Spontaneous and Elicited Production of European Portuguese Clefts

Ana Lúcia Santos, Maria Lobo, and Carla Soares-Jesel

1. Introduction

Clefts have not received much attention in the acquisition literature, even though they raise relevant questions: most clefts imply the projection of CP (in some cases, an embedded CP); in most cases they imply A’-movement; and the production of a cleft is subject to semantic-pragmatic constraints to the extent that clefted material is generally interpreted as focused. In particular, European Portuguese clefts may be of special interest for acquisition, given their diversity. In what follows, we show that European Portuguese displays a wide variety of cleft structures involving different degrees of syntactic complexity (defined according to a specific Complexity Metric) and we make explicit the predictions that the analysis of these different cleft types makes for the acquisition path.

Based on the analysis of spontaneous production (6 children 1;2-4;6, 37376 child utterances) and on an elicited production task eliciting subject, direct object, indirect object and adjunct clefts, applied to 22 adult, 14 3-year-olds (mean 3;6), 20 4-year-olds (mean 4;6), 17 5-year-olds (mean 5;6), we discuss the emergence of different types of clefts. Earlier emergence of é que clefts confirms both Soares (2006) prediction based on a scale of complexity (Derivational Complexity Metric) and an analysis of é que clefts as monoclausal structures. Moreover, the asymmetry found in clefts is not a simple subject/object asymmetry, it is instead a subject vs. (direct/indirect) object, adjunct asymmetry, a fact weakening an intervention explanation in Friedmann, Belletti & Rizzi’s (2009) terms. Finally, the data also support the analysis of certain types of fragments as partially elided clefts, both in adult and in child speech.

Authors: Ana Lúcia Santos, Universidade de Lisboa, als@fl.ul.pt; Maria Lobo, Universidade Nova de Lisboa, maria.lobo@fcsh.unl.pt ; Carla Soares-Jesel, Université Paris-Diderot, carla.soaresjesel@linguist.jussieu.fr.

We thank Mariana Silva, Stéphanie Vaz and Anaísa Gordino for help with data coding, and the audience at BUCLD 2012 for discussion of some of the questions raised by the current results. We thank Jardim Infantil Pestalozzi and Jardim de Infância Letrinhas, where the data were collected. All remaining errors are, of course, our own.
2. The syntax of (some) European Portuguese clefts

In standard European Portuguese (EP), we can find the following clefting patterns (Casteleiro, 1979; Ambar, 1999; Costa & Duarte, 2001, a.o.):

1. a. *A Branca de Neve é que o príncipe beijou.*
   - the White of Snow is that the prince kissed
   - ‘It was Snow White that the prince kissed.’

   b. *Foi a Branca de Neve que o príncipe beijou.*
   - was the White of Snow that the prince kissed
   - ‘It was Snow White that the prince kissed.’

   c. *Foi a Branca de Neve quem o príncipe beijou.*
   - was the White of Snow who the prince kissed
   - ‘It was Snow White who the prince kissed.’

   d. *Quem o príncipe beijou foi a Branca de Neve.*
   - who the prince kissed was the White of Snow
   - ‘Who the prince kissed was Snow White.’

   e. *A Branca de Neve foi quem o príncipe beijou.*
   - the White of Snow was who the prince kissed
   - ‘Snow White was who the prince kissed.’

   f. *O príncipe beijou foi a Branca de Neve.*
   - the prince kissed was the White of Snow
   - ‘It was Snow White who the prince kissed.’

In this paper we will focus on *é que* clefts and standard clefts only, which, as we will show, are the types both children and adults produce more frequently. There is some controversy concerning the analysis of the different types of clefts. One of the aspects about which there is disagreement concerns the status of *é que* clefts. Some authors treat these structures as complex sentences (Costa & Duarte 2001; Ambar 1999; 2005), whereas others claim that they are simple sentences, in which *é que* behaves as a lexicalized expression that fills the C position (Soares 2006; Lobo 2006 a.o.). Two empirical facts support the claim that *é que* in clefts is a lexicalized expression: in *é que* clefts the copula does not share phi-features with the clefted constituent nor does it share tense features with the matrix verb, it is invariable (see (2)); in *é que* clefts, nothing can intervene between *é* ‘is’ and the complementizer *que* ‘that’ (Ambar 1999; Costa & Duarte, 2001; Ambar, 2005; a.o.) – see the example in (3).

