The Acquisition of French Ambiguous Embedded Structures
Introduced by ce que

Ingrid Konrad, Yair Haendler, and Caterina Donati

1. Introduction

French exhibits an interesting ambiguity that is often neglected in the literature and requires a syntactic explanation. A fragment like (1) can be interpreted in two ways: either as a relative clause (RC) as in (2), or as an indirect question (IQ) as in (3).

(1) [DP ce [CP que ta sœur a cuisiné]]
   this that your sister has cooked
   ‘What your sister cooked.’

(2) Je mangerai [DP ce [CP que ta sœur a cuisiné]]
    I eat-FUT.1SG this that your sister has cooked
    ‘I will eat what your sister cooked.’

(3) Je me demande [ce [CP que ta sœur a cuisiné]]
    I me ask this that your sister has cooked
    ‘I wonder what your sister cooked.’

In (1), we can see that the complementizer que, which introduces a subordinate clause containing a gap, is preceded by the weak demonstrative ce. This structure is perfectly compatible with the analysis in (2), where the subordinate CP is embedded under a DP headed by the demonstrative itself. This is a light headed relative clause (Citko, 2004). As in the case of full headed relative clauses, the complementizer alternates between qui and que for subject and object gaps, respectively (Kayne, 1976). In (2), the gap is in object position, and this explains why que is required. However, the fact that the same sequence ce que can also be used as an indirect question like (3) is somehow unexpected: ce is not a wh-element, at least prima facie. Moreover, verba interrogandi (like Je me demande ‘I wonder’) select for a clause, not for a nominal argument.

*Ingrid Konrad, Yair Haendler, Caterina Donati, Université Paris Diderot, Laboratoire de Linguistique Formelle. Correspondence: konrad.ingrid4@gmail.com. We thank all the children who participated and their parents. We are also grateful to Carla Soares-Jesel for her help. This work is part of the program ‘Investissements d’Avenir’ overseen by the French National Research Agency, ANR-10-LABX-0083 (Labex EFL).

One possible explanation for the ambiguity of *ce que* is that this sequence has been reanalyzed as a wh-element, resulting in what we could transcribe as *səkə* (based on the pronunciation). The ambiguity between RCs and IQs would be no more surprising given this hypothesis, since bare wh-structures are systematically ambiguous along these lines. This is the case, for example, of sentences introduced by *what* in English: the same sequence of words, that of a wh-word introducing a clause (4), can be interpreted either as a relative clause (5) or as an interrogative clause (6).

(4) [what your brother bought]

(5) I like [DP what your brother bought].

(6) I wonder [CP what your brother bought].

This ambiguity of wh-structures can be captured by the Labeling Theory (Cecchetto & Donati, 2015). In (4), either the determiner *what* labels the structure, which is then a DP (5), or the hosting clause provides the label, which is thus a CP (6). This ambiguity holds also in French, for example with structures introduced by the wh-element *qui*.

(7) [qui tu voudras inviter]
   ‘who you want invite.INF
   ‘who you want to invite’

(8) J’inviterai [DP qui tu voudras inviter].
   ‘I will invite who you want to invite.’

(9) Je me demande [CP qui tu voudras inviter].
   ‘I wonder who you want to invite.’

Going back to French and to the reanalysis hypothesis, we know that a D+C incorporation is attested in a number of Romance varieties (Munaro, 2001).

(10) Sen-che t’ a-t feit ?
    ‘What have you done?’

(11) Chel-ca fiv adess ?
    ‘What are you doing now?’

(10) is an example of a root question introduced by *sen-che* in the Southern Valdotain variety spoken in Châtillon; (11) is an example of a root question...
introduced by *chel-ca* in the Northern Lombard variety spoken in Albosaggia in Valtellina. These complex elements are originally composed by a demonstrative (*sen* or *chel*) and the complementizer (*che* or *ca*) and have been reanalyzed as new wh-elements, which can indeed introduce root questions. The same is probably true in European Portuguese, where *o que* can introduce a root question as in (12).

(12) O que queres fazer?
   The that want.2SG do.INF
   ‘What do you want to do?’

The presence of these grammaticalization processes make the *səkə* hypothesis plausible for French.

