The Emergence of Full and Reduced Clefts in French L1
Karen Lahousse and Morgane Jourdain

1. Intro and background

This article is about the interaction between syntax and Information Structure (henceforth IS) in the emergence and development of cleft sentences in French first language acquisition (L1).

The linguistic literature on cleft sentences in adult French (1) is very rich; both the syntax and IS of these constructions have been extensively analyzed (see Destruel, 2012; Doetjes, Rebuschi, & Rialland, 2004; Karssenberg & Lahousse, 2018; Lahousse & Borremans, 2014; Lambrecht, 2001 for an overview).

(1) Qui a mangé le gâteau ? – C’est Jean qui l’a fait.
Who has eaten the cake – It-is John who it-has done
‘Who ate the cake? – It’s John who did it.’

However, almost nothing is known on the emergence of clefts in French L1, be it full clefts (2), reduced clefts (3) or cleft attempts (4).

(2) C’est toi qui m(e) fait les tortues.
It-is you who me does the turtles
‘It’s you who to me does the turtles.’
(Marie, 2;9, corpus of Lyon)

(3) C’est Maya.
It-is Maya
‘It’s Maya.’
(Héloïse, 2;10.5, corpus TCOF)

(4) Non ! C’est moi mets.
No It-is me put
‘No! It’s me who puts.’
(Anaïs, 2;5.25, corpus of Lyon)

* Karen Lahousse, KU Leuven (Belgium), karen.lahousse@kuleuven.be Morgane Jourdain, KU Leuven (Belgium) and University of Lille (France), morgane.jourdain@kuleuven.be.
1 For a description of the corpora we used, see section 2 below.

It has been observed that clefts in French L1 appear around age 2 (Belletti, 2005; Canut, 2014; De Cat, 2007; Labelle, 1990; Soares-Jesel & Lobo, 2019), which confirms research on the acquisition of clefts in other languages (Lobo, Santos, & Soares-Jesel, 2016; Pivi, Del Puppo, & Cardinaletti, 2016; Santos, 2006). On the basis of an elicited production experiment, Hupet and Tilmant (1989) show that French-speaking children from ages 4 and older correctly produce contrastive clefts, and that subject clefts are more frequent than object clefts. The results of Soares-Jesel and Lobo’s (2019) and Soares-Jesel, Lobo & Santos’ (in prep.) elicited production experiment – with children from age 3;3 and older – confirm that subject clefts are more frequent than object clefts. Moreover, these authors show that the production of full clefts increases significantly with age, whereas the proportion of reduced clefts, which are more frequent than full clefts at age 3;7, decreases as from age 4.

Since the existing experiments on the acquisition of clefts in French L1 are with children of ages 3 and more, the most important years to describe the emergence and early development of clefts are “missed”. Hence, there is almost no data (and no analysis) of the syntactic development and the discourse (i.e. IS) properties of clefts in early L1 of French (ages 1-3).

The main goals of this article are (i) to analyze the formal development of the main clause (c’est ‘it is’) and the cleft relative clause; (ii) to determine if clefts produced by children and by adults have the same IS and (iii) whether these IS-functions are present from the onset of language production or gradually emerge.

In what follows, we first present our methodology (section 2) and provide an overview of the syntactic development of clefts in our French L1 corpora (section 3). In sections 4 and 5 we use child data to test some specific hypotheses about the structural analysis of clefts. Section 6 analyzes the information structure (IS) of early clefts, and argues in favor of the hypothesis that IS is acquired before (rather than with or after) syntax.

2. Methodology
2.1. Intro

We conducted corpus analysis rather than an experiment, for the following reasons. First, in order to study the emergence of clefts, data is needed from the onset of language production, i.e. from very young children of ages 1 and 2. At this age, it may be quite difficult to elicit clefts via a controlled production experiment. Second, with respect to syntax, we want to describe the full developmental path of clefts, at different moments, from the onset of language production. This is harder to achieve by experiments, unless several follow-up experiments are planned. Thirdly, we consider corpus research as a heuristic: in corpus research, one often finds constructions (see section 5 below) and discourse uses of constructions (see section 6) which would probably not have been incorporated in the setup of an experiment.

