The Rivista di Grammatica Generativa/Research in Generative Grammar (RGG) is a journal devoted to the dissemination of research within the generative paradigm. It is an open access journal, hosted by the linguistic archive LEAR (http://lear.unive.it/jspui/handle/11707/593) of the Center for Language Sciences of Ca’ Foscari University, Venice. All articles published in the journal are subject to an anonymous peer review process.

Editors in Chief
Guglielmo Cinque (Università Ca’ Foscari,Venezia)
Luigi Rizzi (Università di Siena, Université de Genève)

Associate Editorial Board
Manuela Ambar (Universidade de Lisboa), Paola Benincà (Università di Padova), Adriana Belletti (Università di Siena), Luciana Brandi (Università di Firenze), Luigi Burzio (The Johns Hopkins University), Noam Chomsky (MIT), Patrizia Cordin (Università di Trento), Violeta Demonte (Universidad Autónoma de Madrid), Alessandra Giorgi (Università Ca’ Foscari, Venezia), Giorgio Graffi (Università di Verona), Richard Kayne (New York University), Michael Kenstowicz (MIT), Giulio Lepschy (UCL, London and Cambridge University), Giuseppe Longobardi (Università di Trieste), Maria Rita Manzini (Università di Firenze), Joan Mascaró (Universitat Autònoma de Barcelona), Nicola Munaro (Università Ca’ Foscari, Venezia), Marina Nespor (Università di Milano-Bicocca), Jean-Yves Pollock (Université Paris-Est Marne-la-Vallée), Annarita Puglielli (Università di Roma III), Andrew Radford (University of Essex), Lorenzo Renzi (Università di Padova), Alain Rouveret (Université de Paris VIII), Leonardo Savoia (Università di Firenze), Sergio Scalise (Università di Bologna), Laura Vanelli (Università di Padova).

Editorial Assistants
Laura Bortolotto (Università Ca’ Foscari,Venezia), Loretta Manzato (Università Ca’ Foscari, Venezia), Karen Martini (Université de Genève), Giuseppe Samo (Université de Genève)

Editorial Manager
Loretta Manzato (Università Ca’ Foscari, Venezia) manzato@unive.it
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adriana Belletti &amp; Cristiano Chesi</td>
<td>A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production</td>
<td>1</td>
</tr>
<tr>
<td>Tamirand De Lisser, Stephanie Durrleman, Luigi Rizzi &amp; Ur Shlonsky</td>
<td>The acquisition of Jamaican creole: a research project</td>
<td>29</td>
</tr>
<tr>
<td>Cornelia Hamann &amp; Laurice Tuller</td>
<td>Genuine versus superficial relatives in French: the depth of embedding factor</td>
<td>47</td>
</tr>
<tr>
<td>M. Rita Manzini &amp; Leonardo Savoia</td>
<td>Linkers in Aromanian in comparison to Albanian and Romanian</td>
<td>83</td>
</tr>
<tr>
<td>Salvador Mascarenhas</td>
<td>Complementizer doubling in European Portuguese</td>
<td>105</td>
</tr>
<tr>
<td>Chiara Zanini, Paola Benincà &amp; Carlo Semenza</td>
<td>The crucial role of the event structure in the retrieval of nominalizations in aphasia</td>
<td>117</td>
</tr>
</tbody>
</table>
A SYNTACTIC APPROACH TOWARD
THE INTERPRETATION OF SOME DISTRIBUTIONAL
FREQUENCIES: COMPARING RELATIVE CLAUSES
IN ITALIAN CORPORAS 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.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

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).](image)

<table>
<thead>
<tr>
<th>Type of relative clause</th>
<th>BNC</th>
<th>BNC spoken</th>
<th>Brown</th>
<th>Switchboard</th>
<th>WSJ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject relative</td>
<td>14182</td>
<td>9851</td>
<td>15024</td>
<td>9548</td>
<td>18229</td>
</tr>
<tr>
<td>Object relative</td>
<td>2943</td>
<td>3863</td>
<td>1976</td>
<td>5616</td>
<td>1802</td>
</tr>
<tr>
<td>Object relative (reduced)</td>
<td>5455</td>
<td>14423</td>
<td>4746</td>
<td>5314</td>
<td>3385</td>
</tr>
<tr>
<td>Passive relative</td>
<td>3118</td>
<td>1729</td>
<td>2867</td>
<td>302</td>
<td>1224</td>
</tr>
<tr>
<td>Passive relative (reduced)</td>
<td>10730</td>
<td>2886</td>
<td>10733</td>
<td>779</td>
<td>12788</td>
</tr>
</tbody>
</table>

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 University 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

\[ X \ldots Z \ldots Y \ldots \]

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. 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):

(1)  
\[
\begin{array}{cccc}
\text{X} & \text{Z} & \text{Y} \\
+R & +NP & +NP & +R +NP \\
\text{il bambino} & \text{che} & \text{il nonno cerca/trova} & <\text{il bambino}> \\
\text{the child} & \text{that} & \text{the grandpa seeks/finds} & <\text{the child}> \\
\end{array}
\]

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).

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).
A syntactic approach toward the interpretation of some distributional frequencies:
comparing relative clauses in Italian corpora and in elicited production

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.

<table>
<thead>
<tr>
<th>Corpus Name</th>
<th>References</th>
<th>Size (in words)</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHILDES</td>
<td>MacWhinney &amp; Snow (1985)</td>
<td>132 files (390.511 words: 115.357 produced by children, 275.154 produced by adults)</td>
<td>chat format</td>
</tr>
<tr>
<td>SUT</td>
<td>Chesi et al. (2008)</td>
<td>29 TGs (17.981 words)</td>
<td>SUT (specific constituency/ dependency format, XML)</td>
</tr>
<tr>
<td>CIT</td>
<td>Spina (2005)</td>
<td>7 TV programs (42.668 words)</td>
<td>morphologically tagged text</td>
</tr>
</tbody>
</table>

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).
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

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):

\begin{equation}
\text{(3) Regular expressions using “grep”:}
\text{\qquad grep -i -n -E ”TIER:([:space:]]|[:punct:]]|[:alpha:]][[:space:]]*che[:space:]]”}
\end{equation}

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):

\begin{equation}
\text{(4) \quad a. \quad tgrep ‘NP.rel < C.rel’}
\text{\qquad b. \quad tgrep ‘NP.rel-obj.anim, NP-subj.anim’}
\end{equation}

---

\(^5\) 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 abbain)” 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.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Tool used</th>
<th># of analyzed words</th>
<th># of “che” (%)</th>
<th># of RCs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI A</td>
<td>Keyword [che]</td>
<td>275.154</td>
<td>5.580 (2,03)</td>
<td>677 (0,25)</td>
</tr>
<tr>
<td>CHI C</td>
<td>Keyword [che]</td>
<td>115.357</td>
<td>747 (0,65)</td>
<td>94 (0,08)</td>
</tr>
<tr>
<td>CIT</td>
<td>Tag [POS=&quot;pro:rela&quot;]</td>
<td>42.668</td>
<td>1027 (2,4)</td>
<td>477 (1,1)</td>
</tr>
<tr>
<td>SUT</td>
<td>Tag [C.rel.pro]</td>
<td>17.981</td>
<td>210 (1,17)</td>
<td>174 (0,9)</td>
</tr>
</tbody>
</table>

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.

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

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”

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:

<table>
<thead>
<tr>
<th>Corpus</th>
<th># of Rs</th>
<th># SRs (%)</th>
<th># ORs (%)</th>
<th># PORs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIT</td>
<td>477</td>
<td>295 (62%)</td>
<td>117 (25%)</td>
<td>19 (4%)</td>
</tr>
<tr>
<td>CHI A</td>
<td>677</td>
<td>440 (65%)</td>
<td>228 (34%)</td>
<td>1 (0.1%)</td>
</tr>
<tr>
<td>SUT</td>
<td>174</td>
<td>159 (91%)</td>
<td>12 (7%)</td>
<td>3 (2%)</td>
</tr>
<tr>
<td>CHI C</td>
<td>94</td>
<td>83 (88%)</td>
<td>11 (11%)</td>
<td>-</td>
</tr>
</tbody>
</table>

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):

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

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\textsuperscript{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.

<table>
<thead>
<tr>
<th>Corpus</th>
<th># of verbs</th>
<th># trans (%)</th>
<th># ditrans (%)</th>
<th># pass (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUT</td>
<td>872</td>
<td>645 (74%)</td>
<td>50 (6%)</td>
<td>177 (20%)</td>
</tr>
</tbody>
</table>

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:

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

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.

\textsuperscript{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:

<table>
<thead>
<tr>
<th>Corpus</th>
<th># of ORs</th>
<th># animate heads (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CIT</td>
<td>117</td>
<td>6 (5%)</td>
</tr>
<tr>
<td>CHI A</td>
<td>228</td>
<td>33 (14%)</td>
</tr>
<tr>
<td>SUT</td>
<td>12</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>CHI C</td>
<td>11</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>Verb class</th>
<th># SR</th>
<th># OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unacc.+Unerg.+be</td>
<td>231</td>
<td>0</td>
</tr>
<tr>
<td>Transitive</td>
<td>161</td>
<td>193</td>
</tr>
<tr>
<td>Di-transitive</td>
<td>22</td>
<td>35</td>
</tr>
</tbody>
</table>

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.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

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\(^\text{14}\) 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:

<table>
<thead>
<tr>
<th>Cond.</th>
<th>RC head</th>
<th>Subj</th>
<th>cue sentence</th>
<th>elicitation sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+anim</td>
<td>+anim</td>
<td>I poliziotti salutano un ragazzo (the policemen greet a child)</td>
<td>tu quale ragazzo vorresti incontrare? (Which child would you rather meet?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I poliziotti rincorrono un ragazzo (the policemen chase a child)</td>
<td>“vorrei incontrare il ragazzo...”</td>
</tr>
<tr>
<td>2</td>
<td>+anim</td>
<td>–anim</td>
<td>I secchi sbilanciano un imbianchino (The buckets unbalance a decorator)</td>
<td>Tu quale imbianchino vorresti aiutare? (Which decorator would you rather help?)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>I secchi sporcano un imbianchino (The buckets dirty a decorator)</td>
<td>“vorrei aiutare l’imbianchino...”</td>
</tr>
</tbody>
</table>

\(^{14}\) 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.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

<table>
<thead>
<tr>
<th>Cond.</th>
<th>RC head</th>
<th>Subj</th>
<th>cue sentence</th>
<th>elicitation sentence</th>
</tr>
</thead>
</table>
| 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?  
“vorrei incontrare il ragazzo...”  
I would rather meet the child... |
|       | +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?  
“vorrei aiutare l’imbianchino...”  
I would rather help the decorator... |
| 2     | -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... |
| 3     | -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 11. Experiment 1, verb change. 4 conditions.

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

17
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:

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

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?
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

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.

<table>
<thead>
<tr>
<th></th>
<th>H+anim S+anim</th>
<th>H+anim S-anim</th>
<th>H-anim S+anim</th>
<th>H-anim S-anim</th>
</tr>
</thead>
<tbody>
<tr>
<td>POR all</td>
<td>57 (79%)</td>
<td>60 (83%)</td>
<td>65 (90%)</td>
<td>63 (87%)</td>
</tr>
<tr>
<td>POR</td>
<td>11 (15%)</td>
<td>20 (28%)</td>
<td>5 (7%)</td>
<td>5 (7%)</td>
</tr>
<tr>
<td>POR r.</td>
<td>37 (51%)</td>
<td>37 (51%)</td>
<td>50 (69%)</td>
<td>55 (76%)</td>
</tr>
<tr>
<td>POR r. by</td>
<td>6 (9%)</td>
<td>1 (1%)</td>
<td>9 (12%)</td>
<td>3 (3%)</td>
</tr>
<tr>
<td>POR by</td>
<td>3 (4%)</td>
<td>2 (3%)</td>
<td>1 (1%)</td>
<td>0</td>
</tr>
<tr>
<td>OR all</td>
<td>14 (20%)</td>
<td>4 (6%)</td>
<td>7 (10%)</td>
<td>8 (11%)</td>
</tr>
<tr>
<td>OR</td>
<td>2 (3%)</td>
<td>0</td>
<td>2 (3%)</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>OR VS</td>
<td>4 (6%)</td>
<td>1 (1%)</td>
<td>1 (1%)</td>
<td>2 (3%)</td>
</tr>
<tr>
<td>OR pro</td>
<td>8 (11%)</td>
<td>3 (4%)</td>
<td>4 (6%)</td>
<td>4 (6%)</td>
</tr>
<tr>
<td>ALT</td>
<td>1 (1%)</td>
<td>8 (11%)</td>
<td>0</td>
<td>1 (1%)</td>
</tr>
<tr>
<td>ALT SR</td>
<td>1 (1%)</td>
<td>7 (10%)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALT PP</td>
<td>0</td>
<td>1 (1%)</td>
<td>0</td>
<td>1 (1%)</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th></th>
<th>H+anim S+anim</th>
<th>H+anim S-anim</th>
<th>H-anim S+anim</th>
<th>H-anim S-anim</th>
</tr>
</thead>
<tbody>
<tr>
<td>POR all</td>
<td>64 (76%)</td>
<td>64 (76%)</td>
<td>50 (60%)</td>
<td>59 (70%)</td>
</tr>
<tr>
<td>POR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>POR r.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>POR r. by</td>
<td>52 (62%)</td>
<td>52 (62%)</td>
<td>45 (54%)</td>
<td>52 (62%)</td>
</tr>
<tr>
<td>POR by</td>
<td>12 (14%)</td>
<td>12 (14%)</td>
<td>5 (6%)</td>
<td>7 (8%)</td>
</tr>
<tr>
<td>OR all</td>
<td>9 (11%)</td>
<td>3 (4%)</td>
<td>5 (6%)</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>OR</td>
<td>1 (1%)</td>
<td>2 (3%)</td>
<td>2 (2%)</td>
<td>0</td>
</tr>
<tr>
<td>OR VS</td>
<td>8 (10%)</td>
<td>1 (1%)</td>
<td>3 (4%)</td>
<td>3 (4%)</td>
</tr>
<tr>
<td>OR pro</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ALT</td>
<td>11 (13%)</td>
<td>17 (20%)</td>
<td>29 (34%)</td>
<td>22 (26%)</td>
</tr>
<tr>
<td>ALT SR</td>
<td>0</td>
<td>6 (7%)</td>
<td>0</td>
<td>22 (26%)</td>
</tr>
<tr>
<td>ALT PP</td>
<td>11 (13%)</td>
<td>11 (13%)</td>
<td>29 (34%)</td>
<td>0</td>
</tr>
</tbody>
</table>

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](image)

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.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

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:

```
[CP D NP [TP proexpl [V - ] by [vP DP <[vP V - ]> ]]]
```

2. OR with null subject:

```
[CP D NP [TP proref [vP < > [vP V < > ]]]]
```

3. OR with post-verbal subject through smuggling:

```
[CP D NP [TP proexpl [V - ] [vP DP <[vP V - ]> ]]]
```

4. OR with pre-verbal subject:

```
[CP D NP [TP DP [vP < > [vP V < > ]]]]
```
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

On one extreme, 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, 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 and ORs with a post-verbal subject. 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 is it 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.
\section*{Appendix}

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Camilla</th>
<th>Diana</th>
<th>Guglielmo</th>
<th>Marco</th>
<th>Martina</th>
<th>Raffaella</th>
<th>Rosa</th>
<th>Sabrina</th>
<th>Viola</th>
</tr>
</thead>
<tbody>
<tr>
<td>1;5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1;11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;0</td>
<td></td>
<td>2 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;1</td>
<td>1 - 0</td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;2</td>
<td>3 - 0</td>
<td>5 - 0</td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;3</td>
<td>3 - 0</td>
<td>3 - 0</td>
<td>0 - 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;4</td>
<td>5 - 0</td>
<td>2 - 0</td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;5</td>
<td></td>
<td>2 - 1</td>
<td></td>
<td></td>
<td></td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;6</td>
<td>2 - 0</td>
<td></td>
<td></td>
<td>0 - 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;7</td>
<td>1 - 0</td>
<td>2 - 0</td>
<td>9 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;8</td>
<td></td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;9</td>
<td>1 - 4</td>
<td>2 - 0</td>
<td>1 - 1</td>
<td></td>
<td></td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;10</td>
<td></td>
<td>1 - 0</td>
<td>3 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2;11</td>
<td>2 - 1</td>
<td>3 - 2</td>
<td>5 - 0</td>
<td>10 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;1</td>
<td>1 - 0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 - 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;4</td>
<td>6 - 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3;5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production

References

Figure A1. Experimental screenshot with all components displayed.
A syntactic approach toward the interpretation of some distributional frequencies: comparing relative clauses in Italian corpora and in elicited production


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


THE ACQUISITION OF JAMAICAN CREOLE:
A RESEARCH PROJECT

Tamirand De Lisser
University of Geneva
Stephanie Durrleman
University of Geneva
Luigi Rizzi
University of Geneva
University of Siena
Ur Shlonsky
University of Geneva

Abstract: This article describes a research project aimed at filling the gap in syntactic research on language acquisition, in the area of creole languages. For too long acquisitionists have ignored the domain of creole languages, and as such the time is ripe for the present research. The purpose of this paper is to present an outline of the research project entitled 'The Acquisition of Jamaican Creole Syntax: A corpus-based study of early parameter setting', funded by the Swiss National Science Foundation grant 100015_131793/1. The goal of this project is to provide an extensive descriptive analysis of early production in the development of grammatical representations of children acquiring Jamaican Creole (JC). In addition, not only will this project contribute to scientific knowledge, but it may be applicable to well needed developments in early childhood education and language remediation in Jamaica and its diaspora. Additionally, the present project will provide an accessible database for further study of the Jamaican language.

Keywords: Jamaican Creole, language acquisition, research methods, syntax

1. Introduction

Over the last 40 years, there has been increasing research in the domain of language acquisition; but this has been predominantly on European Languages and only more recently has the sphere been opened to non-Indo-European languages. Still, the acquisition of other types of languages, such as creoles, has remained largely unexplored. In this paper, we introduce a new research project entitled ‘The Acquisition of Jamaican Creole Syntax: A corpus-based study of early parameter setting’, funded by the Swiss National Science Foundation (100015_131793/1) from 2011 until 2015. It represents the first longitudinal exploration of the acquisition of Jamaican Creole syntax and the only longitudinal study of Creole syntactic development. The focus of the research project was to explore the emergence and transformation of both target-consistent and target-inconsistent syntactic
developments in children acquiring JC, and to offer a theory-driven analysis on phrase-
structure building.

The paper is divided into four sections. The first section situates the research project in
the field of language acquisition and articulates the main research questions. Section two
describes the language situation in Jamaica and its impact on the study. Section three
outlines the research methodology, the selection of participants and the general analysis
methods which were employed. The paper ends with a brief summary and outlook.

1.1. Background

This project focuses on properties of syntactic systems of early JC and investigates the
development of grammatical properties in the course of the first years of life. Most research
on first language acquisition, despite differences in methodologies, converges on the fact
that developing children acquire the language of their linguistic community effortlessly,
under varying circumstances, in a limited amount of time. This process is normally
achieved uniformly, notwithstanding cross-linguistic variations. In addition, certain
developmental patterns have emerged that are cross-linguistically uniform. There is an on-
going debate in the literature as to whether target-inconsistent production in early child
language is a result of parameter (mis)-setting, immaturity of computational or memory
capacities or other aspects of cognitive development (see Hyams 1986, 1992; Valian 1990;
others). Additionally, there is the controversial view that the parametric choices in creole
languages may directly express default settings, as a consequence of the special conditions
holding during the process of creolization (Bickerton, 1984, 1999; Degraff, 1999). Various
authors have argued against the proposal that Creoles constitute exceptional languages (e.g.
Degraff 2003, 2004; Mufwene 2000, 2001; among others), but little research has explored
these predictions in terms of acquisition: Are creoles more like themselves and less like
other languages, more directly mirroring properties of Universal Grammar and as such
containing little or no target-inconsistency in L1 development? Our current understanding
of the grammatical development of creole children is still extremely limited. While the
acquisition of syntax is a vastly studied area, the acquisition of creole syntax remains a
largely unexplored domain (with the exception of work on Mauritian Creole by Adone

1.2. Significance and Aims

Against this background, the research project contributes to these discussions by
focusing on the acquisition of JC. JC being an analytic language, the overt realization of
various syntactic elements is particularly suited for mapping incremental syntactic
development of child grammar. The study sheds light of a number of controversial topics
related to the nature of target-inconsistent phenomena in early languages and fixation of
parameters. In particular, it examines word order patterns, null subjects, root infinitives,
topicalization, focalization, interrogation, tense, mood and aspect, double negation, verb
serialization, among other phenomena. We seek to address the following questions:

1. Is there a developmental order in the acquisition of lexical and functional
structures, with the latter globally delayed with respect to the former (Radford 1990)? Or do
lexical and functional structures co-occur at all levels of acquisition?

2. Does structure emerge incrementally in line with the incremental structure
building approach to development (Radford 1990, and subsequent works) or are all
structures available when significant production starts in line with the full competence approach (Poeppel & Wexler 1993)?

3. Does the grammar of children acquiring JC replicate the highly structured cartographically-coherent pattern of the target language?

4. Is the acquisition of JC exceptional or is it acquired just like other well-studied languages? Do learners of JC go through a root null subject phase? Does their grammar reflect Root Infinitives?

5. More generally, do maturation factors affect the linguistic development (e.g. as in Borer and Wexler, 1987), or are target-inconsistencies in early productions fully reducible to the mechanism of parameter setting, and/or to an incomplete lexical acquisition?

6. Is the gradual character of the changes in early grammatical systems consistent with a parametric approach, or does it favor item-based approaches (Tomasello 2003) or approaches based on grammar competition (Yang 2002; Roeper 2007).

Moreover, the current research provides an accessible and usable corpus of natural production of Jamaican Creole. This corpus will be archived in the CHILDES (Child Language Data Exchange System) repository. In order to provide a precise and systematic description of the acquisition of JC, the study is couched in the Principles and Parameters / minimalist framework of Generative Grammar. Nonetheless, alternative views of the construction of grammatical knowledge were considered whenever relevant.

2. Language Situation

Jamaica has a population of 2.7 million inhabitants, making it the largest English-speaking Island of the Caribbean. Standard Jamaican English (English) is the official language, i.e. the language used is schools, parliament and the media. JC is the national language, and is spoken by the majority of the population. JC is, for the most part, the ambient language used in the home and is the first language of most Jamaicans. JC is acquired mainly through parent, sibling and extended family interaction while English appears to be primarily acquired from school interactions in the classroom (Carpenter, 2009). The Ministry of Education has however adopted an approach in which teachers “promote basic communication through the oral use of the home language in the early years (e.g. K-3) while facilitating the development of literacy in English” (Bryan 2001, 23 in Lewis 2010, 13).

The Jamaican language situation is described as a Creole Continuum (Decamp 1971) with speakers varying across the continuum from basilect to acrolect. The basilectal end is also referred to as the ‘deep creole’. Speakers of this variety tend to be located in the more rural areas and manifest the highest degree of substratum influence (i.e. influence from West African languages as transmitted during slavery). This variety is farthest from the ‘local standard’. Speakers at the other end of the continuum (the acrolect) are mainly associated with the urban areas and generally speak the ‘local standard’, which is the prestigious variety, containing the most ‘superstrate’ (British English) influences. Situated in between the two extremes are the mesolectal varieties, which share features of both extremes to varying degrees. Speakers of the basilect and the acrolect varieties may be mutually unintelligible; however this is very rare as most people can adjust their production upward or downward on the continuum (Durleman-Tame 2008). The distinction between mesolect and basilect is not quite clear-cut as due to the relatively fluid social structure, rural varieties are becoming more and more urbanized, making more overlap between the
two varieties (Winford 1993). There are however, speakers who command only one of the varieties (monolinguals) and others who command both varieties (bilinguals). The Language Competence Survey of Jamaica (2007) reports that 46.4% of its sample demonstrated bilingualism; however while only 17.1% were monolingual English speakers, 36.5% were monolingual JC speakers. The majority of the monolingual English speakers was located in the eastern and urban areas and was concentrated in the highly skilled and professional groups.

The project concentrates primarily on the variety found at the basilectal extreme of the continuum (which we have been referring to as JC). The choice for this selection is based on the fact that it is the variety with the least influence from Standard English, and therefore offering the most syntactic novelty (in line with Durrleman-Tame 2008 and Bailey 1966). JC has been considered a canonical example of an Atlantic Creole (Patrick 2004), since it is characterized by a cluster of grammatical properties typically found in such creoles. Features that are characteristic of JC, which make it quite distinct from English, include serial verb constructions, double negation, lack of subject-auxiliary inversion, lack of case morphology or gender distinction on pronouns. To find monolingual speakers whose linguistic repertoire contains only these features and absolutely no English influence is however challenging due to the continuum situation. In the next section, we detail the procedures in ensuring the selection of the most appropriate informants for inclusion in the research project.

3. Research Methodology

In order to investigate the emergence of the early syntactic systems of JC, six children, age ranging from 18 to 23 months at the beginning of recordings, were recorded for a period of 18 months. This age-range corresponds to the period in which syntax typically emerges in children and during which target-inconsistent productions have been documented in other languages. In addition, it is the period in which the methodology we have adopted can be most fruitfully utilized. The linguistic production of children younger than 18 months is often too poor and too dispersed to provide coherent data. Above 36 months, children are ‘talking machines’ and data collection based on longitudinal recordings is generally much less informative than research employing experimental methods. As the acquisition of JC is an understudied domain, a longitudinal corpus study is an excellent starting point to provide a general overview of the relevant phenomena.

3.1. Participants

For the research project, informants were strategically selected from households where JC was the primary language spoken, and as such the interference from English in the child’s linguistic environment was minimal. Given the existence of the creole continuum, various factors were considered in identifying and selecting the participants for inclusion in the study. Primary consideration was given to the area of residence and the level of education of the primary care-giver. More specifically, speakers from rural communities with less education were ranked closer to the basilectal end of the continuum (Meade 2001). In light of this observation, in the search for children to be included in our study, we targeted Southfield and neighboring communities, located in the parish of St. Elizabeth.
This area was selected based on the socio-demographic profile of the residents (Francis 2012) and general opinions on where the most conservative JC could be found.

In order to find participants in the desired age group, we consulted the local Health Centre, where all children of the community and surrounding areas are expected to be registered. A letter was sent to the Head Nurse, explaining the research objectives and the rationale behind seeking participants. A list of prospective participants and their contact details was received.

Preliminary interviews, guided by ethical principles, were conducted with care-givers of the prospective participants. We were mindful of the observers’ paradox (Labov 1972). This is where a researcher tries to observe naturalistic speech, however his/her presence as an observer creates a situation in which speakers are highly conscious of their speech and will therefore modify it. In order to minimize this, the interviews were informal and took place in the homes of the prospective participants. The language used by the interviewer was JC. These initial interviews allowed us to analyze the language used by the caregivers and members of the household for typical creole features (as described in Bailey 1966; Patrick 2004; Durulleman-Tame 2008). If these features were sufficiently present, then children in such households would be eligible subjects for participation in the research project. Notwithstanding this, further selection criteria were applied, involving the willingness and availability of the informants and the level of speech production by the children.

One participant was immediately ruled out, as despite their overwhelming interest in the research, the language of the household contained many mesolectal and acrolectal features, and as such did not conform to the basilectal criteria necessary for participation. After 3 sessions of recording, another informant was not producing any words in contrast to what his mother had reported. To continue recording him was not profitable for the research project, and as such he was subsequently replaced by another informant. A third informant was clearly not interested in participating in the research project and was also removed. Additional informants were included in the study based on references received from participants. At the end of the selection process the following table represents the participants included in the study (participants are referred to via pseudonyms).

<table>
<thead>
<tr>
<th>Name</th>
<th>Age(^1) at 1(^{st}) recording</th>
<th>Gender</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL</td>
<td>1;6,11</td>
<td>Male</td>
<td>Back Flagaman</td>
</tr>
<tr>
<td>ALA</td>
<td>1;7,19</td>
<td>Female</td>
<td>Southfield</td>
</tr>
<tr>
<td>RJU</td>
<td>1;7,28</td>
<td>Male</td>
<td>Back Flagaman</td>
</tr>
<tr>
<td>TYA</td>
<td>1;9,18</td>
<td>Female</td>
<td>Round Hill</td>
</tr>
<tr>
<td>KEM</td>
<td>1;11,3</td>
<td>Male</td>
<td>Round Hill</td>
</tr>
<tr>
<td>SHU</td>
<td>1;11,25</td>
<td>Female</td>
<td>Back Flagaman</td>
</tr>
</tbody>
</table>

Table 1: Research Participants

We present in the next section individual profiles of each participant.

3.1.1. Participants’ Individual Profiles

COL (age range: 1;6,11 – 2;11,7): COL was the youngest informant in the study. He had a Mean Length of Utterance (MLU) of 1.6 at 20 months and 3.81 at 35 months. He was

\(^1\) Age is presented in Year; Month, Day format.
a single child living with both parents in the community of Back Flagaman. They lived in a ‘nestled’ area where five other houses were a stone's throw away. These houses had other children, one of whom was also a participant of this study. In addition, his paternal grandmother and cousins were immediate neighbors, and as such the yard was normally active. His maternal grandmother lived in the same community, before migrating abroad, 9 months into the study. Some of his recordings were conducted at her house; however the majority was done in his home. His father was a farmer, who habitually fished, while his mother was a housewife. After starting kindergarten at 28 months old, his parents constructed a small shop in their yard, where the mother worked as a shopkeeper.

ALA (age range: 1;7,19 – 3;0,15): At 21 months, with an MLU value at 1.48 and 5.66 at 36 months, she was one of the most vocal participants in the study. She lived in the community of Southfield with her parents in a family house where they occupied a room at the back. She was a single child for her parents but had cousins, aunts, uncles and grandparents in the extended household. Her mother held a clerical/administrative position in a governmental agency in the parish, while her father was a driver distributing goods for a furniture company. As such, when both parents were at work, the child would stay with relatives in the neighboring community of Seaview. Recordings were therefore conducted at Seaview and at the child’s home. ALA started attending school at 33 months old.

RJU (age range: 1;7,28 – 3;0,25): RJU lived in an extended family household with his parents, paternal grandmother, aunts, uncles, and cousin. His cousin, a girl aged 6 years, appeared to be his best friend with whom he interacted the most. His mother was unemployed while his father worked as a farmer. Throughout the 18 month period, RJU visited his paternal grandfather in a district located about 10 miles away from his home community; where some of his recordings were conducted. During the last four months of the recording sessions, RJU and his mother relocated to the neighboring community of Crossroads, where they lived with other relatives. RJU later returned home under the care of his paternal grandmother and aunt. The majority of his recordings were conducted in his home in the presence of his cousin. At 22 months, he had an MLU of 1.39 and ending with an MLU of 4.86 at 36 months.

TYA (age range: 1;9,18 – 3;2,15): TYA lived in the community of Roundhill with her parents and two siblings. Her mother was a housewife while her father was a taxi-driver. Her immediate neighbors were her grandmother, aunts and cousins. She had an MLU of 1.22 at 23 month and 4.86 at 38 months. She started school at 35 months; at this point we saw a very rapid development in her speech, moving from an MLU of 2.16 at 34 months of age to a high of 5.38 at 37 months. Her recordings were conducted mainly in the comfort of her home.

KEM (age range: 1;11,3 – 3;3,11): KEM’s MLU was 2.1 and 5.47 at 24 and 39 months respectively, peaking at 6.46 at 37 months. He started school at 32 months. He lived with his mother, maternal grandparents, aunt and uncle in the community of Roundhill. His mother was unemployed, his grandfather was a farmer and his grandmother operated a shop in the yard. KEM sometimes had playmates with whom he ran about freely in his large yard space and neighboring taverns. All of his recordings were conducted at his home.

SHU (age range: 1;11,25 – 3;4,13): SHU was the eldest participant in the research project. She had an MLU of 2.88 at 25 months and 5.02 at 40 months. She lived in the community of Back Flagaman with her mother and her brother and enjoyed a visiting relationship with her father and his family who lived just a few meters away. She was the only child for her father. Upon the passing of her father at 28 months, she lived with her paternal grandparents, aunt and uncle, and then had a visiting relationship with her mother.
Her mother was generally unemployed but worked occasionally as a store clerk. Her grandfather operated a shop and her grandmother was a housewife. She started school at 33 months of age. Recordings were conducted mainly at the home of her mother or paternal grandparents, and on a few occasions at the home of her maternal grandmother.

MLU values are plotted in Figures 1 – 6 showing a steady increase of utterance length against age for all participants in the study. On the x-axis we present the MLU values and on the y-axis the informants’ age in year; months, days format. The complete list of MLU values are given in Appendix 1, Tables (1) – (6). Note that this does not include the 2 initial months of data collection and only two recordings per month are represented. Justification for this is presented in section 3.4. The MLUs were calculated automatically with the help of CLAN (Computerized Language Analysis). They are all word based, as opposed to morpheme based. Being an isolating language, all lexical and functional elements are counted as independent words. The MLU presented for JC may therefore not be immediately comparable to that in languages with morphologically complex words for which a morpheme-based count is adopted.

Fig. 1: COL’s MLU
Fig. 2: ALA’s MLU

Fig. 3: RJU’s MLU
Fig. 4: TYA's MLU

Fig. 5: KEM's MLU
3.2. Recordings

Preliminary interviews and recordings were conducted by Tamirand De Lisser, the main researcher, at the end of October 2011 and recording sessions began in November 2011. Nickesha Dawkins, a graduate student from the University of the West Indies, was employed as a Research Assistant to the project in December 2011. Along with the main researcher, she carried out the recordings and transcription of the data.