1 In non-standard varieties of European Portuguese, there are other clefting types (see Costa & Lobo, 2009; Vercauteren, 2010).
We will thus take as a working hypothesis that *é que* ‘is-that’ clefts (unlike standard clefts, wh-clefts and pseudoclefts) are simple clauses, in which *é que* is a lexicalized expression that occupies a functional category in the left periphery of the clause, in the C-domain (Lobo, 2006; Soares, 2006). The clefted constituent is in the specifier position of a higher functional category and was extracted from a position internal to the embedded IP.

Assuming this type of analysis for *é que* and for standard clefts allows specific predictions for acquisition, as noticed by Soares (2006). Soares (2006) claims that the emergence of clefts in EP children’s production is constrained by syntactic complexity. Her claim stems from Jakubowicz’s (2004, 2005, 2011) proposal that language acquisition is affected by developmental constraints such as the capacity of working memory, which are sensitive to the computational complexity of the derivation. The general hypothesis, the Derivational Complexity Hypothesis (henceforth DCH) is that less complex derivations are input convergent (i.e. correctly spelled out at the interfaces) before more complex ones. A Complexity Metric defines complexity:
(6) Derivational Complexity Metric (DCM) (Jakubowicz, 2011)
A. Merging $\alpha$ $n$ times gives rise to a less complex derivation than merging $\alpha$ $(n +1)$ times.
B. Internal Merge of $\alpha$ gives rise to a less complex derivation than Internal Merge of $\alpha + \beta$.

The characterization in (6) above was refined by Soares (2006), who claims that the reason why embedding gives rise to more complex derivations is the fact that it entails dependence relations and not because it involves merging more syntactic units in the structure (see (7)).

(7) The computation of a dependent head gives rise to a more complex derivation than the computation of a non-dependent head.

According to Soares (2006), the derivation of é que clefts, which are simple sentences, is less complex than the derivation of standard clefts, which involve embedding. In what follows, we discuss whether both spontaneous and elicited production confirm this prediction.

Apart from the set of structures already mentioned, there are also fragment structures that have been analysed as truncated or reduced clefts (Matos, 1992; Santos, 2004; Belletti, 2005, 2008; Soares, 2006; Santos, 2009b). These are discourse fragments (we will call them be-fragments) in which the verb be is followed by a constituent interpreted as focus (see 8). Santos (2009b) shows that these fragments display movement effects, differently from simple fragments (a simple fragment is presented in 8c). We thus expect to relate the acquisition of this type of fragments to the acquisition of (standard) clefts.

(8)   a. - O João chegou atrasado. / - Quem chegou atrasado? /
     the João came late. / who came late?
     - João came late.' /   ‘- Who came late?’
 b. (Nâo.) Foi o Rui que chegou atrasado. (Be fragment)
     no was the Rui that arrived late
     ‘(No.) It was Rui who arrived late.’
 c. O Rui.
     the Rui

3. Previous results on the acquisition of clefts

Studies on the acquisition of cleft structures are scarce. We report here the experimental results we took as background, as well as Soares’ (2006) analysis of early spontaneous production in European Portuguese. We also show that the analysis of Santos’ (2006) corpus allows us to confirm previous generalizations put forward by Soares.
3.1. Previous experimental results

One of the first studies that we are aware of is a comprehension task (act out) carried out by Lempert & Kinsbourne (1980). Their results show that English speaking children from 2:5 to 6:3 perform significantly better in the case of subject clefts than in the case of object clefts. Hupet & Tilmant (1989) report similar results for French and show a subject / verb asymmetry, this time in production. Their results show that French speaking children from 4 to 10 years produce significantly more subject clefts than object clefts. To highlight the object, children prefer to produce other structures in which the focus interpretation of the object is determined by contrastive stress.