Nevertheless, we are convinced that our analysis of spontaneously produced clefts by very young children can serve as input for further experimental research:
if it is determined in which precise contexts children produce clefts, these “natural” discourse conditions will be more easily integrated in an experiment.

2.2. Corpora

Given that our goal is to study the emergence of clefts, and that children as from age 4 produce adult-like clefts, we restricted our analysis to children no older than 3. Our analysis is based on data from two corpora of spontaneous speech production: (i) part of the cross-sectional TCOF corpus (subcorpus of ATILF 2018, www.ortolang.fr/market/corpora/tcof, see (André & Canut, 2010)) and (ii) the longitudinal Lyon corpus (Demuth & Tremblay, 2008), available on the CHILDES database (MacWhinney & Snow, 1990), the details of which can be found in Table 1.

<table>
<thead>
<tr>
<th>Table 1: Description of the corpora</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type of file</strong></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>TCOF (2005-2018)</td>
</tr>
<tr>
<td>Lyon (2002-2005)</td>
</tr>
</tbody>
</table>

2.3. Data extraction and selection

We first extracted all sentences containing *c’est* ‘it is’ from the two corpora. We then made a manual section excluding all examples with referential *c’, i.e. which refers to an entity given in the discourse context (5), e.g. when the child is using *c’est* clauses to describe something that is physically present in the context, including examples in which *c’* resumes a dislocated constituent (6). In a second step, we removed all examples in which the relative clause is clearly restrictive, such as (7). Both cases, i.e. *c’est-*clauses with referential *c’* and *c’est-*clauses with a restrictive relative clause, cannot be considered as clefts (see Karssenberg, 2018 for an overview of ways to distinguish clefts from cleft-lookalikes).

(5) Father: *C’est qui là ?*  
  It-is who there  
  ‘Who is it there?’  

Child: *C’est Amtaro.*  
  It-is Amtaro  
  ‘It’s Amtaro.’

(Anaïs, 2;9.29, corpus of Lyon)
(6) [Context: the child is giving a toy to the adult]

Ça c’est pour toi.
That it is for you
‘That’s for you.’

(Anaïs, 2;8, corpus Lyon)

(7) Adult: C’est quoi ?
It is what
‘What’s that?’

Child: C’est un bouchon qui est très énervé.
It is a cork which is very angry
‘It’s a cork which is very angry.’

(Marie, 3;2.8, corpus of Lyon)

Table 2 gives an overview of our final dataset. This is quite a large dataset for spontaneous speech production by very young children (ages 1 – 3).

Table 2: overview of the data

<table>
<thead>
<tr>
<th></th>
<th>Reduced clefts</th>
<th>Cleft attempts</th>
<th>Full clefts</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCOF</td>
<td>22</td>
<td>12</td>
<td>9</td>
<td>43</td>
</tr>
<tr>
<td>Lyon</td>
<td>120</td>
<td>73</td>
<td>59</td>
<td>252</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>295</td>
</tr>
</tbody>
</table>

3. Development of syntax: general view

An analysis of the syntactic form of the clefts in our dataset reveals the following developmental path of the syntax of clefts, with only some months between the 1st occurrence of each type:

I. The first attestations of REFERENTIAL C’EST X CLAUSES (with referential c’)
(5-6) appear between age 1;8 and 1;10 in our corpora.

II. Adult-like REDUCED CLEFTS consisting of non-referential c’+est+X (with cleft prosody) (8) are first attested in our corpora between age 2;0 and 2;53.

(8) Adult: Et après on va remettre par-dessus.
And then we go to-put-back above
‘And then, we will put it on top of it.’

Child: Non, c’est Nathan !
No it is Nathan
‘No, it’s Nathan!’

(Nathan, 2;5.1, corpus of Lyon)

2 We follow Hamann and Tuller (2014, p. 52) in taking the “order of acquisition in spontaneous language production to be manifestations of the effect of syntactic complexity”.

3 Soares-Jesel and Lobo (2019) even mention that reduced clefts are already produced by children at age 1;10.
III. Non adult-like CLEFT ATTEMPTS show up between ages 2;0 and 2;7\(^4\). These combine *c’est* \(X\) and an isolated word (NP, infinitive, adjective, participle) (9), and later a VP with inflected verb (sometimes followed by a complement) (10)\(^5\). As far as we can tell, the production of cleft attempts in child language has only been reported for Portuguese L1 by Lobo et al. (2016).