Each participant was recorded every 10 days in separate 60 minutes sessions, for the initial five months, and subsequently every 15 days for the remaining duration of the data collection phase of the project. Not accounting for the first two months when selection process was still underway, we ended up with a database of 204 hours of recordings. Recordings were conducted in JC, in the presence of the interviewers, care-providers and in some instance other relatives and friends. The format of the recording sessions was quite informal and participants, becoming close friends with the researchers, spoke freely on various subject matters. The sessions were usually interactively built upon games and story books. The researchers would take a gift for each child at every recording session.

Based on the close proximities of the communities in which the children reside, recording sessions were usually all conducted in one day. The researchers would get in the field around 9 a.m. (sometimes earlier depending on the availability of the participants) and recordings would begin immediately. Both researchers would interview ALA, as she was the only participant of her community, located approximately 15 minutes by car from the community where the other informants reside. TYA and KEM lived in the same community and as such the researchers conducted parallel recording sessions. Parallel recording sessions were usually conducted with SHU and COL, and a joint session for RJU. This schedule was not always maintained as the sessions were flexible depending on the availability of the participants. On occasions, where informants were unavailable on the
scheduled recording day, another visit would be arranged where the recording session would be conducted. As some participants started school during the course of the research project, recordings were subsequently scheduled to take place on the weekends, and in some instances after-school.

Hand-held digital voice recorders were used as the main tool for data collection. Initial recordings were conducted with the recorder attached to the child bearing a mike, however this proved to be problematic as, not only was it a major distraction for the child, but the use of a single input mike allowed only for the audible processing of the child’s data and not the surrounding interlocutors. For subsequent recordings, the recorder’s built-in microphone was used, with the recording device strategically positioned or held by the researcher, so as to effectively capture the required data. Where necessary, notes were recorded after the sessions.

3.3. Transcription and Orthography

JC is mainly an oral language. Many of the lexemes are English-based but their phonology is quite different. We have adopted the JLU (Jamaica Language Unit) modified Cassidy-LePage orthography for all transcriptions.

All data was transcribed in CHAT (Codes for the Human Analysis of Transcripts) format, following the standard guidelines of the CHILDES Database. The transcription procedure proved to be very time consuming as it took approximately 10 hours to transcribe one hour of data. For transcribing the data, the recordings were transferred from the recording device to the computer. High quality earphones were used and the researcher transcribes the exact production as uttered. In some instances, due to surrounding noise including overlapping speech, recordings had to be listened repeatedly in order to ensure accurate transcription. Inaudible speech was transcribed as xxx. A backed-up copy of all recordings and transcriptions is stored on the University of Geneva Database for safe keeping.

3.4. Coding

Based on the time-frame for the completion of the research project, a decision was taken to initially code only the utterances of the child. The approximate time to code one transcription was 6 hours. This time however could not be fixed as it depended largely on the number and length of child utterances in the transcription. Tahirah Charles, Patrice Clarke, Shenell Ellis and Danielle Smith, all final-year students from the University of the West Indies, were employed to carry out the coding of the data, under the supervision of the main researcher.

Based on the non-standard conventions in transcribing JC, coding of the data had to be done manually. A list of codes was developed for conveying the morpho-syntactic relations of the data. Despite this comprehensive list, coding of the data did not prove to be unproblematic as there are some instances where a particular lexical item could lead to different interpretations or yield different codings in the same context. To deal with these occurrences, native speakers’ judgments were employed where applicable, or the word in question coded as unknown.

In dealing with issues regarding the coding of single word utterances, we adopted the method employed in comparative syntax, assuming on grounds of continuity and uniformity, that child language approximates adult grammar (in line with Bates et.al. 1994;
Gillette et al. 1999, among others). We acknowledge that this is not necessarily true in general but it is the necessary initial assumption for comparative research.

Several meetings were conducted with the coding team to ‘iron out’ all issues. Two recordings were coded per child for the period starting January 2012 to March 2013 and one coding done for each child in April 2013. The decision to start coding the data for analysis as of January 2012 was based on:

i. the maximal use of funds available
ii. the final confirmation date of all the research participants
iii. the initial two months involved familiarization of the participants with the researchers thereby maximizing their language production levels

All completed codings, amounting to a total of 186 files, were duly checked for verification of accuracy, and for inclusion of additional details as required for the analysis.

3.5. Analysis

The analysis of the production data was based mainly on age and developmental stages in line with Radford (1990). Nonetheless, where necessary the participants’ production was classified and compared by their MLU (see Brown 1973; Miller 1981; Miller and Chapman 1981).

Some utterances were excluded from the data analysis, these include:
- utterances in which any unintelligible portions (coded an UNK) could be critical for the analysis
- utterances where the meaning was unclear based on the context of the discourse
- the child’s stuttering or self-repetitions without the production of contentful utterances in-between
- repetitions of memorized materials, e.g. songs and nursery rhymes
- immediate repetitions of adult’s exact utterance

The analysis was based on automatic computing of the morpho-syntactic coding using CLAN, NotePad++ and other software where necessary. Nonetheless, manual analysis was inevitable for certain computations.

3.6. Limitations

Manual transcription and coding are extremely time consuming. Our initial choice was to code only the child speech; this made the project feasible within the assigned temporal and financial constraints, but also limited the possible comparison between child production and child-directed adult speech. It would be desirable to also code adult utterances in a future development of the project.

4. Summary and Outlook

This research project contributes to filling gaps that exist in acquisition research in the area of creole linguistics. The data on both target-consistencies and target-inconsistencies in the acquisition of JC offers a basis for studying the acquisition of JC against the background of the established results in comparative acquisition studies. This rich empirical data has in fact already given rise to the fine-tuning of existing theoretical analyses of
language acquisition, such as truncation theory (Rizzi 1993) (see De Lisser et al. 2015). By providing the only longitudinal corpus of Creole acquisition, we further facilitate future studies concerning grammatical development in Creoles. Finally, the impact goes beyond the realm of linguistics: by indicating, for the first time, syntactic milestones in the early speech of typically developing Creole children, our work should be useful for detecting atypical Creole development and thus lead to earlier implementation of language remediation. Indeed assessments will no longer need to be based on the acquisition of English, which is likely not to be the child’s target language, but rather on the language most frequently targeted, namely creole, a medium also now encouraged in the classroom since 2001 (as advised by the Ministry of Education). In addition, the project could help writers of books for children to be informed of what grammatical level a Creole-speaking child is expected to master at a certain age, which would in turn contribute to advancing literacy in Creole. In sum, findings from the current research project should be of interest to acquisitionists, creolists, teachers, speech and language therapists, developmental psychologists and writers of children’s books. We hope that our research will set the stage for more comparable research into the acquisition of other creole languages.
### Appendix: Participants’ MLU

<table>
<thead>
<tr>
<th>COL RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>1;8,17</td>
<td>160</td>
<td>256</td>
<td>1.6</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>1;8,27</td>
<td>181</td>
<td>325</td>
<td>1.796</td>
</tr>
<tr>
<td>16.02.2012</td>
<td>1;9,17</td>
<td>219</td>
<td>392</td>
<td>1.79</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>1;9,28</td>
<td>191</td>
<td>287</td>
<td>1.503</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>1;10,8</td>
<td>261</td>
<td>401</td>
<td>1.536</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>1;11,1</td>
<td>251</td>
<td>449</td>
<td>1.789</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>1;11,12</td>
<td>194</td>
<td>389</td>
<td>2.005</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>1;11,28</td>
<td>230</td>
<td>421</td>
<td>1.83</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;0,12</td>
<td>138</td>
<td>343</td>
<td>2.486</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;0,28</td>
<td>162</td>
<td>410</td>
<td>2.531</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;1,14</td>
<td>240</td>
<td>497</td>
<td>2.071</td>
</tr>
<tr>
<td>30.06.2012</td>
<td>2;2,0</td>
<td>127</td>
<td>279</td>
<td>2.197</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;2,16</td>
<td>354</td>
<td>813</td>
<td>2.297</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;3,1</td>
<td>196</td>
<td>503</td>
<td>2.566</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;3,16</td>
<td>215</td>
<td>615</td>
<td>2.86</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;3,30</td>
<td>239</td>
<td>678</td>
<td>2.837</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;4,15</td>
<td>238</td>
<td>679</td>
<td>2.853</td>
</tr>
<tr>
<td>30.09.2012</td>
<td>2;5,0</td>
<td>200</td>
<td>551</td>
<td>2.755</td>
</tr>
<tr>
<td>14.10.2012</td>
<td>2;5,14</td>
<td>193</td>
<td>606</td>
<td>3.14</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;5,27</td>
<td>401</td>
<td>1110</td>
<td>2.768</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;6,10</td>
<td>218</td>
<td>555</td>
<td>2.546</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>2;6,25</td>
<td>413</td>
<td>1407</td>
<td>3.407</td>
</tr>
<tr>
<td>08.12.2012</td>
<td>2;7,8</td>
<td>341</td>
<td>1190</td>
<td>3.49</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>2;7,22</td>
<td>266</td>
<td>866</td>
<td>3.256</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>2;8,6</td>
<td>339</td>
<td>1235</td>
<td>3.643</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>2;8,20</td>
<td>343</td>
<td>1429</td>
<td>4.166</td>
</tr>
<tr>
<td>10.02.2013</td>
<td>2;9,11</td>
<td>163</td>
<td>611</td>
<td>3.748</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>2;9,24</td>
<td>271</td>
<td>907</td>
<td>3.347</td>
</tr>
<tr>
<td>12.03.2013</td>
<td>2;10,10</td>
<td>275</td>
<td>1042</td>
<td>3.789</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>2;10,21</td>
<td>271</td>
<td>933</td>
<td>3.443</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>2;11,7</td>
<td>267</td>
<td>1019</td>
<td>3.816</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ALA RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>1;9,25</td>
<td>271</td>
<td>401</td>
<td>1.48</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>1;10,4</td>
<td>208</td>
<td>254</td>
<td>1.221</td>
</tr>
<tr>
<td>16.02.2012</td>
<td>1;10,25</td>
<td>385</td>
<td>634</td>
<td>1.647</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>1;11,5</td>
<td>349</td>
<td>551</td>
<td>1.579</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>1;11,16</td>
<td>341</td>
<td>632</td>
<td>1.853</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>2;0,9</td>
<td>381</td>
<td>856</td>
<td>2.247</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>2;0,20</td>
<td>254</td>
<td>458</td>
<td>1.803</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>2;1,5</td>
<td>395</td>
<td>794</td>
<td>2.01</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;1,20</td>
<td>353</td>
<td>852</td>
<td>2.414</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;2,6</td>
<td>377</td>
<td>1012</td>
<td>2.684</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;2,12</td>
<td>370</td>
<td>1140</td>
<td>3.081</td>
</tr>
<tr>
<td>30.06.2012</td>
<td>2;3,8</td>
<td>117</td>
<td>291</td>
<td>2.487</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;3,24</td>
<td>397</td>
<td>1266</td>
<td>3.189</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;4,9</td>
<td>283</td>
<td>933</td>
<td>3.297</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;4,24</td>
<td>388</td>
<td>1237</td>
<td>3.188</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;5,7</td>
<td>294</td>
<td>1025</td>
<td>3.486</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;5,23</td>
<td>493</td>
<td>2111</td>
<td>4.282</td>
</tr>
<tr>
<td>04.10.2012</td>
<td>2;6,12</td>
<td>193</td>
<td>623</td>
<td>3.228</td>
</tr>
<tr>
<td>14.10.2012</td>
<td>2;6,22</td>
<td>217</td>
<td>937</td>
<td>4.128</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;7,5</td>
<td>336</td>
<td>1382</td>
<td>4.113</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;7,18</td>
<td>361</td>
<td>1448</td>
<td>4.011</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>2;8,2</td>
<td>467</td>
<td>2547</td>
<td>5.454</td>
</tr>
<tr>
<td>08.12.2012</td>
<td>2;8,16</td>
<td>313</td>
<td>1729</td>
<td>5.524</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>2;9,0</td>
<td>294</td>
<td>1251</td>
<td>4.255</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>2;9,14</td>
<td>285</td>
<td>1336</td>
<td>4.688</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>2;9,28</td>
<td>249</td>
<td>1259</td>
<td>5.056</td>
</tr>
<tr>
<td>16.02.2013</td>
<td>2;10,25</td>
<td>297</td>
<td>1351</td>
<td>4.549</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>2;11,1</td>
<td>246</td>
<td>1078</td>
<td>4.382</td>
</tr>
<tr>
<td>12.03.2013</td>
<td>2;11,18</td>
<td>264</td>
<td>1292</td>
<td>4.894</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>3;0,1</td>
<td>277</td>
<td>1353</td>
<td>4.884</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>3;0,15</td>
<td>341</td>
<td>1930</td>
<td>5.66</td>
</tr>
</tbody>
</table>

Table 1: COL’s MLU  
Table 2: ALA’s MLU
### Table 3: RJU’s MLU

<table>
<thead>
<tr>
<th>RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>1;10,4</td>
<td>51</td>
<td>71</td>
<td>1.392</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>1;10,14</td>
<td>80</td>
<td>137</td>
<td>1.712</td>
</tr>
<tr>
<td>16.02.2012</td>
<td>1;11,4</td>
<td>198</td>
<td>440</td>
<td>2.222</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>1;11,15</td>
<td>39</td>
<td>65</td>
<td>1.667</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>1;11,26</td>
<td>66</td>
<td>105</td>
<td>1.591</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>2;0,19</td>
<td>135</td>
<td>168</td>
<td>1.985</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>2;0,30</td>
<td>147</td>
<td>332</td>
<td>2.259</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>2;1,15</td>
<td>207</td>
<td>692</td>
<td>3.343</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;2,0</td>
<td>202</td>
<td>528</td>
<td>2.614</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;2,16</td>
<td>173</td>
<td>361</td>
<td>2.087</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;3,1</td>
<td>315</td>
<td>950</td>
<td>3.016</td>
</tr>
<tr>
<td>30.06.2012</td>
<td>2;3,18</td>
<td>206</td>
<td>604</td>
<td>2.932</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;4,4</td>
<td>207</td>
<td>605</td>
<td>2.923</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;4,19</td>
<td>212</td>
<td>674</td>
<td>3.179</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;5,3</td>
<td>136</td>
<td>435</td>
<td>3.199</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;5,17</td>
<td>220</td>
<td>570</td>
<td>2.591</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;6,2</td>
<td>206</td>
<td>720</td>
<td>3.495</td>
</tr>
<tr>
<td>30.09.2012</td>
<td>2;6,18</td>
<td>268</td>
<td>762</td>
<td>2.843</td>
</tr>
<tr>
<td>17.10.2012</td>
<td>2;7,5</td>
<td>249</td>
<td>1071</td>
<td>4.301</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;7,15</td>
<td>257</td>
<td>1031</td>
<td>4.012</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;7,28</td>
<td>209</td>
<td>919</td>
<td>4.397</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>2;8,12</td>
<td>276</td>
<td>1040</td>
<td>3.768</td>
</tr>
<tr>
<td>08.12.2012</td>
<td>2;8,26</td>
<td>233</td>
<td>859</td>
<td>3.687</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>2;9,10</td>
<td>236</td>
<td>773</td>
<td>3.275</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>2;9,24</td>
<td>226</td>
<td>780</td>
<td>3.451</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>2;10,7</td>
<td>187</td>
<td>730</td>
<td>3.904</td>
</tr>
<tr>
<td>10.02.2013</td>
<td>2;10,29</td>
<td>142</td>
<td>452</td>
<td>3.183</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>2;11,11</td>
<td>173</td>
<td>626</td>
<td>3.618</td>
</tr>
<tr>
<td>09.03.2013</td>
<td>2;11,25</td>
<td>328</td>
<td>1662</td>
<td>5.067</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>3;0,11</td>
<td>206</td>
<td>1079</td>
<td>5.238</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>3;0,25</td>
<td>277</td>
<td>1346</td>
<td>4.859</td>
</tr>
</tbody>
</table>

### Table 4: TYA’s MLU

<table>
<thead>
<tr>
<th>RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>1;11,25</td>
<td>126</td>
<td>154</td>
<td>1.222</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>2;0,4</td>
<td>130</td>
<td>157</td>
<td>1.208</td>
</tr>
<tr>
<td>16.02.2012</td>
<td>2;0,25</td>
<td>113</td>
<td>153</td>
<td>1.354</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>2;1,5</td>
<td>257</td>
<td>324</td>
<td>1.261</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>2;1,16</td>
<td>108</td>
<td>152</td>
<td>1.407</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>2;2,9</td>
<td>121</td>
<td>162</td>
<td>1.339</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>2;2,20</td>
<td>174</td>
<td>246</td>
<td>1.414</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>2;3,5</td>
<td>101</td>
<td>149</td>
<td>1.475</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;3,20</td>
<td>218</td>
<td>334</td>
<td>1.532</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;4,6</td>
<td>162</td>
<td>235</td>
<td>1.451</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;4,22</td>
<td>128</td>
<td>169</td>
<td>1.32</td>
</tr>
<tr>
<td>30.06.2012</td>
<td>2;5,8</td>
<td>62</td>
<td>99</td>
<td>1.597</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;5,24</td>
<td>103</td>
<td>152</td>
<td>1.476</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;6,9</td>
<td>58</td>
<td>111</td>
<td>1.914</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;6,24</td>
<td>205</td>
<td>411</td>
<td>2.005</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;7,7</td>
<td>22</td>
<td>52</td>
<td>2.364</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;7,23</td>
<td>240</td>
<td>467</td>
<td>1.946</td>
</tr>
<tr>
<td>30.09.2012</td>
<td>2;8,8</td>
<td>302</td>
<td>926</td>
<td>3.066</td>
</tr>
<tr>
<td>14.10.2012</td>
<td>2;8,22</td>
<td>143</td>
<td>401</td>
<td>2.804</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;9,5</td>
<td>233</td>
<td>696</td>
<td>2.987</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;9,18</td>
<td>53</td>
<td>99</td>
<td>1.868</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>2;10,2</td>
<td>260</td>
<td>635</td>
<td>2.442</td>
</tr>
<tr>
<td>08.12.2012</td>
<td>2;10,16</td>
<td>178</td>
<td>385</td>
<td>2.163</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>2;11,0</td>
<td>137</td>
<td>510</td>
<td>3.723</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>2;11,14</td>
<td>271</td>
<td>914</td>
<td>3.373</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>2;11,28</td>
<td>261</td>
<td>1168</td>
<td>4.475</td>
</tr>
<tr>
<td>10.02.2013</td>
<td>3;0,19</td>
<td>42</td>
<td>194</td>
<td>4.619</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>3;1,1</td>
<td>268</td>
<td>1442</td>
<td>5.381</td>
</tr>
<tr>
<td>09.03.2013</td>
<td>3;1,15</td>
<td>325</td>
<td>1722</td>
<td>5.298</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>3;2,1</td>
<td>177</td>
<td>923</td>
<td>5.215</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>3;2,15</td>
<td>175</td>
<td>851</td>
<td>4.863</td>
</tr>
</tbody>
</table>

The Acquisition of Jamaican Creole: A Research Project
Table 5: KEM's MLU

<table>
<thead>
<tr>
<th>RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>2;1,21</td>
<td>280</td>
<td>589</td>
<td>2.104</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>2;1,0</td>
<td>348</td>
<td>654</td>
<td>1.879</td>
</tr>
<tr>
<td>16.02.2012</td>
<td>2;1,21</td>
<td>309</td>
<td>552</td>
<td>1.786</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>2;2,1</td>
<td>297</td>
<td>574</td>
<td>1.933</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>2;2,12</td>
<td>423</td>
<td>821</td>
<td>1.941</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>2;3,5</td>
<td>305</td>
<td>600</td>
<td>1.967</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>2;3,16</td>
<td>347</td>
<td>725</td>
<td>2.089</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>2;4,1</td>
<td>392</td>
<td>933</td>
<td>2.38</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;4,16</td>
<td>337</td>
<td>767</td>
<td>2.276</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;5,2</td>
<td>279</td>
<td>753</td>
<td>2.699</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;5,18</td>
<td>386</td>
<td>929</td>
<td>2.407</td>
</tr>
<tr>
<td>30.06.2012</td>
<td>2;6,4</td>
<td>303</td>
<td>882</td>
<td>2.911</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;6,20</td>
<td>376</td>
<td>1148</td>
<td>3.053</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;7,5</td>
<td>373</td>
<td>1513</td>
<td>4.056</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;7,20</td>
<td>311</td>
<td>1322</td>
<td>4.251</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;8,3</td>
<td>259</td>
<td>1103</td>
<td>4.275</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;8,19</td>
<td>276</td>
<td>1287</td>
<td>4.663</td>
</tr>
<tr>
<td>30.09.2012</td>
<td>2;9,4</td>
<td>341</td>
<td>1607</td>
<td>4.713</td>
</tr>
<tr>
<td>14.10.2012</td>
<td>2;9,18</td>
<td>261</td>
<td>1170</td>
<td>4.483</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;10,1</td>
<td>159</td>
<td>683</td>
<td>4.296</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;10,14</td>
<td>254</td>
<td>1141</td>
<td>4.492</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>2;10,29</td>
<td>404</td>
<td>1884</td>
<td>4.663</td>
</tr>
<tr>
<td>08.12.2012</td>
<td>2;11,12</td>
<td>265</td>
<td>1160</td>
<td>4.377</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>2;11,26</td>
<td>355</td>
<td>2203</td>
<td>6.206</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>3;0,10</td>
<td>285</td>
<td>1606</td>
<td>5.635</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>3;0,24</td>
<td>333</td>
<td>2151</td>
<td>6.459</td>
</tr>
<tr>
<td>10.02.2013</td>
<td>3;1,15</td>
<td>261</td>
<td>1640</td>
<td>6.284</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>3;1,28</td>
<td>339</td>
<td>1937</td>
<td>5.714</td>
</tr>
<tr>
<td>09.03.2013</td>
<td>3;2,11</td>
<td>384</td>
<td>2397</td>
<td>6.242</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>3;2,25</td>
<td>384</td>
<td>2278</td>
<td>5.932</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>3;3,11</td>
<td>295</td>
<td>1614</td>
<td>5.471</td>
</tr>
</tbody>
</table>

Table 6: SHU's MLU

<table>
<thead>
<tr>
<th>RECDATE</th>
<th>AGE(Y;M,D)</th>
<th>UTT</th>
<th>WORDS</th>
<th>MLU</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.01.2012</td>
<td>2;1,23</td>
<td>177</td>
<td>511</td>
<td>2.887</td>
</tr>
<tr>
<td>26.01.2012</td>
<td>2;2,2</td>
<td>233</td>
<td>705</td>
<td>3.026</td>
</tr>
<tr>
<td>04.02.2012</td>
<td>2;2,11</td>
<td>191</td>
<td>517</td>
<td>2.707</td>
</tr>
<tr>
<td>27.02.2012</td>
<td>2;3,3</td>
<td>307</td>
<td>1059</td>
<td>3.45</td>
</tr>
<tr>
<td>09.03.2012</td>
<td>2;3,14</td>
<td>287</td>
<td>1147</td>
<td>3.997</td>
</tr>
<tr>
<td>31.03.2012</td>
<td>2;4,7</td>
<td>91</td>
<td>287</td>
<td>3.154</td>
</tr>
<tr>
<td>11.04.2012</td>
<td>2;4,18</td>
<td>133</td>
<td>368</td>
<td>2.767</td>
</tr>
<tr>
<td>27.04.2012</td>
<td>2;5,3</td>
<td>175</td>
<td>604</td>
<td>3.451</td>
</tr>
<tr>
<td>12.05.2012</td>
<td>2;5,18</td>
<td>372</td>
<td>1084</td>
<td>2.914</td>
</tr>
<tr>
<td>28.05.2012</td>
<td>2;6,4</td>
<td>239</td>
<td>893</td>
<td>3.736</td>
</tr>
<tr>
<td>13.06.2012</td>
<td>2;6,20</td>
<td>234</td>
<td>755</td>
<td>3.226</td>
</tr>
<tr>
<td>10.07.2012</td>
<td>2;7,16</td>
<td>332</td>
<td>1157</td>
<td>3.485</td>
</tr>
<tr>
<td>16.07.2012</td>
<td>2;7,22</td>
<td>244</td>
<td>894</td>
<td>3.664</td>
</tr>
<tr>
<td>31.07.2012</td>
<td>2;8,7</td>
<td>244</td>
<td>1080</td>
<td>4.426</td>
</tr>
<tr>
<td>15.08.2012</td>
<td>2;8,22</td>
<td>171</td>
<td>663</td>
<td>3.877</td>
</tr>
<tr>
<td>29.08.2012</td>
<td>2;9,5</td>
<td>281</td>
<td>921</td>
<td>3.278</td>
</tr>
<tr>
<td>14.09.2012</td>
<td>2;9,21</td>
<td>333</td>
<td>1387</td>
<td>4.165</td>
</tr>
<tr>
<td>30.09.2012</td>
<td>2;10,6</td>
<td>195</td>
<td>695</td>
<td>3.564</td>
</tr>
<tr>
<td>14.10.2012</td>
<td>2;10,20</td>
<td>58</td>
<td>198</td>
<td>3.414</td>
</tr>
<tr>
<td>27.10.2012</td>
<td>2;11,3</td>
<td>284</td>
<td>1001</td>
<td>3.525</td>
</tr>
<tr>
<td>09.11.2012</td>
<td>2;11,16</td>
<td>156</td>
<td>524</td>
<td>3.359</td>
</tr>
<tr>
<td>24.11.2012</td>
<td>3;0,0</td>
<td>299</td>
<td>1340</td>
<td>4.482</td>
</tr>
<tr>
<td>13.12.2012</td>
<td>3;0,19</td>
<td>283</td>
<td>1440</td>
<td>5.088</td>
</tr>
<tr>
<td>22.12.2012</td>
<td>3;0,28</td>
<td>375</td>
<td>1900</td>
<td>5.067</td>
</tr>
<tr>
<td>05.01.2013</td>
<td>3;1,12</td>
<td>285</td>
<td>1366</td>
<td>4.799</td>
</tr>
<tr>
<td>19.01.2013</td>
<td>3;1,26</td>
<td>509</td>
<td>2883</td>
<td>5.664</td>
</tr>
<tr>
<td>10.02.2013</td>
<td>3;2,17</td>
<td>193</td>
<td>942</td>
<td>4.881</td>
</tr>
<tr>
<td>23.02.2013</td>
<td>3;2,30</td>
<td>226</td>
<td>1122</td>
<td>4.965</td>
</tr>
<tr>
<td>09.03.2013</td>
<td>3;3,13</td>
<td>191</td>
<td>623</td>
<td>3.262</td>
</tr>
<tr>
<td>23.03.2013</td>
<td>3;3,27</td>
<td>306</td>
<td>1626</td>
<td>5.314</td>
</tr>
<tr>
<td>06.04.2013</td>
<td>3;4,13</td>
<td>331</td>
<td>1661</td>
<td>5.018</td>
</tr>
</tbody>
</table>

References


The Acquisition of Jamaican Creole: A Research Project


Mufwene, Salikoko. 2000. “Creolization is a social, not a structural, process”. In *Degrees of restructuring in creole languages*, ed. by Ingrid Neumann-Holzchuh, and Edgar Schneider, 65-84. Amsterdam: John Benjamins.


GENUINE VERSUS SUPERFICIAL RELATIVES
IN FRENCH: THE DEPTH OF EMBEDDING FACTOR

Cornelia Hamann
Carl von Ossietzky Universität Oldenburg

Laurice Tuller
Université François-Rabelais Tours

Abstract: This article takes up economy/complexity of linguistic derivations, as classically illustrated in relative clauses. Relative clauses in French are of particular interest in this regard in that the syntax of surface relative clauses in fact varies in complexity, some of them corresponding to real relative clauses and others involving a flatter structure. We propose a syntactic analysis of superficial relatives, taken to include various kinds of cleft and presentational constructions such as c’est and il y a, arguing that they are “flatter” than genuine relatives in structurally definable ways. In particular, we demonstrate that these structures do not involve genuine recursion, of either DP or of a full CP. We argue that study of spontaneous language samples of typically developing (TD) children and adolescents and children and adolescents with Specific Language Impairment (SLI) provides crucial evidence for the fundamental nature of this key aspect of complexity, depth of embedding, especially when related factors such as intervention in dependencies are controlled for. The acquisition of relative clauses in French was examined in the following groups: 12 TD 6-year-olds, 12 TD 8-year-olds, 12 TD 11-year-olds, 12 TD 14-year-olds) and 16 10- to 12-year-olds with SLI.

Keywords: relative clauses, embedding, French Clefts

1. Complexity Effects in Relative Clauses

The notion of economy/complexity of linguistic derivation has recently come to the forefront of work on language acquisition and impairment. Inherent in this notion is the idea that elements and operations manipulated in linguistic computation have an impact on the efficiency of linguistic performance and provide an explanation for the order of emergence and frequency of occurrence of different constructions in child language. Simplistically put, the idea is that an increase in the number of elements manipulated and/or the number of manipulations is linked to a decrease in linguistic performance. Different ways of instantiating this basic intuition have been proposed, notably within generative approaches which, in recent formulations, permit close parallels between syntactic derivation, processing, and limitations in working memory. These instantiations have targeted various aspects of syntactic computation which appear to be relevant. This article will focus on one of these—depth of embedding.

RGG (2014) 36: 47-82
1.1. Theoretical and empirical background

One classical locus of complexity effects is relative clauses (RC). These show that, in the case of center embedding, the factor of depth of embedding further exacerbates processing difficulty even in adults, as is illustrated in (1b), which is exceedingly difficult to interpret, compared to (1a).

(1) a. The administrator who the intern supervised lost the medical reports.
   b. #The administrator who the intern who the nurse supervised had bothered lost the medical reports. (adapted from Gibson 1998)

Kimball (1973) proposed that complexity be calculated in terms of the number of incomplete clauses that must be stored in memory at any given moment in parsing, a characterization which would seem to capture well the intuition behind the idea that depth of embedding affects processing. This might be translated in phase theory (Chomsky 2005), as the activation of phase edges which are needed to compute cross-phase dependencies in the course of a derivation. Related to the notion of depth of embedding is the fact that RCs involve recursion, a property which is particularly important here, and to which we will return.

In addition to the factor of depth of embedding, RCs also display a dissociation between subject- and object-extracted relatives (SR, OR), as in (2a) – (2b), with the latter giving rise to longer reaction times, more comprehension errors, etc.

(2) a. The cook that helped the plumber quit work after a month
   b. The cook that the plumber helped quit work after a month (Gordon et al. 2001)

The factor illustrated in (2b) has received considerable recent attention. What is at issue here is the notion of distance and how this is computed. In the processing literature, this is taken to be due to “interveners”. Gibson (1998, 2000) advanced the idea that dependencies which contain intervening new referents make computation difficult, and Gordon et al. (2001) referring to Bever (1970) and Crowder (1976) emphasized that it is the similarity of the intervener to the antecedent which is at the base of this intervention effect. A structural counterpart of these insights appeared in Rizzi (1990, 2004, 2013) as a particular instance of the syntactic locality constraints unified by Relativized Minimality (RM). 1

---

1 The basic idea is that in a configuration like X…Z…Y, “a local structural relation cannot hold between X and Y if Z is a potential bearer of the relevant relation and Z intervenes between X and Y” (Rizzi 2004). Potential bearers Z of the relevant relation are characterized as being of the same structural type as X. Defining constituents, or more precisely specifiers, as being of the same structural type, Rizzi proposes four such types with specific features: argumental specifiers (person, number, gender, case); quantificational specifiers (Wh, Neg, measure, focus…); modifiers (evaluative, epistemic, Neg, frequentive…manner…); topics. Starke (2001) observed that potential interveners which only contain a subset of the features of the antecedent, do not act as interveners, so that a configuration such as (i) +A+B…..A ….<&A+B> (where A and B are sets of features) does not block the relevant structural relation. This latter configuration has been exploited in the explanation of difficulties in the acquisition of relative clauses in recent discussion (Friedmann et al. 2009). More recently Beletti et al. (2012) highlighted the crucial impact on intervention of similarity of linguistic features attracting movement as opposed to features involved in cognitive computation which have an impact on working memory or theta-role assignment. Note also that the feature of animacy has received much attention in attempts to determine its role in intervention effects (Mak et al. 2002; Kidd et al. 2007; Grillo 2008; Arosio et al. 2011) and that the purely formal characterization of the intervention effect as “NP-restriction” has been critically evaluated by Goodluck (2010).
The SR-OR dissociation is well known in the acquisition literature. Friedmann et al. (2009), taking up the particular sensitivity children show for this dissociation, start from the idea that RM seems to operate in a stricter fashion in children than in adults, much as in Grillo’s (2008) work on adults with aphasia. In other words, the difficulties children experience with ORs are due to intervention of the subject. In adult grammar the DP subject does not interfere in the strict sense of blocking the derivation since the relevant feature shared by the relative operator and its trace is the wh-feature of the relative operator, which is not a feature of the subject. Gordon et al. (2001) showed, however, that subjects, and full DP subjects in particular, have an effect on processing cost, making reading times and interpretative errors increase in typical adults. Investigating Italian and Hebrew children’s production and comprehension of SRs and ORs, Friedmann et al. argued that it is a grammatical feature, the NP-restriction on the DPs, that makes it difficult for adults to relate the gap to its antecedent and, in child grammar, blocks this relation altogether, as illustrated in (3).

(3) a. Show me the cow that the cow is kissing the chicken [not problematic]
b. Show me the elephant that the lion is wetting the elephant [problematic or blocked]

Apart from the purely structural factors of depth of embedding and interveners between filler and gap, information structure also plays a role in performance on relatives: Topichood in particular, has been shown to greatly reduce the asymmetry in the production (Fox and Thompson 1990; Zubin 1979) and comprehension of subject and object extracted relatives (Frauenfelder et al. 1980, for early findings, and Mak et al., 2002, 2006, 2008 for recent discussion). It is likewise remarkable that, as will be shown below, some of the less deeply embedded structures employed in French relativization are tied to specific discourse functions such as introduction of new discourse topics (Jisa & Kern 1998; Lambrecht 1988).