This asymmetry is commonly found in other constructions involving A’-movement, such as relatives (Friedmann & Novogrodsky 2004, a.o). Friedmann, Belletti & Rizzi (2009) (following Grillo 2005, 2008) offer an explanation for this asymmetry in terms of intervention. The children aged 3;7-5;0 they tested find more difficulties in comprehension of object relatives that present an A’-moved objet that shares features with an intervening lexically filled subject – according to their proposal, the relevant feature is characterized as a lexical NP restriction. The authors consider that children, but not adults, follow a stricter version of Relativized Minimality.

3.2. Spontaneous production of European Portuguese clefts: previous and new results

Soares (2006) analysed a corpus including 18884 utterances produced by three children (1;2.0-2;2.17, 2;6.3-3;5.17 and 3;6.24-4;6.18). She shows that é que clefts emerge slightly before standard clefts, in agreement with the idea that é que clefts may be less complex than standard clefts. Trying to confirm Soares’ observations, we coded Santos’ (2006) corpus, composed by data from children in a comparable age range and of similar extension: three children (INI: 1;6.6 – 3;11.12, TOM: 1;6.18-2;9.7, INM: 1;5.9-2;7.24), 18492 utterances produced by the children (see Santos 2006, 2009a, for details concerning data collection and transcription). All the clefts or attempts to produce clefts were coded. Two of the three children in this corpus indeed show a developmental pattern comparable to what has been suggested by Soares: production of é que clefts precedes production of standard clefts – in the case of INI, an é que cleft is produced at 2;1, closely followed by a standard cleft at 2;3 (and only at 3;0 do we find an attempt to produce other types of cleft structures); in the case of INM, the child produces a few proto-clefts (non-target clefts with omitted complementizers) and only at 2;3 does the child produce a cleft structure with an overt complementizer, i. e., an é que cleft (no other clefts have been registered for this child). The case of TOM in the same corpus is, however, different: in TOM’s production, standard clefts emerge slightly before (at 2;1) é que clefts (which emerge at 2;3). TOM also produces another type of clefts, be-clefts, very early (at 2;4). TOM had a very fast linguistic development (see
Santos, 2006) and it is possible that the point of emergence of \( é que \) clefts was missed. However, if we take into account the majority of the children, spontaneous production suggests that \( é que \) clefts generally emerge before standard clefts, even though the delay between the two may be small. The idea that \( é que \) clefts emerge earlier is in agreement with the idea that the computation of an embedded clause is more complex than the computation of a root clause, if we maintain that \( é que \) clefts are mono-clausal structures.

The other relevant result coming from the analysis of the two corpora concerns the type of constituents clefted: in both corpora clefted subjects largely outnumber clefted objects or adjuncts (merging the data, it is possible to say that 148 clefted subjects were produced, against 27 clefted objects and 30 clefted adjuncts). Moreover, with the exception of TOM in Santos’ corpus, all the children produce subject clefts before object clefts. This may be seen as the result of well-known subject-object asymmetries (see section 3.1), but it might as well receive other types of explanation: children may prefer other strategies different from clefting to focus objects or adjuncts; the relevant context for the production of object or adjunct clefts might have occurred less often in the data. We were therefore left with the question that justifies the experiment that we report in the next section: would children produce other types of clefts if they found the relevant contexts?

The research reported in the next section tries to answer the following questions:

If provided with relevant contexts:
(i) Do children equally produce subject, object or adjunct clefts or do they show asymmetries in the production of the different cleft structures?
(ii) If such asymmetries are found, is the production of direct object clefts more difficult than the production of indirect object or adjunct clefts?
(iii) If children do not produce clefts in certain conditions, to what strategies do they resort when focusing the relevant constituents?
(iv) Is the production of \( é que \) clefts more frequent earlier than the production of standard clefts, as predicted by the complexity metric presented in (6) and (7)?

4. Elicited production of clefts

4.1. Methodology

To elicit the production of clefts, we used a task similar to a truth-value judgment task but with a different goal: a puppet describes images and has to be corrected. As in Hupet & Tilmant (1989), we have offered children an appropriate discourse context for cleft structures: the context creates in children the necessity to “contrast their own belief or knowledge with that of their addressee” (Hupet & Tilmant, 1989: 251). The test includes four conditions, defined according to the adjunct / argument status or type of argument contrasted: subject, direct object, indirect object and adjunct. The indirect object
condition included 4 test trials, all the others included 10 trials each. The test also included at least two training items and 10 distractors, i.e. cases in which the puppet’s utterance is a correct description of the picture. In figure 1 and example (9), we present a case of a test item built to elicit object clefts.