(9) *C’est Marie avoir.*
   It-is Marie to-have
   ‘It’s Marie have.’
   (Marie, 2;0.28, corpus of Lyon)

(10) Child: *C’est moi fais.*
   It-is me do
   ‘It’s me who does it.’
   (Marie, 2;5.1, corpus of Lyon)

IV. Clefts of the form *C’EST X + JUXTAPOSED SENTENCE* (with cleft prosody but without complementizer), such as (11), appear between ages 2;5 and 2;9. In these cases, the clefted element can (11a) but does not have to be (11b) coreferential with the subject of the juxtaposed sentence. As far as we can tell, the existence of this type of clefts in child language has not been reported before. Interestingly, native speakers moreover confirm that this type of clefts is adult-like in very informal French\(^6\).

(11) a. *C’est Tigrou il est pas content.*
   It-is Tiger he is not happy
   ‘It’s Tiger he is not happy.’
   (Nathan, 2;9.7, corpus of Lyon)

b. *C’est ça on montre.*
   It-is that we show
   ‘It’s that we show.’
   (Marie, 2;5.16, corpus of Lyon)

---

\(^4\) Note that all examples mentioned here come from the Lyon corpus. The children of the TCOF corpus are older and produce only one example with a VP. All the other examples are instances of adult-like (full or reduced) clefts.

\(^5\) Note that these attempts of cleft relative clauses are inherently predicative, which confirms that cleft relative clauses are pseudo-relative clauses (see Casalicchio, 2016 for an overview). For reasons of space, we cannot go deeper into this issue.

\(^6\) In the literature on *c’est* clefts in adult French, we did not find any mention of such complementizerless clefts introduced by *c’est*. However, the existence of such clefts introduced by *il y a* ‘there is’ has been reported by Willems and Meullemans (2010):

(i) *Il y a des gens ils viennent acheter de l’aspirine pour faire de l’eau gazeuse* ‘There are people they come buy aspirin to make sparkly water.’
V. Adult-like FULL CLEFTS with a complete cleft relative clause (c’est X qui/que ‘it’s X who/that’ + clause) appear between ages 2;6 and 2;9 (12).

(12) Adult: Oui c’est un petit peu compliqué mais qu’est-ce qu’on peut faire d’autre ?
Yes it-is a little bit complicated but what-we can to-do else
‘Yes it’s a bit complicated but what else can we do?’

Child: C’est toi qui m(e) fait les tortues.
It-is you who me does the turtles
‘It’s you who does the turtles for me.’

(Marie, 2;9.1, corpus of Lyon)

In this section, we have analyzed, on the basis of corpus analysis, the acquisition trajectory of different types of clefts in French L1 acquisition. In the next two sections we will provide some more specific data and ‘accidental’ findings in our corpora. These will allow us to zoom in on two existing claims and hypotheses on adult clefts, concerning the acquisition and syntactic analysis of reduced clefts vs full clefts (section 4), and the structural position of the clefted element (section 5). The background question here is: given that children master clefts so early, do child data offer extra arguments for specific claims on the syntactic analysis of (adult) clefts?

4. The acquisition and syntactic analysis of reduced vs full clefts

The following graphs show that, in the Lyon corpus, at age 2, reduced clefts (II) are first much more frequent than other types of clefts (III, IV, V), but that the difference between the frequency of both types of clefts reduces between age 2 and age 3 (the child Marie, who is in general ahead of the other children in the corpora, is an exception).

![Figure 1: Frequencies of clefts and reduced clefts in the corpus of Lyon](image-url)
In the TCOF corpus, the proportion of reduced and other types of clefts inverses between ages 2 and 3:

![Figure 2: Frequencies of clefts and reduced clefts in the corpus TCOF](image)

There are two theoretically possible ways to explain how children acquire full and reduced clefts:

**OPTION 1:** Children either acquire full clefts (type V: *c’est X qui/que* ‘it’s X who/that’ + clause) from the onset of the acquisition. The production of reduced clefts (type II) is then a by-product of complexity and/or processing.