1.2. Levels of embedding of relative clauses

As we have reviewed above, multiple embedding in center-embedded relatives is a source of difficulty even for typical adults (see (1) above). Hamann et al. (2007), Delage et al. (2008), and Tuller et al. (2012) suggest that depth of embedding is an important variable to consider in defining the complexity of subordinate clauses in general, and RCs in particular. The hypothesis that embedding is part of the definition of syntactic complexity is also developed and supported by Soares (2006) on the basis of data from child Portuguese which shows in particular that clauses without embedding emerge before clauses with embedding. Intuitively, an RC in a DP that is a constituent of the matrix clause is less complex than an RC in a DP that is a constituent of another subordinate clause, in that the former (which we will refer to as a Level-1 RC, as in (4a)) is less deeply embedded than the latter (which would be of Level-2 or higher, as in (4b)). And, of course, if the subordinate clause is itself an RC, as in (4c), then this is relative recursion.

(4) a. Lea knows the man [CP who lives in Paris ] Level-1 RC
b. Lea knows [CP that we saw the man [CP who lives in Paris ]] Level-2 RC
c. Lea knows the man [CP who lives in the city [CP that is the capital of France ]] Level-2 RC
As mentioned above, in Kimball’s (1973) calculation of complexity, it is the number of incomplete clauses that must be stored in memory. If clauses are the important notion, it follows that an RC in a DP which forms an utterance by itself, and therefore is not a constituent of a matrix clause, will be simpler than a Level-1 RC. Such an RC (which we refer to as a 0-level RC) is, in a certain sense mono-clausal, in that it is not embedded within another clause, though it does display DP recursion (see Goodluck & Tavakolian 1982). DP utterances of this type (as in (5)) are fully discourse-felicitous, for instance, as answers to questions.

(5) (Who’s that?) The guy [CP who lives in Paris] 0-Level RC

There are other syntactic configurations in which embedding of an RC might also be considered to be shallower, following this simple definition crucially involving embedding of an RC inside a clause. One such case might be that of a left- or right-dislocated DP with an RC, since in these cases, the DP is not embedded within the matrix clause, but rather adjoined to it (see Kiparsky 1994 with similar criteria for the historical development of subordinate clauses in English from adjunction to embedding). In this sense, an RC in a dislocated DP would also be a 0-level RC.

(6) [[DP That guy [CP who lives in Paris,]] [IP he never comes to Tours anymore]]

We will be arguing in this paper, that, in French, depth of embedding can be observed by applying the above scale (0-level, level-1, level-2, etc.), but also in a number of superficial relative constructions entailing a flatter structure in that either there is no embedding in a DP or no embedding in a clause, i.e. an IP or CP, or both.

1.3. Acquisition of relative clauses

The aforementioned factors entering into the complexity of RCs – depth of embedding, intervention and discourse – have all been addressed in the psycholinguistic literature, but usually from the perspective of comprehension and/or processing. Focusing on child language acquisition, it is generally claimed that production of RCs starts as early as three years of age (Crain, McKee and Emiliani 1990; de Villiers et al. 1994; De Cat 2002 and others), or even earlier, as reported in Håkansson and Hansson (2000), for Swedish, but that ORs are difficult for typical children (see Hamburger and Crain 1979, for elicited production). This observation raised the question of whether it was children’s grammar that did not allow ORs or whether it was a performance problem. This general question led to experimental exploration of a number of factors which might be pertinent in explaining order of emergence and frequency in spontaneous language, as we have seen. Goodluck and Tavakolian (1982) argued that the grammar of 4-year-old children includes true relatives, i.e. embedding under DPs. They showed that the tendency to choose the wrong antecedent in comprehension experiments interacts with the animacy of constituents within the relative clause, but that no such interaction occurred if the embedded clause was an infinitival complement. Since the characteristic of RCs is their recursiveness (DPs within DPs), Goodluck and Tavakolian (1982, 2) concluded that children’s errors arise when complexity (recursion) “overloads the language processor”. In the spirit of this line of enquiry, Corrêa (1995) investigated the comprehension of several types of RCs, corroborating the asymmetry in comprehension of center-embedded to right-branching relatives as well as the
subject-object asymmetry. Kidd et al. (2007) pointed to a strong influence of animacy, specifically of the relative head, in both the frequency of ORs in spontaneous discourse and in the successful repetition of RC sentences, in English and German children. Citing work by Diessel and Tommasello (2005), they also emphasize the possible importance of the fact that early relatives found in spontaneous language of English-speaking children are predominantly found in presentational constructions, illustrated in (7).

(7) This is the sugar that goes in there.

This observation had already been made by Labelle (1990, 1996) for French children on the basis of elicited production data. She proposed that these constructions are in fact not genuine relatives, arguing that they do not involve movement, and claiming that early relatives in French children’s productions lack movement. This approach neatly fits into the complexity metric based on movement as proposed by Jakubowicz (2004). However, it remains highly controversial whether early French relatives involve movement or not (see Guasti & Shlonsky 1995; Guasti & Cardinaletti 2003).

Questions about complexity of structure, availability of movement, and, additionally, theta-role assignment in non-canonical structures have likewise been pursued in research on atypical development (notably, children with Specific Language Impairment, and children with hearing loss). The results complement and corroborate results about typical development, and often allow the isolation of particular complexity factors (for work on relatives, see Delage 2008; Friedmann and Novogrodsky 2004; Friedmann and Szterman 2006; Håkansson and Hansson 2000; Novogrodsky and Friedmann 2006; Stavrakaki 2001; Tuller et al. 2012).

We will argue, on the basis of our analysis of RCs in French, that movement is not the only relevant factor for earliness or frequency of occurrence. Other factors like depth of embedding and the nature of the interveners seem to be more likely candidates for increasing complexity and thus processing load. In other words, we will in fact be pursuing the line of investigation initiated by Goodluck and Tavakolian (1982), in that DP recursion, which they highlight, is fundamentally about depth of embedding. Summarizing, the important factors highlighted in studies of typical and atypical child acquisition of RCs correspond to the components of syntactic complexity described here: depth of embedding, movement/intervention, and information structure. These studies share the following premise: complexity of linguistic computation provides an explanation for the order of emergence and the frequency of production of different syntactic constructions in children.

1.4. Aim and structure of the article

The aim of this article is to contribute to theoretical discussion on syntactic complexity. In particular, we will propose a unified analysis of superficial RCs in French, which argues, on syntactic grounds, that these have a comparatively flat structure compared to genuine RCs, which show classical recursion. In the course of this demonstration, we will also make explicit what the syntactic properties are that render these structures flatter. The specific syntactic analysis we propose takes into account associated discourse effects, in the spirit of Belletti (2008). We will put forward empirical support for the idea that depth of embedding

---

2 Jakubowicz’s Derivational Complexity Metric (DCM) (Jakubowicz 2004, 2005) is given in (i):

(i) a. Merging $\alpha$ n times gives rise to a less complex derivation than merging $\alpha$, (n + 1) times.
   b. Merge of $\alpha$ gives rise to a less complex derivation than Merge of $\alpha + \beta$. 
is a decisive factor in complexity in the development of French RCs, both genuine and “flat,” in typically developing children of various ages and children with specific language impairment. We will follow here the strategy of using child language acquisition, typical and atypical, as a potential source of evidence in support for characterization of syntactic complexity, with atypical acquisition providing sharper insight into these questions due to how highly affected by this property these children are, even at advanced ages. In line with considerable work in the field, we will take appropriately calculated frequency of occurrence and order of acquisition in spontaneous language production to be manifestations of the effects of syntactic complexity.

Frequency and avoidance in spontaneous speech are counterparts of each other in the sense that infrequent use of some constructions is interpreted to be a reflection of these constructions being avoided, causing other constructions to be more frequent. This view of economy would seem to involve a “look-ahead” problem. This potential problem is not specifically related to production. We assume here that just as the parser develops preferences in comprehension (for example, attachment preferences, as proposed in Grillo 2012), in production, strategies develop for ease of processing, leading to avoidance and more frequent production of preferred structures.3

Summarizing, our premise is that syntactic complexity explains order of emergence and frequency of occurrence in language acquisition. Our hypothesis is that depth of embedding is an essential ingredient in the calculation of syntactic complexity. We offer an analysis – on purely syntactic and semantic grounds – of various superficial relative constructions in French which distinguishes them in terms of depth of embedding. We then present evidence from child language which shows that frequency, order of acquisition and avoidance/preference are sensitive to depth of embedding. Our conclusion is that these data provide support for the syntactic analysis we propose and for the hypothesis that depth of embedding is a relevant factor in linguistic complexity. It will furthermore emerge that the type of methodology used offers a relevant complement to experimental evidence, in that information structure effects are naturally controlled for.

Section 2.1 reviews the syntactic and discursive properties of French RCs, introducing the notion that these can be divided into genuine and superficial relatives. Section 2.2 reviews analyses which have been proposed to account for the syntactic and discourse properties of clefts and presentational constructions (contrastive clefts, avoir ‘have’-clefts, il y a ‘there is/are’ presentational), which will collectively be referred to as superficial relatives, in that they superficially resemble RCs. In Section 2.3, we propose to unify these analyses in a way that brings out their fundamental hierarchical similarity. Section 2.4 summarizes the findings on the acquisition of genuine and superficial relatives in French. Supporting evidence for this analysis of superficial relatives, which we claim to be flatter and thus less complex than genuine relatives, is put forward in section 3 and comes from investigation of spontaneous productions of typically developing French children and French children with SLI. Finally, section 4 discusses the findings and interprets them in light of our syntactic analysis, and the hypothesis that depth of embedding plays a crucial role in the determination of linguistic complexity.

3 Evidence for such development of preferences of this type can be seen in a trend observed in children and adolescents with SLI. Hamann et al. (2007) found that adolescents with SLI do not produce embedded clauses more frequently than children with SLI (contrary to what is observed in TD children and adolescents), yet they do produce over-all fewer errors. See also Tuller et al. (2012).
2. The Syntax of Superficial Relatives in French

2.1. Types of relative clauses

RCs in French display the typical operator-gap dependency, illustrated in (8a) with an adjunct relative (AR). Subject (8b) and object (8c) relatives are morphologically distinguished via the form of the (obligatory) relative pronoun, which is classically analyzed (Kayne 1975; Pesetsky 1982; Rizzi 1982) as SPEC-head agreement of the complementizer *que* ‘that’ which is realized as *qui* in the case of subject extraction:

(8)  a. le café où j'ai vu Max

the café where I have seen Max
‘The café where I saw Max’

b. la femme qui connaît Max

the woman QUI knows Max
‘The woman who knows Max’

c. la femme que Max connaît

the woman QUE Max knows
‘The woman that Max knows’

French also has a variety of what are variously termed “pseudo-relatives,” “apparent subordinate relatives,” “superficial relatives,” etc. On the one hand, these constructions resemble RCs in that they contain a clause headed by a relative pronoun linked to a gap which refers to a DP head to the immediate left of the clause. On the other hand, there are compelling reasons for not assimilating these constructions to RCs, in terms of their interpretation, their syntax, and even their prosody. Among these are various cleft constructions (introduced by *c'est* ‘it/that’s’), illustrated in (9a-b), various presentational constructions (introduced by *c'est, (il) y a* ‘there is/are’, and *avoir* ‘have’), illustrated in (9c-e), as well as complements of perception verbs.

(9)  a. C'est Max que Léa a vu (pas Pierre) [contrastive cleft]

that’s Max that Lea has seen not Lea 
‘Lea saw Max (not Pierre)’

b. C'est Max qui est venu [presentational cleft—focus on DP]

that’s Max that is come
‘Max came’

c. C'est le petit qui est tombé dans l’escalier [presentational cleft—focus on event]

that’s the little that is fallen in the stair
‘The kid fell down the stairs’

d. Y a Jean qu’ a téléphoné [presentational (il) y a]

there has Jean that has phoned
‘John called’

e. Elle a son père qui est malade [presentational avoir]

she has her father who is sick

4 In recent generative literature the term “pseudo-relatives” is reserved for structures involving perception verbs (*J’ai vu Marie qui courrait* – I saw Marie who ran – I saw Marie run – I saw Marie running), whereas it is used more liberally by other authors. In order to avoid confusion with the more specific use and contrary to the use in earlier versions of this article, we opt for the term “superficial relatives.”
‘Her father’s sick’

These structures in French have been studied in different grammatical traditions, which have revealed pertinent discourse and syntactic properties. We propose to first of all review these properties here, but also to present data on them from acquisition studies. We will then take up the question of explicit syntactic analyses attempting to account for these properties in the following section (Section 3).

Lambrecht (1988), pursuing observations made by Blanche-Benveniste (1983) on “pseudo-subordinates” (and notably on “apparent” or “false” relatives) in spoken French, presents arguments that superficial relatives introduced by the verb *avoir* ‘have’ entail neither the syntax nor the semantics of restrictive RCs. This construction, which he refers to as the “presentational cleft” or “avoir-cleft” construction consists of superficial relatives introduced by the existential (*il y a*, as in (9d) and (10a), and those introduced by a ‘possessive’ *avoir*, as in (9e) and (10b).

(10) a. Y a le téléphone qui sonne !
    there-has the phone that rings
    ‘The phone’s ringing’

   b. J’ai les yeux qui m’ont mal.
    I-have the eyes that me-DAT-do bad
    ‘My eyes hurt’

   (Lambrecht 1988, 136-7)

Noting that corresponding sentences with a monoclausal SVO order (*Jean a téléphoné*, *Le téléphone sonne*, *Les yeux m’ont mal*) are not felicitous in most discourse contexts in spoken French, because of a constraint entailing that subjects cannot be focused, 5 Lambrecht shows that *avoir*-clefts are used either to introduce new discourse referents or for ‘event-reporting’, both of which involve presentational focus, of a DP, as in (9d) or of an event, as in (10a, b). The superficial RC introduced by *avoir* is argued to differ from restrictive relatives in several ways. It does not function as a modifier of the preceding DP, as is evidenced by the fact that this ‘antecedent’ can be a proper name. It is not presupposed, but rather its predicate corresponds to the main assertion of the sentence. Lambrecht argues that it is non-compositional, and thus there is no felicitous sentence corresponding to (10b) like *j’ai les yeux* ‘I have the/my eyes’. In other words, the semantics are that of an event. He suggests that tense-marking in *avoir* presentations provides an additional argument for the non-compositionality of these constructions. As Blanche-Benveniste also pointed out, *avoir* ‘have’ is either “frozen” in the present tense, as in (11a), or is dependent on the tense of the superficial relative, as in (11b), where the brother-in-law in question is not presumed to be deceased:

(11) a. il y a la tante qu’elle demandait des nouvelles
    there has the aunt that she asked some news
    ‘My aunt would ask for news’

   b. j’ai eu mon beau-frère qui a fait un eh Paris-Nice
    I have had my brother-in-law who has done a uh Paris-Nice
    ‘My brother-in-law did a uh Paris-Nice’

5 See also Zubizaretta (1998).
Lambrecht takes all of these facts to confirm the observation that the superficial relative in these constructions is “not ‘subordinate’ in any clear sense, but expresses the main predication of the two-clause sequence” (Lambrecht 1988, 158).

Blanche-Benveniste (1983), in addition to discussing this special use of avoir, distinguishes two uses of c’est ‘it/that is’. One is a marker of contrastive focus (see (12a)), which cannot alternate with a full deictic subject plus copula (*Cela est à lui que je pense ‘That’s him that I’m thinking of’), and which she thus considers not to have the status of a matrix verb. The other, illustrated in (12b), she argues, is a true matrix verb in a presentational construction, as the subject c’ ‘that’ can alternate with deictics (cela ‘that’ and ceci ‘this’).

(12)  a. c’est à lui que je pense
     it’s to him that I think
     ‘I’m thinking of him’ / ‘It’s him I’m thinking of’
     b. c’est le chocolat que je voulais (= ceci est le chocolat que je voulais)
        that’s the chocolate that I wanted (that right there is the chocolate that I wanted)

Again, her conclusion is that in (12a), as in the special avoir constructions, an analysis taking the superficial relative clause to be subordinate to a clause headed by the verb être ‘be’ would be ill-founded.

2.2. Analyses of clefts and presentational constructions in French

The properties of French clefts and presentational constructions, pointed out by authors such as Lambrecht and Blanche-Benveniste, have been complemented and analyzed formally. We present in turn previous analyses of contrastive/corrective clefts (section 2.2.1), avoir-clefts, il y a constructions (section 2.2.2), and contrastive clefts versus new information clefts (2.2.3). As we will see, these analyses tend to be focused on a particular subset of these constructions (clefts or il y a).

2.2.1. Contrastive/corrective clefts

Clech-Darbon & al. (1999) present both syntactic and prosodic evidence that contrastive focus clefts do not have the syntax of restrictive RCs. They argue for an analysis in which the superficial RC is in fact a CP which is right-adjoined to an identificational IP, as illustrated in (13):

(13)  A: Ta fille est tombée dans l’escalier?
     ‘Your daughter fell down the stairs?’
     B: Non. [IP [IP c’est le petit] [CP qui est tombé dans l’escalier]]
     ‘No. It’s the little boy who fell down the stairs’

This analysis (see also De Cat 2002) captures the basic insight that various superficial RCs do not function as restrictive RCs (intonationally or semantically) and therefore do not have their syntax either. In particular, they do not involve embedding inside a DP.

* See also Kayne and Pollock (2009).
2.2.2. *Avoir*-clefts and *il y a* constructions

A construction which also involves an apparent DP containing an embedded clause is that of complements to perception verbs analyzed by Cinque (1992) and others as being in fact propositional with a DP followed by a gerund clause. In his analysis of these, Cinque (1992) points out that the motivation for a propositional analysis is found in several other constructions as well. Among these constructions, interestingly (and based on Moro 1989), small clauses in existential contexts are also listed, illustrated in (14). This type of analysis, thus, converges with Blanche-Benveniste’s and Lambrecht’s observations about presentational *avoir*-clefts with *il y a* in French, see (15).

(14) There is somebody climbing the stairs

(15) *Y a Jean qu’a téléphoné*  
there is John that has telephoned

The same analysis, involving a small clause (SC), is explicitly proposed by Côté (1999), on the basis of data from Quebec French, see (16); Côté notes that the relevant properties hold for spoken French in general, citing Lambrecht (1988), among others. Côté argues that the small clause analysis of existential RCs that have an event-reading accounts not only for their interpretation, but also for the fact that this reading is available only if the DP following the existential corresponds to the subject of the embedded predicate as in (17b).

(16) *Y a*  
there is  

(17) a. Qu’est-ce qui s’est passé ?  
‘What happened?’  
b. *Y a Jean qui est venu*  
‘John came’  
c. *Y a Jean que j’ai appelé*  
‘I called John’

These facts follow, she argues, because the small clause analysis of such complements, following Haïk (1985), Cinque (1992), Guasti (1994), entails that the DP, as subject of the SC, is an A-position, and therefore that movement to this position is A-movement. This means that movement of a non-subject is prevented by Relativized Minimality (Rizzi 1990) or Minimal Configuration (Rizzi 2004), because this movement would encounter the subject of the CP predicate, which is also an A-position with argumental features, and thus constitutes an intervener, blocking A-movement of the object as shown in (18):

(18)  

\[
Y \quad [CP_{(ec)} \quad Jean \quad [CP \text{ que j’ai appelé } \text{<Jean> } ] ]
\]

These facts follow, she argues, because the small clause analysis of such complements, following Haïk (1985), Cinque (1992), Guasti (1994), entails that the DP, as subject of the SC, is an A-position, and therefore that movement to this position is A-movement. This means that movement of a non-subject is prevented by Relativized Minimality (Rizzi 1990) or Minimal Configuration (Rizzi 2004), because this movement would encounter the subject of the CP predicate, which is also an A-position with argumental features, and thus constitutes an intervener, blocking A-movement of the object as shown in (18):

(18)  

\[
Y \quad [CP_{(ec)} \quad Jean \quad [CP \text{ que j’ai appelé } \text{<Jean> } ] ]
\]

5 Côté also entertains an analysis à la Guasti (1994), in which the subject of the small clause is base-generated and there is a *pro* in the subject position of the predicate CP.
The subject-object asymmetry noted by Blanche-Benveniste and by Lambrecht is thus accounted for, as is Lambrecht’s observation that the apparent object of *avoir* in these constructions is really a semantic subject. In the Côté analysis, the DP following *y a* is indeed a subject—of the propositional complement to *y a*. The apparent object status stems only from the fact that this DP is a small clause subject, and thus is case-marked from outside, by *avoir*. Côté notes that her analysis applies to other constructions in Quebec French: presentational clefts (*C’est Marie qui est venue - pas Jean qui a appelé* ‘It’s that Marie came - not that John called’), deictic constructions (*V’là Jean qui arrive * ‘Here’s Jean arriving’), and possessive constructions (*J’ai l’auto qui est en panne* ‘I’ve got a car that’s broken down’), all of which display the same subject-object asymmetry, all of which are restricted to stage-level predicates, and all of which are also discussed by Blanche-Benveniste and by Lambrecht, for spoken French.

Summarizing, in *c’est* cleft constructions, as analyzed by Clech-Darbon et al. (1999) and by Côté (1998), who extends the analysis to *y a*, we find a small clause CP containing another CP as a predicate. As predicates, notice that such CPs are not embedded within a DP, making them fundamentally different from genuine relative clauses.8

### 2.2.3. Contrastive clefts vs. new information clefts

Belletti (2008, 2009, 2015) proposed an analysis of cleft constructions, with specific reference to French, which, in fact, derives syntactically the important observations about information structure made by Blanche-Benveniste and by Lambrecht (see Section 2.1 above). Her analysis distinguished new information clefts, which, as we have seen, are restricted to subjects, from contrastive focus clefts (which are not restricted to subjects). She argued that these correspond to two separate focus positions – a higher position, which is in the left periphery of the sentence, and corresponds to contrastive focus, and a lower position, which is in the topic-focus field just above vP.

Contrastive focus clefts are taken to involve a CP complement to the copula and movement of the focused constituent to the left-peripheral SPEC, FocP position. Since CP complements to the copula are not declarative CPs, but rather are predicates, Belletti (2013 and to appear/2014) reasoned that their C does not express Force and thus does not occur in the left-most periphery of CP, but rather in a low position, to the right of the FocP as illustrated in (19a,b).

(19) a. être [CP [FocP Marie [que Jean a embrassé <Marie> ]]] [contrastive focus of object] be that has kissed
   b. être [CP [FocP Marie [ qui <Marie> a parlé à Jean ]]] [contrastive focus of subject] be who has spoken to

Belletti points out that Relativized Minimality is not violated in (23a), since, although movement has crossed the subject position, this movement of the object is to an A’-position which hosts operator features, whereas the subject occupies an A-position. Notice also that in the structure argued for by Belletti, as in the Clech-Darbon & Rebuschi proposal presented above, the CP cannot be a modifier of the DP, since it is not embedded inside of DP.

---

8 Very recently, *il y a* constructions have come to the notice of linguistic research; see Karssenberg & Lahousse (2015), who still call this construction “under-researched”.

57
New information (= presentational) clefts, on the other hand, are taken to have a different structure, in which the CP is the predicate of a CP small clause complement to the copula as shown in (20).

\[
\begin{align*}
(20) \quad [TP \text{ ce être} & \quad [\text{FocP} \text{ Jean} \quad [vP <\text{être}> \quad [\text{CP (sc) <Jean> [\text{CP qui <Jean> a parlé } ]]} ]]] \\
\text{it is} & \quad \text{Jean} \quad [+\text{EPP}] \quad \text{who} \quad \text{has spoken}
\end{align*}
\]

This CP has the special property of bearing an EPP feature, which entails that the subject of the CP predicate must move to the SPEC position of the small clause. The reason only the subject can fulfill this requirement is because this movement is A-movement, and a direct object would have to cross the predicate clause subject, invoking an RM violation. Belletti, like Côté, thus follows Guasti (1994) (see also Haïk 1985; Cinque 1992) in supposing that the subject of the CP predicate is an A-position, and therefore that new information clefts can only be subject clefts. Thus, the facts in (21) and (22) regarding possible answers to wh-questions, mirror the facts reviewed above regarding possible answers to the broad focus question ‘What’s going on?’ in that only a subject cleft is permitted, see (21b,c).

(21)  
\begin{align*}
a. \quad \text{Qu’est-ce que tu as acheté ?} & \quad \text{‘What did you buy?’} \\
b. \quad *\text{C’est un livre (que j’ai acheté) ‘It’s a book (that I bought)’}
\end{align*}

(22)  
\begin{align*}
a. \quad \text{Qui a parlé ?} & \quad \text{‘Who spoke?’} \\
b. \quad \text{C’est Jean (qui a parlé) ‘It’s John (that spoke)’}
\end{align*}

Belletti also notes the similarity between this way of looking at new information clefts and Guasti’s (1994) analysis of complements of perception verbs as in (23). This structure accounts for the fact that both of these constructions involve predication and thus allow an event-reading, as is evidenced by the fact that (23) is a possible answer to a question such as \text{Qu’est-ce que t’as vu?} ‘What did you see?’

(23)  
\text{J’ai vu [[Marie [qui parlait avec Jean]]}

‘I saw Mary who was speaking with John’

Summarizing, the analyses we have reviewed here have in common two essential, and related, properties. They capture the fact that relative-like constructions in fact do not entail DP-modification, but are rather event descriptions and thus clauses (CPs) that are predicated of the DP. This in turn means that the CP is not in fact embedded inside the DP and thus that, structurally, these relative-like constructions are flatter than genuine RCs (where a full CP is embedded inside a DP). In addition, the DP is the subject of a small clause CP which, as Belletti argues, is truncated. This implies that these constructions are not cases of a full CP directly embedded inside a full CP. Instead, they are constituents of a truncated CP which itself is a complement of the copula in the matrix CP. Thus the term

\footnote{See also a more recent formulation involving a Predication Phrase suggested by Belletti (to appear/2014).}
\footnote{See also Haïk (1985).}
\footnote{Minimally it is the ForceP of the left periphery that is missing in a small clause CP, see Belletti (to appear).}
superficial relative seems particularly appropriate for these structures without DP (or CP) recursion.

The structure proposed by Belletti (2008, 2009) has the additional advantage of providing structural representation for the discourse properties observed by Lambrech and others. We will therefore adopt significant aspects of this analysis, adapting it to take into account Côté’s analysis of existential relatives.

2.3. A syntactic analysis of c’est and il y a a superficial relatives

Building on our previous work, we will propose an analysis in which various types of apparent relatives which appear as complements to c’est or avoir in fact involve not a DP complement which contains a CP (the structure of a genuine relative clause) but rather a CP complement. While complements to ‘be’ in French can entail either contrastive or presentational focus, and thus, adopting Belletti’s analysis, can involve either the high or the low focus position, focus constructions with ‘have’ entail only presentational focus, and, thus should only involve the low focus position. In other words, ‘have’ selects either a DP or a small clause complement.

A genuine RC (a restrictive relative) can occur in a construction with ‘have’ when it is embedded inside the DP complement of ‘have’, as in (24a-b). In these cases, it is the entire DP which moves to the low focus position.

(24) a. [Max a [DP un frère [CP que Marie ne supporte pas]]] [genuine RC]
   ‘Max has a brother that Mary can’t stand’
   b. Il y a [DP un yaourt [CP que Max a acheté hier]] [genuine RC]
   ‘There’s a yoghurt that Max bought yesterday’

When a small clause complement is selected by avoir, it is the event that has new information focus, and thus it is the entire CP small clause, which moves to the low focus position. These constructions do not contain genuine RCs, but rather CP predicates of a small clause, illustrated in (25).

(25) a. y a [CP (sc) Jean [CP qui <Jean> est venu]] [superficial rel]
   ‘There’s John who came’
   b. j’ai [CP (sc) ma voiture [CP qui <ma voiture> est au garage]] [superficial rel]
   ‘I have my car that’s in the garage’

These superficial relatives are restricted to SRs, because of the interaction between the small clause analysis and Relativized Minimality, as we have seen, while genuine relatives are not restricted in this way, as they involve A’-movement only.

Returning to complements of ‘be’, we adopt essentially Belletti’s (2008, 2009) analysis in which the copula selects for either a full CP complement or a small clause CP. In the first case, any constituent of the full CP complement may move up to the left peripheral
Cornelia Hamann & Laurice Tuller

contrastive focus position, (26). In the second case, the subject of the CP small clause complement necessarily corresponds to the subject of the CP predicate (again, because of Relativized Minimality). This subject may then move to the lower focus position, (27).

(26) c’… est [CP … [FocP Max … que [TP j’ai vu <Max>] ] ] (pas Pierre)  
      it is Max that I have seen not Pierre  
      [superficial rel, contrastive focus]

(27) [TP c’est … [FocP Jean … [vP [CP [sc] [DP <Jean> [CP qui <Jean> a parlé]]]]]]  
      it is John who has spoken  
      [superficial rel, new information focus]

Rialland et al. (2002) show that cases like (26) must be distinguished both semantically and prosodically from focus constructions involving *c’est* and an RC in which an entire clause is focused, and which are answers to the question ‘What’s going on?’. They argue that these properties can be derived from a reduced cleft analysis, where there is a truncated predicate *qui se passe* ‘that is going on’. Belletti suggests that the RC in these constructions is a genuine restrictive relative inside the subject of the small clause whose predicate is ‘that is going on’ as in (28).

(28) c’est [le petit [qui est tombé dans l’escalier]] qui se passe  
      it is the boy who is fallen in the stairs that CL-REF happens

Notice, however, that this analysis means that the subject of *qui se passe* is a DP, which semantically is odd (*The boy is happening). This is because the subject of ‘happen/go on’ must be an event and clearly a boy is not an event. What is happening is not a boy, but rather that the boy fell down the stairs. If the subject should in fact be an event, we expect to have a clause. We therefore suggest the analysis in (29), in which the subject of this small clause is in fact itself a small clause CP.

(29) [CP (sc) [CP (sc) [DP le petit [qui <le petit> est tombé dans l’escalier]] [ ( [qui se passe] )  
      the small who is fallen in the stairs that CL-REF happens

It is then the lower small clause, the subject of ‘is happening’, which moves to the focus position. Recall that the subject of this lower small clause has been moved out of its CP small clause predicate. This predicts that these constructions also display the subject-object asymmetry typical of new information focus, which is correct, as shown in (30a–d). If the RC is a restrictive relative embedded within the subject DP, as Belletti (2008) suggested, then (34d) is wrongly predicted to be grammatical. We will therefore assume the structure in (29), which gives rise to movement of the CP(SC) subject *le petit qui est tombé dans l’escalier* to the low focus position, correctly representing the fact that focus is on the event.

(30) a. Qu’est-ce qui se passe ? ‘What’s happening?’
   b. Léa a vu Max dans l’escalier ‘Lea saw Max in the stairs’

13 We are aware that the analysis may involve more complex mechanisms in order to overcome Criterial Freezing (Rizzi 2006) and allow movement to the edge of a phase.
Genuine versus superficial relatives in French: The depth of embedding factor

c. C’est Léa qui a vu Max dans l’escalier ‘It’s Lea who saw Max in the stairs’
d. *C’est Max que Léa a vu dans l’escalier ‘It’s Max that Lea saw in the stairs’

To summarize, we have distinguished genuine relatives, RCs which are modifiers restricting their antecedent, from superficial relatives, a term we are using in a very general sense to indicate CPs which are deceptively similar to RCs, but involve predication and thus, syntactically, are not embedded within a DP. The crucial difference, we have argued, is the depth of embedding of the subordinate small clause CP.

(31) a. Genuine relatives (CP inside DP, inside IP)
   b. Contrastive clefts (XP in left-peripheral focus position, not inside a DP)
   c. Presentational focus (DP/small clause CP in low focus position, not inside a DP)

In superficial relatives, both contrastive focus and presentational focus, the CP is a small clause and, crucially, is not inside the DP, at any point in the derivation. In genuine relatives, on the other hand, the CP is a modifier of the head of the relative clause, and as such, is merged inside the DP.

Thus, RCs are not differentiated by the presence or absence of wh-movement. Rather, they can be distinguished by the presence of an intervener to that movement (as argued by Gibson 1998; Gordon 2001; Friedmann et al. 2009), or by the depth of embedding (see example (28) vs. the examples (29)-(33) and the summary in (35)). In this light, the fact that clefts and other presentational constructions involve movement to a focus position does not make them as complex as genuine RCs. They are in fact less complex precisely because they involve a predicational, less deeply embedded CP, in the sense that it is not embedded in a DP. Notice also that this analysis means that contrastive focus clefts are very much like presentational clefts in terms of depth of embedding of the CP. Their difference lies in whether the DP is the small clause subject of a predicate CP, the case of presentational focus, or not, the case of contrastive focus. However, we do expect that object (contrastive focus) clefts might be sensitive to intervention effects, which separates them from subject clefts, contrastive or presentational (see in particular the findings of Gordon et al. 2001). This is because contrastive focus entails a configuration which is reminiscent of the one investigated by Friedmann et al. (2009) for object relatives, and which is potentially subject to relativized minimality effects.

2.4. The acquisition of relatives in French

Many authors have reported that superficial relatives are produced very early on by children acquiring French, appearing prior to genuine RCs. Labelle (1989, 1990) found that superficial relatives (and specifically contrastive focus clefts and presentational c’est constructions) were mastered earlier than RCs in a study of elicited production to 3- to 6-year-olds, and she also cites other work arriving at the same conclusion on the basis of spontaneous production data (see, for example, Rondal 1979). De Cat’s (2002) longitudinal spontaneous language study of four young French-speaking children (2 Belgian, 1 Canadian, and 1 French) found that attempts at clefts (for example, non, c’est moi [qui l’ai] le pingouin ‘No, I’m the one (who has) the penguin’, De Cat 2002, 265) occurred before the first embedded finite clauses, and that these latter were (subject) clefts, generally with complementizer omission (c’est Mamy [qui l’] a mangé ‘It’s Mamy (who) has eaten (it)’). Jisa & Kern’s (1998) study of relative constructions in Frog story narratives also found that children learning French favor certain types of relative(-like) structures, long into
childhood. Comparing 5-year-olds, 7-year-olds, and 10-year-olds to adults, and looking at both internal structure and narrative function of RCs, they observed that the children of these three age groups generally performed similarly and differently from the adults. The proportion of SRs compared to non-SRs among children remained at 90% or higher, compared to 79% in adults; only adults had more than a negligible proportion of center-embedded relatives; children used about twice as many intransitive relative constructions as transitive ones, whereas transitivity was equally distributed in adults’ relative constructions. Children’s RCs displayed frequent use of presentational constructions (with *il y a* or with *c’est*), and thus they performed well on the discourse function of introducing new referents. However, their preference for structurally simpler RCs (intransitive), it is argued, was also responsible for their infrequent use of RCs with narrative advancing functions (as transitivity in the RC is crucial for advancing the story plot – a transitive RC promotes a new referent, but also sets an agent-patient relationship between the new referent and the other story participants). Protracted development of the internal complexity of RCs thus was taken to influence narrative development.