Figure 1 – Picture for test item in (9)

(9) Puppet: A mãe está a pentear a menina.
the mum is PREP comb the girl
‘Mum is combing the young girl.’

Four groups of speakers were tested: a control group of 22 adults with no background in linguistics (age ranging from 18 to 48), 14 3 year-olds (age range 3;2-3;11, mean 3;6), 20 4 year-olds (age range 4;0-4;11, mean 4;6) and 17 5/6 year-olds (age range 5;0-6;2, mean 5;6), all monolingual speakers of European Portuguese. Responses were both annotated during the sessions and audiotaped.

4.2. Results

The results obtained in the subject condition are presented in table 1. One of the main results obtained concerns the type of cleft produced: although all the groups produce both é que clefts and standard clefts, the production of standard clefts among 3 year-olds is scarce; the preference for é que clefts over standard clefts is maintained in all the groups, including the adult control group. However, clefts are not the most frequent type of answer obtained: in the case of adults, simple sentences, typically produced with prosodic stress on the subject, were the most frequent answer; as far as children are concerned, especially in the 3 and 4 year-old groups, fragments were the most frequent answer – in this case, only the focused material is produced, whereas given material is elided. It is worth noticing that be-fragments, the type of fragments that have been argued to be elided (standard) clefts, attain their highest frequency among 4 year-olds. When the child offered a fragment as an answer, the experimenter often insisted, trying to obtain a full sentence. These second answers are not counted in the
table, but it is interesting to notice that among 4 year-olds, 13 standard clefts were produced as a paraphrase of a be-fragment (see 10).

Table 1 – Responses obtained in the subject condition

<table>
<thead>
<tr>
<th></th>
<th>é que SU</th>
<th>Standard SU</th>
<th>Cleft 0that SU</th>
<th>Pseudo-cleft</th>
<th>Wh-cleft</th>
<th>BE fragment</th>
<th>Simple fragment</th>
<th>Simple sentence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>15,7</td>
<td>1,4</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>15,0</td>
<td>40,0</td>
<td>22,9</td>
<td>5,0</td>
</tr>
<tr>
<td>4</td>
<td>1,0</td>
<td>3,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>57,9</td>
<td>23,8</td>
<td>11,9</td>
<td>2,5</td>
</tr>
<tr>
<td>5</td>
<td>20,0</td>
<td>11,2</td>
<td>0,6</td>
<td>0,0</td>
<td>0,0</td>
<td>29,4</td>
<td>10,6</td>
<td>28,2</td>
<td>0,0</td>
</tr>
<tr>
<td>Adults</td>
<td>26,5</td>
<td>13,9</td>
<td>0,0</td>
<td>2,2</td>
<td>5,4</td>
<td>1,3</td>
<td>0,4</td>
<td>49,3</td>
<td>0,9</td>
</tr>
</tbody>
</table>

(10) Child: É o gato.
       is the cat
Frog: É o gato o quê?
       is the cat what
Child: É o gato que está a morder a bola.
       is the cat that is PREP bite the ball

The results obtained in the object condition are presented in table 2.

Table 2 – Responses obtained in the object condition

<table>
<thead>
<tr>
<th></th>
<th>é que O</th>
<th>Standard O</th>
<th>é que SU</th>
<th>Standard SU</th>
<th>Ambiguous (null SU)</th>
<th>Bare cleft</th>
<th>BE fragment</th>
<th>Simple fragment</th>
<th>Simple sentence</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0,0</td>
<td>0,0</td>
<td>1,4</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>57,9</td>
<td>34,3</td>
</tr>
<tr>
<td>4</td>
<td>0,0</td>
<td>1,5</td>
<td>0,0</td>
<td>0,0</td>
<td>1,5</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>27,0</td>
<td>44,5</td>
</tr>
<tr>
<td>5</td>
<td>0,0</td>
<td>0,0</td>
<td>1,8</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>0,0</td>
<td>4,7</td>
<td>30,0</td>
<td>62,9</td>
</tr>
<tr>
<td>Adults</td>
<td>0,0</td>
<td>0,0</td>
<td>0,5</td>
<td>0,0</td>
<td>0,5</td>
<td>0,0</td>
<td>0,0</td>
<td>0,5</td>
<td>97,3</td>
<td>0,9</td>
</tr>
</tbody>
</table>