**OPTION 2:** Children acquire clefts in a compositional way, by combining a non-referential *c’est* main clause (i.e. a reduced cleft, type II) and a relative clause.

Option 2 predicts that relative clauses (outside clefts) in L1 appear before clefts with a full CRC. This prediction is however not borne out in our corpus data: in the Lyon corpus, the first relative clause outside the cleft construction appears at age 2;8 (13a). This is slightly later than the first cleft relative clause, which is produced at 2;6 (13b).

(13) a. Mais mais à la fille *qui fait bubu.*
But but to the girl who does bubu
‘But but to the girl who does bubu.’
(Marie, 2;8.14, corpus of Lyon)

b. *C’est moi qu’est fatiguée toute seule.*
It-is me who-is tired all alone
‘It’s me who is tired all alone.’
(Marie, 2;6.2, corpus of Lyon)

This is independently confirmed by previous literature according to which all first embedded finite clauses appear inside the cleft construction (De Cat, 2002 and Labelle 1990 on French; see also Diessel, 2004 and Diessel & Tomasello, 2005 on English).
Hence, these data argue in favor of option 1: children immediately acquire full clefts but are not always capable of producing the cleft relative clause (CRC). For reasons of processing (by hypothesis), the CRC is either (i) omitted, giving rise to reduced clefts (as is argued for by Belletti, 2013 for adult clefts) or (ii) partially produced, in cleft attempts.

5. The ‘high’ vs ‘low’ structural position of the clefted element

With respect to the structural analysis of clefts, two cartographic analyses have been proposed (see Haegeman, Meinunger, & Vercauteren, 2013 for a detailed overview of both). In the ‘high’ or ‘monoclusal’ analysis (Frascarelli & Ramaglia, 2013; Kiss, 1998; Meinunger, 1998; Sleeman, 2011), the clefted element (CE) is in the FocP position in the clausal left periphery, which is also supposed to host preposed foci and the fronted wh-phrase in root questions (Benincà, 1988, p. 144; Rizzi, 1997, 2001):

\[
(14) [\text{Ground/TopP it is } [\text{FocP the dog } [\text{TopP that Mary saw } [\text{TP . . . }]]] ]
\]

In the ‘low’ analysis proposed by Belletti (2013), the CE is in the vP periphery or the periphery of the CRC, depending on its interpretation:

(15) a. CE = narrow contrastive focus (focus-background cleft)

\[
[\text{vP be } \text{CP/FocPcorr/contr Gianni } \text{FocP che } [\text{TP t Gianni ha parlato}]]
\]

b. CE = narrow new information focus (focus-background cleft)

\[
[\text{FocP/NewInfo Gianni } \text{vP be } \text{CP } \text{FocP che } [\text{TP t Gianni ha parlato}]]
\]

c. CE = discourse-given (given-new clefts)

\[
[\text{TopP Gianni } \text{vP be } \text{CP } \text{FocP che } [\text{TP t Gianni ha parlato}]]
\]

(adapted from Belletti, 2013, details omitted)

Haegeman et al. (2013) provide arguments against the high analysis. According to the authors, the grammaticality of examples in which a CE has undergone further wh-movement (16) or focus-movement (17) is unexpected (i) if the clausal left periphery contains only one focal position (Rizzi, 2001) and (ii) if SpecFocP in the clausal left periphery (in the high analysis in (14)) hosts CEs, preposed foci and the wh-moved phrase in root questions (cf. supra).

---

7 See Reeve (2012) for an alternative analysis. For reasons of space, we do not present other syntactic analyses of clefts than the cartographic analysis.
(16) a. *What* was it ___ that you saw?
   b. *Who* was it _ that you were going to invite? (Haegeman et al., 2013, p. 13)

(17) a. *Was it Sue* who polished off the cookies?  
   No, *Pat* it was _ who ate them. (Ward, Birner, & Huddleston, 2002, p. 1420)
   
   b. What was it ___ that you saw?  
   *John* it was ___ that Mary saw. (Reeve, 2011: (94a))
   (Haegeman et al., 2013, p. 16)

Our corpora also contain clefts in which the CE is *wh*-moved:

(18) *Qui c’est ___ a mis scotch ?*  
Who it-is ___ has put scotch  
‘Who is it that put scotch?’  
(Marie, 2;2.17, corpus of Lyon)

The examples such as (18) are incompatible with a view in which the focus position in the clausal left periphery hosts CEs and *wh*-words in roots, and the clausal left periphery contains only one focus position. Hence, they are arguments against a high structural analysis for clefts such as (14).

