (2): 81-120.

Contemori, Carla, and Adriana Belletti. 2013. “Relatives and passive object relatives in Italian-speaking children and adults: Intervention in production and comprehension”. *Applied Psycholinguistics* 35 (6): 1021-1053.

Contemori, Carla, and Maria Garraffa. 2010. “Comparison of modalities in SLI syntax: A study on the comprehension and production of non-canonical sentences”. *Lingua* 120 (8): 1940-1955.

de Villiers, Jill G., Peter A. de Villiers, and Esme Hoban. 1994. “The central problem of functional categories in the English syntax of oral deaf children”. In *Constraints on Language Acquisition: Studies of Atypical Children*, ed. by Helen Tager-Flusberg, 9-47. Hillsdale, NJ: Lawrence Erlbaum.

De Vincenzi, Marica. 1991. *Syntactic Parsing Strategies in Italian*. Dordrecht: Kluwer.

Franck, Julie, Glenda Lassi, Urlich H. Fraunefelder, and Luigi Rizzi. 2006. “Agreement and movement: A syntactic analysis of attraction”. *Cognition* 101 (1): 173-216.

Friedmann, Naama, Adriana Belletti, and Luigi Rizzi. 2009. “Relativized relatives: Types of intervention in the acquisition of A-bar dependencies”. *Lingua* 119 (1): 67-88.

Friedmann, Naama, Adriana Belletti, Laurice Tuller, João Costa, Maria Teresa Guasti, Ewa Haman, and Spyridoula Varlokosta. 2010. *The acquisition of relative clauses in sixteen languages*. Ms. Cost Action-33.

Gennari, Silvia P., and Maryellen C. MacDonald. 2009. “Linking production and comprehension processes: The case of relative clauses”. *Cognition* 111 (1): 1-23.

Gordon, Peter C., Randall Hendrick, and Marcus Johnson. 2004. “Effects of noun phrase type on sentence complexity”. *Journal of Memory and Language* 51: 97-114.

Grillo, Nino. 2008. “Generalized Minimality”. PhD dissertation, Utrecht Institute of Linguistics, OTS.

Guasti, Maria Teresa, and Anna Cardinaletti. 2003. “Relative clause formation in Romance child production”. *Probus* 15: 47-8.

Guasti, Maria Teresa, Chiara Branchini, and Fabrizio Arosio. 2012. “Interference in the production of Italian subject and object wh-questions. *Applied Psycholinguistics* 33: 185-223.

Hamann, Cornelia and Laurice Tuller. 2010. *Relative Clause Production in French Children and Adolescents*. University of Oldenburg-Université de Tours.

Kidd, Evan, Silke Brandt, Elena Lieven and Michael Tomasello. 2007. "Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses". *Language and Cognitive Processes* 22 (6): 860-897.

Lin, Chien-Jer Charles, and Thomas G. Bever. 2006. "Subject Preference in the Processing of Relative Clauses in Chinese". In *Proceedings of the 25th WCCFL*, ed. by Donald Baumer, David Montero, and Michael Scanlon, 254-260. Somerville, MA: Cascadilla.

MacDonald, Maryellen C., Neal J. Pearlmutter, and Mark S. Seidenberg. 1994. "The lexical nature of syntactic ambiguity resolution". *Psychological review* 101 (4): 676-703.

MacWhinney, Brian, and Elizabeth Bates. 1989. *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

MacWhinney, Brian, and Catherine E. Snow. 1985. "The child language data exchange system". *Journal of child language* 12 (2): 271-95.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. Massachusetts: MIT press.

Matteini, Simona. 2011. *Sabrina Corpus*. Ms. CISCL.

Novogrodsky, Rama, and Naama Friedmann. 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.

Reali, Florencia, and Morten H. Christiansen. 2007. "Processing of relative clauses is made easier by frequency of occurrence". *Journal of Memory and Language* 57 (1): 1-23.

Rizzi, Luigi. 1990. *Relativized Minimality*. Massachusetts: MIT Press.

Rizzi, Luigi. 2004. "Locality and the left periphery". In *Structures and Beyond: The cartography of syntactic structures 3*, ed. by Adriana Belletti, 223-251. Oxford: Oxford University Press.

Rizzi, Luigi, and Ur Shlonsky. 2007. "Strategies of subject extraction". In *Interfaces + Recursion = Language?*, ed. by Hans Martin Gärtner and Uli Sauerland, 115-160. Berlin: Mouton de Gruyter.

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

Roland, Douglas, Frederic Dick, and Jeffrey L. Elman. 2007. "Frequency of basic English grammatical structures: A corpus analysis". *Journal of Memory and Language* 57 (3): 348-379.

Spina, Stefania. 2005. "Il Corpus di Italiano Televisivo (CiT): struttura e annotazione". In *Tradizione & Innovazione. Il parlato: teoria - corpora - linguistica dei corpora: Proceedings of the VI SILFI Conference*, ed. by Elisabeth Burr, 413-426. Firenze: Franco Cesati.

Starke, Michal. 2001. "Move Dissolves into Merge: A Theory of Locality". PhD dissertation, University of Geneva.

Tavakolian, Susan L. 1981. "The conjoined-clause analysis of relative clauses". In *Language Acquisition and Linguistic Theory*, ed. by Susan L. Tavakolian, 167-87. Massachusetts: MIT Press.

Tomasello, Michael. 2009. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Massachusetts: Harvard University Press.

Traxler, Matthew J., Robin K. Morris, and Rachel E. Seely. 2002. "Processing subject and object relative clauses: Evidence from eye movements". *Journal of Memory and Language* 47 (1): 69-90.

Traxler, Matthew J., Rihana S. Williams, Shelley A. Blozis, and Robin K. Morris. 2005. "Working memory, animacy, and verb class in the processing of relative clauses". *Journal of Memory and Language* 53 (2): 204-224.

Warren, Tessa, and Edward Gibson. 2002. "The influence of referential processing on sentence complexity". *Cognition* 85: 79–112.

Adriana Belletti
Università degli Studi di Siena
adriana.belletti@unisi.it

Crisitiano Chesi
Istituto Universitario di Studi Superiori di Pavia
cristiano.chesi@iusspavia.it