2 Clefts coded as “Cleft 0that” are clefts with an omitted complementizer.
The most relevant result is the fact that there is almost no production of object clefts. Adult speakers and 5 year-olds prefer to produce simple sentences, in which focus is marked prosodically; 3 and 4 year-olds prefer to produce fragments, with be-fragments more frequently produced by 4 year-olds. As it sometimes happens when other structures involving A’-movement are elicited (Novogrodsky & Friedmann 2006; Costa, Lobo & Silva, 2011), child speakers sometimes produce a non-target subject cleft when the object is focused (see 11). Adults do produce subject clefts in this context, but these are passive structures and thus congruent with the context.

(11) Frog: O menino está a pintar a mãe.
        the boy is PREP paint the mum

Child: O menino é que (es)tá a pintar o bebé.
        the boy is that is PREP paint the baby

The results obtained in the indirect object and the adjunct conditions were very similar to those obtained in the direct object condition. For reasons of space, we cannot present them here in a complete form and we will thus only report the main results obtained (we refer to Lobo, Santos & Soares, in prep. for a detailed presentation of the results). Both in the indirect object and in the adjunct condition, there is almost no production of clefts: adults prefer to produce simple sentences (96.6% in the indirect object condition and 93.6% in the adjunct condition), whereas 3 and 4 year-olds produce higher rates of fragments. As in the direct object condition, be-fragments are a frequent type of answer only among 4 year-olds (23.8% in the indirect object condition, 23% in the adjunct condition). As in the direct object condition, we obtained non-target subject clefts in contexts where the indirect object or an adjunct is contrasted:

(12) Frog: O pai está a abrir a porta ao gato.
          the dad is PREP open the door to+the cat

Child: Não. É o pai que (es)tá a abrir a porta à mãe.
         no is the dad that is PREP open the door to+the mum

5. Discussion

The results presented in the preceding section contribute to our understanding of the acquisition of clefts as well as to the discussion concerning the syntactic structure of different cleft structures.

In the first place, the data obtained are in agreement with the predictions made by Soares (2006), based on a Complexity Metric defined as in (6) and (7) (section 2). As expected, é que clefts are more frequent and, especially, more frequent earlier than standard clefts in the elicited data (see the evolution of production of é que and standard clefts in the different age groups in the subject condition). Of course, the background given by the spontaneous production data makes it clear that we are not talking about emergence of standard clefts when
looking at the elicited production of 3 year-olds, since most children in Santos’ and Soares’ corpora produce both *é que* and standard clefts before 3:0. Nevertheless, at 3 years of age we may be not far from that point: one of the children in Santos’ corpus (INM) does not produce any target-like standard cleft until 2:7. Now it is relevant to stress the fact that the Derivational Complexity Hypothesis such as it was put forward by Jakubowicz’s (2004, 2005, 2011) is based on the idea that working memory is sensitive to computational complexity. This means that the visible effects of processing limitations in child production may be of different kinds in different developmental stages. To this extent, this hypothesis may explain not only the emergence of different structures but also the frequency of different structures at a given period: different frequency rates may also be an effect of computational complexity.

The other relevant conclusion that can be taken from the results concerns the syntactic analysis of European Portuguese cleft structures. Soares’ prediction can only be maintained if *é que* clefts are taken as mono-clausal structures whereas standard clefts are taken as bi-clausal. The earlier frequency of *é que* clefts may actually be taken as an argument favoring that analysis.