6. The “order” of acquisition of syntax & IS
6.1. Three competing hypotheses

Our corpus data also sheds new light on three competing hypotheses with respect to the acquisition order of syntax and information structure (IS).

**HYPO 1: SYNTAX PRECEDES IS:** “Children may first acquire the syntax of cleft sentences, initially not being fully sensitive to the pragmatic conditions on their use” (Aravind et al 2018:284). This hypothesis predicts occurrences in child speech of syntactically adult-like clefts which do not have an adult-like IS.

**HYPO 2: SYNTAX & IS TOGETHER:** syntax and IS are acquired simultaneously. The prediction from this hypothesis is that children produce clefts with an adult-like IS immediately when adult-like syntax of the construction is acquired.

**HYPO 3: IS PRECEDES SYNTAX:** this hypothesis is in line with the cognitive hypothesis put forward by Lambrecht (1994) and psycholinguists such as Levelt (1989), according to whom IS is fundamental in language production and prior to (rather than simultaneous with) purely linguistic processes concerning form and meaning. If IS is at the basis of syntax and prior to linguistic processes, rather than being computed together with or after linguistic processing, then children should “show sensitivity to pragmatic aspects of clefts before they can build an adult-like syntax for these sentences” (Aravind, Hackl, & Wexler, 2018, p. 284).

---

8 Note that focus-preposing is very rare in French and limited to very specific registers and contexts (Lahousse, 2015).
Hence, this hypothesis predicts that children, as soon as they start to produce clefts do so with an adult-like IS. We believe that cleft attempts may be an ideal test case, to the extent that they have non-adult-like syntax.

6.2. IS-articulations of early clefts

C’est clefts in adult French can have a wide range of IS-articulations (for an overview, see Dufter, 2009; Karssenberg & Lahousse, 2018; Lahousse & Borremans, 2014). They prototypically have a (i) focus-background IS-articulation in which the CE is a (contrastive, corrective or a new information) focus, or (ii) topic-comment articulation9, in which the CE is an aboutness-topic (“what the sentence is about”, Reinhart, 1981) or a stage topic (Chafe, 1976; Erteschik-Shir, 1997). They can also (iii) display a double contrast and (iv) in very specific registers and discourse contexts, even an all-focus interpretation.

In our corpora, all children’s early c’est clefts display the same discourse interpretations as adult clefts, in all syntactic types of CRC. For instance, in the reduced clefts in (19-21) below, the CE is a new information focus (19), a contrastive focus (20) or a corrective focus (21).

(19) Adult: Oui et qui est-ce qui te l’a offert ?
   Yes and who you it-has offered
   ‘Who offered it to you?’
   Child: Euh c’est Maya.
   Erm it-is Maya
   ‘It’s Maya.’
   (Héloïse, 2;10.5, corpus TCOF)

(20) Adult: C’est qui qui fait une maison ? C’est moi ou c’est toi ?
   It-is who who makes a house it-is me or it-is you
   ‘It’s who who makes a house? It’s you or it’s me?’
   Child: C’est toi.
   It-is you
   ‘It’s you.’
   (Mélanie, 3;9.19, corpus TCOF)

(21) Adult: C’est toi qui l’as acheté ?
   It-is you who it-has bought
   ‘It’s you who bought it?’
   Child: Non, c’est maman.
   No it-is mommy
   ‘No, it’s mommy.’
   (Clara, 3;1.5, corpus TCOF)

9 In our corpora, such clefts show up at age 4. This relatively late appearance could however be due to the fact that such clefts are typical of narrative contexts.
The CE in clefts attempts in our data may also be a corrective focus or a new information focus (22):10

(22) Father: Ah oui je l’ai abîmé.
Ah yes I it-have damaged
‘Ah yes, I damaged it.’
Child: Ici. C’est toi a cassé ?
Here it-is you has broken
‘Here. It’s you who broke?’

