

RELATIVE CLAUSES FROM THE INPUT. SYNTACTIC CONSIDERATIONS FROM A CORPUS-BASED ANALYSIS OF ITALIAN *Adriana Belletti & Cristiano Chesì - University of Siena*

Our point of departure will be the well-known classical finding from both acquisition and processing, that Object relatives (ORs) are harder than Subject relatives (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 over a long period of time). 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 some recent experimental findings, and the syntactic account that has been proposed in terms of a featural approach to locality (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 syntax in grammar on the one side, and frequency in input on the other (Gennari & Mac Donald 2009, Tomasello 2003).

1. 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. One crucial finding concerns adults: in an elicited production task, 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 able to preserve the same intended meaning. One privileged such alternative is offered by the use of passive (utilized up to almost 90% in the different groups of adults investigated in the experiments; see also Belletti 2008 for related findings). The results have indicated that the production of what we refer to as *Passive Object Relatives* (PORs), becomes the preferred option for children as well, as soon as passive becomes available to them, around age 5. PORs have also been recently tested in comprehension (Contemori & Belletti 2010), 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 2005 on Mandarin Chinese). These somewhat surprising results, which also open up the issue of a comparison of the complexity of different syntactic computations such as passive and (active) object relatives, can find a direct account in terms of locality, specifically in terms of a featural approach to Relativized Minimality, as developed in Starke 2001, Rizzi 2004, and in Friedmann et al. 2009 (see also Grillo 2005, for a related approach to agrammatism). As also proposed in Belletti 2010, use of passive can be seen as an optimal way to overcome the intervention effect which inevitably arises in the relativization of a direct object across an intervening lexical subject; this syntactic computation, although possible for adults to process, is nevertheless harder than one where no such intervention occurs (cfr. Friedmann et al. 2009, for a conjecture on the comparison between children and adults in this domain). Assuming a derivation of what we call *passive* in the terms developed in Collins 2005, as involving 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* - 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) that the experimental results have so clearly revealed. The assumed derivation is schematically illustrated in (1) for the Italian POR “il bambino che è abbracciato dalla mamma” (the child that is hugged by the mom):

(1) Il bambino che è [VP abbracciato <il bambino>] da [VP la mamma <[VP abbracciato il bambino]>]

the child that is [hugged <the child>] by [VP the mom <[VP hugged the child]>]

A natural question to ask is: to what extent are POR found in naturalist corpora?

2. Corpora analyzed. We considered the following corpora: the Italian Television Corpus (CIT; transcribed television programs in standard Italian, Spina et al. 2003); the Siena University Treebank (SUT; transcribed television news simplified for the simultaneous translation Italian - Italian Sign Language (LIS), Chesi et al. 2008); all the Italian files in the CHILDES database, integrated with 19 files from a child Italian corpus, gathered and transcribed in CHAT format (Matteini 2010), both adults (CHI A) and children (CHI C, age range 1:5-3:4). As shown in Table 1, ORs are overall disfavored with respect to SRs of about 1 to 2, a ratio interestingly very close to the one also reported for other diverse languages (Carreiras et al. 2010 for Basque, Rohde-Gibson 2003 for English).

Corpus	# rRC	SR %	OR %	POR %
CIT	477	61,6%	25%	4,4%
SUT	174	93%	2%	-
CHI A	678	65%	34%	0,1%
CHI C	40	95%	5%	-

Table 1. Restrictive Relative Clauses (rRCs) types (SRs, ORs, and PORs only)

# OR	# pro + pron V (%)	# V S (%)	# S V (%)
117	72+15 (74%)	16 (13%)	14 (13%)
4	3+0 (75%)	1 (25%)	-
228	138+3 (62%)	80 (35%)	7 (3%)
2	-	2 (100%)	-

Table 2. ORs types: pre-verb. null + pronom. subj. (pro + pron V), post-verb. lexical subj. (V S), pre-verb. lexical subj. (S V)

As for the main question raised above, our pilot investigation indicates that PORs are poorly present in the analyzed corpora. Specifically, only one POR is present in the CHI A corpus (no passive at all in the children's data as is expected given their young age), vs 34% ORs; only 21 out of 303 SRs are PORs (both verbal and reduced relatives; repetitions included) in the CIT corpus, corresponding to 4,4% of the total number of relative clauses, vs 25% ORs. A further finding concerns the shape of the distribution of the ORs present in the corpus, as to the nature (null vs overt; pronominal vs lexical) and the position of the Subject within the relative clause, summarized in Table 2.

3. Outline of discussion. The finding on the very limited presence of PORs in the corpus study contrasts with the production and comprehension experimental results reviewed in §1, for both adults and children. This could be partly due to the limited presence, in the analyzed data, of naturalistic exchanges comparable to those present in the eliciting design: the head of the relative is mostly inanimate in the corpora (only 6% of animate heads in ORs) and always animate in the experimental task. We propose that the crucial factor is syntactic in nature and it is to be recognized in the intervention of the subject in the establishment of the dependency between the relative head and its merge/gap position in the relative clause, which inevitably occurs in ORs. The comparison between the experimental results and the corpus study is revealing in this respect. On the one hand, the experimental results from adults indicate that the syntactic computation optimally eliminating intervention - passive - is favorably selected. On the other hand, the developmental results, which show that children approach the adults' behavior, neatly indicate that PORs are favored in production and comprehension, no matter how infrequent the structure turns out to be in the naturalistic input. Indeed, the naturalistic data from the corpus further support the assumed locality interpretation as they indicate that in spontaneous production the favored ORs produced are those where intervention expressed in featural terms, is either arguably weaker, as when the subject is a null or overt pronominal (74%, CIT; 62% CHI A), or arguably absent, as when (to a lesser extent) the subject is a post-verbal subject (13% CIT; 35% CHI A), a type of structure also possibly amenable to a *smuggling* analysis. The ORs which are less present are those where intervention is stronger, as in the case of a preverbal (lexical) subject (13%, CIT; 3%, CHI A; 0 in both CHI C and the SUT corpus, with a remarkable overall similarity of CHI C and SUT in the SR/OR ratio, Table1, and the nature and position of subject, Table 2). We will conclude by illustrating ways of modulating intervention in ORs, which are revealed by both the corpus analysis and the experimental results, in which manipulations of the animacy feature and other features such as e.g. gender, number in the relative head and the intervening subject may facilitate, to different degrees, the processing of these structures.