Gayraud & Martinie (2004) studied the development of the complexity of French RCs in older children (9-, 12-, and 15-year-olds), compared to adults, in spoken and written narratives. They found that, in the spoken narratives, the proportion of what they refer to as “segmented” RCs (clefts, pseudo-clefts, and presentational *(il) y a* and presentational *avoir*) compared to other RCs decreased with age to a very significant degree (65% in 9-year-olds, 34% in 12-year-olds, and around 25% in 15-year-olds and in adults), and these were also significantly more prevalent than other relatives in oral narratives compared to written narratives (except for the youngest children, who used high proportions of “segmented” relatives in both modes). These results mirror the developmental progression found in younger children, though Gayraud & Martinie propose that superficial relatives (their “segmented” relatives) are the most complex (for reasons they do not make clear). This is problematic (as they point out), as these constructions are precisely those which are frequent in younger speakers; their suggestion is that this high frequency is an example of a complex structure having been schematized.

Contrary to Gayraud & Martinie, Hamann et al. (2007), Delage et al. (2008), and Tuller et al. (2012) have argued that order of emergence and relative frequency in younger speakers are in fact directly linked to the complexity of the syntax. Thus the early and frequent occurrence of presentational *y a* and *avoir* constructions is explained by the idea that some apparent RCs in fact involve simpler syntax. This work has also found that younger typically-developing children, as well as children acquiring French in atypical contexts (SLI, hearing loss, and childhood epilepsy) show clear signs of preference for SRs over non-SRs, and that clefs, presentational *c’est / y a*, and, RCs used without a root clause (0-level relatives) occur frequently. The studies by Hamann, Tuller and colleagues showed that the bulk of the relatives produced by a group of TD 6-year-olds and by a group of children with SLI were of the following type: 0-level, left dislocation, presentational *y a*, and *c’est* constructions. They suggested that the generalization underlying these more frequent constructions was that contrary to genuine relatives, which involve the embedding of a CP within a DP within an IP, in each of these other constructions one of these layers is missing, IP or DP. In the preceding section, a syntactic analysis has been developed on

14 Note that a formulation as “a CP within a DP within a vP” would exclude relative clauses with head nouns that are constituents of adjuncts such as *John plays tennis on a court that has never been raked*. The crucial property we aim to capture here is that the modified DP is a constituent of a clause, minimally an IP.
theoretical grounds, providing a unified analysis in particular for *il y a* and *c’est* constructions and thus a foundation for the intuition that superficial relatives, as a group that includes *il y a* constructions, are simpler than genuine relatives. This analysis constitutes an explanation for previous results. In the cited previous studies, all superficial relatives were compared to all genuine relatives. However, given what is known about the effects of intervention in non-SRs clauses, the role of depth of embedding can be ascertained only if these effects can be controlled for. We will report on results of a study which explores in greater depth the ways in which the different syntactic properties of superficial relatives and genuine relatives have an effect on their occurrence in early child production and in the production of children with SLI.

3. Supporting Evidence from (A)typical Child Spontaneous Production Data

The view of relativization in French that we are advocating here draws thus on a large body of empirical and analytical work suggesting that so-called RCs are not equally complex, in terms of the level of embedding that their syntax entails. We offer support for this syntactic analysis from our own results on child spontaneous production data, some of which have appeared (Delage et al. 2008; Tuller et al. 2012) and some of which are new. This research program has sought to examine (inter alia) whether the factor of depth of embedding plays a role in determining what kinds of RCs are predominant in young children compared to older children, and whether an even sharper pattern can be detected in production of children with SLI. This program is pursued in the present data analysis.

After looking at the overall frequency of RCs produced by TD children and children with SLI, we present results about avoidance strategies and errors, and evidence for the subject-direct object asymmetry, manifested in terms of relative frequency of types of RCs (all kinds, genuine and superficial relatives). We broaden the SR-OR dissociation to include adjunct relative (AR) clauses, which we suggest should behave like ORs under the Friedmann et al. (2009) analysis of intervention as an RM effect (since in both cases there is a potential intervener in the subject position).

Once these basic properties of RCs have been reported on, we turn to evidence for the relevance of depth of embedding. Simply determining level of embedding with respect to the matrix clause is a first indication of production sensitivity to depth of embedding. Do younger children and children with SLI produce a greater proportion of the lowest level of embedding (0-level)? The flatter structure argued for in contrastive focus clefts and presentational constructions also means that the proportion of these in spontaneous production should decrease with age in TD development, in accordance with the finding that they are the first relatives to appear (Labelle 1990, 1996; De Cat 2002), that they have been found to decrease in TD children after age 9 (Gayraud & Martinie 2004), and that they have been observed to predominate among RCs in children with SLI (Delage et al. 2008).

We propose to control for possible cumulative effects of both intervention and depth of embedding by alternatively looking at constructions in which depth of embedding is the same, comparing SRs to non-SRs, and then only at SR constructions, genuine and superficial relatives, for which intervention does not play a role.

3.1. Method

We studied spontaneous language samples taken from five groups of monolingual French-speaking children: typically developing six-year-olds (TD6), eight-year-olds (TD8),
eleven-year-olds (TD11), and fourteen-year-olds (TD14), and a group of ten- to twelve-year-old children with SLI (SLI 10-12) not previously reported on. The children with SLI were all recruited at the same university teaching hospital language reference center, on the basis of a diagnosis for expressive SLI (all subjects had Performance IQs of 85 or greater).

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean Age</th>
<th>Sex</th>
<th>Age Range</th>
<th>MLU (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD6 (N = 12)</td>
<td>6;4</td>
<td>7 M 5 F</td>
<td>6;1 - 6;7</td>
<td>6.97 (.88)</td>
</tr>
<tr>
<td>TD8 (N = 12)</td>
<td>8;2</td>
<td>6 M 6 F</td>
<td>7;9 - 8;7</td>
<td>7.59 (1.28)</td>
</tr>
<tr>
<td>TD11 (N = 12)</td>
<td>11;4</td>
<td>6 M 6 F</td>
<td>11;1 - 11;9</td>
<td>7.90 (.83)</td>
</tr>
<tr>
<td>TD14 (N = 12)</td>
<td>14;5</td>
<td>6 M 6 F</td>
<td>13;7 – 14;10</td>
<td>8.22 (.99)</td>
</tr>
<tr>
<td>SLI 10-12 (N = 16)</td>
<td>11;7</td>
<td>11 M 5 F</td>
<td>10;1 – 12;10</td>
<td>6.38 (.77)</td>
</tr>
</tbody>
</table>

Table 1. Characteristics of TD Groups and SLI Group

The four TD groups, which range in age from 6 to 14, allowed us to look at developmental trends over a broad age-range, an important source of evidence since we are hypothesizing that frequency of complex RCs increases with age, well after childhood and into adolescence. In addition, the TD6 group served as a rough language-match for the children with SLI, and the TD11 group served as an age-match. The mean MLU for the SLI 10-12 did not differ significantly from the TD6 group (U = 60; p = .094), while it was significantly lower not only than that of their age controls, the TD 11-year-olds (U = 10; p < .01), but also than that of the TD 8-year-olds (U= 39.5; p < .01).16

Spontaneous language samples were collected via fifteen minute digital audio recordings of conversation with a student investigator, which followed an identical protocol. Participants were first requested to provide a story for a series of drawings and then questioned about their interpretation of the drawings. A free conversation followed in which the investigator asked about school and extra-curricular activities and asked the participant to tell the story of a book/video game/movie/television show that they knew well. Transcription, which began at minute five of the recording, to make sure the child felt comfortable with the situation, continued until the sample contained 60 to 70 utterances produced by the subject (yes/no answers and repetitions were excluded).

RCs were coded by type (subject, object, adjunct, as in (32)), but also so that the role of depth of embedding could be assessed. Embedding was coded for level: 0-level corresponding to an RC occurring in an utterance with no matrix clause, level-1

15 See also Monjauze (2007), Delage (2008), Delage et al. (2008), Tuller et al. (2012) for other analyses of the language samples from these TD groups, in comparison with different atypical populations (SLI, mild-to-moderate hearing loss, and Rolandic Epilepsy). These analyses did not include the in-depth analyses of relative clauses of the type proposed here.

16 It can also be seen that MLU increases with age in the TD groups; these differences are significant only between TD6 and TD11 (U = 32.0; p < .05) and between TD6 and TD14 (U = 25.0; p < .05). This result mirrors those found by other studies of spontaneous language in older children (see Tuller et al. 2012).
corresponding to a embedded in a DP of the matrix clause, level-2 corresponding to an RC embedded in a DP of a clause immediately embedded within the matrix clause. For example, (33a, c) are level-1 adjunct RCs: REL1-A. Likewise, clefts (contrastive and new information) were coded (see CLE as in (33a)), as were relatives introduced by presentational *avoir* or *c’est*, (see PRES in (33b)). Object and adjunct relatives and clefts were further coded for the nature of the RC subject and of the head of the relative clause, in order to control for intervention effects, as illustrated in (33b, c).

(32) a. j’ai eu le costume de Cendrillon qui est réversible. [REL1-S] (TD6)  
   ‘I got the Cinderella costume that is reversible’

b. puis après j(e) vais regarder un DVD qu’on m’a prêté [REL1-O] (TD11)  
   ‘then after I’m going to watch a DVD that somebody loaned me’

c. et son mari est tombé sur son journal intime où elle avait tout marqué [REL1-A] (TD14)  
   ‘and her husband found her diary where she had noted everything’

(33) a. et puis c’est là qu’elle travaille [REL1-A] [CLE] (TED, SLI)  
   ‘It’s a boy that the girl locked up because she has robots’

b. c’est un garçon que la fille a enfermé [REL1-O][PRES] parce qu’elle a des robots   
   Head = animate DP  Subject = animate DP

   ‘It’s a boy that the girl locked up because she has robots’

   (JUL, SLI)

c. oui c’est ça qu’il m’a dit [REL1-O] [PRES] (TD14)  
   ‘Yep, that’s what he told me’

Finally, all sentences were analyzed for errors, as illustrated in (34), an utterance produced by a child with SLI, which contains errors both inside the RC itself and elsewhere in the utterance. Furthermore, all constructions which arguably could have surfaced as an RC were also coded. These included juxtaposition of two root clauses (where one could have been an RC inside the other, see (35a)), as well as utterances containing self-interruptions inside the RC (and thus production of the RC was abandoned, see (35b)), and utterances containing an RC, but in which the matrix clause was missing a verb (see (35c)), and thus embedding was not really complete.

(34) et elle retrouve quelqu'un qui... qu'ils [*] se [*] connaissent [*] [REL1-S] bien car il a fait [CIR1] une course contre lui [*] (ROM, SLI)
   ‘And she met somebody who th.. that they know each other well because he raced against him (=her)’

(35) a. comment on dit [PR] &euh. Ben 0*y 0*a deux: groupes. Un groupe de &f filles et un groupe de gars. I(l)s vient [*] dans ... &euh avec &euh ... (CHL, SLI)
   ‘How do you say um. Well (there are) two groups. A group of two girls and a group of guys. They comes in ... um with um...’

b. et du coup y a une fille qui voudrait [REL1-S] [PRES] de pour euh f ...  
   ‘So there’s a girl who would like... to... for.. um f...’ (ROM, SLI)

c. mais *(c’est) un ami du monsieur qu'elle veut [REL1-O] tuer (JUL, SLI)
   ‘But (he’s) a friend of the man that she wants to kill’

65
3.2. Results and analysis

Analysis of the 64 transcribed and coded spontaneous language files gave rise to a corpus of 395 RCs (of all types). The measures described in the preceding section allowed us to examine depth of embedding as a complexity factor, compared to other complexity factors, as revealed in inter-group relative frequencies. We proceeded in the following manner: first of all, looking at all types of RC’s (genuine and superficial, subject, object, and adjunct), which have in common the factor of A’-movement, we measured, for each group of participants, relative frequency, proportion of erroneous RC’s, and “active avoidance” of RC’s. Next, we compared SR’s to OR’s and AR’s (all types) in each group to ascertain whether Intervention affects frequency. Depth of embedding was then examined, in two ways: first frequencies of levels of embedding of all types of RC’s were compared (0-level, level-1, etc.), and, second, frequencies of genuine relatives were compared to frequencies of superficial relatives, since, by hypothesis, the former involve deeper embedding compared to the latter. Finally, we attempted to cross potential intervention effects with depth of embedding effects by looking at the relative frequencies of genuine and superficial relatives only among SR’s (where there is no potential intervention), and then examining the proportion of SR/OR+AR only in genuine relatives.

Taking into account all RCs produced, both genuine and superficial and both subject and non-subject, the proportion of subordinate clauses that are RCs was basically the same in each of the groups of participants (about 1/3). Figure 1 shows that the groups are also similar, in fact, for the mean number of RCs produced per child. In the samples studied the differences between the TD groups are not significant, nor is that between the SLI group and each of the TD groups (except for SLI /TD-14 comparison, U = 46.0; p < .05). This result is a priori surprising if RCs, in the broad sense, are difficult because they involve movement/internal merge. In particular, why do children with SLI appear to have no more difficulty than children their age in producing RCs? This result suggests that the nature of the relatives produced needs to be examined.

![Figure 1. Frequency of all embedded relative clauses (mean N)](image)

17 The mean % of embedded RCs over total embedded subordinate clauses is 33.9% (20.1) for the SLI 10-12, 32% (16.7) for TD-6, 30.6% (12.6) for the TD-8, 32.3% (11.0) for the TD-11, and 29.7% (8.7) for the TD-14.
We expected that RC production might cause children with language impairment to make errors (either in the RC itself or elsewhere in the sentence), and we also expected that the difficulty involved in the production of RCs might have as a consequence that these children would express the same information by using other, simpler syntactic means, and, finally, we expected that these children might actually show signs of abandoning RCs, for example, through self-interruptions.

On average, over \( \frac{1}{4} \) of utterances (26.6\%) containing an RC produced by SLI participants were indeed erroneous. As can be seen in Figure 2, this rate was under 10\% in the TD6 (6.2\%) and TD8 (9.2\%) groups, and amounted to only 1 or 2\% in the TD11 (1.2\%) and TD14 (2.1\%) groups. We see a developmental trend in the TD groups (TD6 and TD8 versus TD11 and TD14), and the SLI group, not surprisingly, was worse than even its language matched group (TD6).\(^{18}\)

In order to assess “active” avoidance of RCs, we added the total number of RCs produced to the number of ATTEMPTS at RCs, those which did not surface as RCs because either the clause was abandoned via a self-interruption or because two juxtaposed clauses were produced instead. This sum was termed POTENTIAL RCS. Out of this total mean number of potential RCs, we then calculated the mean percentage of attempts. This calculation yielded the fact that nearly 1/4 of the potential RCs in the SLI group were attempts (M 23.5\%, SD 19.3\%), whereas this rate was well below 10\% in each of the TD groups (3\%, 4\%, 6\%, and 3\%, respectively, for TD6, TD8, TD11, and TD14). A Kruskal-

---

\(^{18}\) The extremely low error rates in the TD groups did not lend themselves to qualitative analysis. The rarity of genuine RCs and of non-subject RCs in the SLI group makes it impossible to pursue meaningful comparative error analyses. Genuine relatives in the SLI group amounted to only 13 total, and half of the 16 children did not produce a single one, and thus it was impossible to compare errors in genuine relatives to those in superficial relatives, though it was clear that errors occurred on the same elements (affecting complementizers, tense morphemes, number and gender agreement, etc. and consisting of both omissions and substitutions). Likewise, comparing error types between SRs and non-SRs, qualitatively or quantitatively, is very hard to do, because of the small number of non-subject relatives produced in the SLI group (a total of 10 ORs and 8 ARs, and 6 of the 16 children did not produce any non-subject RCs).
Wallis Anova confirmed a significant group effect ($H = 15.9, p < .01$). The rate of attempts at RCs in the SLI 10-12 was significantly higher than that in each of the other groups (TD6 $U = 36.5, p < .01$; TD8 $U = 39.5, p < .05$; TD11 $U = 43.5, p < .05$; TD14 $U = 38.0, p < .05$). Figure 3 presents the proportions of produced RCs, self-interruptions of RCs, and cases of two juxtaposed clauses, the latter two constituting what we are calling attempts.

![Figure 3. Mean % of RCs, Self-Interruptions of RCs, and Juxtapositions](image)

Examples of attempts at RCs were given in (35) above. The examples in (36) are particularly illustrative of the specific difficulty that embedding poses. (36a) displays first a self-interruption of the RC (She found a black horse that was at...) and then a second pass consisting of juxtaposition of two root clauses (It was a magic horse. She found it). In (36b), we see self-interruption of an RC (After/then there’s a, a friend of... the woman...), followed by a 0-level RC in the second pass (The old woman who found him), which is in turn followed by a juxtaposition (She, he knows his first name), avoiding thus production of two RCs dependent on the same head noun (There’s a friend of the old woman who found him who knows his first name). Similary, (36c) displays an (erroneous) utterance consisting of a juxtaposition which gets around production of a RC which would include a level-5 complement clause (For example, that there is one of his best childhood friends he has he is missing because he went to go to see if there weren’t any Germans around).

(36)  

a. ça raconte. ben ça raconte qu'elle a trouvé un le... Je sais plus comment ça s'appelle. elle a trouvé une un cheval noir qu'était à ... Ca fait un cheval magique. Elle l'a trouvé.

b. après y en a un... un ami à la dame la vieille dame qui l'avait trouvé bah. elle il sait son prénom.

(JUL, SLI)
Genuine versus superficial relatives in French: The depth of embedding factor

She he knows her first name’ (MAR, SLI)
c. par exemple que y a un d(e) son [*] meilleurs copains d'enfance il a il est porté disparu car il est allé faire un tour pour voir si y a pas d'Allemands autour.
‘for example, that there is one of his best friends from childhood he has he is missing because he went to go see if there are any Germans around’ (ROM, SLI)

As expected, SRs of all kinds, which do not have a potential intervener, are generally more frequent than ORs, as can be seen in Figure 4. We also see that ARs are similar in frequency to ORs, and not to SRs, supporting the idea that intervention effects are not limited to ORs, as we have suggested (but see footnote 24). The mean number of SR is higher than that of non-SRs (OR + AR) (in TD-6, Z = 2.27 p < .05; in TD-8, Z = 2.8 with p < .01; in TD-14, Z = 2.55, p < .05; in SLI, Z = 2.98, p < .01), except for the TD-11 (Z = .56, p = .575).19

Figure 4. Subject, Object, and Adjunct Relatives (Mean N)

The results reported in the above figures have all concerned general RC production of all kinds. We argued above, on syntactic grounds, that superficial RCs in French in fact can be classified according to depth of embedding. Turning then to results which target specifically the embedding factor, when 0-Level RCs (RCs not embedded within an IP) are compared to embedded RCs (Level-1, Level-2, etc.), as in Figure 5, it was found that participants with SLI produced a larger proportion of 0-Level RCs (23%), significantly more than each of the TD groups.

19 Regarding the SR/OR asymmetry, it is important to note that no evidence was found that ORs are avoided through use of subject relatives with a passive or a reflexive causative (se-faire) construction, contrary to what has been found in elicited production experiments (Manetti and Belletti 2015; Belletti and Contemori 2010; Delage 2008). Such constructions were not found at all in the samples from the SLI, the TD6 and the TD8, and only one token each was found in the TD11 and the TD14 samples. We argue elsewhere (Hamann and Tuller 2015) that the discourse situation in elicited production of object relative clauses favors production of a passive.
The mean rate of 0-level RCs over total RCs was 22.9% (SD 14.0) in the SLI 10-12 group, significantly higher than in the TD6 group (M = 7.3%, SD 10.8; U = 38.5, p < .01), than in the TD8 group (M = 10.6%, SD 13.4; U = 50.5, p < .05), than in the TD11 group (M = 4.9%, SD 8.3; U = 32.0, p < .01), and also than in the TD14 (M = 9.6%, SD 11.9; U = 46.0, p < .05). Inter-group differences in the mean numbers of RCs per participant (both for 0-level and for embedded RCs) were not significant.

We turn now from the least embedded RCs (0-Level) to deeply embedded RCs, those RCs which are embedded deeper than Level-1, illustrated in (37), which contains two RCs, the first of which is a Level-3 RC [REL3-S], since it is embedded in a non-finite complement clause [NFC2], which is itself embedded in an adverbal clause [ADV1], itself embedded within the matrix clause [ROOT].

\[(37)\] et euh donc comme ils voudraient [ADV1] trouver [NFC2] un une espèce de &bah> une personne qui pourrait &m qui pourrait [REL3-S] m le ramener ils ont ils ont vu [ROOT] cette jeune fille là et … qui ressemble [REL1-S] à Cléopâtre (TD14)

‘And um so since they would like to find a a sort of a person who could um who could um bring him back, they they saw this girl there and who looked like Cleopatra’

While the SLI 10-12 and the three younger TD groups produced similar proportions of these RCs (< 10% of all RCs produced), TD-14 produced nearly twice as many (M = 17.7%, SD 12.7), as is shown in Figure 6. Nearly all TD-14 participants produced at least one deeply embedded RC (10/12 participants), whereas only one third to one half of participants in the other groups did so. A Kruskal-Wallis ANOVA showed a trend toward a group effect (H = 4, p = .065), and paired comparisons showed that the TD-14 produced
Genuine versus superficial relatives in French: The depth of embedding factor

significantly more deeply embedded RCs than the TD-6 (U = 35.5, p < .05), the TD-8 (U = 38.0, p < .05), and the SLI (U = 43.0, p < .01) (no other differences were significant).

When we examined all embedded RCs, those which are deeply embedded and those which are embedded merely at Level-1 (and thus putting aside 0-level RCs, which are, by definition not embedded in a clause), the embedding factor manifested itself in the relative frequency of genuine relatives versus superficial relatives (clefts, presententials, *avoir* with event reading, dislocated relatives). Figure 7 shows a rather sharp developmental profile among TD groups, TD-6 children producing proportionally fewer genuine RCs (M 17.1%, SD 21.0) compared to the other age groups (TD8 M 44.0%, SD 31.2; TD11 with a M of 41.7%, SD 19.8, and TD14 with a M of 55.8%, SD 20.7), and demonstrates that the SLI children, at 17.9% (SD 22.5), produced RCs like their language matches (TD-6) did, not like their age matches (TD11). A Kruskal-Wallis ANOVA showed a significant effect for group (H = 24.369, p < .0001), and paired comparisons showed that both TD6 and SLI 10-12 produced significantly lower rates of genuine RCs than each of the other groups (SLI vs. TD8, U = 49.5, p < .05; SLI vs. TD11, U = 43.0, p < .05; SLI vs. TD14, U = 20, p < .001; ). The difference between SLI 10-12 and TD6 was not significant (U = 68.5, p = .299), nor was that between TD8, TD11, and TD14. Chi-2 analysis showed that there were indeed significantly more superficial RCs than genuine-RCs in the SLI (chi-2 = 15.61, p < .001) and the TD6 (chi-2 = 16.1, p < .001), but not in any of the other groups.
In order to isolate the depth of embedding factor from the intervention factor, it is useful to look exclusively at SRs, as these do not also involve a potential RM configuration. Figure 8 shows once again the SLI 10-12 group patterned with the TD-6: both of these groups displayed a low rate for genuine SRs (with mean rates of 9.6 and 5.8%, respectively) compared to superficial SRs. The rate of genuine relatives was much higher in the other TD groups, nearly reaching half of all SRs in the oldest group (with means of 37%, 39.4%, and 47%, respectively, in TD8, TD11, and TD14). A Kruskall Wallis ANOVA showed a significant effect for group ($H = 25.40625$, $p < .001$). As was the case for overall rates of genuine RCs, differences between each of SLI 10-12 and TD6 were significantly lower than those in each of TD8, TD11, and TD14, though they did not differ from each other, nor did any of the rates between the other three groups (SLI vs. TD8: $U = 49.5$, $p < .05$; SLI vs. TD11: $U = 43.0$, $p < .05$; SLI vs. TD14: $U = 6.5$, $p < .001$; TD6 vs. TD8: $U = 25.5$, $p < .01$; TD6 vs. TD11: $U = 17.5$, $p < .01$; TD6 vs. TD14: $U = 6.5$, $p < .001$; SLI vs. TD6: $U = 68.8$, $p = .294$; TD8 vs. TD11: $U = 68.0$, $p = .807$; TD8 vs. TD14: $U = 68.0$, $p = .10$; TD11 vs. TD14: $U = 45.0$, $p = 1.56$).
Likewise, when we control for depth of embedding by considering only genuine relatives, as in Table 2, it is interesting to note that the subject-object/adjunct asymmetry illustrated in Figure 4 above becomes less obvious. As we have just seen, in Figure 7, genuine relatives were extremely rare in both the SLI group and the MLU-matched TD 6: This was true for both SRs (mean N of 0.3 in each group) and non-SRs (mean N of 0.3 for ORs in each group, and no TD6 tokens of genuine ARs, and a mean of 0.2 in the SLI group). Although the TD8 and the TD14 did produce more genuine SRs (means of 1.4 and 2.5, respectively) than genuine ORs/ARs (means of 0.3/0.1 and 0.8/0.3), no such asymmetry was found in the TD11 (19 occurrences of genuine SRs and 16 occurrences of non-S RCs).

![Figure 8. Genuine Subject RCs versus Superficial Subject RCs (Mean Proportion)](image)

Table 2. Subject, Object, and Adjunct Genuine Relatives: Tokens and Mean N

<table>
<thead>
<tr>
<th>N Tokens in each group</th>
<th>Mean N per participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
</tr>
<tr>
<td>SLI</td>
<td>5</td>
</tr>
<tr>
<td>TD6</td>
<td>3</td>
</tr>
<tr>
<td>TD8</td>
<td>17</td>
</tr>
<tr>
<td>TD11</td>
<td>19</td>
</tr>
<tr>
<td>TD14</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 2. Subject, Object, and Adjunct Genuine Relatives: Tokens and Mean N

If only transitive verbs are considered, the picture also changes (see Belletti & Chesi 2011, on adult Italian): Jisa & Kern (1998) argued that transitivity of relative clauses is a factor of complexity that blocks children’s use of RCs for narrative advancing. In their study, 5-, 7-, and 10-year-olds produced only roughly 1/3 RCs with a transitive verb, compared to ½ of adults RCs. The logic behind restricting counts of SR to transitive verbs would also seem to imply that ARs should be calculated in some way over total number of adjuncts produced.
Although this result suggests that a deeper analysis of the nature of the occurring interveners in genuine ORs is necessary (see Belletti & Chesi 2011; Arosio et al. 2011; Hamann & Tuller 2010), it is noteworthy that the subject/non-subject asymmetry is visible in clefts, precisely in the younger TD groups and in the SLI group. The lack of intervention effects in genuine relatives and the presence of these effects in clefts go together, we would argue. Genuine relatives are avoided so much that it is difficult to see the S/O asymmetry. Clefts don’t have the added depth of embedding factor and thus are not avoided to the same degree and therefore the S/O asymmetry is visible, in the SLI, TD6 and TD8 groups, but not really in the older TD groups (TD11 and TD14), as can be seen in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>N Clefts in each group</th>
<th>Mean N per participant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SR</td>
<td>OR</td>
</tr>
<tr>
<td>SLI</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>TD6</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>TD8</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>TD11</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>TD14</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 3. Subject, Object, Adjunct Clefts: Tokens and Mean N

Summarizing the empirical evidence provided here, we have seen that language samples of the children with SLI, besides showing more erroneous RCs (Fig. 2) and more active avoidance of RC production (Fig. 3), display the following properties: 1) high rates of 0-level RCs (Fig. 5), 2) low rates of deeply embedded RCs (Fig. 6), and 3) high rates of superficial RCs compared to genuine RCs (Fig. 7). In the TD-groups, higher levels of deep embedding are only observed in the group of adolescents (Fig. 6), whereas the rate of genuine relatives in the TD-6 is comparable to that of the SLI-group and significantly lower than the rates in the groups with older children and adolescents (Fig. 7). The language samples analyzed also revealed a general SR-OR/AR dissociation across all groups (except TD-11, Fig. 4), which becomes less clear when only genuine relatives are examined (Table 2). On the other hand, when the SR-OR/AR dissociation is neutralized, and we concentrate only on RCs that do not also display intervention (i.e. SRs), the embedding factor sharply emerges (Fig. 8).

4. Discussion and Conclusion

The aim of this article was to explore the nature of syntactic complexity through close examination of depth of embedding. French has several constructions which resemble RCs, but which are actually flatter in structure: these include clefts and presentational constructions introduced by *c’est* and *il y a*. We have proposed here an analysis of presentational *c’est* and *il y a* which identified a common underlying structural property in these constructions. *Avoir* ‘have’ in these cases selects a small clause CP complement just as in presentational *c’est* so that the apparent relative actually expresses predication. In the case of *il y a*, which can only have event readings, it is the whole small clause which is subsequently fronted to the position reserved for new information focus in the lower topic/focus field above vP. For event readings of *c’est*, we proposed the same analysis,
whereas movement of the DP into this low focus position allows new information focus on this DP alone. Contrastive clefts, on the other hand, are derived by moving the DP to the high focus position in the left periphery of the CP. It is the fact that all these so-called “superficial relatives” involve movement of the DP to a focus position from a CP which functions as a small clause predicate which makes them flatter and less complex than genuine RCs. In this light, contrastive clefts and presentational constructions are alike in terms of depth of embedding since they do not involve embedding within a DP. The difference between these constructions was located in the fact that, in the case of presentential constructions, the DP is the subject of the embedded small clause, whereas this is not the case for contrastive focus. The analysis also captured the fact that contrastive clefts can focus the subject or the object of the predicational CP, whereas presentational structures can only focus the subject (or the whole event). This difference was derived from Relativized Minimality and predicts that object clefts can be subject to intervention effects as described for ORs (see Gordon et al. 2001; Del Puppo 2016). Making use of the higher and lower focus fields, in the vein of Belletti (2008), and additionally postulating movement of the whole small clause allowed us to capture the event semantics as well as the discourse effects apparent in these structures.

Our analysis entails in particular that complexity in RCs should not be solely characterized by the presence of wh-movement or the presence or absence of interveners between the target and the original position of the moved constituent, but that a distinction should be based on the depth of embedding. Our classification of RC’s according to depth of embedding, summarized in (31) and repeated here as (38) for convenience, thus highlights the fact that in superficial relatives (38b, c) there is in fact no DP recursion, in other words, no embedding of a CP (containing DP constituents) within a DP. Likewise we do not find true CP recursion since the embedded CP in these constructions is a truncated small clause CP.

(38) a. Genuine relatives (CP inside DP, inside IP; CP inside CP)
   b. Contrastive clefts (XP in left-peripheral focus position, not inside DP; truncated CP)
   c. Presentational focus (DP/small clause CP in low focus position, not inside DP)

Following up on earlier observations of French child language (Labelle 1990, 1996; De Cat 2002; Hamann et al. 2007; Delage et al. 2008; Tuller et al. 2012), we have reported on results of systematic investigation of the notion of depth of embedding in the production of RCs in TD children and also in children with SLI. Taking seriously our analysis of superficial relatives, broadly interpreted to include all those RCs whose structure is in a definable way flatter than a genuine RC embedded within a sentence, we looked at the relative proportions of these types of RCs in child language. The evidence is quite overwhelming in the case of children with SLI, who are particularly sensitive to derivational complexity (as many studies have shown). Our results showed this sensitivity to the combined factors involved in complexity (movement, intervention and depth of embedding) in the SLI group in the high rate of erroneous RCs and also in the high rate of active avoidance, neither of which were found in any of the TD groups. In addition to these general measures of complexity, the children in the SLI group manifested specific sensitivity to the depth of embedding factor. They produced many more unembedded RCs (0-level RCs) than the children in each of the TD groups, including both younger, language-matched children, and age-matched children. Along with the youngest TD
children (the 6-year-olds), the SLI children, who were aged 10 to 12, had far lower rates of genuine RCs compared to all of the other TD groups (including the 8-year-olds). The TD14, on the other hand, were the only group displaying high rates of very deeply embedded RCs (> level-1), significantly more than each of the other groups. As expected, the SLI group provided the clearest picture of the effect of computational complexity in general, and depth of embedding in particular, on syntactic processing, and a developmental trend is found in the TD groups.

We would like to emphasize that the impairment effect and the developmental trend are thus observed in the data on genuine relatives versus superficial relatives, which we uniformly analyzed as not involving DP (or CP) recursion. This trend was even sharper when potential intervention effects, possible in non-SRs, are factored out: SLI and TD-6 once again patterned together, with very low production rates for genuine SRs (9.6% and 5.6% respectively) compared to a significant rise in production of genuine SRs in the older TD groups, which almost reaches the 50% level in the TD-14.

While the factor of intervention (see Friedmann et al. 2009; Belletti and Contemori 2010; Belletti and Rizzi 2012; Belletti et al. 2012) was corroborated by the rarity of non-subject relatives and non-subject contrastive clefts in spontaneous production in our groups, we also found that the Subject/Non-subject dissociation could no longer be clearly established when depth of embedding was factored out.