(Anaïs, 2;9.16, corpus of Lyon)

The next example is a cleft with a juxtaposed clause, which has a doubly contrastive interpretation:

(23) Adult: tu veux les deux ?
You want the two
‘you want both of them?’
Child: ouais
‘yes’
Adult : et puis Geoffrey on lui prend quoi ?
and then Geoffrey we for-him take what
Child: c’est Jéré- c’est Jérémy il veut avoir celle-ci
It-is Jéré- it-is Jérémy he wants to-have this-one
‘It’s Jéré- it’s Jérémy who wants to have this one’

(Jordan, 2;8.25, corpus TCOF)

Hence, in all the examples of early clefts in our corpus the construction has an adult-like IS-articulation. This shows that IS is present at the onset of language production, which goes against HYPOTHESIS 1 (“syntax precedes IS”). Furthermore, the fact that cleft attempts (22-23), which do not have adult-like syntax, do have adult-like IS, argues in favor of HYPOTHESIS 3 (“IS precedes syntax”) rather than HYPOTHESIS 2 (“syntax and IS together”).

7. Conclusion

The main aims of this article were (i) to analyze the formal development of the main clause and the cleft relative clause; (ii) to determine if clefts produced by children and by adults have the same IS and (iii) if these IS-functions gradually emerge and develop. The main findings are the following.

The developmental path of the syntax of clefts is (with only some months between the 1st occurrence of each type): I. c’est X clauses with referential c ’; II. reduced clefts of the type c’est X; III. cleft attempts: c’est X + an isolated word

---

10 In this example, the function of the interrogative cleft is to reconfirm the previous statement. Hence, the CE can be considered new information focus.
(infinitive, adjective, participle) and later a VP (with inflected verb, without complementizer) (see Lobo et al. 2016 on proto-clefts in Portuguese L1); IV. clefts of the form c’est X + juxtaposed sentence (without complementizer); V. full clefts with a complete cleft relative clause (c’est X qui/que + clause).

At age 2, reduced clefts are the most frequent, but their proportion decreases between age 2 and 4. It is inside the cleft construction that the first relative clauses are produced in the corpus, which confirms that relative clauses emerge in contexts with a “light” main verb (Diessel & Tomasello 2005). The production of reduced clefts, cleft attempts, full clefts and relative clauses in French L1 acquisition argues in favor of an analysis of reduced clefts in child L1 as full clefts with an elided cleft relative clause (see Belletti 2013 on adult clefts).

Young children perform complex syntactic operations (wh-movement) on the clefted element. This indicates that the clefted element is not in the left periphery, but in a low, clause-internal position (see Haegeman et al. 2013; Belletti 2013).

With respect to Information Structure, we have shown that all discourse interpretations of adult c’est-clefts (corrective / contrastive / new information focus-background clefts, doubly contrastive clefts) occur in our French L1 corpora (ages 1.5 – 3), in all syntactic types of clefts. The fact that children seem to have access to all discourse features of adult clefts before full (adult-like) syntax of clefts is acquired, indicates that syntax does not “precede” IS in L1. The fact that syntactically non-adult-like cleft attempts have adult-like IS, suggests that IS comes “before” syntax, which confirms cognitive hypotheses put forward by Lambrecht (1994) and psycholinguists such as Levelt (1989).

References


Casalicchio, Jan. (2016). Pseudo-relatives and their left-periphery. In Ernestina Carrilho, Alexandra Fiéis, Maria Lobo, & Sandra Pereira (Eds.), Romance Languages and


Pivi, Margherita, Del Puppo, Giorgia, & Cardinaletti, Anna. (2016). The elicited oral production of Italian restrictive relative clauses and cleft sentences in typically developing children and children with developmental dyslexia. In Pedro Guijarro Fuentes, Maria Juan-Garau, & Pilar Larrañaga (Eds.), Acquisition of Romance Languages (pp. 231-262). Berlin: De Gruyter.


Sleeman, Petra (2011). [Quantifier-focalization in French and Italian].