The second set of results obtained concerns clefting of different constituents. In general, in the elicited production task, children (as well as adults) produced subject clefts but not object clefts. Importantly, the difference between subjects and other constituents is even more radical than what we have obtained in spontaneous production. This fact may be interpreted in different ways. It may be that our experiment is not as appropriate to elicit other clefts as it is for subject clefts. But it may also be the case that the type of object clefts that we are eliciting here are more difficult for children than the type of object clefts that we obtained in spontaneous production. In our experimental setting, the speaker was forced to produce a cleft where a lexical object has crossed a lexical subject (i.e. both are full D+NP structure) (see 13). However, most object clefts that were spontaneously produced are cases in which either the subject or the object are pronouns (see 14) – we leave more detailed discussion of these data for Lobo, Santos & Soares (in prep.). This means that in (14) the intervelner and the crossing element are of different types in terms of lexical NP restriction and therefore no strong intervention effects are expected (see Friedmann, Belletti & Rizzi’s 2009 account of free object relatives and headed object relatives with a null subject pronoun).

\[13\] É o bebé que a mãe está a pentear.
*is the baby that the mother is PREP comb*

‘It is the baby that Mummy is combing.’

\[14\] SAN: É o pequenino que eu quero. 3:0.21
*is the small that I want*

‘It is the small one that I want.’

---

3 We thank Bonnie Schwartz for making this point.
Nevertheless, this discussion must include the fact that almost no clefts were produced not only in the direct object but also in the indirect object and the adjunct conditions. In this case, the present data may suggest that the asymmetry is not between subject and direct object but between subject and other constituents. The fact that indirect objects may generate intervention effects, in agreement with the idea that the preposition in an indirect object is not a true preposition but only a Case-marker, has already been pointed out by Abalada (2011), who tested comprehension of topicalized objects in European Portuguese. However, it is not clear how we should explain the absence of adjunct clefts.

Another alternative explanation has to do with the nature of clefts and their pragmatic function. Clefts are used to contrast information (and this was the situation created in the experiment), but they are not the only possible syntactic structure used to convey contrast. It is thus reasonable to think that in each case the speaker makes a choice concerning which strategy he uses to convey the relevant meaning. According to most views on information structure and focus, post-verbal material is more likely to be interpreted as default focus for prosodic reasons (see the NSR, Cinque 1993), so it is conceivable that speakers resort more to prosodic strategies when the contrasted material is post-verbal.

The last set of relevant results produced by this experiment concerns be-fragments, its nature and also its relevance as a structure among others available to convey contrast. First, increase of be-fragments among 4 year-olds coincides with the ability to paraphrase them as a standard cleft. This means that lack of production of standard clefts is not necessarily due to inability to produce them but possibly due to the fact that the speaker makes other choices, led by processing constraints. Therefore, the paraphrases may be taken by themselves as an argument for the analysis of be-clefts as partially elided standard clefts (Matos 1992, Soares 2006, Santos 2009b). At the same time, if children prefer to produce partially elided clefts rather than clefts this may be in agreement with the idea that the production of an elided structure, if possible in the target grammar, is probably easier for an immature processing system (Santos 2006, 2009a).

6. Conclusions

In this paper we have provided results regarding child production data showing that: (i) the order of emergence of clefts in spontaneous production is predicted by the DCH; also, derivational complexity may determine the frequency of different constructions in a given developmental period; (ii) a mono-clausal syntactic analysis of é que clefts should be favored; (iii) the preferred way of focusing post verbal constituents is prosodic marking; (iv) the production of elided clefts is most likely easier for an immature processing system.
References


Casteleiro, João Malaca (1979) Sintaxe e Semântica das Construções Enfáticas com “é que”. Boletim de Filologia, XXV.


Costa, João, Maria Lobo & Carolina Silva (2011) Subject-object asymmetries in the acquisition of Portuguese relative clauses: adults vs. children. Lingua 121.6; 987-1158


Jakubowicz, Celia (2005). The Language Faculty: (Ab)normal development and interface constraints. Paper presented to GALA, Siena, 8-10 Sept.
Lobo, Maria, Ana Lúcia Santos & Carla Soares-Jesel (in prep.) Syntactic structure and information structure: the acquisition of Portuguese clefts and fragment answers.
Santos, Ana Lúcia (2009b) The problem of fragment answers. Iberia 1.1; 115-142.