Analysis of spontaneous production has thus provided significant evidence for the influence of depth of embedding on production, and for the syntactic analysis this was based on, the analysis of superficial RCs as lacking the degree of embedding found in genuine RCs. As predicted, the fact that they are less complex (since they do not involve true recursion) had an impact on frequency of their production in spontaneous speech.

Besides looking at the relation between depth of embedding and intervener and frequency of types of RCs, our study looked at more “active” avoidance of RCs and found children with SLI to behave differently from all of the TD groups, including their age-matches, the TD6. Not only were their sentences containing RCs much more error-prone than those of even the youngest TD children, they made significantly more failed attempts at RC production than TD children (more juxtaposed root clauses, self-interruptions of RCs). Thus, although the children in the SLI 10-12 group produced on average very few RCs (M = 4.1), they seem to have set out to produce more of them, and, when mean number of attempts at RCs (M = 1.3) are added to the RCs which actually surface, the SLI 10-12 group in fact resembles more closely their language matches (TD6 group) for production of RCs (M = 5.4). We believe that these are valuable data, not readily available in elicited production.

Elicited production data (such as those presented in Friedmann et al. 2009), on the other hand, typically give rise to production of 0-level RCs, and thus study intervention effects at this level only or without taking into account whether the RC is embedded or not within a sentence. Study of spontaneous language production provides the possibility of comparing production of 0-level relatives with RCs which are embedded within a matrix clause. Our data have demonstrated a developmental trend in this regard: 14-year-olds produced significantly more RC’s which were deeply embedded, whereas the children with SLI produced significantly more 0-level RCs than the children in the other groups. We have argued that these data show that depth of embedding figures prominently in the calculation of derivational complexity.

Our study of genuine and superficial RCs in spontaneous production, joining a long tradition in child language studies, has analyzed the frequency of occurrence of these
different constructions as a result of the complexity of the linguistic calculation in their derivation. It was suggested that younger children and children with language impairment avoid constructions entailing greater complexity and thus the relative frequency of the least complex constructions is greater in these children. Developmentally, avoidance declines as children are more able to cope with linguistic complexity, and thus the frequency of more complex constructions increases with age. Evidence for this scenario has been provided here, including data from children with SLI which illustrates “active avoidance” of complexity in the form of attempts at RCs, which are abandoned in favor of juxtaposed clauses, etc. This scenario raises the question of look-ahead. How does the child know what to avoid without “looking ahead” to see what would happen if a particular derivation is pursued? This question would seem to be identical to the question of parsing preference in comprehension. Grillo (2012) argued that locality in parsing is a universal principle of processing, and showed that adults are sensitive to complexity of syntactic and semantic structure in comprehension of sentences in which attachment of an RC is potentially ambiguous. He presents cross-linguistic experimental evidence showing that adults in languages in which a pseudo-relative analysis is possible, prefer that interpretation because it is less complex. Here we have argued that children, and particularly those with language impairment, are sensitive to these same considerations in production, preferring pseudo-relatives over genuine relatives because the former entail derivations with less complex syntactic (and semantic) calculation.

We conclude that depth of embedding is a factor which is clearly visible in spontaneous production as greater frequency of occurrence of less deeply embedded structures in younger TD children and in children with language impairment. This factor has particularly sharp consequences in constructions where intervention does not play a role (i.e., SRs), a result which might indicate that in spontaneous production depth of embedding contributes to processing load at least as much as the presence of potential interveners. Our findings concerning depth of embedding confirm the importance of the embedding factor, already pointed out by Kimball (1973). Kimball established a relation between depth of embedding and working memory: the number of incomplete clauses that must be stored in memory at any given moment crucially influences processing load. In more recent terminology, it is the notion of cycle or phase that is pertinent here, in that the difficulty with embedding arises because the calculation cannot be closed as long as the cycle or phase is still open. In particular, it is antecedent-gap relations which cannot be resolved within one clause/cycle and thus have to activate the respective phase edges in order for pertinent features to remain visible for derivation of the higher phases. In these open dependencies the second factor influencing computational complexity, the factor of intervention due to similarity of elements in grammatical features, becomes relevant since depth of embedding and intervention both entail keeping in mind some structural relation/features while another calculation is being undertaken. Deep embedding requires keeping clauses in mind, and, potential intervention, as Friedmann et al. suggest, requires a calculation of which DP goes with the gap via comparison of (sub)sets of features. Clearly, these two factors interact, as we have argued: both crucially impact the processing of unresolved dependencies. In the case of RCs and superficial relatives, which we have analyzed as involving true DP and CP recursion in the first, but not the latter case, there may be another property which these factors share. Intervention and ultimately Relativized Minimality effects are due to

21 Whereas intervention has been prominent in processing theories, depth of embedding has not. How existing theories, such as Gibson’s (1998), could eventually incorporate this notion is beyond the scope of this article.
structural similarity which makes it difficult to keep elements differentiated during computation. Genuine RCs involve true recursion, which by definition requires structurally identical elements (DPs and CPs). Under the assumption that other cognitive components such as working memory interface with grammar, storage during computation may be difficult because of similarity. Superficial relatives, as we have argued, do not involve this kind of true recursion, i.e. similarity, and are easier to process. Depth of embedding, and recursion in particular, i.e. the presence of dependencies across similar cycles and the occurrence of similar syntactic objects in these dependencies, can therefore be identified as the factors pertinent in the computation of RCs.

References
Genuine versus superficial relatives in French: The depth of embedding factor


Crain, Stephen, Cecile McKee, and Maria Emiliani. 1990. “Visiting relatives in Italy”. In Language Processing and Language Acquisition, ed. by Lyn Frazier, and Jill de Villiers, 335-356. Dordrecht: Kluwer.


Goodluck, Helen. 2010. “Object extraction is not subject to child relativized minimality”. Lingua 120(6): 1516-1521.


Prévost, Phillipe, Laurice Tuller, Maruene Scheidnes, Sandrine Ferré, and Martin Haiden. 2010. “Computational complexity effects in the acquisition of *Wh*-questions in child L2 French”. In *Selected proceedings of the romance Turn III*, ed. by L.
Cornelia Hamann & Laurice Tuller

Soares, Carla. 2006. La Syntaxe de la périphérie gauche en portugais européen et son acquisition. Doctoral dissertation, Université de Paris 8, Saint-Denis.
Abstract: We present data from Aromanian varieties spoken in South Albania, including the towns of Divjakë and Fier. Unlike Romanian and like Albanian, Aromanian has preadjectival linkers. Furthermore, Aromanian has linkers in front of both datives and genitives and agreeing with the latter. These configurations are absent from Albanian and Romanian, which have linkers in front of genitives, but not of datives, and agreeing with the head noun. The fact that the same elements that appear as linkers also occur as demonstratives/articles leads us to categorize them as Ds. We propose that the pre-adjectival linker provides a (partial) saturation for the argument of adjectival predicates, to be ultimately satisfied by the head DP. We characterize oblique case as introducing an inclusion/part-whole relation, which takes the oblique DP as one of its arguments (i.e. the whole, or possessor, etc.). The linker provides a (partial) lexicalization of the second argument (i.e. the part, or possessee etc.).

Keywords: linker, genitive, dative, agreement, Determiner

1. Introduction

In many languages a linker element is inserted between a noun and an adjective that modifies it or a (genitive) complement that the noun embeds or a relative clause (not considered here). Among Indo-European languages, the Iranian ezafe is generally taken to be such an element. According to traditional descriptions (Lazard 1992), the ezafe indicates nothing about the precise semantic or syntactic nature of the relation holding between the modifier/complement and the head-noun. While in Persian the ezafe is invariable (-e), in Kurdish varieties, the ezafe agrees with the head noun (Holmberg and Odden 2008; Haig 2011); thus, any account of linkers must encompass a certain amount of variation.

How much variation is admissible, and what kind, depends on the theory. For instance den Dikken and Singhapreecha (2004, fn. 31) explicitly exclude that Greek polydefiniteness counts as an instantiation of linker structure. On the other hand, Androutsopoulou and Espanol-Echevarria (2007) start from Greek in their survey of the phenomenon, and Larson and Yamakido (2008) also include Greek in theirs. Manzini et al. (2014), Franco et al. (2015) discuss the relation between another Balkan language, namely Albanian, and Kurmanji Kurdish. The Albanian article (as it is called in traditional grammars) has the same distribution observed for the ezafe (and the Greek article), namely before adjectives
and genitives\(^1\). Albanian has a specialized series of nominal endings, inflected for definiteness as well as for phi-features and case\(^2\); the pre-adjectival/pre-genitival articles are related to the definite endings with which they often coincide. This is illustrated in (1) for pre-adjectival contexts and in (2) for pre-genitival contexts\(^3\).

(1) a. ěrđi  dial-i  i  mað  
    came  boy-nom.m.def  Lkr.m  big  
    'The big boy came'

b. ěrđi  vaiz-a  e  mað-e  
    came  girl-nom.f.def  Lkr.f  big-f  
    'The big girl came'

c. ěrđən  dirn-t  tə  maðin-t  
    came  boy-pl.def  Lkr.pl  big-pl  
    'The big boys came'

(2) a. libr-i
    book-nom.m.def  Lkr.m  brother-m.obl.def  
    'the book of the brother'

b. putr-a  e  cεn-it  
    leg-nom.f.def  Lkr.f  dog-m.obl.def  
    'the leg of the dog'

In Romanian, no linker needs to appear between a noun and an adjective – though the so-called strong (i.e. non-clitic) form of the definite article (cel etc.) may appear in linker position, as in (3). Importantly, cel is mutually exclusive with demonstratives, pointing to an operator-like content for it, denoting familiarity (Cornilescu and Giurgea 2013), which seems to be missing from, say, the Albanian article. Only genitives, as in (4), are generally introduced by a linker agreeing with the head noun (al etc.) – which can be left out only under adjacency with a definite head noun. As in Albanian, al is a form of the definite article (Lat. ille) (Giurgea 2013).

(3) maşin-a (cea)  nouă  
    car-the.f  (the.f)  new.f  
    'the new car'

\(^1\)To be precise, it is a lexically defined subset of adjectives that takes the article (Camaj 1984; Solano 1972; Turano 2004; Campos 2008). We have nothing to say on those adjectives that do not take it. If uniformity of structures is desired, then we must conclude that apparently article-less (post-nominal) adjectives have an empty article.

\(^2\)The formal literature treats these endings as post-nominal articles derived via movement of N to D (Dimitrova-Vulchanova and Giusti 1998; Turano 2002, 2003; cf. also Dobrovie-Sorin 1994 on Romanian). However, Albanian has a system of prenominal articles (e.g. with kinship terms) which can combine with definite inflections. This makes the implementation of a movement analysis difficult. Therefore we assume direct Merge in inflectional position.

\(^3\)Data reflecting standard Albanian are taken from an informant of Gjirokastër, in South Albania and transcribed in a broad IPA to facilitate morphological parsing. The same broad IPA transcription will be used for the Aromanian data.
Linkers in Aromanian in comparison to Albanian (and Romanian)

(4) două kâmăș-i ale băiat-ul-ui
two shirts-fpl the.fpl boy-the-obl
‘two shirts of the boy’

Against this background, we focus on the discussion of linkers in Aromanian, explored *per se* and in comparison with the distribution of linkers in the cognate language Romanian and in the language in contact, Albanian; our data refer to Aromanian varieties spoken in South Albanian areas, including the towns of Divjakë and Fier.

2. The Aromanian of Divjakë and Fier

As a preliminary to the discussion to follow, we illustrate the system of nominal inflections in Aromanian. The direct vs. oblique Case distinction in Aromanian is not preserved in the masculine singular, except for the 3rd person pronoun; the definite inflection –u contrasts with the zero inflection for indefinites, as shown in (5). Note that the oblique covers both the dative (5b) and the genitive (5c).

(5) a. ari vənitə/am vəzutə fitʃor-u/ un fitʃor/ atse-u
    has come/I have seen boy-msg/a boy/ that-msg
    ‘The/a boy/he has come’/’I have seen the/a boy/him’

b. i  o am datə o fitʃor-u/ o un fitʃor/ ots-ui
    him it I have given Lkr boy-msg/ Lkr a boy /that-obl.msg
    ‘I gave it to the/a boy/him’

c. libr-a o fitʃor-u/ ots-ui
    the book Lkr boy-msg/ that-obl.msg
    ‘the boy’s/his book’

The feminine singular presents case distinctions (direct vs. oblique) and definiteness distinctions (at least in the direct case), as in (6).

(6) a. ari vənitə/ am vazutə fet-a/
    has come/I have seen girl-def.fsg/ a girl/ that-fsg
    ‘The/a girl/she has come’/’I have seen the/a girl/her’

b. i  o am datə ali fet-i/
    them it I have given Lkr girl-oblfsg/ Lkr a girl-oblfsg/that-obl.fsg
    ‘I gave it to the/a girl/ her’

c. libr-a ali fet-i/ ots-jei
    book-def.fsg Lkr girl-oblfsg/ that-obl.fsg
    ‘the girl’s/ her book’

Case and definiteness distinctions are present in the plural both for the masculine (7) and for the feminine (8). Note that in the oblique plural there is a single form for the masculine and feminine pronoun. In order to process the data it is useful to keep in mind that –γ is a phonological alternant of –l-.

(7) a. ari vənitə/ am vazutə fitʃor-ja/ ndoi fitʃor/ atse-i
    have come/I have seen boy-mpl/ some boy/ that-mpl
Preadjectival linkers, in the sense briefly defined in section 1, are generally present in Aromanian 4, unlike in Romanian and mimicking closely Albanian. The demonstrative series, seen in (5)-(8) in its pronominal usage, is also deployed as a linker. The linker agrees with the head noun in gender, number, and case as shown in (9)-(10) for the masculine (singular and plural). It is generally excluded in contexts with an indefinite noun, cf. (9a"), recalling the ‘polydefiniteness’ distribution of Greek. It is also excluded in the presence of the comparative element ka-ma ‘lit: how-more’, as in (9a’), (10a’), though the linker may optionally combine with the simple form ma ‘more’, preceding it, as in (9a’). Finally, the adjective agrees in gender and number – and to some extent in case, cf. the oblique plural in (10b’ ). It is also sensitive to the definite or indefinite nature of the head DP. Thus in definite direct contexts the adjective combines, say, with –u in the singular masculine in the definite (9a), but is inflectionless in the indefinite (9a”) . This seems to indicate that it agrees with the head noun in definiteness as well.

(9) a. fitfor-u (a)tse-u mar-u/ yung-u/ fkurt-u
    boy-msg Lkr-msg big-msg/ tall-msg/ short-msg
    ‘the big/tall/short boy’
   a’. fitfor-u kama mari/ fitfor-u atse-u ma mar-u
    boy-msg more big/ boy-msg Lkr-msg more big-msg
    ‘the bigger boy’
 a”. un fitfor mari
    a boy big
   b. o fitfor-u ats(-uy)ui mar-u/ yung-u/ fkurt-u
    Lkr boy-msg Lkr-obl.msg big-msg/ tall-msg/ short-msg
Linkers in Aromanian in comparison to Albanian (and Romanian)

‘to the big/tall/short boy’

(10) a. fitxor-jə (a)tse-jə mar-jə
    boys-mpl Lkr-mpl big-mpl
    ‘the big boys’
    a’. fitxor-jə kama mari
    boy-mpl more big
    ‘the bigger boys’
    b. o fitxor-ju ots-uyor mar-jə
    Lkr boy-obl.mpl Lkr-obl.pl big-mpl
    ‘to the big boys’
    b.’ o fitxor-ayu otsa-yoru mar-uyu
    Lkr boy-obl.mpl Lkr-obl.pl big-obl.mpl
    ‘to the big boys’

Similar conditions are found in the feminine, as illustrated in (11)-(12) for the singular and plural respectively. The example in (13) shows that the reduced declension (generally only phi-features) realized on adjectives does not correspond to a morphological limitation. Note the limited extent to which case is present on adjectives, namely optionally in (11b); by contrast, the nominalized adjective in (13) is obligatorily inflected for case.

(11) a. fɛt-a ats-ɛ mar-ɛ/ yung-a / jkurt-a
    girl-fsg Lkr-fsg big-fsg / tall-fsg / short-fsg
    ‘the big/tall/short girl’
    a’. fɛt-a kama mari
    girl-fsg more big
    ‘the bigger girl’
    a”, un fɛt-ə mari
    a girl-fsg big
    ‘a big girl’
    b. ali fɛt-i ats-jei mari / yung(-i)
    Lkr girl-obl.fsg Lkr-obl.fsg big / tall-obl.fsg
    ‘to the big/tall girl’

(12) a. fɛt-ə-li atse-li marə-li
    girl-fpl Lkr-fpl big-fpl
    ‘the big girls’
    b. o fɛt-uyu ots-uyor marə-li
    Lkr girl-obl.fpl Lkr-obl.pl big-fpl
    ‘to the big girls’

(13) ali jkurt-i
    Lkr-fsg short-obl.fsg
    ‘to the short one’

5 In contexts demonstrative-noun the definite form of the noun is possible, though not necessary, as in Albanian. Thus, we can find atse-li məʃer next to atse-li məʃer-li ‘those women’. This may have some relevance in the present connection.
The schema in (14) shows a summary of phi-features, definiteness and case inflections in Aromanian, limited to the forms that we have chosen to illustrate here\(^6\).

(14) a. Aromanian Definite nominal inflection

<table>
<thead>
<tr>
<th></th>
<th>ms</th>
<th>fs</th>
<th>mp</th>
<th>fp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Acc</td>
<td>u</td>
<td>a</td>
<td>jə</td>
<td>li</td>
</tr>
<tr>
<td>Dat/Gen</td>
<td>u</td>
<td>-i</td>
<td>ju/υγυ</td>
<td>υγυ</td>
</tr>
</tbody>
</table>

b. Aromanian Indefinite nominal inflection

<table>
<thead>
<tr>
<th></th>
<th>ms</th>
<th>fs</th>
<th>mp</th>
<th>fp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Acc</td>
<td>∅</td>
<td>ϕ</td>
<td>∅</td>
<td>fp</td>
</tr>
<tr>
<td>Dat/Gen</td>
<td>∅</td>
<td>i</td>
<td>∅</td>
<td>i</td>
</tr>
</tbody>
</table>

Now, from the discussion that precedes it emerges that there are in fact two candidates for linker status in Aromanian. In (15) we schematize oblique introducers, found in front of genitives and of datives. The \( o \) introducer of genitive/ dative coincides with the object clitic ‘him/her’; the element \(-li\) coincides not only with the plural inflection but also with the plural object clitic ‘them’. They appear to coincide with the Romanian morphological series – but differ from Romanian in two important respects. First in Romanian, as seen in (4) the pre-genitival linker agrees with the head noun; in Aromanian the linker agrees with the genitive, as seen in the (c) examples in (5)-(8). The other difference is that in Romanian, the linker series related to (15) only appears in genitive contexts of the type in (5c), (6c) etc.; dative complements are externalized by the oblique DP without any preposed linker. In Aromanian, on the contrary, linkers are obligatory in front of datives, as illustrated in the (b) examples of (5)-(8) and of (9)-(12).

(15) Aromanian pre-genitive/dative linkers

<table>
<thead>
<tr>
<th></th>
<th>ms</th>
<th>fs</th>
<th>mp</th>
<th>fp</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>ali</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
</tbody>
</table>

The second candidate for the role of linkers are pre-adjectival ones, lexicalized by the demonstrative, whose declension we summarize in (16). An interesting property of the demonstrative is that it seems to lexicalize dative contexts without need for an introducer of the series in (15). However we may consider that the change in colour of the vocalic initial from \( a-\) in the direct cases to \( o-\) in the oblique is due to the fact that the element \( o-\) is incorporated in the oblique.

(16) Aromanian pre-adjective linkers

<table>
<thead>
<tr>
<th></th>
<th>ms</th>
<th>fs</th>
<th>mp</th>
<th>fp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom/Acc</td>
<td>atse-u</td>
<td>atse(-a)</td>
<td>atse-jə</td>
<td>atse-li</td>
</tr>
<tr>
<td>Dat/Gen</td>
<td>ots-(υγυ)ui</td>
<td>ots-jei</td>
<td>ots-υγυor</td>
<td></td>
</tr>
</tbody>
</table>

Morphologically, the declension of \( ats-\) seen in (16) is the same when it plays a referential role, for instance as the 3rd person pronoun in (5)-(8), and when it functions as a pre-adjectival linker, for instance in (b)-(c) examples in (5)-(8). It can also occur in a

\(^6\) For instance we have not illustrated the forms in \(-i\) such as \( kənɪ \) ‘dog’.

88
demonstrative function proper, as in (17) and combine with a pre-adjectival linker; of the two occurrences of *ats-* is the first one that determines the demonstrative reading of the DP; the linker does not.

(17) u o m datə ats-ɔyor doi fitʃor tso/ats-ɔyor marə
to.them it I.have given those-obl.pl two boy-mpl Lkr-obl.pl big
‘I have given it to those two boys’

The data in (17) once again differentiate Aromanian from Romanian. As discussed by Cornilescu and Giurgea (2013, 408), the *cel* element that optionally appears in pre-adjectival position in Romanian is in complementary distribution with other demonstratives, as in (18), pointing to the fact that it maintains (part of) the D force of the demonstrative.

(18) *acele case  cele vechi
those houses the old

Romanian, then, lacks pre-adjectival linkers, like other Romance languages. On the contrary, the Aromanian distribution parallels closely that of Albanian. In (1) we have already illustrated the distribution of the definite inflections of the noun and of the pre-adjectival linker in the nominative. In (19) we provide two examples in the accusative. The comparison between (1) and (19) shows that the form of the liker is sensitive to the case of the head noun, while comparison between (19a) and (19b) shows that it is sensitive to the definiteness of the head noun. In (20) we exemplify an oblique context.

(19) a. patʃ dial-in / vaiz-ən ɛ mað/mað-ɛ
I.saw boy-acc.m.def / girl-acc.f.def Lkr big/big-f
‘I saw the big boy/girl’

b. patʃ ɲə dial/ vaiz ɛ mað/mað-ɛ
I.saw a boy / girl Lkr big/big-f
‘I saw a big boy/girl’

(20) j-a δatʃ dial-ıt / vaz-əs to mað/mað-ɛ
him/her-ıt I.gave boy-obl.m.def / girl-obl.f.def Lkr big/big-f
‘I gave it to the big boy/girl’

As seen in (1), the pre-adjectival linker takes the form *i* for the masculine singular, *ɛ* for the feminine singular and *tə* for the plural in the context of a nominative noun. An accusative definite noun is followed by *ɛ* in the singular; an indefinite selects *tə*, as in (19). If the noun is oblique the linker is *tə* for masculine singular, as in (20) and (21a). The form of the linker, *sə*, for feminine singular oblique is illustrated in (21b) with a pre-genitival context. The same linker paradigm characters pre-adjectival and pre-genitival contexts.

(21) a. j-a δatʃ dial-ıt to mətr-əs
him-ıt I.gave boy-obl.m.def Lkr sister-obl.f.def
‘I gave it to the child of the sister’

b. para putr-əs ə cen-ıt
before leg-obl.f.def Lkr dog-obl.m.def
‘before the leg of the dog’

The linker of Albanian is schematized in (22) (Franco et al. 2015). The pronominal clitic forms, correspond to a subset of the linkers, namely ē for the accusative singular (‘him/her/it’) and I for the oblique singular and the accusative plural (‘them/to him/to her’).

(22) a. Albanian linkers with definite head nouns

<table>
<thead>
<tr>
<th>Case</th>
<th>ms</th>
<th>fs</th>
<th>pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>i</td>
<td>ē</td>
<td>tə</td>
</tr>
<tr>
<td>Acc</td>
<td>ē</td>
<td>ē</td>
<td>tə</td>
</tr>
<tr>
<td>Obl</td>
<td>tə</td>
<td>sə</td>
<td>tə</td>
</tr>
</tbody>
</table>

b. Albanian linkers with indefinite head nouns

<table>
<thead>
<tr>
<th>Case</th>
<th>ms</th>
<th>fs</th>
<th>Pl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nom</td>
<td>i</td>
<td>ē</td>
<td>tə</td>
</tr>
<tr>
<td>Acc</td>
<td>tə</td>
<td>tə</td>
<td>tə</td>
</tr>
<tr>
<td>Obl</td>
<td>tə</td>
<td>tə</td>
<td>tə</td>
</tr>
</tbody>
</table>

In short, in both Aromanian and Albanian, pre-adjectival linkers agree with the head noun in phi-features and case. In both languages they are sensitive to the definiteness of the head noun, since in Romanian only definite head nouns admit of linkers and in Albanian the definite and indefinite linkers paradigms differ along the lines of (22). In both languages the adjective takes on nominal class (gender) and number inflection agreeing with the head noun, though in Aromanian it also marginally displays case.

On the basis of the morphological evidence seen so far, but also of syntactic and interpretive evidence to be analyzed in later sections, linkers (at least in the languages considered) are close to what is usually called agreement. If linkers are agreement heads, the key theoretical question is why they would surface in the form of definiteness morphology, namely articles in Albanian and even demonstratives in Aromanian.

3. Previous analyses of linkers

The generative literature on linkers is deeply influenced by the model of nominal embedding provided by the of Insertion rule (Chomsky 1981). In English, given the lexical items red and ball, syntactic merger can take place without any extra material being inserted, yielding red ball. Yet merger of the book and John does require an extra element to be inserted, namely of, as in the book *(of) John. For Chomsky (1981) of is inserted in order to assign case on John when governed by an N head (cf. Vergnaud 2008 [1978]). In later literature, the occurrences of of in contexts such as that idiot of a student (roughly ‘that idiotic student’) are taken to parallel that of the copula in sentential domains (Hoekstra 1999; den Dikken 2006). Other scholars emphasize the role of of in identity avoidance (breaking an *N-N string, Richards 2010). The range of theoretical proposals on linkers closely reproduces the range of theories on of Insertion, as case assigners, as copulas, as means for identity avoidance.

There are good reasons for rejecting these various proposals. For example, the construal of linkers as copulas, proposed by den Dikken and Singhapreecha (2004) is undermined by the observation that in Albanian the pre-adjectival linker is not restricted to DP-internal
Linkers in Aromanian in comparison to Albanian (and Romanian)

contexts, but appears in copular constructions, as in (23), where the copula is independently lexicalized. The same is true of the pre-genitival linker, as illustrated in (23’) with an example from the Arbëresh (Italo-Albanian) variety of Vena (Manzini and Savoia 2007).

(23) a. ɐʃt i kuc/ɛ kuc-ɛ
    s/he.is Lkr.m red/ Lkr.f red-f
    ‘S/he is big’

b. jan tə kuc / kuc-ɛ
    they.are Lkr.pl red-mpl/ fpl
    ‘They are fat’

(23’)

kjɔ tʃt tə neri-utɔ
this is Lkr man-Obl.msg
‘This is of the man’s’

The conclusions from Albanian are confirmed by Iranian languages, often taken as paradigmatic examples of linker languages. In the Bahdînî dialect of Kurmanjî Kurdish in (24) a linker je(t)/ja: agreeing with the subject precedes the sequence adjective - enclitic copula. The fact that in predicative contexts the linkers are not in complementary distribution with the copula but combine with it, is again an argument in favour of separating linkers from copulas.

(24) a. au je / ja: mazən-e
    3sg Lkr.m / Lkr.f big-is
    ‘(s)he is big’

b. au jet sur-ən
    3pl Lkr.pl red-are
    ‘they are red’

(Kurmanjî Kurdish, Bahdînî dialect)

Note that in Aromanian adjectives are not preceded by the demonstrative linker in copular context, though linkers precede post-copular genitives, as in (25).

(25) atse esti o fitʃor-u/ ali majer-i
    this is Lkr.obl boy-def/ Lkr-obl.f woman-obl.f
    ‘This belongs to the boy/the woman’

In the context of the present discussion it is of particular relevance that Campos (2005), Campos and Stavrou (2005) propose a construal of linkers as copulas for Greek and Aromanian. For them, each modifier of N is introduced as part of a small clause PredP. The article in Greek is a lexicalization of the Pred head; according to Campos (2008) at least the pre-genitival linker of Albanian follows the same model, as schematized in (26a) for example (2a) above. Unfortunately this idea clashes with the fact that the linker co-occurs

---

7 Our informants give us the same forms as Haig’s (2011) for feminine singular and for plural; in the case of the masculine singular we obtained the form jet, as reported in some examples, differently from the only form (y)e recorded by Haig.
with a true copula in sentences like (23'). For Campos and Stavrou, on the other hand, the demonstrative in Aromanian is the subject of the predication, as in (26b) – which seems a much more natural role for a D element.

(26) a. $[\text{DP} \text{libri} \ [\text{PredP} \ [\text{Pred} \ i \ [\text{DP} \ vəða-it}}$
    b. $[\text{PredP atseau} \ [\text{Pred} \ φ \ [\text{AP maru}}$

Another line of work takes linkers to semantically licence the possession relation. For Larson and Yamakido (2008), linkers are necessary to case licence $+N$ complements of N heads, including adjectives. The data of Albanian suggests a different conclusion, namely that the oblique case morphology of Albanian is sufficient to support the possession relation, as shown by the fact it is sufficient to introduce the possessor in dative contexts, for instance in (21a) or in (27). Datives are connected to possession in the formal literature at least since Kayne (1984). Furthermore, the Albanian linker reproduces the agreement features of the head noun, and indeed in (27) it replicates exactly the inflection of the head noun (non-ambiguously an oblique feminine definite). We may wonder why the linker would solve any problem with $+N$ embedding that the nominal inflection couldn’t itself solve.

(27) ja $\delta\alpha\tau\acute{f}$ vaiz-əs $(s\circ \ ma\delta-ɛ)$
    to.her-it I.gave girl-obl.f,def Lkr.obl.f big-f
    ‘I gave it to the (big) girl’

Contrary to Albanian, Larson and Yamakido’s construal of linkers as case licencers has a certain prima facie plausibility for Persian, where there is no overt case morphology. Nevertheless, in Kurmanji Kurdish a direct vs. oblique case distinction is available and the possessor is invariably marked oblique; despite this, it is introduced by the ezafe, as in (28a). This is true, notwithstanding the fact that the oblique inflection alone is able to lexicalize the possessor in dative environments, as in (28b).

(28) a. dest-ɛ kurk-i / ketʃk-ɛ
    hand-Lkr,m boy-obl,m/girl-obl,f
    ‘the hand of the boy/girl’
    b. de qalam-ak-i dama ketʃk-ɛ / kurk-i
    progr pen-one-obl give-1sg girl-obl,f/ boy-obl,m
    ‘I give a pen to the girl/boy’

Campos (2008, 1027) argues that “in spite of the parallelism between Greek and Albanian … Albanian constructions with adjectival articles cannot be analysed as polydefinite constructions and should be better analysed as containing a complex adjectival head”. However this forces him to invoke a process of grammaticalization to relate pre-adjectival linkers to pregenitival ones: “adjectival articles could have originated as polydefinite constructions, parallel to the structures in Greek and Aromanian … where the (adjectival) article later got grammaticalized … This would explain why the same set of adjectival articles is used with possessives and why the same restrictions that apply to adjectival articles are also applicable to possessive articles” (1029). For him, as a consequence of the grammaticalization process “the adjectival article and the adjective form a complex adjectival head A in modern Albanian” (1026). This set of additional assumptions is unnecessary under the present treatment, which is therefore simpler in this respect.
As for Aromanian, in (29a-a’) we reproduce examples from section 2 showing that the pre-adjectival demonstrative has the same inflectional properties as the head noun – which can further be duplicated on the adjective. In turn, the linker in dative and genitive contexts introduces a DP endowed with rich case inflections; a relevant example is reproduced in (29b).

(29) a. o fitjor-o-yu otsa-yor u mar-u-yu
   to the boys Lkr-obl big-obl.pl
   ‘to the big boys’

a’. o fit-u-yu ots-u-yor mara-li
   to the girls Lkr-obl big-pl
   ‘to the big girls’

b. i o ded o fitfor-ju/ o fit-u-yu
   him/her it I gave Lkr boys-obl/Lkr-obl girls-obl
   ‘I gave it to the boys/to the girls’

A final family of accounts for linkers takes them to be means for identity avoidance. This approach has recently been revived by Richards (2010) as part of a more general account of identity avoidance/ syntactic haplology in morphosyntax (Yip 1998; Neeleman and van de Koot 2006; van Riemsdijk 2008; Manzini 2014). Empirical reasons lead us to doubt that linkers are part of this phenomenon. Linkers occur in copular context, cf. (23)-(25) above, where they do not avoid any type of N-N identity.

We believe that much of the theoretical literature about linkers provides important insights into the nature of the elementary components that enter into adjectival modification and predication and into possessor embedding. However we conclude that the linker is not a copula, nor a case assigner, nor does it introduce the possession predicate, nor is it an identity avoidance device.

4. Analysis of Albanian linkers

As we saw in section 3, the Albanian the linker-adjective sequence is not restricted to noun phrase internal contexts, but appears in predicative contexts with an overt copular ‘be’. Importantly, copular sentences provide us with a straightforward argument for constituency. The linker that appears in front of the adjective, following the copula, is part of the structure of the AP, as in (30). Hence in complex nominals as well, it is not a functional projection of the head noun, but rather of the modifier AP (or of the genitive DP). Indeed the formal literature on Albanian concludes – much as we do here – that the article is part of the adjectival constituent (Dimitrova-Vulchanova and Giusti 1998; Turano 2002, 2003; Giusti and Turano 2007). Following Manzini and Savoia (2011a, b), in (30) we further assign the linker head to the D category, based on the morpholexical identity of linker elements with clitic pronouns (ɛ, i – cf. the discussion surrounding table (22)) and with definite nominal inflections (often analysed as postposed definite articles, cf. fn.2). Following Manzini and Savoia, we categorize the adjectival inflection as an N exponent (for Nominal class/gender).

(30) \([D \varepsilon [A m\alpha [N -\varepsilon]]]\)
An analogous structure and categorization of linker material can be proposed for Iranian languages, as shown in (31) for Bahdiñ Kurmanji, cf. example (24). This is consistent with occurrences of the so-called ezafe as a ‘stand-alone’ element (a demonstrative) and as a ‘tense’ element (Haig 2011), analysed by Franco et al. (2015) as a subject clitic.

(31) $[D\ A: \ [\lambda \ mazen]]$

The case of Aromanian is also telling, since it recruits the demonstrative as an adjectival linker, i.e. an element standardly associated with definite denotation and with the D position of the DP. The analysis in (30)-(31) then extends to Aromanian, as in (32).

(32) $[D\ (a)tse\ [N\ -u]]\ [\lambda \ yuNg\ [N\ -u]]$

Summarizing our conclusions so far, linkers most often vary according to the phi-features, case and definiteness properties of the head noun being modified (section 2). Second, the same elements that appear as linkers/agreement also occur with demonstrative/determiner interpretation, as we have just seen. The second fact has led us to categorize them as Ds; the first fact suggests that, whatever else they may be, they are agreement elements. The theoretical question we are faced with is what a referential category like D may have to do with agreement morphology, which is taken not to contribute to interpretation in traditional approaches, and in recent generative ones (Chomsky 1995) alike.

According to Dimitrova-Vulchanova and Giusti (1998), writing on Albanian, the pre-adjectival article is just a ‘redundant’ agreement. One of the central tenets of current minimalist theory is that agreement results from an uninterpretable set of features (a probe) seeking a matching interpretable set of features (a goal) for checking (i.e. deletion or valuation of the uninterpretable set). According to Toosarvandani and van Urk (2012), writing on the Iranian language Zazaki, linkers are probes, i.e. they are associated with uninterpretable phi-features. This captures the connection between linkers and agreement morphology, in terms of the notion of probe. However, this may not be the right way to go if we want to explain why linkers overlap with Ds, i.e. definite determiners and pronouns. It is true that clitic pronouns have been treated in the minimalist literature as pure bundles of phi-features (i.e. as ϕPs, cf. Roberts 2010), but apart from any other problem, demonstratives, as in Aromanian, seem unlikely candidates for such a status. The other logical option is to start from the D, hence presumably interpretable, status of linkers and see whether the continuity of linkers (deteminers, demonstratives) with agreement can be captured this way.

Following Manzini and Savoia (2011a, 2011b), we take it that Albanian linkers have at least one important semantic property in common with D determiners, for instance in English. We apply the analysis, fairly standard in the literature (cf. Higginbotham 1985; Williams 1994), whereby Ns, even non-eventive ones, are predicates and have an argument slot (called the R-role). In English the determiner D saturates the argument of N according

---

9 Tomić (2006), quoted by Campos (2008, 1009), characterizes the Albanian pre-adjectival article as ‘agreement clitic’.

10 In Zazaki the ezafe is identical to the demonstrative and to the third person singular agreement marker, as in Kurmanji Kurdish.
to Higginbotham (1985)—and we can assume that the same role is played by definite nominal inflections in Albanian and Aromanian. In the same way, the adjectival predicate must be satisfied by an argument, which is provided by the D element in (30), i.e. the linker, in a language like Albanian. In other words, in (30) the linker/D element ɛ provides a (partial) lexicalization for the argument of the predicate mað ‘big’ to be further fixed by the subject of a copular sentence of by the head noun of the DP. This also lays the bases for the common lexicalization with pronominal clitics, i.e. D arguments saturating verbal predicates – eventually doubled by full DPs in so-called clitic doubling.

At the same time, there are also differences between determiner Ds and linker Ds. Distributional differences are particularly easy to detect. To take just English, the D determiner precedes all material with which it can co-occur, including quantifiers, as in the three/many/few children; alternatively it is in complementary distribution with other quantifiers, as in the/every/no child. On the contrary, in Albanian, elements quantifying over the adjective precede the D linker, as in (33). Therefore (33) suggests that the linker D is inserted within the AP in a position lower than the one the determiner D fills within DPs.

(33) me/fum ɛ mað-ɛ
more/much Lkr big-f
‘bigger/very big’

More evidence on the low position of the linker D comes from instances where the same lexical bases that we have considered so far as adjectives are nominalized. As other nouns, they are inflected for case and definiteness, displaying the full system of nominal inflections. At the same time, they are also preceded by the determiner. The latter is embedded under quantifiers of the noun, including the indefinite article, as in (34). The structure of a DP like (34) can then be schematized as in (35), where the linker D and the determiner D co-occur – the linker in a lower position and the determiner in a higher position.

The question is how a structure where two Ds are present, as in (34)-(35), is to be interpreted. The higher D, i.e. the determiner, is interpreted in the standard way – namely as indicating that there is an individual (or a set of individuals, or a unique/familiar/etc. individual) on which the properties of the NP predicate and those of the sentential predicate overlap, i.e. as a quantifier. On the other hand, the lower D, i.e. the linker, simply values the argument slot of A, but does not provide a quantificational closure, nor lead to a referential interpretation – the latter is provided by the higher D.

(34) crð pa ɛ maðɛ
he.came a Lkr.f big.f
‘A big one came’

(35)

\[ \text{DP} \]
\[ \text{AP} \]
\[ \text{D} \]
\[ \text{A} \]
\[ ɛ \]
\[ maðɛ \]
The relation of determiner and linker Ds is essentially the same as between pronominal clitics and doubling clitics within the sentential domain. Indeed we noticed that in Albanian not only articles are a subset of nominal inflections – but pronominal clitics are a subset (i, ε) of linkers. Two interpretations are available to pronominal clitics. In non-doubling contexts the clitic has referential import, and is capable of deictic or anaphoric pronominal reference. On the other hand when a doubling DP is present the clitic is interpreted as a bound variable of it.

At this point of the discussion we are ready to define a linker (or at least the Albanian linker). What a linker D and a determiner D have in common is that they are both able to satisfy argument slots. What they do not share depends on their different position of merger. A D closing off the DP is an operator, establishing a relation between a restrictor (the NP) and a domain of quantification (a VP). A linker D is a bound variable of the higher D – it provides a satisfaction for a theta-role ultimately bound by the higher D. In other words, it has the meaning of a bound pronominal that satisfies the adjectival role, prior to the introduction of higher operators.

For completeness, let us consider the embedding of an AP under a larger DP, for example in (1b), with the structure in (36). According to the discussion that precedes, in (36) the adjective mað- ‘big’ is a property, i.e. has a single, obligatory argument position, suggested in (36) by the λx notation (cf. Adger and Ramchand 2005 on the Λ feature); the pre-adjectival linker ε provides a satisfaction of the argument slot of the predicate. A fortiori, the same is true of the –a definite inflection of the noun, satisfying the R-role of the noun (here λy). Following Higginbotham (1985), we assume that adjectival modification involves the identification of the theta-role of the adjective with the R-role of the noun. In other words, in (36) there is ultimately a single argument, satisfying both the predicate ‘girl’ and the predicate ‘big’; the referent denoted by the complex DP correspondingly lies at the intersection of the ‘big’ and ‘girl’ properties.

A formal possibility that we further suggest in the structure in (36), is that the N class inflection -ε on the adjective is like the linker in that it provides a (partial) saturation of the argument slot of the nominal predicate. In other words, the connection between linkers and agreement is not that linkers are uninterpretable heads i.e. the head counterpart to uninterpretable phi-features inflections in minimalist theory (see Philip 2012). On the contrary, phi-features inflections are endowed with elementary interpretive content, which

As for DP-internal word order, Dimitrova-Vulchanova and Giusti (1998) generate the order Noun-Adjective by movement of the Noun to a Focus position. Turano (2002, 2003), following Cinque (1999), derives the Noun-Adjective order by movement of N to D. However following Abels and Neeleman (2012), generating the noun-adjective order does not require Cinque-type movement.
Linkers in Aromanian in comparison to Albanian (and Romanian)

This idea is developed in more detail in Manzini and Savoia (2011a, 2011b). It is worth remarking that it is compatible with the Minimal Search and Match conception of Agree in Chomsky (1995, 2001) – except that in the absence of uninterpretable inflections, Agree can no longer be triggered by the need to delete/value uninterpretable features before LF and the application of Full Interpretation. Rather, Manzini and Savoia suggest that the trigger is Full Interpretation, in so far as it forces inflections and free standing elements that concur to the satisfaction of the same argument slot to be identified as picking up a single referent.

4.1 Pregenitival linkers

In order to understand the role of pre-genitival linkers, it is necessary to consider the nature of genitive case – or of oblique case, which in languages like Albanian subsumes both genitive and dative contexts. ‘Possessor’ is the traditional characterization of genitives. It is also natural to construe ditransitive verbs as events causing a possession to hold; in other words, ‘I give the book to John’ translates as ‘I cause the book to be in John’s possession’ (Kayne 1984). We take this to be the origin of the widespread syncretism between genitive and dative – holding in Albanian as well as in Romanian, Aromanian, and in those Iranian languages (e.g. Kurmanji Kurdish) which still have a case declension.

Following Belvin and den Dikken (1997), we take the relevant characterization of possession to be an ‘inclusion’ one, that we notate as \( \subseteq \), as in Manzini and Savoia (2011a, 2011b). Under this proposal, and adopting for pre-genitival linkers the same position and structure as for pre-adjectival ones, the representation of an Albanian Noun-genitive DP structure, for instance (2b), is as in (37).

(37)

\[
\begin{array}{c}
\text{DP} \\
\text{N} \quad \text{D} \quad \text{D} \\
\text{putr} \quad a \quad e_x \\
\text{N} \quad (\subseteq) \\
\text{ceni}_y \\
\end{array}
\]

The genitive noun is formed by the N base \textit{ceni}-(the predicate \textit{ceni} ‘dog’ followed by the N class ending \textit{-i}) merged with the \( \subseteq \) ending \textit{-t}. The latter is an elementary two-place predicate \( \lambda x, \lambda y \), establishing a possessor/inclusion relation between the noun to which it attaches and the head noun, so that ‘the dog’ possesses/zonally includes ‘the leg’.

As indicated in (37), the role of the pre-genitival linker is essentially the same as the pre-adjectival linker, namely to provide a partial saturation of an argument slot. In this instance, the predicate of which it satisfies an argument is ‘inclusion’ \( \subseteq \). Though the external argument is ultimately supplied by the head noun \textit{putra} ‘leg’, in Albanian it is necessary to provide a saturation of the external argument of \( \subseteq \) within the embedded complement DP, namely by the linker \( e \), which agrees with the head noun in the sense that they concur to the lexicalization of the same argument slot.
5. Linkers in Aromanian

The obligatory presence of pre-adjectival linkers in Aromanian definite DPs is a contact phenomenon with linker languages – and in particular with Albanian. As briefly indicated in section 1, standard Romanian may position what is traditionally called a strong article (*cel*) between a noun and a modifying adjective, cf. example (3). However this element is not obligatory and it is in complementary distribution with prenominal determiners, showing that it partakes of their nature – and is not a linker.

At the same time, Aromanian presents a differences with respect to Albanian. In Aromanian, pre-adjectival demonstratives either precede quantifiers like *ma* or they are in complementary distribution with them, as shown in (9a’). Therefore the demonstrative linker appears to lexicalize a higher D position in the AP than the Albanian article linker, as shown for Aromanian in (38).

\[(38)\quad \text{DP}(x=y)\]

\[\text{DP}\]

\[\text{D}\]

\[\text{QP}\]

\[\text{A}\]

\[\text{N}\]

\[\text{fitfor}_x\]

\[\text{atseu}_x\]

\[\text{mar}_y\]

\[\text{ma}\]

As far as we can tell, the interpretation remains unchanged. In (28) two predicative bases are present, namely the adjective *mar-*’big’ and the head noun *fitfor-* ‘boy’. They both have an argument slot and the linker is necessary to provide a satisfaction for the argument of the adjective prior to theta-unification with the argument of the noun. It is interesting to note that (38) converges with Campos (2005) in construing the Aromanian linker as the subject of a predication (cf. (26b) above). This point of contact serves to better highlight the differences. In the predicational structure adopted by Campos, encoding interpretation structurally, a head like the Albanian article can only play the role of copula (cf. (26a) above). In our approach, where structure is projected on the basis of morpho-lexical properties of the elements involved, the Aromanian and Albanian linker are seen to have parallel structures, and the predicative interpretation is not mediated by any abstract category.\(^{12}\)

In short, it appears that a higher position (lower in any event than that of the D determiner of the whole DP) is equally compatible with a linker reading. The difference between Albanian and Aromanian may be connected to the different lexical properties of the linker involved. In particular, we suggest that the clitic nature of the Albanian

\(^{12}\) Campos (2008) uses the contrast between the position of quantifiers in Albanian, e.g. (34) and in Greek or Aromanian, as in (38) to argue that in Albanian the article is incorporated into the adjective. The discussion in the texts shows that this conclusion is not necessary. See also fn. 8 for an argument that the present theory is simpler overall.
Linkers in Aromanian in comparison to Albanian (and Romanian)

determiner/pronoun/linker may allow for attachment in the inflectional domain or of the sentence (clitic string) which are not open to the Aromanian demonstrative/linker.

A second difference between the linker in Albanian and Aromanian is that in Aromanian the linker cooccurs only with definite head nouns, behaving in this respect like Greek polydefiniteness (cf. Lekakou and Szendroi 2012 for a recent analysis, based on somewhat similar assumptions to the present ones). Recall that in Albanian, as summarized in table (22), linkers cooccur both with definite and with indefinite head nouns. This means that the linker of Albanian does not contribute definiteness to the larger AP/DP that embeds it. At the same time, table (22) shows that a partially different series of linkers is instantiated depending on the definite or indefinite nature of the head noun. We may see the definiteness restriction holding of the linker construction in Aromanian as a consequence of the fact that the linker (the demonstrative) agrees in definiteness with the DP – and is therefore only compatible with definite inflections on the head noun. Recall that in (17) we have already shown how demonstrative linkers are not excluded from combining with prenominal demonstratives, while in Romanian (18), the preadjectival cel demonstrative is barred from co-occurring with a prenominal demonstrative. This contrast confirms that in Aromanian the linker does not determine reference.

Summing up so far, there is an alignment of Aromanian on Albanian (or Greek) in what concerns pre-adjectival linkers, which are not present in the cognate language Romanian, or in fact in other Romance languages. At the same time, this alignment of Aromanian on contact languages makes use of existing lexical resources, i.e. demonstrative, as well as existing structural possibilities, i.e. the alternation between the prenominal and preadjectival position of the demonstrative in Romanian.13

5.1 Pre-genitival linkers

Linkers introducing genitives and datives in Aromanian are different not only from the pre-genitival linkers of Romanian (despite their lexical relatedness) – but also from the pre-genitival linkers of Albanian. Cross-linguistically pre-genitival linkers normally agree with the head noun of the DP embedding the genitive; for instance, this generalization is central to Philip’s (2012) understanding of linkers14. The generalization holds in Albanian; in the structure in (37), agreement between the linker and the head noun corresponds to the fact that the linker satisfies the external argument of the (⊆) elementary predicate introduced by genitive case, whose ultimate lexicalization is provided by the head noun.

Essentially the same can be said of the pre-genitival linker in Romanian, as illustrated in (4), with the structure in (39). Oblique case introduces the (⊆) possession/inclusion predicate. The ale linker provides a lexicalization of the possessee (external) argument of (⊆) within the complement structure, acting essentially as a bound variable (a ‘doubling clitic’) of the head noun. The internal argument of (⊆) is the DP to which the oblique case attaches, i.e. the possessor.

13 On the position of the Romanian demonstrative, see Giusti (1995, 2002) who locates it in an AgrP position. Our analysis supports a different solution for Aromanian, based also on the comparison with Romanian.

14 Split agreement also appears to be a possibility. For instance, for Zazaki, Toosarvandani and van Urk (2012) argue that the pregenitival linker agrees in case with the embedded N and in phi-features with the head N.
Aromanian differs from Romanian, but also Albanian, Kurdish, etc. in that pre-genitival linkers agree with the genitive DP. At least in the feminine singular it appears that the linker includes both a D constituent (al-) and an inflection (-i) exactly reproducing that of the genitive DP. In other words, the linker doubles the phi-features of the genitive DP and its case properties, as schematized in (40), cf. example (6c). In the masculine singular where the –u inflection only lexicalizes N class properties, it is the o linker that introduces the oblique relation.

(40)  
```
  DP
    libra_, (⊆)P
      (⊆) ali_{x,y} N
        N (⊆) fet_{x,y}
```

The constituent structure assigned to the Aromanian linker in (40) is the same as in Romanian (39) or in Albanian (37) – or for that matter in Aromanian (38). Interpretively, on the other hand, the linkers seen so far provides a lower level satisfaction for the external argument slot of (⊆), ultimately bound by higher material (the head DP). The pre-oblique particle of Aromanian, by contrast, helps introducing the oblique case (⊆) itself. This also helps us understand why the same linker material introduces not just the genitive (i.e. the adnominal possessor), but also the dative (i.e. the possessor in a sentential context). Furthermore, in so far as it lexicalizes (⊆), the linker can be merged in a superordinate position to a determiner/quantifier head, as can most clearly be seen with indefinites, for instance (6b) – to which we assign the structure in (41).

(41)  
```
  (⊆)P
    (⊆) ali QP
      Q (⊆) N
        N (⊆) fet i
```

By contrast, in Romanian, linkers agree with the head noun (the possessee) and depend on the absence of determiners preceding the genitive, as in (42a). (42b) shows that indefinite genitives are introduced by the preposition a (Giurgea 2012 and literature quoted there)\(^\text{15}\).

\(^{15}\) Giurgea argues in favour of the same constituent structure adopted here, where the linker is a projection of the genitive DP and not a functional category of the head DP. He entertains the two
In Albanian and in Romanian, the part-whole relation (the oblique case on the embedded noun) and an inflectional level lexicalization of its external argument (the linker embedding the oblique) are separately merged in the syntactic tree. However in Aromanian only the first component appears to be lexicalized. Its similarity to linkers of the more canonical type consists in the fact that it is a clitic double of an element otherwise realized in the structure. Specifically the pre-oblique linker of Aromanian doubles the \( (\subseteq) \) elementary predicate in a position where it has in its local (Minimal Search) domain both of its arguments, namely the possessor (its complement) and the possessee (its Spec).

6. Conclusions

Aromanian displays alignment phenomena (Gumperz and Wilson 1971) with Albanian, with which it is in contact – specifically the use of the Romance-type demonstrative as a preadjectival linker. In theoretical terms, existing lexical resources (i.e. the demonstrative) find a new structural collocation in Aromanian (‘reanalysis’), in conditions of systematic bilingualism with Albanian (‘contact’).

For pre-genitival linkers, Romanian has a separate lexical series which Aromanian shares. At the same time, structures that are present neither in Romanian nor in the contact language Albanian also emerge – namely linkers in front of datives as well as of genitives, and agreeing with the genitive, rather than with the head noun. Thus the pressures of language contact (the potential tension between structural alignment and existing lexical resources in the language) brings about a result that is different from both Romanian and Albanian, even in respects in which the latter two coincide (agreement of the linker with the head noun). In theoretical terms, these outcomes are especially interesting, to the extent that they are attributable directly to UG.

References


formal possibilities discussed here, namely that the linker is an agreement or a case marker. He concludes that it is a case marker in Romanian as well. We believe that the contrast between Aromanian and Romanian argues against such a conclusion.

101


Turano, Giuseppina. 2002. “On modifiers preceded by the article in Albanian DPs”. 103


COMPLEMENTIZER DOUBLING
IN EUROPEAN PORTUGUESE

Salvador Mascarenhas*
St Catherine’s College, Oxford

1. Introduction

Since the establishment of the “cartographic” enterprise in generative syntax (Rizzi 1997; Cinque 1999), left periphery phenomena in the Romance languages have been particularly influential in motivating theoretical proposals about the structure of the C-domain. For example, the intricate structure of topocalized and focused phrases in Romance languages was one of the main motivations for Rizzi’s (1997) proposal as well Benincà and Poletto’s (2004) proposed revision.

Romance languages might prima facie appear to be far from an ideal testing ground for cartographic explorations, given that they typically have a relatively small catalog of complementizers. These sorts of functional items are naturally of special significance to the cartographer: complementizers typically populate the heads of the functional projections the theorist is proposing, and thus can be instrumental in motivating novel projections and investigating the boundaries between them. For example, work on the left periphery of, in particular, isolating languages (see, for example, Aboh 2004, 2006) has provided further motivation for the split-C hypothesis, as opposed to, say, Chomsky’s (1995) multiple specifiers proposal. Romance languages, though they may appear to be less exciting than isolating languages in this regard, can display a number of phenomena involving functional heads in the left periphery, most notably complementizer deletion, (what seem to be) doubly filled COMP violations, and complementizer doubling. This article is concerned with the latter.

C-doubling (sometimes referred to, especially in the literature on European Portuguese, as “recomplementation”) is a characteristic of, among others, the Northern Italian dialects Turinese and Ligurian (Paoli 2007), spoken (Castilian) Spanish (Demonte and Soriano 2007), Galician, and certain variants of European Portuguese. The most well-known data are surely those pertaining to the Italian dialects, and it is fair to say that the Iberian data, and the Portuguese data in particular, have been largely overlooked, with some noteworthy exceptions (Uriagereka 1995; Barbosa 2000). This paper aims to contribute to filling that gap.

* This article is a minimally updated version of an unpublished manuscript from 2007, which has been available for download from my webpage since 2008. The changes with respect to the 2007 version are numerous, but almost exclusively concern matters of style. I thank Guglielmo Cinque and Luigi Rizzi for the opportunity to revisit this work, and Guglielmo Cinque for discussion of the earlier manuscript. I am very grateful to Manuela Ambar and João Peres for extensive discussions of this work back in 2007.

RGG (2014) 36: 105-116
As it turns out, C-doubling in European Portuguese (henceforth, EP) has some rather surprising and interesting characteristics that have remained, to the best of my knowledge, unnoticed by the literature, and that seem to set the phenomenon apart from its correlates in other Romance languages. Most notably, in EP there seems to be no syntactic constraint on the number of Cs doubled (provided that appropriate phonologically realized material is between every two instances of C), as can be seen in (1), and doubling is possible with the interrogative complementizer ‘se’ (if/whether) — see (2).

(1) Acho que amanhã que Ana que vai conseguir acabar o trabalho.
   I think that tomorrow that Ana that will manage to finish the assignment
   ‘I think tomorrow Ana will manage to finish the assignment.’

(2) Não sei se o João se já chegou.
    not I know if João if already arrived
    ‘I don’t know if João already arrived.’

I begin by focusing on declarative C-doubling, and noting some of the most relevant differences with respect to Paoli’s (2007) Northern Italian data, so as to offer a comparison with better-known C-doubling data. I then discuss the two properties that single out the EP system from other Romance languages: the possibility of having more than two complementizers in a doubling construction, and the possibility of doubling the interrogative complementizer. While the aim of this article is primarily descriptive, I conclude with some theoretical remarks.

2. C-doubling in European Portuguese

2.1 EP declarative C-doubling

Unlike the case of Turinese and Ligurian, discussed at length by Paoli (2007), C-doubling in EP is insensitive to mood in the embedded verb. In those Northern Italian dialects, doubling is only possible if the embedded verb is in subjunctive mood, and the author argues that the type of matrix verb has no influence whatsoever in legitimizing or barring the construction. The examples in (3)–(5), with an indicative, subjunctive and conditional embedded verb, show that embedded verb mood has no clear influence in the Portuguese construction.

(3) O João disse que a Maria que vai chegar atrasada.
    João said that Maria that will arrive late
    ‘João said that Maria will arrive late.’

---

1 More recent scholarship (after the time of writing of the original version of this article, in 2007) has steadily contributed toward reversing this trend, with several articles focusing on complementizer doubling in Iberian Romance. See for example Villa-García (2012) for Spanish and Ribeiro and Torres Morais (2012) for Old Portuguese.
Complementizer doubling in European Portuguese

(4) Duvido que a Ana que goste de ópera.
    I doubt that Ana that likes-SUBJ opera
    ‘I doubt Ana likes opera.’

(5) Acho que se lhe ligasses que tudo se resolveria.
    I think if him/her called that everything REFL solve-COND
    ‘I think if you called him/her everything would turn out fine.’

It is not entirely clear to what extent, if at all, the class of matrix verb selecting the CP plays a role in allowing for doubling. It is uncontroversial that matrix epistemic verbs, such as the ones in the three sentences above, are perfectly compatible with doubling, and the same seems to be true of non-epistemic predicates such as ‘esperar’ (to hope), the deontic verb ‘exigir’ (to demand), both subcategorizing subjunctive mood in the embedded verb, or ‘prometer’ (to promise), with indicative:

(6) Espero que a Ana que traga o livro.
    I hope that Ana that bring-SUBJ the book
    ‘I hope Ana brings the book.’

(7) Prometo que a carta que chega amanhã.
    I promise that the letter that arrives-IND tomorrow
    ‘I promise the letter will arrive tomorrow.’

As for factive verbs, the situation appears more complex. ‘Lamentar’ (to regret) seems to reject C-doubling quite strongly (all my informants deemed sentences like (8) quite bad), while sentences with ‘reparar’ (to notice) are perfectly fine.

(8) ?? Lamento que o Filipe que tenha chumbado o exame.
    I regret that Filipe that have-SUBJ failed the exam
    ‘I’m sorry that Filipe failed the exam.

(9) Reparei a semana passada que os miúdos que chegam sempre atrasados.
    I noticed last week that the kids that arrive-IND always late
    ‘I noticed last week that the kids always come in late.’

Interestingly, a factive verb that carries evaluative content such as ‘lamentar’, requiring a subjunctive embedded verb, rejects C-doubling, while non-evaluative factives such as ‘reparar’, ‘observar’ (to observe) and so on, which call for indicative mood in the embedded verb, accept doubling. Notice that both dimensions, evaluation and factivity, seem to play a role. A verb like ‘esperar’ (to hope) clearly carries evaluative information regarding the embedded proposition and selects subjunctive mood, but it does not pattern with ‘lamentar’, plausibly because it is not factive.

For the time being, I can do no more than to notice this contrast, so I will disregard factive verbs in general for the remainder of this article, and will proceed as if the matrix verb had no
influence on whether C-doubling is allowed. I now turn to the issue of what can occur between the complementizers. This matter the existing literature on EP does consider, though not in sufficient detail, as I hope to show.

In examples (3)–(9) I have shown that DP subjects, adverbs and if-clauses can all occur in the relevant position. We can also observe direct and indirect objects in that same position, with or without a resumptive clitic:

(10) Acho que esse livro que a Ana já (o) leu.
‘I think that DET book that Ana already CL-ACC read
‘I think that book, Ana already read.’

(11) Disseram-me que ao João que o professor (lhe) deu um dezoito.
‘I heard the professor have João an A.’

Data similar to these prompted Uriagereka (1995) and Barbosa (2000) to argue that topics, left dislocated (LD) and clitic left dislocated (CLLD), can occur in this position. As these authors remark, quantified phrases cannot occur between the two complementizers:

\[\text{Tenho a certeza de que a Maria que vai chegar a horas.}\]
‘I’m sure Maria will arrive on time.’

The same sentence with the preposition ‘de’ governing the embedded sentence (that is, “Tenho a certeza de que a Maria que vai chegar a horas”) is extremely odd. It seems very plausible that that is due to the fact that C-doubling is a characteristic of colloquial EP and of conservative EP dialects spoken mainly in rural areas, while the use of the preposition ‘de’ in these constructions rings very formal. The sentence as a whole thus becomes incoherent with both strongly formal and strongly colloquial elements.

One final remark concerning selection is in order. Ambar (1992, 2005) noticed that ‘que’-headed sentences with a subjunctive verb are quite acceptable as (at least apparently) matrix sentences under a strong exclamative / imperative reading:

\[\text{Que venham as chuvas!}\]
‘May the rain come!’

As Ambar notes, these kinds of utterances are far from common, and I for one do not use them productively at all. Moreover, I find them very hard to interpret if the verb is not an unaccusative or unergative, and even with these kinds of verbs the sentences become considerably worse with a pre-verbal subject (?? “Que as chuvas venham!”)

In any event, C-doubling seems to be completely out, turning what to most speakers my age think is an odd sentence into an unintelligible one. This point further distances the EP data from the Italian case considered in the previous section. In Turinese and Ligurian, constructions exist that are perfectly analogous to (ii), and C-doubling is indeed allowed in those contexts.
Complementizer doubling in European Portuguese

(12) * Eu acho que ninguém que leu esse livro.
    I think that no one that read that book

(13) * Eu acho que muitas pessoas que leram esse livro.
    I think that many people that read that book

(14) * Eu acho que alguém que leu esse livro.
    I think that someone that read that book

This fact again separates EP C-doubling from Turinese and Ligurian, where quantified phrases, including negative quantifiers such as no one, can occur, under certain readings, between the complementizers. In EP, such constructions are sharply ungrammatical.

Uriagereka (1995) goes a step further than Barbosa (2000), proposing that only topics can occur between the two complementizers. He argues that the second ‘que’ is the head of a functional projection, which he calls FP, and he stipulates that, in languages such as Portuguese, the head F0 lacks the focus feature necessary to probe for and attract a focus phrase to its specifier. Thus, under minimalistic assumptions, derivations where a focused constituent is moved to the position between the complementizers are barred.

To the best of my knowledge, the impossibility of quantified expressions to occur between the complementizers is the only reason presented in the literature to conclude that focused constituents cannot appear in that position. This argument seems too weak, although what it purports to show is I believe correct.

In fact, the kinds of quantifiers in the sentences Uriagereka experiments with could not be interpreted as contrastive focus, (see among others, Szabolcsi 1981, and Kiss 1998), though they might be interpreted as informatively focused. The distinction between these two categories of focus has semantic motivations (most notably, that contrastive focus entails exhaustivity, whereas informational focus merely presents new, non-presupposed information) as well as syntactic (contrastive focus displays weak cross over effects, which informational focus does not). Benincà and Poletto (2004) show that contrastive focus must precede informational focus, under a split-C hypothesis, and revise Rizzi’s (1997) proposal concerning the position of focus, rejecting the freely recursive character of Topic phrases. Thus, if Benincà and Poletto’s proposal is right, Uriagereka has only shown that the constituent between complementizers cannot be an informational focus. There might still theoretically be a contrastive focus position that non-quantified DPs could occupy, to the left of a doubling ‘que’.

To discard this possibility, we check for exhaustive interpretation in an element between the two complementizers.

---

3 Barbosa (2000) makes no claims about the syntactic structure behind C-doubling, or the locus of the lower complementizer. Her discussion of C-doubling is directed towards using some of its characteristics as a motivation for her proposal that pre-verbal (non-quantificational) subjects in EP are CLLDed.

4 The actual examples Uriagereka (1995) discussed are not contemporary Portuguese, but it is clear from the text that his proposal is meant to apply to EP.
(15) Eu acho que UM LIVRO que a Ana comprou.
    I think that a book that Ana bought
    ‘I think A BOOK, Ana bought.’

In a dialogue situation, the hearer cannot reply to (15) with (16). In (16), the hearer tries to
deny an exhaustivity implication that is not there, hence the statement is infelicitous. This shows
that a contrastive focus interpretation for the element between complementizers is not available.

(16) # Não, ela também comprou uma revista.
    ‘No, she also bought a magazine.’

We have established that LDed and CLLDed topics, and not focused elements of any kind,
can occur between the complementizers. In terms of Benincà and Poletto (2004), this suggests
that C-doubling in EP is intimately associated with what they call the topic field. A natural
question arises, concerning hanging topics, what they show to be the outermost category in the
topic field: can they occur between the complementizers as well? The answer is yes, as can be
seen from the following examples:

(17) Acho que, este livro, que a Ana não gostou dele.
    I think that this book that Ana not liked of it
    ‘I think that this book, Ana didn’t like.’

(18) * Acho, este livro, que a Ana não gostou dele.
    I think this book that Ana not liked of it

The sentence in (17) shows what is clearly a hanging topic (HT): between the
complementizers is a DP with a resumptive pronoun governed by the appropriate preposition,
given the selection properties of ‘gostar’ (to like). Furthermore, unlike the Italian data presented
by Benincà and Poletto (2004), European Portuguese does not accept a HT scoping over the
embedded sentence to the left of the topmost ‘que’, as evidenced by (18). This piece of data is
particularly interesting, because it suggests that the complementizers ‘que’ and ‘che’ show
different characteristics even in simplex (non-doubled) contexts. Given the Italian data, Benincà
and Poletto (2004) suggest that ‘che’ be located in a position lower than that of HT, between HT
and LD, but they do not elaborate on the proposal. At the very least, the EP data show that such
a configuration is by no means universal for Romance declarative complementizers.

To sum up what has been established so far, we have seen that complementizer doubling,
modulo certain restrictions on the selection by matrix verbs (the ‘lamentar’ cases I alluded to
above), is possible in embedded declarative sentences, under the condition that the material
between the complementizers is part of the topic field, in the terms of Benincà and Poletto
(2004).
2.2 The puzzles of C-tripling and interrogative C-doubling

Thus far, I have been working under the tacit assumptions that only one topic can occur between the complementizers and that a third complementizer position is not available. The former will prove true, the latter, surprisingly, will not. Consider the following sentences:

(19) *Acho que amanhã(,) a Ana(,) que vai conseguir acabar o trabalho.
I think that tomorrow Ana that will manage to finish the assignment

(20) Acho que amanhã(,) que a Ana(,) que vai conseguir acabar o trabalho.
I think that tomorrow that Ana that will manage to finish the assignment

In (19) we have two topical elements between the complementizers, clearly two distinct maximal projections, and, under the assumptions we are working with (namely, that multiple specifiers are not an option), two corresponding T0 nodes. The sentence is quite bad, according to all my informants, regardless of whether one observes the pauses indicated or not.

Sentence (19) contrasts sharply with (20), where we have what seems to be essentially the same sentence, but with two instances of doubling, one after each topic. Notice that no pauses are required after each topic constituent. In fact, EP speakers prefer little to no pause in the spots indicated with commas, and some informants actually report ungrammaticality judgments with prominent pauses. This indicates that we are dealing with a phenomenon that is syntactic in nature, not a result of performance-related issues (hesitance pauses, fillers, and so on)5.

5 João Peres (p.c.) pointed out to me an interesting set of data, for which I am very grateful. Witness the contrasts in (i)-(iv):

(i) Acho que o Pedro e a Ana foram a Roma.
I think that Pedro and Ana went to Rome
‘I think Pedro and Ana went to Rome.’

(ii) Acho que o Pedro e a Ana que foram a Roma.
I think that Pedro and Ana that went to Rome

(iii) Acho que o Pedro e que a Ana que foram a Roma.
I think that Pedro and that Ana that went to Rome

(iv) *Acho que o Pedro e que a Ana foram a Roma.
I think that Pedro and that Ana went to Rome.

Interestingly, heading the second conjunct in the subject DP with the complementizer forces C-doubling before the IP. Without committing to a any particular analysis, I note that sentences like (iii) seem to require a distributive interpretation of the predicate over the subject, as the following contrast with a collective predicate shows (my own judgments, not checked with other informants):

(v) Acho que o Pedro e a Ana que reuniram ontem.
I think that Pedro and Ana that met yesterday
‘I think Pedro and Ana met yesterday.’

(vi) *Acho que o Pedro e que a Ana que reuniram ontem.
I think that Pedro and that Ana that met yesterday
There does not seem to be a syntactic reason to disallow four complementizers, but the judgments become rather degraded (cf. (21)). This is unsurprising, for presumably the same performance restrictions that make (22), a sentence with the same topic constituents as (21) but no C-doubling, quite odd will be at play in (21) as well.

(21)  ?? Duvido que ontem que o Pedro que à Ana que lhe tenha telefonado.  
I doubt that yesterday that Pedro that to Ana that her have called  
‘I doubt that yesterday Pedro called Ana.’

(22)  ? Duvido que ontem o Pedro à Ana lhe tenha telefonado.  
I doubt that yesterday Pedro to Ana her have called  
‘I doubt that yesterday Pedro called Ana.’

I now turn to the other oddity in the EP complementizer-doubling phenomenon, interrogative C doubling. In EP, it is possible to double the complementizer of embedded interrogatives, ‘se’ (if / whether), as the following sentences show:

(23)  Não sei se o João vai chegar a horas.  
not know if João will arrive at hours  
‘I don’t know if João will arrive on time.’

(24)  Não sei se o João se vai chegar a horas.  
not know if João if will arrive at hours  
‘I don’t know if João will arrive on time.’

These sentences sound perfectly natural to all my informants. In fact, even speakers who cringed at the sound of a declarative C doubling were more than happy to accept ‘se’-doubling as in (24). Somehow, these constructions are available to a larger number of standard EP speakers than ‘que’-doubling constructions. These facts are remarkable: to the best of my knowledge, interrogative complementizer doubling has not been described in other Romance languages.

As expected, the same restrictions as with ‘que’-doubling apply to the elements between complementizers, namely, only topics can occupy in that position:

(25)  * Pergunto-me se à MARIA se o João deu um presente.  
I wonder if to Maria if João gave a present  
(intended: ‘I wonder if John gave MARY a present.’)  
(no contrastively focused phrases between ‘se’-complementizers)

Constructions such as (vi) are possibly elliptical, for they seem essentially equivalent in meaning to (vii), and are similarly deviant.

(vii)  *Acho que o Pedro reuniu ontem e que a Ana reuniu ontem.  
I think that Pedro met yesterday and that Ana met yesterday
(26) Não sei se o Filipe se o Pedro gosta dele.
   ‘I don’t know about Filipe whether Pedro likes him.’

Finally, the issue of recursive ‘se’ doubling (‘se’-tripping) is a bit less clear than that of
‘que’. Somehow grammaticality judgments on sentences like (27) are not as strong as the ones
concerning ‘que’. No informant thought they were totally out, but most considered them a bit
unnatural. In any event, (28) is, as one would expect if the judgments from ‘que’ sentences were
to carry over here, uniformly considered quite bad:

(27) ? Não sei se amanhã se o Pedro se consegue entregar o trabalho.
   ‘I don’t know if tomorrow Pedro will manage to hand in the assignment.’

(28) ?* Não sei se amanhã o Pedro se consegue entregar o trabalho.

3. Concluding remarks

So far, I have concentrated on giving an accurate description of European Portuguese
complementizer doubling phenomena, only sporadically mentioning the theoretical implications
the data might have. I recapitulate the main points of what came before:

1. Embedded sentences with overt complementizers (declarative ‘que’ and interrogative
   ‘se’) have the possibility of displaying more than one instance of their C-head.
2. These (phonologically identical) copies must each follow exactly one topical maximal
   projection, subject to restrictions on the topic field in the C-domain.
3. Recursion of this process is possibly not bound by syntactic reasons, but by performance
   related issues.

Or, in a somewhat weaker formulation: The doubling process can apply at least twice.

In connection with the Italian data mentioned briefly in the previous section, it is clear that
an analysis along the lines proposed by Paoli for Turinese and Ligurian cannot be applied to EP.
Because the lower ‘che’ (che₂) is only licensed in subjunctive contexts, and given that
subjunctive mood is by and large morphologically unmarked in those languages, Paoli argues
that che₂ contains mood and finiteness features, and that its locus is Fin⁰.

The lower Cs in EP doubling constructions do not seem to instantiate any sort of mood
category such as subjunctive (or indicative for that matter), since C-doubling in EP appears to be
insensitive to the mood of the embedded verb. Moreover, the position Paoli argues for ‘che’ in
these Northern Italian dialects, namely FinP, is much too low to account for the EP data: if a
doubling ‘que’ is in FinP, then what could be the reason why focused elements cannot precede
doubling ‘que’s? Finally, Paoli’s proposal cannot account for point 3. above, recursion of the process.

In general, it seems clear that an analysis under the cartographic hypothesis that argues that the identical complementizers delimit the C-domain is bound to fail for EP. If one were to commit to that analysis, one would be forced to either assume that three ‘que’s in a Portuguese triple-C construction occupy three categorially different functional positions, or that the whole domain is recursive. Both options are very undesirable.

On the former, one would have to explain why the presence of a ‘que’ in Fin allows only one topic in the higher Top phrases, as well as why focused phrases are not allowed to appear between complementizers. Moreover, the triple complementizer case would have to be seen as the instantiation of three complementizers in three different functional positions. This multiplication of diverse functional heads, when there seems to be no reason to assume that the complementizers have different functions, is contrived at least.

The other logical hypothesis, that the whole C-domain is recursive, is far from minimalistic. To stick to the idea that ‘que’s only occur in For and Fin, one would have to say that the left periphery can be doubled, so as to house the three complementizer case. This is extreme, and I know of no independent motivation for it.

It seems that, to account for the EP data, we have to drop the assumption that the complementizers sit in For and Fin, an assumption which, incidentally, can hardly even be formulated coherently for the case of ‘se’-doubling.

This article does not have a complete alternative analysis to present. The data surveyed here are novel and puzzling, apparently suggesting that complementizers such as ‘que’ and ‘se’, in Portuguese at least, can occupy recursive Top positions, yielding representations such as  for sentence (20), repeated here as (30).

\footnote{Roughly the same objections apply to Demonte and Soriano (2007) on Spanish ‘que’ doubling, which does not allow recursion, as it too seems to have a lower locus than that of Portuguese ‘que’.}
(29) Acho que amanhã(?,) que a Ana(?,) que vai conseguir acabar o trabalho.
I think that tomorrow that Ana that will manage to finish the assignment
‘I think tomorrow Ana will manage to finish the assignment.’

Although it is the simplest hypothesis, further investigation is required to understand whether [??] is a tenable route toward an analysis. Be that as it may, I hope to have argued convincingly that consideration of complementizer doubling in Iberian Romance raises novel and exciting puzzles concerning the structure of the left periphery and the nature of complementizers.

References


THE CRUCIAL ROLE OF THE EVENT STRUCTURE IN THE RETRIEVAL OF NOMINALIZATIONS IN APHASIA

Chiara Zanini
Università degli Studi di Padova

Paola Benincà
Università degli Studi di Padova

Carlo Semenza
Università degli Studi di Padova

Abstract: In the aphasic production, the verb-noun (V-N) dissociation (a condition whereby brain damage selectively affects one of the two categories, while sparing the other) may have a different nature in different cases, reflecting semantic, syntactic or grammatical class effects. The observation that the V-N dissociations have a different nature goes in hand with the fact that in Linguistics it is hard to spot out the differences between nouns and verbs in a clear-cut way. In this regard, the nominalization phenomenon is explicative since nominalizations share both nominal and verbal properties. In the literature on aphasia, only few have studies tested how verbal and nominal selective deficits impact on closely related pairs involving verbs and the corresponding nominalizations. The present study aims at verifying whether the aphasic production of nominalizations can be affected by class selective impairments or by semantic/syntactic deficits. It will be showed that the aphasic speakers’ errors enlighten the crucial role played by the event structure, the Aktionsart and the grammatical aspect in the retrieval of the nominalizations. Such conclusions provide neurolinguistic evidence in favour of the idea that some features and properties can cross the boundaries of the verb-noun class distinction.

Keywords: nominalization, aphasia, verb-noun dissociation, aspect, event structure

1. Introduction
1.1 Verbs and Nouns in Linguistics

In Linguistics the nature of verbs and nouns has been long debated and this concern is still crucial for many modern linguistic approaches such as the Generative framework. That nouns and verbs consist of two distinct well-cut classes is a quite common thought dating back at least from Dionysius Thrax’s grammar (c. 100 BC) and Vico’s and Condillac’s works in the 16th century, just to mention a few. If it is out of question that in human languages some words behave as nouns and others as verbs, nonetheless a clear definition of these two lexical categories is far from being achieved. Prototypically, nouns have referential functions and denote objects while verbs have predicative functions, take
arguments and denote events. However, some words hardly fit in one of these two definitions. For example, nominalizations have the same syntactic distribution of a noun, but can denote events and take arguments as verbs (1).

(1) Colombo’s discovery of America was in 1492.

The nominalization puzzle can be taken as a piece of evidence for arguing in favour of the idea that nouns and verbs are not really two polar entities. In this sense, the paper on category squish by John Ross in 1972 laid the basis for an analysis of verbs and nouns in terms of a *gradatum*. He took into account a number of linguistic facts pointing out that the distinction between verbs and nouns (and adjectives as well) is not a discrete one, but is spread in a “category space” as illustrated below:

(2) verb > present participle > perfect participle > passive participle > adjective > preposition > adjectival noun > noun

Prototypical nouns and prototypical verbs are set at the opposite edges of this category space, while in the middle some mixed categories are listed. More relevant for the purposes of the present work, in 1973 Ross elaborated the following “nouniness scale”:

(3) a. That clauses: John said [that Bill gave Mary a book]
   b. For to clauses: He would resent (it) [for me to go out with Mary]
   c. Embedded questions: I wonder [how long time I have to wait here]
   d. Acc-ing (complements): He resented [me going out with Mary]
   e. Poss-ing (complements): He resented [my going out with Mary]
   f. ing-of (Action nominals): He resented [my careless examining of the body]
   g. Derived nominals: He resented [my careless examination of the body]
   h. Underived nominals: He resented [the daughter of Bill]

The idea that noun and verb are not atoms, i.e., primitive entities, has been formalized in different ways (Borer 2005; Clark & Clark 1979; Marantz 1997 and following works). In particular, the Distributed Morphology approach (Alexiadou 2001; Marantz 1997) posits that syntax creates complex objects out of two different types of morphemes, roots and functional morphemes. Roots are assumed to be category-neutral and they are categorized in syntactic structures by functional heads such as v and n in order to yield verbs and nouns. On this account, there would be two scales: a verbal one (occurrence of nominative and accusative cases, modal and auxiliaries, morphological aspect and argument structure) and a nominal one (occurrence of determiners, gender and plural markers, genitive case). If the two scales can interact with one another it follows that the distinction between verbs and nouns is not polar, but gradual in nature since the verb-noun cut-off point can be set at various levels.
1.2 Verbs and nouns in Neurolinguistics

For an insight from a different perspective about the nature of the two lexical categories, it is possible to look at the neurolinguistic literature. At a first glance, both studies on people affected by aphasia (an acquired disorder following brain damage) and neuroimaging studies seem to support the mental and neural representation of verbs and nouns in terms of two distinct well-cut categories. In fact, the verb-noun (V-N) dissociation (i.e. a condition whereby brain damage selectively affects one of the two categories, while sparing the other) is one of the most frequently described dissociations in the literature on aphasia (for a review: Crepaldi et al. 2010, 2011; Druks 2002; Luzzatti, Mondini & Semenza 2001; Mätzig et al. 2009). In their review of the literature, Mätzig et al. (2009) calculated that aphasic people showing problems in picture naming tasks suffer selective verbal deficits about seven times more frequently than selective noun deficits. This estimate may be biased, however. Batteries evaluating verbs and nouns could be unbalanced in terms of difficulty: verbs are less easy to picture and many items are thus ambiguous. Normal control may perform at ceiling, while the unbalance in difficulty, penalizing verb naming, may appear only in patients. Besides problems with picturability of items in specific batteries, the higher frequency of the relative verb deficit could be accounted for in terms of linguistic complexity, that is, verbs are generally more complex than nouns.

Problems for patients may raise at both the morphosyntactic and the semantic level. At the morphosyntactic level, verb processing would be more demanding since verbs are generally associated with a greater number of morphemes (Caramazza & Berndt 1985). However, this hypothesis is to some extent problematic. In fact, in some natural languages, such as Chinese, both nouns and verbs are modified by no inflectional or derivational morphemes at all while in others, as Hungarian, nouns and verbs share the same number of morphemes. A clarifying example comes from Nuu-chah-nulth, a polysynthetic VSO language of the Wakashan family spoken by the Nootka in the Vancouver Island. In this language, nouns and verbs share not only the same number but also the same kind of morphemes with the exception of the last one that marks the categorial status (4):

(4) a. ʔinkw – iɬ – minh – ʔis – it
   fire/burn – at home – pl. – diminutive – past

   b. ʔinkw – iɬ – minh – ʔis – it – ʔi
      determinant
      ‘the little fires [we had] once at home’

   c. ʔinkw – iɬ – minh – ʔis – it – a
      3a Sg. indicative
      ‘[some fires] burned at home’

(Sapir 1921)

Indeed, it is worth noticing that some studies on aphasia have reported V-N dissociations even in patients speaking languages in which nouns and verbs share the same morphosyntactic complexity as Chinese (Chen & Bates 1998).

According to the semantic hypothesis, verbs would have more complex representations, since they require an argument structure and a specific number of participants implied in the described event, that is, an agent, a theme, a goal, etc. (Gentner 1981). Even in this case, there is linguistic evidence in favour of the idea that nouns can be associated to a thematic and argument structure as well (Alexiadou 2001; Chomsky 1970; Grimshaw 1990). As
already mentioned, nominalizations, for example, are substantives referring to a verbal predicate and sharing some verbal properties.

At the end, all the hypotheses concerning an assumed greater verbal complexity have to face with a cognitive issue and a linguistic one, however. Firstly, if verbs processing is more demanding than nouns processing, the selective nominal deficits reported in the literature (Berndt et al. 2002; Caramazza & Hillis 1991; Damasio & Tranel 1993; Menichelli & Semenza 2006; Miceli et al. 1999; Zingeser & Berndt 1990) remain unexplained. Secondly, it is not clear where the linguistic boundary between verbs and nouns has to be set. In this respect, it would be interesting to test not only prototypical nouns and verbs, but also hybrid phenomena as compounds (see Semenza & Mondini 2010 on this issue) and nominalizations.

More recently, cognitive neuropsychologists agree that nouns and verbs are represented and processed by means of different functional as well as anatomical ways. In this respect, it becomes crucial to understand which linguistic levels are relevant for the functional organization of the mental lexicon. Some studies have argued that the organization of the mental lexicon would reflect a grammatical class effect (Badecker & Caramazza 1991; Caramazza & Hillis 1991; Rapp & Caramazza 2002; Shapiro & Caramazza 2003; Shapiro, Moo & Caramazza 2006). If this were true, patients with selective verbal deficits would not exhibit problems with nouns describing events, that is nominalizations. A study conducted by Collina, Marangolo & Tabossi (2001), reported below, seems to invalidate this conclusion. In addition, a study by Menichelli & Semenza (2006) conducted on an amnesic patient, reported nominal selective impairment that cannot be explained in terms of a grammatical class deficit nor in terms of a semantic deficit. In this sense, these findings are problematic for those hypotheses based on a semantic perspective as well. According to these hypotheses (Bird, Franklin & Howard 2000, 2003; Mätzig et al. 2009), the functional organization of the lexicon would be semantic in nature and would reflect the prototypical meaning of nouns and verbs. On the one hand, verbs would mainly refer to actions and would be less imaginable; on the other, nouns would mainly denote objects and would be more imaginable. As a consequence, the V-N dissociation would not reflect a grammatical class effect, but it would be a by-product of the imageability feature. If the nouns and the verbs of a task are matched for this variable, the V-N dissociation will be not observable anymore. This hypothesis, though, is not much fine-grained from a linguistic point of view. It is not true that generally verbs describe actions; in fact, some verbs describe states as well and in these cases they show nominal and adjectival properties (Vendler 1967). Moreover, nouns can depict not only objects, but even abstract entities or events, as nominalizations do. In support of this linguistic analysis, some neurolinguistic studies have reported cases of V-N dissociations even if the items of the administered tests had been matched for imageability. In particular, Luzzatti and coworkers (2002) noted that the V-N dissociation was still observable in one third of the cases after ruling out the imageability effect. They noted that imageability mainly affects verbs retrieval while frequency mainly affects nouns retrieval. Luzzatti & Chierchia (2002) have argued that these findings cannot be a mere consequence of an imageability effect, but depend on the fact that retrieving a verb is a qualitative different task from retrieving a noun (see also: Chiarello, Shears & Lund 1999). Luzzatti et al.’s (2002) work suggests that the V-N dissociations described in the literature may be determined by the interplay of different variables in the mental organization of the lexicon.
The crucial role of the event structure in the retrieval of nominalizations in aphasia

1.3 The nominalization puzzle

Thus it seems that the V-N dissociations reported after brain damage may have different a nature in different cases. In turn, this supposition goes in hand with the fact that in Linguistics it is hard to spot out the differences between nouns and verbs in a clear-cut way. Many variables and linguistic levels are involved in the definition of such categories and the set of nouns and the set of verbs change their elements depending on the variable and on the level considered.

In this sense, derived nominalizations (DNs) are explicative, since they share both nominal and verbal properties (Alexiadou 2001; Chomsky 1970; Grimshaw 1990; Marantz 1997). As already mentioned, from a morphosyntactic point of view, they are nominal elements since they have the same syntactic distribution of the nouns and, at least in Italian, they take prepositional complements. From a semantic point of view, though, they are verb-like since they can denote events and can be associated with a thematic and argument structure. In (5) the noun *giacca* ‘jacket’ takes a prepositional complement, ‘di Simone’ ‘Simone’s’, that can indicate the owner of the jacket, the person who created that jacket, a person who likes that jacket, etc. On the contrary, in (5b) the DN *mangiata* ‘eating’ takes the same prepositional complement, *di Simone*, that can indicate only the entity ‘x’ of the event ‘x eats y’. The relations between a nominalizations and its complements are less ambiguous than those between a noun and its complements because a nominalization assigns the same thematic and argument roles that the corresponding verb (*‘to eat’*) would assign.

(5) a. *La giacca di Simone*
   ‘Simone’s jacket’
   b. *La mangiata di Simone*
   ‘Simone’s eating’

However, at a first glance, there seem to be some differences between verbal arguments and nominal arguments. A verb assigns nominative and accusative case in such a way that its arguments are always recognizable as agents, causes, themes, etc. For instance, in (6a) it is not possible to interpret the constituent ‘Romans’ as the agent of the event ‘destroy’. In other words, verbal heads are associated to an argument structure that must be always satisfied otherwise the sentence will be agrammatical (6b).

(6) a. Barbarians destroyed Romans.
   *Destroyed.

Moreover, since a sentence must be always interpreted as the union of a subject and a predicate, it follows that the syntactic subject of a sentence must be always realized, in agreement with the EPP (Extended Projection Principle). Throughout the literature it has been noted that, differently from verbs, nominalizations have not to obligatorily carry out their arguments. In (7a) there are no arguments and the phrase is grammatical. It could be speculated, though, that in (7a) the DN takes null arguments in a parallel way an inflected verb can take null subject and null object in a pro-drop language (as Italian). Besides this, it remains the fact that, differently from verb arguments, if one of the DN arguments is omitted, the interpretation of the remaining one is ambiguous, at least in Italian. In (7c) it is
no longer possible to decide if the noun ‘barbarians’ is the agent or the theme of the event ‘destroy’.

(7) a. *La distruzione*  
‘The destruction’

b. *La distruzione dei Romani da parte dei barbari*  
‘Barbarians’ destruction of Romans’

c. *La distruzione dei barbari*  
‘The destruction of Barbarians’

Interestingly, in Italian there is a particular kind of nominalization, called infinito sostantivato (IS; see Zucchi 1993), that permits an unambiguous interpretation of the cases illustrated in (7c). The IS is an infinitive that have the same distribution of a noun, can be modified by a determinant and by an adjective. Differently from DNs, the IS is applicable to all verbal routes (productivity), cannot be pluralized, can assign accusative case (and sometimes nominative case as well), can be modified by adverbs, auxiliaries, modals, can be turned into passive and can be negated. Briefly, IS shows a more verb-like syntax with respect to DNs. In a certain sense, IS is similar to English –ing nominals.

(8) *Il distruggere case da parte dei vandali.*  
‘Vandal’s destroying houses.’

Given these peculiarities, it becomes clear the reason why DNs and IS can be informative for the study of the V-N dissociations and the linguistic pathologies in general. The phenomenon of the nominalization is the prism through which the nature of the V-N dissociation can be observed. Selective deficits and problems with the nominalizations can be explained as the consequence of three possible effects: (i) event vs. non-event; (ii) obligatory realization of the arguments vs. optional realization of the argument; (iii) noun syntax vs. verb syntax. In a first case the V-N dissociation would be of a conceptual-semantic nature. In this condition, both verbal selective deficits and problems with nominalizations should be observable. In a second case, the deficit would involve the semantic-syntactic interface and patients with verbal selective damage should do well with nominalizations. In a third case, the deficit would be syntactic in nature and verbal selective problems would display a specific kind of nominalization showing a more verbal syntax, that is, IS in Italian.

Moreover, if patients suffering selective nominal deficits do bad with DNs and patients suffering selective verbal deficits do well with DNs, it will be the case for an analysis of verbs and nouns as two distinct well-cut classes. On the contrary, if patients suffering selective verbal deficits do bad with DNs or at least with particular kinds of nominalizations, it should be the case that there exist some features and properties that cross the boundaries of the class distinction.

1.3.1 Nominalizations in aphasia

Despite the great amount of studies devoted to the analysis of the V-N dissociations in aphasia, only in few cases (Collina, Marangolo & Tabossi 2001; Crepaldi et al. 2006; Siri et
al. 2008) tests have been designed to verify how verbal and nominal selective deficits impact on closely related pairs involving verbs and the corresponding nominalizations.

The study by Collina and co-workers is one of the first studies on the production of nominalizations in aphasia and it has the merit of testing the DNs production in order to have a better comprehension of the V-N dissociation nature. The authors administered a picture naming task in which the target items were made up of simple nouns, DNs, one-argument and two-arguments verbs. The three Italian aphasic patients tested did better (i) in the production of nouns than verbs; (ii) in the production of nouns than DNs and (iii) in the production of one-argument verbs than two-arguments verbs. According to the authors, these results would be in agreement with the argument complexity hypothesis (i.a.: Thompson 2003). The V-N dissociations would not be due to an imageability effect or to a grammatical class effect, but to a statistic effect according to which verbs and not nouns are generally associated to an argument structure. The production of the three patients tested would show that the complexity argument effect would take place at both verbal and nominal level.

However, some remarks can be moved. A first doubt concerns the items of the tasks. It is not clear why DNs like *scontro* ‘crash’ and *divisone* ‘division’ were listed with the simple nouns. As for the verbs lists, a reason cannot be easily found for having considered one-argument verbs causative verbs as *bollire* ‘to boil’ and transitive verbs as *giurare* ‘to swear’, *stirare* ‘to iron’ and *suonare* ‘to play’. As far as the DNs list is concerned, some targets cannot phonologically express their arguments, at least in Italian (*il pattinaggio/nuoto di Gianni*, lit. ‘the skating/swimming of John’). If the aim of the study is to compare the production of verbs and nouns on the basis of the argument structure it seems natural considering only those DNs that clearly take arguments. Moreover, the DNs targets do not reflect the complexity of the nominalization phenomenon in Italian. In fact, the nouns derived by highly productive nominalization processes were excluded, that is, the DNs ended in –*mento*, –*zione*, –*(a)ta* and the IS.

Secondly, the analysis of errors is sometimes obscure. The authors claim that the errors are due to a syntactic deficit and not to a loss of semantic information, but their argumentation is not completely convincing. For instance, semantic paraphasia as *pugnala*re ‘to stab’ > *uccidere* ‘to kill’ seems depending on the kind of the picture showed during the test or on a lexical impairment, but not on a syntactic deficit. In fact, a syntactic deficit would have caused a shift even as far as the argument structure of the target is concerned (i.e. *pugnala*re: two-arguments verb > *morire* ‘to die’: one-argument verb). A similar objection can be moved as for the analysis of the DNs production. Inexplicably, paraphasic productions such as *pugnalata* ‘stab’ > *uccisione* ‘killing’, *scoperta* ‘discovery’ > *sorpresa* ‘surprise’, *sparatoria* ‘shooting’ > *litigio* ‘quarrel’ have been considered syntactic and not semantic or lexical errors. It is worth noticing that these answers counted as errors are DNs that can phonologically take arguments.

In brief, Collina and co-workers’ study has undoubtedly the merit of first testing the DNs production in aphasia, but for the moment there is not enough evidence to agree with the authors’ conclusions, that is, with the claim that patients’ difficulties in naming the DNs would reflect an argument complexity effect and a syntactic deficit. The main remark on Collina, Marangolo & Tabossi’s work concerns the fact that the authors did not face with the question of the comparability between the argument structure of a verb and that of a noun. As already illustrated, a verb must obligatorily carry its arguments while a DN does not. The focus is not only on the possibility for some constituents (and not for others)
to have an argument structure, but also on the way that this possibility is given at the semantic-syntactic interface. Does this difference play a role in our mind/brain?

In the next section a neuropsychological study conducted on three Italian aphasic speakers will be described. This study aims at testing whether the production of nominalizations can be affected by verbal selective impairments, argument complexity, and the kind of syntax that nominalizations display, noun syntax (DNs) vs. verb syntax (ISs).

2. The neurolinguistic study

2.1 Participants

Three Italian aphasic patients (ED, LI and ZA) and three control subjects took part to this study. The control subjects matched the aphasic patients as for age and education. Both patients and control subjects were born in Veneto, in the North-East of Italy. The patients were selected by the speech-therapists of Clinica Neurologica I, Padua and IRCCS Ospedale San Camillo, Venice.

At the time of test administration, ED was a 76-years-old right-handed retired bank clerk with eight years of education. His L1 was Italian as spoken in the North-East areas near Padua. He suffered a vascular accident in the left frontal lobe. His aphasia was accompanied by dysgraphia, dyslexia and bucco-facial apraxia, but not by constructional and ideomotor apraxia. The AAT (Aachener Aphasia Test, Italian version) lead to a diagnosis of non-fluent Broca’s aphasia. The BADA (Batteria per l’analisi dei deficit afasici) revealed a V-N dissociation, verbs being significantly more impaired than nouns (V: 5/28; N: 16/30). The tasks of the present study were administered 15 months post-onset, where the verb-noun dissociation persisted. A screening test highlighted in ED’s production some difficulties with the number morphology (singular: 15/20; plural: 15/20) and with verb argument structure. More precisely, ED performed well with the transitive verbs in their active form (12/12) and with the unergative verbs (10/10), while he performed worse with the unaccusative verbs (7/10) and the passive verbs both with realized by-phrase (0/9) and not realized by-phrase (1/9).

LI was a 36-years-old right-handed employee with 18 years of education. Her L1 was Italian as spoken in the North-East areas near Padua. She suffered a vascular lesion involving the left temporal lobe. The AAT lead to a diagnosis of a mild fluent Wernicke’s aphasia. The BADA did not reveal a significant V-N dissociation (V: 15/28; N: 18/30). A screening test highlighted in LI’s production no problems with morpho-syntactic features of the noun, but, unlike the other patients in this study, problems with verbal negation (12/20). On the other hand, LI performed relatively well with verb argument structure (transitive verbs in their active form: 12/12; unergative verbs: 10/10; unaccusative verbs: 10/10; passive verbs both with realized by-phrase: 8/9 and not realized by-phrase: 7/9).

ZA was a 61-year-old right-handed worker with 8 years of education. His L1 was Italian as spoken in the North-East areas near Belluno. He suffered a vascular accident in the left frontal lobe. The AAT lead to a diagnosis of a non-fluent Broca’s aphasia. BADA’s picture naming task highlighted a significant V-N dissociation, verbs being more impaired than nouns (V: 13/28; N: 25/30). The tasks of the present study were administered four months post-onset, where the verb-noun dissociation persisted. A screening test highlighted in ZA’s

---

1 Here-hence all the results and the percentages reported indicate the correct production.
production some difficulties with the number morphology (singular: 13/18; plural: 17/17) and with verb argument structure. More precisely, ZA performed relatively well with the transitive verbs in their active form (6/12) while he performed worse with the unaccusative verbs (4/10) and the passive verbs both with realized by-phrase (1/9) and not realized by-phrase (0/9).

2.2 General methods

The items of the present study were presented one at a time, in both oral and written modality, in random order to minimize perseveration errors. The tasks were administered by the speech-therapists in different sessions, thus minimizing the effects of tiredness. Before each task participants were trained in order to make sure that they had well understood the assignment. They were required to answer orally and no time limits were set. Patients’ answers were transcribed by the speech-therapists and tape recorded. Word frequency data were collected from Colfis (Bertinetto et al. 2005).

2.3 Task 1: the derived nominalizations (DNs) and the infinito sostantivato (IS)

Hypothesis. Consistently with their V-N dissociation, ED and ZA were expected to perform better with target DNs involving noun syntax than with target ISs (‘the + infinitive’). In fact, as already illustrated (see section 1.3), ISs are a particular kind of nominalizations that involve verb syntax and can assign accusative case.

Methods. A completion task required to derive the target nominalization starting from a neutral verb form (infinitive form or 3rd singular, simple present in the case of target DNs and 3rd singular, simple present in the case of target ISs). Twenty targets involved DNs (noun syntax; 9a) and 20 targets involved ISs (verb syntax; 9b):

(9) a. La ____ di una bambola da parte dello zio
   ‘the(-S.) ____ of a doll by the uncle’
   Input: to promise Target: (the) promise
   b. Il __ favori da parte del ministro
   ‘the __ favours by the minister’
   Input: promise-3rd s. Target: (the) promising

Target words could not be fully balanced for word frequency since DNs (mean frequency: 220) are marginally more frequent than ISs (mean frequency: 121). However the difference was not significant (t = 1.63, p < .11).

Results. In this task, all the control subjects performed at ceiling with the exception of one who totalized 19/20 in the IS condition. The three patients’ results are reported in (10) and a clear trend in favour of DNs (noun syntax) can be noted in each case.

(10) DNs/noun syntax ISs/verb syntax
    ED:13/20 (65%) ED: 2/20 (10%)
    ZA: 9/20 (45%) ZA: 6/20 (30%)
    LI: 16/20 (80%) LI: 12/20 (60%)

ED and ZA did clearly better with nominalizations involving noun syntax (13/20; 9/20) than with nominalizations involving verb syntax (2/20; 6/20; z = 3.233, p < .001). As for
LI’s performance, instead, the difference was not significant (16/20 vs. 12/20; z = 1.38, p < .1).

Errors were computed separately for each patient. Most errors consisted of substitutions, although ED committed four omissions. For both types, nominal targets were most frequently substituted with either a verb, or, less frequently, with a DN by both ED and ZA. Unexpectedly, because of the nature of ED and ZA’s problems, most frequent substitutions consisted of a 3rd person singular present or past, past participle, and only rarely with a non-target DN. In (11) ED and ZA’s errors direction analysis is summarized:

(11) DNs (noun syntax)
ED: substitutions with verb = 6/7 substitutions with noun = 1/7
ZA: substitutions with verb = 11/11 substitutions with noun = 0/11

ISs (verb syntax)
ED: substitutions with verb = 8/12 substitutions with noun = 4/12
ZA: substitutions with verb = 9/14 substitutions with noun = 5/14

ED and ZA substituted 17 times out of 26 with a verb and 9 times out of 26 with a noun; the difference is significant (z = 2.219, p < .02). Substituting words tended to be less frequent with respect to the target for both patients, however the difference is not significant (ED: t = 1.86, p < .07; ZA: t = 1.87, p < .07).

Differently from ED and ZA, LI’s substitutions differ with respect to the kind of nominalization involved. In the case of target DNs (noun syntax), her substitutions consisted of infinitive and a compound. In the case of target ISs (verb syntax), her substitutions consisted of gerundive, present participle and non-target DNs.

2 Target: ‘la SCOPERTA (DN) dell’America da parte di Colombo’ (‘Colombo’s DISCOVERY of America’) → patient’s substitution: ‘*la SCOPRÌ dell’America da parte di Colombo’ (‘Colombo’s DISCOVERED of America’); Target: ‘il DISTRUGGERE (IS) città da parte dei barbari’ (‘barbarians’ DESTROYING cities’) → patient’s substitution: ‘*il DISTRUSSE città da parte dei barbari’ (‘barbarians’ DESTROYED cities’).

3 Target: ‘il LAVAGGIO (DN) dell’auto da parte del meccanico’ (‘mechanic’s WASHING of the car’) → patient’s substitution: ‘*il LAVATO dell’auto da parte del meccanico’ (‘mechanic’s WASHED of the car’); Target: ‘il RACCOGLIERE (IS) giochi da parte dell’UNICEF’ (‘UNICEF’s COLLECTING toys’) → patient’s substitution: ‘*il RACCOLTO giochi da parte dell’UNICEF’ (‘UNICEF’s COLLECTED toys’).

4 Target: ‘la CAMMINATA (DN) di Maria nel bosco’ (‘Maria’s WALK in the wood’) → patient’s substitution: ‘*la CAMMINO di Maria nel bosco’ (‘Maria’s PATH in the wood’).

5 Target: ‘la CORSA della mamma verso casa’ (‘mum’s RUNNING home’) → LI’s substitution: ‘la CORRERE della mamma verso casa’ (‘mum’s RUN home’).

6 Target: ‘il LAVAGGIO dell’auto da parte del meccanico’ (‘mechanic’s WASHING of the car’) → LI’s substitution: ‘*il LAVASCIUGA dell’auto da parte del meccanico’ (‘mechanic’s WASH-DRY of the car’).

7 Target: ‘il CATTURARE i ladri da parte della polizia’ (‘police’s ARRESTING the thieves’) → LI’s substitution: ‘*il CATTURANDO i ladri da parte della polizia’ (‘police’s ARRESTING the thieves’).

8 Target: ‘il COMBATTERE i giganti da parte di Ercole’ (‘Hercules’ FIGHTING the giants’) → LI’s substitution: ‘*il COMBATTENTE i giganti da parte di Ercole’ (‘Hercules’ FIGHTER the giants’).
Interim discussion. As expected, given their V-N dissociation, ED and ZA performed better with DNs involving noun syntax than with ISs involving verb syntax. These results are not due to a frequency effect; moreover, the patients showed a trend in substituting the targets with less frequent words. Thereby, a morphosyntactic deficit can be argued for both patients. Such a conclusion matches the predictions of those linguistic studies embracing a morphosyntactic approach to the nominalization puzzle (even if from slightly different perspective, see: Alexiadou 2001; Alexiadou, Iordachioaia & Schäfer 2011; Borer 2003; Chomsky 1970; Marantz 1997; Ramchand 2008). The crucial point is that nominalizations do not form an homogeneous category. Some nominalizations can assign accusative case and be modified by aspectual adverbs, modals and auxiliaries (as the Italian IS does) while others can take of-complements and be modified by adjectives (as the Italian DNs do). A possible account that seems to be in line with the results of the present study posits that such differences rely on the number and the type of verbal and nominal layers entering the nominalization structures. If a DP and a vP projections must be always postulated since the former is responsible for the presence of a determiner and the latter for the argument structure/Aktionsart properties, the occurrence of other projections is restricted to the particular kind of nominal considered. In terms of Alexiadou (2001) and following works, a nominalization to be more verb-like must project AspectP (which determines the morphological aspect) while a nominalization to be more noun-like must project ClassP and eventually NumberP (which respectively determine the mass/count properties and the morphological number).

(12) a.  [DP ([NumberP [ClassP [nP [VoiceP [vP) [Root
b.  [DP [AspectP [VoiceP [vP [Root
(from Alexiadou 2010b)

If ED and ZA’s relatively better performance with noun-like nominalizations was somehow expected, their substitution errors were not. Both patients significantly substituted a target nominalization with the correspondent fully inflected verb, frequently a past tensed verb. This is a surprising result since in the literature (Bürchert et al. 2005; Friedmann & Grodzinsky 1997; Yarabay Duman & Bastiaanse 2009) it is well known that aphasic people (especially agrammatic people) find it difficult to produce past tensed verbs.

It could be argued that this kind of substitutions is a consequence of the lack of a full verb in the stimulus. A further task was developed in order to compare the production of nominalizations in phrase (DP) context with the production of DNs and ISs in full sentence context. This task was given following the same methodologies and using the same targets of the previous one. The only difference consisted in having given a full sentence, and not a phrase, as stimulus. Targets consisted of 24 DNs (13a) and 12 ISs (13b):

(13) a. *La ___ dell’America fu nel 1492
   ‘The ___ of America was in the 1492’
   Input: discover-3rd s. Target: discovery
b. *(Il) __ i pranzi fa male alla salute
   ‘(The) ___ the meal is not healthy’
   Input: skip-3rd s. Target: skipping

No significant difference was found (ZA: tot: z = 1.001, p < .3; DNs: z = 0.137, p < .8; ISs: z = 1.58, p < .1; LI: tot: z = 1.167, p < .2; DNs: z = 0.585, p < .6; ISs: z = 0.636, p < .1).
In addition, no significant difference in the kind of substitutions committed in the two tasks was found as well.

It could be further argued that patients’ poor performance with ISs is a consequence of a general impairment on infinitive form as such. In order to verify this hypothesis, another task was given in which the patients had to complete the sentence stimulus with the infinitive form of the verb.

(14) \textit{Il nonno pensa di ___ la maratona}

‘Grandpa thinks of ___ the marathon’

\text{Input: run-3rd s. Target: running}

ZA and LI scored respectively 13/14 and 14/14. Thus, ZA’s failure with the production of ISs is not due to a general problem with the infinitive .

In the last section of this paper a possible explanation of ED and ZA’s substitution errors will be given.

2.4 Task 2: the argument effect (I)

In task 1 the two Broca’s aphasic subjects (ED and ZA) scored better with those nominalizations involving noun syntax (DNs) than with nominalizations involving verb syntax (ISs). On the contrary, the fluent aphasic patient without V-N dissociation (LI) did not perform significantly better in one type with respect to the other. Another task was developed, therefore, in order to verify whether the patients’ production of nominalizations is influenced non only at the syntactic level (i.e., noun syntax vs. verb syntax), but also at the semantic level (or, at least, at the semantic-syntactic interface), in particular as far as the number of the arguments involved is concerned. In this respect, it is worth noticing that, in the screening test, the patients’ production of one-argument verbs was not better than the production of two-arguments verbs (see section 2.1) and that this result is not similar to Collina et al. (2001) findings (see section 1.3.1). ED did significantly better with two-arguments verbs \((z = 2.042; p < .04)\), ZA performed the same with both types of verbs \((z = 0.469, p < .6)\) and LI performed always at ceiling. However, the patients performed relatively better with those verbs that do not involve the movement of the object to the subject position. That is, they scored higher with transitive active verbs than with passive and unaccusative verbs (see section 2.1). In this sense, ED’s and ZA’s verb production seems to be affected at the morphosyntactic level (i.e., active vs. passive verbs) rather than at the semantic level (i.e., one-argument vs. two-argument verbs).

\textit{Hypothesis.} If ED and ZA’s problems merely concern the syntactic level, no significant difference between their production of DNs derived from one argument verbs and their production of DNs derived from two arguments verbs should be observable.

\textit{Methods.} The patients were given a completion task in which they had to derive the required DN starting from a neutral verb form (infinitive). Forty targets consisted of DNs derived from two arguments verbs (15a) and 40 targets consisted of DNs derived from one argument verb (15b):

(15) a. \textit{La ___ dei ladri da parte della polizia}

‘The___ of the thieves by the police’

\text{Input: to arrest Target: arrest}

b. \textit{La ___ della zia per Londra}
‘The ___ of (the) aunt toward London’
Input: to depart Target: departure

Results. In this task the control subjects performed at ceiling. Patients’ results are reported in (16). As expected, no significant difference between the production of DNs derived from one argument verbs and the production of DNs derived from two arguments verbs is observable in ED and ZA’s performance (respectively: $z = 0.709, p < .4$ and $z = 0.905, p < .3$). As for LI’s performance, a trend is to be noted in the direction of DNs derived from one argument verbs, though, the difference is not significant ($z = 1.071, p < .2$).

(16) DNs from two arguments verbs
ED: 15/40
ZA: 19/40
LI: 29/40
DNs from one argument verbs
ED: 12/40
ZA: 15/40
LI: 33/40

Once again, the errors committed by the two patients with V-N dissociation are surprising. As in task 1, errors consisted of substitutions. For both types, DNs targets were most frequently substituted with a verb (3rd person singular present or past; past participle; infinitive) and only rarely with a non-target DN or other nominal elements. Instead, LI’s substitutions mainly consist of non-target DNs or infinitives.

Interim discussion. As expected, no significant difference between the production of DNs derived from one argument verbs and the production of DNs derived from two arguments verbs was observed in ED and ZA’s performance. This result, taken together with the scores obtained by the two Broca’s aphasics in task 1 (i.e. the production of noun-like nominalizations better than that of verb-like nominalizations) and in the screening test (i.e. no difference between the production of one argument verbs and two argument verbs), can be seen as an evidence for a morphosyntactic analysis of ED and ZA’s problems.

This result differs from Collina et al. (2001) findings since the three patients of their study performed better with one-argument verbs than with two-arguments verbs. This performance can be due to a different deficit, more semantic than syntactic in nature as the patients’ substitutions seem to indicate (see section 1.3.1). However, while Collina and co-workers reported an argument complexity effect for verbs production, they did not mention whether the same effect was involved in the DNs production as well, that is, they did not specify whether the patients performed better with DNs derived from one argument verbs than with those derived from two arguments verbs. This is not a trivial issue if one aims at demonstrating that the argument complexity effect plays a crucial role at both verbal and nominal level.

As for the present study, if ED and LI’s production of one argument verbs is compared with their production of DNs derived from one argument verbs the difference is significant (ED: $z = 2.331, p < .01$; LI: $z = 1.991, p < .04$). A similar finding is observable in the case
of two arguments verbs and DNs derived from two arguments verbs (ED: $z = 3.801$, $p < .0002$; LI: $z = 2.046$, $p < .04$). At a first glance, the fact that the participants performed better with the verbs than with the nominalizations, i.e. nominal elements, would be hard to settle with their verbal selective deficit. This result cannot be ascribed to a greater morphological complexity of the DNs, as Collina and co-workers did regarding the performance of the participants at their study. In fact, in the present study, differently from Collina et al.’s study, both verbs and nominalizations had to be derived from a neutral form and, at the end, the participants were asked to produce the required target selecting the correct suffix for the input (an inflectional suffix in the case of verbs, a derivational suffix in the case of DNs). In this regard, it is worth stressing that the three subjects (especially ED and ZA) often substituted the target nominalization with a verb and this could support the observation that nominalizations were more demanding to produce than verbs in this task. Where does the difficulty rely on if the greater complexity of nominalizations with respect to verbs is not merely morphological? In the last section of this paper a possible explanation for the nominalization complexity will be briefly argued, in comparison with other neuropsychological studies.

2.5 Task 3: the argument effect (II)

**Hypothesis.** Even if from different perspectives, some linguistic theories (Alexiadou 2001; Alexiadou & Grimshaw 2008; Grimshaw 1990; Marantz 1997) claim that only DNs carrying out thematic roles denote events and, in this sense, are verb-like, while DNs not carrying out thematic roles do not denote events and behave as simple nouns. If this is correct, the two Broca’s patients with the V-N dissociation were expected to do better with DNs not carrying out thematic roles, since these latter are supposed to be more noun-like.

**Methods.** The patients were given a completion task in which they had to derive the required DN starting from a neutral verb form (infinitive). Ten targets consisted of DNs carrying out thematic roles (17a) and ten targets consisted of DNs not carrying out thematic roles (17b):

17a. *La ___ di Gianni in piscina*  
‘The ___ of John in the swimming pool’
Input: to swim Target: swim

17b. *Il ___ in piscina fa bene*  
‘The ___ in the swimming pool is healthy’
Input: to swim Target: swimming

Target words were balanced for frequency (DNs with thematic roles: mean frequency 101; DNs without thematic roles: mean frequency 171; $t = 0.84$, $p < .09$).

**Results.** Two of the three control subjects scored 9/10 with DNs realizing thematic roles and 10/10 in the other condition. The third control subject performed always at ceiling. Patients’ results are reported in (18):

As for ZA’s performance, the difference is not significant ($z = 0.152$, $p < .8$ and $z = 0.146$, $p < .8$). This can be due to the fact that ZA suffered a more severe grade of aphasia with respect to the other two patients. As a consequence, this result cannot be taken as evidence for the equivalence between the verb argument structure and the DNs argument structure. It seems more reasonable thinking that ZA found it difficult to perform in both conditions since his greater impairment.
The crucial role of the event structure in the retrieval of nominalizations in aphasia

Surprisingly, the two patients with the V-N dissociation (ED and ZA) did better with DNs carrying out thematic roles (verb-like) than DNs not carrying out thematic roles (noun-like). Taken together, ED and ZA’s results are significant \( z = 2.543, p < .01 \); however this effect is mainly due to ZA’s score, while only a trend is observable in ED’s performance. Interestingly, LI’s score goes in the same direction of the other two patients’ scores. A part of one omission by ED, even in this task all errors committed can be classified as substitutions consisting of past participle, infinitive, or another non-target DNs. Substituting words were found to be same frequent with respect to the target (ED: \( t = 1.65, p < .1; \) ZA: \( t = 0.04, p < .9 \)).

**Interim discussion.** The patients’ performance goes in the opposite direction with respect to that predicted by the hypothesis. Their results cannot be explained in terms of a target frequency effect nor in terms of an argument structure complexity effect. A possible reason can be seen in the provided input, that is, the neutral verb form from which the DNs had to be derived. More precisely, it is possible to argue that patients’ performance may be influenced by the stimulus, that was a verb. Thus they might have preferred the event interpretation (since verbs describe events) rather than the non-event interpretation. As a consequence they did better with DNs carrying out thematic roles (that are verb-like since the occurrence of thematic roles is always linked to the event reading) than with DNs not carrying out thematic roles (noun-like according to some theories, as Grimshaw 1990).

Besides this, it could be argued that the occurrence of thematic roles does not constitute a necessary condition in order to interpret a nominalization as an event. In a more recent work, Alexiadou and Schäfer (2010) stress that “while complement structure builds on event structure, the presence of event structure does not necessarily imply the presence of complement structure”. If it is the case, in what sense are the two types of nominalization different from each other? Following Borer (2005) and Mittwoch (2005), Alexiadou & Schäfer (2010) try to demonstrate that “rather the aspectual properties of the constructions are instrumental for the licensing of complement structure”. In the next section the link between aspectual properties and complements licensing will be better illustrated since it represents a crucial point in order to make sense of subjects’ performance and substitution errors.

**2.6 General conclusions**

Nominalizations represent a crucial test for the hypothesis on the nature of the V-N dissociations since they share both nominal and verbal properties at both the semantic and the syntactic level. However, despite the great amount of studies devoted to the analysis of the V-N dissociations in aphasia, only few studies (Collina, Marangolo & Tabossi 2001; Crepaldi et al. 2006; Siri et al. 2008) tested how verbal and nominal selective deficits impact on closely related pairs involving verbs and the corresponding nominalizations. The present study aimed at testing whether the production of nominalizations can be affected by verbal selective impairments, argument complexity, and the kind of syntax that the nominalizations display, noun syntax (DNs) vs. verb syntax (ISs).
Task 1 highlighted a syntactic effect in the two Broca’s aphasic subjects’ production of nominalizations. In fact, ED and ZA performed significantly better with nominalizations involving noun syntax than with nominalizations involving verb syntax. This result is compatible with both ED and ZA’s kind of aphasia and V-N dissociation, verbs being significantly more impaired than nouns in their production. As a consequence, a syntactic deficit can be argued for both patients. On the other hand, the Wernicke’s aphasic subject with no V-N dissociation (LI) performed relatively well in both conditions. Her scores should be considered in the light of her mild grade of aphasia: as illustrated in section 2.1, LI suffered a less severe aphasia with respect to ED and ZA.

Secondly, the patients’ production of nominalizations was not affected at the semantic level by the number of arguments involved (task 2) and the occurrence vs. the lack of the thematic roles (task 3). This result, taken together with the scores obtained by the two Broca’s aphasic subjects in task 1 (i.e., the production of noun-like nominalizations better than that of verb-like nominalizations) and in the screening test (i.e., no difference between the production of one argument verbs and two argument verbs), supports an interpretation of the patients’ errors as stemming from a morphosyntactic deficit. In this respect, it is worth stressing that the participants performed relatively better with noun-like nominalizations than with verb-like ones.

At the same time, though, these findings are problematic, especially because of ED’s performance; despite his verbal selective deficit, he performed relatively better with verbs than with nominalizations, that is, nominal elements. Moreover, the two Broca’s aphasics often substituted a target nominalization with a fully inflected verb and not with a semantically/lexically related noun, as expected and as to some extent done by the Wernicke’s aphasic participant LI. These errors could be partially explained pointing out that the stimulus from which the target had to be derived was a verb: producing a fully inflected verb from a neutral verb form is maybe easier or a more canonical task than producing a related noun or DN. Nevertheless, this way of reasoning leaves the fact that the patients often substituted the required nominalization with a past tensed verb (simple past and present perfect) unexplained. In the literature, past tenses are extensively demonstrated to be demanding for aphasic people, especially Broca’s aphasic speakers (Bürchert et al. 2005; Friedmann & Grodzinsky 1997; Yarabay Duman & Bastiaanse 2009). Given this scenario, ED and ZA’s production of nominalizations does not seem to depend directly on their V-N dissociation, but on a syntactic impairment on the one hand and on the complexity of nominalizations on the other. If correct, this reasoning may lead to the conclusion that DNs are complex nouns, more complex even than verbs. An MRI study conducted by Siri et al. (2008) comes to the same result. Siri and co-workers asked some Italian normal subjects to name a picture by means of an inflected finite verb, an infinitive or a nominalization, according to the provided instruction. They found no specific cortical activation in the case of verbs; however, they found a greater IFG (Inferior Frontal Gyrus, which involves Broca’s area) activation for the production of nominalizations with respect to verbs (both finite verbs and infinitives), that is, an increase of the cognitive resources needed in the nominalization condition. In what does nominalizations’ complexity lie?

3. Regarding the errors: some final considerations looking at linguistic analyses

The last issue concerns a possible account for nominalizations’ complexity. In what sense are nominalizations more demanding than verbs (at least in the tasks of the present
The crucial role of the event structure in the retrieval of nominalizations in aphasia

The greater evidence for nominalizations’ complexity is represented by the fact that the subjects suffering verbal selective deficit often substituted the target DN or IS with an inflected verb. It is surprising that in the most cases these substitutions consisted of past tensed verbs since in the literature it is well know that the production of past tensed verbs is demanding for agrammatic people (Bürchert et al. 2005; Friedmann & Grodzinsky 1997; Yarabay Duman & Bastiaanse 2009). Therefore, it seems reasonable thinking that a deeper, linguistic analysis of subjects’ errors could highlight something interesting about the nominalization phenomena. In the tables below the substitution errors are reported for each participants. The substitutions of target DNs are listed in table 1, those of target ISs are listed in table 2.

<table>
<thead>
<tr>
<th>Finite verbs</th>
<th>Non finite verbs</th>
<th>Non target DNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Present</td>
<td>Perfect</td>
</tr>
<tr>
<td>ED</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ZA</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>LI</td>
<td></td>
<td>✔</td>
</tr>
</tbody>
</table>

Table 1: substitutions of a DN target

<table>
<thead>
<tr>
<th>Finite verbs</th>
<th>Non finite verbs</th>
<th>Non target DNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Present</td>
<td>Perfect</td>
</tr>
<tr>
<td>ED</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>ZA</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>LI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: substitutions of a IS target

First considering table 1, ED and ZA mostly substituted a target DN with a verb in the infinitive, past participle, present tense, present perfect tense and past tense. On the contrary, LI’s substitutions go in the direction that was firstly hypothesized for the other
two patients: besides the substitutions with infinitive, the target DNs were replaced by other non-target DNs. The fact that the stimulus from which the DN had to be derived was a verb in a neutral form is not enough to explain the production of verbs in the past participle, present perfect tense and past tense. It can be noted that in Italian these three verbal tenses share the same kind of aspect, i.e. the perfective aspect. A detailed discussion on aspect does not fall within the aims of this paper. Nonetheless, following Bertinetto (1991) and Salvi & Vanelli (2004), it is possible to argue that the perfective aspect shows an event in its completeness, the endpoint of the event in question being included. On this regard, it is significant that in some cases LI replaced the target DNs with a particular kind of nominalization, the *nomina agentis* ending in *-tore*, similar to English *–er* nominals, which typically denotes the external argument of the underlying predicate. In fact, as argued in Lo Duca (2004), differently from other kind of *nomina agentis*, the *nomina agentis* ending in *-tore* can receive event-reading and bear perfective value. In other words, this kind of *nomina agentis* can refer to non-generic events and to a single occurrence of the relevant event. If it is true that subjects’ errors are not accidental, but are aspect-driven, two observations have to be drawn. Firstly, aphasic data seem to suggest that, at least in theory, every kind of nominalization can receive event-reading and be linked to an event structure (as argued in Alexiadou & Schäfer 2010). In fact, the three participants replaced a target nominalization with a verb even in those conditions where a non-event-reading was supposed (see task 3). Secondly, since subjects’ substitutions can be comparable as far as the perfective aspect is concerned, it follows that the event described in an Italian DN is preferably seen in its completeness. If this reasoning is on the right track, it cannot be a case that the subjects performed relatively better with nominalizations ending in *–(a)ta*, a suffix that is peculiar to Italian past participle, than with nominalizations ending in *–mento e –zione* (19). The fact that DNs in *–(a)ta* mainly refer to single, inherently bound events (Gaeta 2004) goes in hand with the perfective value carried by the Italian past participle suffix. In turn, subjects’ preference for DNs in *–(a)ta* well fits within the hypotheses that analyse subjects’ substitutions in terms of aspect.

(19)  
<table>
<thead>
<tr>
<th><em>–(a)ta</em> DNs</th>
<th><em>–mento/-zione</em> DNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ED: 9/20</td>
<td>ED: 1/10</td>
</tr>
<tr>
<td>ZA: 8/20</td>
<td>ZA: 0/10</td>
</tr>
<tr>
<td>LI: 19/20</td>
<td>LI: 6/10</td>
</tr>
</tbody>
</table>

Considering table 2, ED and ZA committed the same kind of substitutions even in the case of target ISs. Hence they kept on preferring a perfective reading of the event described in the nominalizations in spite of the compatibility of the *infinito sostantivato* with an atelic value and therefore with an imperfective reading (i.a.: Vanvolsem 1983; Zucchi 1993). More interestingly, LI’s substitutions of target ISs are different from those of target DNs. It

10 For example, the *nomina agentis* in sentence in (i) can be paraphrased as “the person who rescued x” referring not to a generic person who habitually rescues people, but to the precise person that in a precise moment has rescued John.

(i)  
Fortunatamente il soccorritore di Gianni è un bravo infermiere.  
Thankfully the rescuer of John is a good nurse.

11 In this respect, Melloni (2007) noted that in Italian affixes such as *–zione* inherit the inner aspect properties of the base verb while others, as *–ata*, modify the inner aspect of the verb.
is worth reminding here that LI does not seem to suffer the same syntactic deficits which affect the other two subjects’ production. It is likely that her errors may reflect a phonological/lexical impairment rather than a syntactic one. Her substitutions of target ISs consisted of gerunds and present participles which are comparable with the IS as for: (i) the unlimited productivity with all verbal roots; (ii) the possibility of assign accusative case and (iii) the [+ durative] and [- telic] values (Benincà & Cinque 1991; Lo Duca 2004). Significantly, LI never substituted target DNs with gerunds or present participles on one hand and target ISs with *nomina agentis* in –tore on the other. Under these circumstances, it should be likely that LI substituted the required nominalizations just with those verbs or those nominals that could maintain the syntactic properties and the semantic values of the targets whenever the retrieval of the targets in question was prevented because of her phonological/lexical problems. LI could take advantage of such a strategy since her milder (and fluent) aphasia with respect to ED and ZA’s more severe (and non-fluent) aphasia.

In conclusion, errors’ analysis enlightens the crucial role played by the Aktionsart, the kind of the event structure and, to some extent, the grammatical aspect in the nominalization phenomena. In this respect, the data from aphasia seem to be in line with those linguistic analyses that explain the variation among the subtypes of nominalizations in terms of different event structures and different aspect projections (Alexiadou 2001 and following works; Borer 2005; Fábregas & Marin 2012 based on Ramchand 2008; Sichel 2010) rather than with those analyses that posit the differences in dichotomies such as the occurrence vs. the lack of the arguments or the event vs. the non-event reading (Grimshaw 1990). For example, capitalizing on Harley & Noyer (2000) and Pesetsky (1995), Sichel (2010) noted that it is not possible to derive a transitive DN from English verbs displaying an inchoative/causative alternation (20) unless the syntactic subject of this kind of verbs is understand not only as the cause of the event (21b), but also as the direct participant to the event (21c).

(20)  a. *Bill’s growth of tomatoes
    Bill grew tomatoes / Tomatoes grew
   b. Bill’s cultivation of the tomatoes
    Bill cultivated the tomatoes/ *The tomatoes cultivated

(21)  a. The weather gradually improved her mood
   b. #The weather’s alteration of her plans
   c. The wind’s alteration of the position of the rock

On Sichel’s account, the restriction on the direct participant is a consequence of the kind of the event a DN can represent. Briefly, an event consists of two sub-events: the first one is made up by the verbal root and its internal argument, the second one is the external argument. In a simple, single event the two sub-events must overlap, that is, (i) the sub-events must have the same spatial and temporal properties and (ii) these properties must be the predicates of an entity that is necessarily a participant in both sub-events. Definitively, “if a simple event includes an external argument, the participation of the argument is co-temporal with the initiation of the event” (Sichel 2010). It follows that “when the participation of the external argument is not co-temporal the event is a complex one”. In this sense, the DNs are compatible with simple events only, while more verbal nominalizations such as the English –*ing* nominals are compatible with complex events as well. In fact, differently from the DNs, –*ing* nominals are grammatical with a non direct
participant external argument, as in the case of –ing nominals derived from verbs displaying an inchoative/causative alternation (22):

(22) a. Bill’s growing of tomatoes  
   b. Bill’s growing tomatoes  
   c. #Bill’s growth of tomatoes

Although the English causative verbs class does not perfectly overlap the Italian one, this kind of analysis seems to hold for Italian nominalizations as well. Italian causative verbs, although less frequent than the causative periphrases, mostly overlap with ergative verbs. Interestingly, Italian nominalizations derived from simple causative verbs behave the same of the corresponding English nominalizations. When the external argument is present, the ND is out while the related IS is perfectly grammatical (23; 24).

(23) a. La crescita dei pomodori (*da parte di Gianni)
   ‘(*John’s) growth of tomatoes’
   b. Il crescere i pomodori da parte di Gianni
   ‘John’s growing tomatoes’

(24) a. La guarigione dei vitelli (*?!da parte del veterinario)
   ‘(*Veterinarian’s) recovery of calves’
   b. Il guarire i vitelli da parte del veterinario
   ‘Veterinarian’s recovering calves’

Thus, as in English, in Italian the DNs cannot host complex events. It is not a case that the above examples on DNs (23a; 24a) turn out grammatical when the prepositional phrase introducing the external argument da parte di ‘by’ is replaced with other phrases, such as per opera di, lit. ‘at the hands of’ or per volontà di, lit. ‘at the wish of’, that can be analysed as further nominalizations. It is to say that two DNs are needed in order to describe a complex event. As Giorgi (1988) noted, the phrase ‘da parte di’ indicates the origin of the event, that is, the direct participant in terms of Sichel (2010). Moreover, when the internal argument is [-animate] and so there is no ambiguity about the origin of the event, the simple preposition di ‘of’ can replace the phrase da parte di (25), but not per volontà/opera di (26). Significantly, in the case of ISs the phrase da parte di can introduce not only the direct participant, but also an external cause. This is to be drawn from IS’ more complex, verbal structure.

(25) a. La mangiata di caramelle gommose da parte di Simone
   b. La mangiata di Simone di caramelle gommose
      ‘Simone’s eating of gumdrops’

(26) a. La crescita dei pomodori per opera di Gianni
   b. #La crescita di Gianni dei pomodori
      ‘The growth of tomatoes (due to the work of John)’

By the end, Sichel’s analysis can account for Italian nominalizations as well. In her view, the different properties of nominalizations have to be derived from the different kinds of event structure a nominalization can host. If this arguing is correct, it is possible to reinterpret subjects’ problems with nominalizations in terms of event structure rather than
The crucial role of the event structure in the retrieval of nominalizations in aphasia

in terms of argument structure or syntax. More precisely, the relatively better performance with DNs than with ISs is not just a matter of difference between noun and verb syntax, but also a matter of difference between the various event structures. In fact, according to Sichel (2010), the more nominal or more verbal properties of a nominalization are firstly to be derived from the kind of event structure (simple vs. complex). Thus, subjects’ performance can be explained as a preference for nominalizations hosting simple events with respect to nominalization hosting complex events and their substitutions errors as morphological strategies in the effort to retrieve the event properties of the target nominalizations.

As the same author claims, Sichel’s analysis can be integrated with other approaches, as that elaborated by Alexiadou (2001 and following works) and above described (see section 2.3). Crucially, both approaches stress that the linguistic data in nominals are better captured above the lexical root level and have not to be derived from the occurrence vs. the lack of arguments or the event vs. the non-event reading. Sichel’s constraint on event structure can be understood as a constraint on the kind of morphological and syntactic properties a nominalization can host. Capitalizing on Alexiadou’s works, it is possible arguing that nominalizations (at least Italian nominalizations) hosting simple events lack some verbal projections such as AspP, while nominalizations hosting complex events (i.e. IS) can enter a syntactic derivation in which AspP is allowed. Alexiadou and Sichel’s analyses, taken together, can account for the aspect-driven substitution errors committed by the subjects of the present study since they enlighten the link between event structure, Aktionsart (or inner aspect in the spirit of Verkuyl 1993), grammatical aspect (or outer aspect) and syntactic properties. On this regard, an analogous proposal is developed in Fabregas & Marín (2012) on the basis of Ramchand (2008). According to Ramchand (2008), in the first phase of the syntactic derivation there are three relevant projections for the introduction of the arguments: (i) initP which introduces the causation event and licences the external argument; (ii) procP which specifies the nature of the change or process and licenses the entity undergoing change or process; (iii) resP which gives the result state of the event and licenses the entity that comes to hold the result state. Fabregas & Marín (2012) claim that nominalizing affixes could spell out any of these three verbal heads giving rise to different interpretations that are sensitive to inner aspect (Aktionsart). For example, the suffix -ing could realize both process and initiator.

In sum, even if from slightly different perspectives, all these approaches stress the crucial role of the inner (and outer) aspect in the nominalizations’ derivation and this is in agreement with the analysis developed for the substitution errors made by the subjects of the present study. If this reasoning is on the right track, a tentative answer to the initial question can be given. In what does nominalizations’ complexity lie? In what sense are nominalizations more demanding than verbs (at least in the tasks of the present study)? The above discussion has emphasized the role of the aspect in nominalizations’ derivation. Capitalizing on the analysis put forward so far, it could be argued that Aktionsart and aspect (and not thematic roles) must necessary show up at a semantic and syntactic levels in order to turn a full referential noun into an event. In syntax, events can be fully realized in the DP or in the CP domain: in the former case they behave as nominalizations, in the latter they behave as verbs. In general, aspect (especially outer aspect) is a property of verbs, but not of full referential nouns. Nominalizations are still referential (Giorgi & Longobardi 1991; Longobardi 1994, 2001; Simone 2008), but they refer to events (i.e. can express Aktionsart and aspect) as well. Reminding Ross’ categorial space (see section 1.1), it is possible arguing that nominalizations are set in an intermediate position between the two edges represented by the canonical verbs and the canonical nouns. This is to say that
nominalizations express aspect in a less canonical way than the verbs do. Therefore, in the
effort of retrieving the aspect –crucial for turning a full referential noun into an event- the
subjects of the study found it easier to produce a verb than the required nominalization
given that the former bears aspect in a more “canonical” way. Ultimately, the subjects did
not replace a target nominalization with a verb as such, but with a verb as a word that can
express aspect and that was available for some reasons in their mind in order to fulfil the
requirements of the task. Further studies are needed in order to confirm these suppositions.

References
Amsterdam/Philadelphia: John Benjamin Publishing Co.
nominalization puzzle”. Language and Linguistics Compass 4: 496-511.
Alexiadou, A. 2010 b. “Nominalizations: a probe into the architecture of grammar. Part II: the
aspectual properties of nominalizations, and the lexicon vs. syntax debate”. Language and Linguistics Compass 4: 512-523.
nominals”. In The syntax of nominalizations across languages and frameworks, ed. by A. Alexiadou, and M. Rathert, 9-38. Berlin: Mouton De Gruyter.
Alexiadou, A., G. Iordachioaia, and F. Schäfer. 2011. “Scaling the variation in Romance
and Germanic nominalizations”. In The noun phrase in Romance and Germanic, ed. by P. Sleeman, and H. Peridon, 25-40. Amsterdam: John Benjamins.
system”. Cognitive Neuropsychology 8: 335-367.
Cambridge University Press.
Berndt, R., A. Haendiges, M. Burton, and C. Mitchum. 2002. “Grammatical class and
imageability in aphasic word production: Their effects are independent”. Journal of Neurolinguistics: 353-371.
Bertinetto, P. M., C. Burani, A. Laudanna, L. Marconi, D. Ratti, C. Rolando, and A. M.
http://linguistica.sns.it/CoLFIS/CoLFIS_home.htm
Grammatical category and semantic category deficits”. Brain and Language 72:
246-309.
The crucial role of the event structure in the retrieval of nominalizations in aphasia


The crucial role of the event structure in the retrieval of nominalizations in aphasia