In this paper, we will try to check this hypothesis against acquisition data, trying to capture how French-speaking children acquire this double usage of the sequence *ce+que* and in particular whether *ce que* is acquired in a parallel way in the two structures (RCs and IQs) or not. Although numerous studies have looked at the acquisition of French wh-questions and relative clauses (Fragman, 1998; Hamann, 2006; Bentea, Durrleman & Rizzi, 2016, among others), none of them has looked at the acquisition of the double usage of *ce que* in French. For instance, Bentea et al., 2016 only looked at the comprehension of object relative clauses introduced by *ce*, comparing them to object relatives introduced by a demonstrative like *celle* ‘this/that.FEM’. To our knowledge, this is the first experimental study that explores how children deal with the double usage of *ce que*, with the goal of determining whether this sequence is analyzed as a new wh-element (*səkə*), or whether it retains a complex nominal structure. Moreover, we will explore what kind of mistakes French-speaking children make when producing these structures, in case they have difficulties with the *ce que* sequence.

To this end, we set up an elicited imitation task, which arguably involves not only memory capacity, but also grammatical knowledge and comprehension of the child (Lust, Flynn & Foley, 1996). Specifically, the child should not be able to reconstruct the sentence she heard, in order to repeat it, if its structure is not part of her grammatical knowledge at that specific stage of acquisition. In our imitation task, children were asked to repeat sentences containing the *ce+que* sequence that was either a relative clause or an indirect question.

1.1. Predictions

If French children perceive *ce+que* as a wh-element *səkə* at the edge of a subordinate clause, we have two possible expectations. As we have seen with the English example in (4-6), these structures are systematically ambiguous in adults’ grammar. If wh-structures are equally ambiguous in children’s grammar, we expect them to behave the same in the two conditions, and display the same accuracy in repeating relative clauses and indirect questions. If, on the other hand, children do not perceive this ambiguity, we expect them to be more accurate on indirect questions, which are directly compatible with the interrogative status of the wh-element. But if they perceive the sequence as involving a determiner
followed by a sentence, we expect better performance on relative clauses (where a nominal structure is selected) than on indirect questions (where a nominal structure should be at odds). We expect substitution errors to happen more frequently in indirect questions (where a wh-element is required) than in relative clauses (where it is not necessarily).

2. An experimental study of ce que clauses in French

2.1. Participants

We tested 52 monolingual French-speaking children, all typically developing, with an age range of 3;0 to 6;0. All participants were tested at the preschool Ecole Maternelle de Nogaro in the French department of Gers. Parental consent has been obtained prior to testing the children. Three children were eventually excluded from the analysis. For two of them, the reason was they had difficulties with fillers too, which were very simple sentences. The third child was excluded because of lack of concentration on the task.

2.2. Material

The experiment consisted of a two-by-two design, yielding 4 experimental conditions, which was obtained by crossing the factor SENTENCE TYPE (relative clauses or indirect questions) and the factor GAP TYPE (subject or object gap). For each condition we constructed 4 items, resulting in a total of 16 experimental items, and 10 fillers. There were also 3 practice trials at the beginning of the experiment, to make sure that children, especially the youngest ones, understood the task.

The items were structured using a vocabulary that refers to the environment of the child (for instance, at school or at home), in order to make sure that children were familiar with the words and events described in the sentences. Moreover, we controlled the length of the items in terms of the number of syllables (each item contained 12 syllables), in order to neutralize potential memory-related effects.

The structure of each item was the following: DP (main subject) – V (main verb) – ce que / ce qui – (DP) (subject of the subordinate clause) – V (verb of the subordinate clause). Some items ended with a modifier after the last verb. The items were randomized, creating as many lists as the number of participants. This way, in each testing session the participating child was exposed to the same items but in a different order.

The following are examples of the four experimental conditions: a subject-gap relative clause (13); a subject-gap indirect question (14); an object-gap relative clause (15); and an object-gap indirect question (16).

(13) Le papa mange ce qui est tombé du placard.
The dad eats this that has fallen from-the cupboard
‘Dad eats what has fallen from the cupboard.’
Papa sait pas ce qui est tombé du placard.
‘Dad doesn’t know what has fallen from the cupboard.’

L’élève attrape ce que la maîtresse a lancé.
‘The pupil catches what the teacher has thrown.’

L’élève demande ce que la maîtresse a lancé.
‘The pupil asks what the teacher has thrown.’

Note that the subordinate clauses are identical across conditions; what changes is the main verb and its selection requirements. A verb like manger ‘eat’ requires a DP and is thus followed by a relative clause. But a verb like demander ‘ask’ selects a CP, so in this case it is followed by an indirect question. Fillers were short, simple sentences, without a subordinate clause (17-18).

La fille adore la peinture.
‘The girl likes painting.’

Marie est amoureuse de Lucas.
‘Marie is in love with Lucas.’

Mamie pense que papi a trop mangé de soupe.
‘Grandma thinks that grandpa ate too much soup.’

2.3. Procedure

Children were tested individually by the first author. The task was to listen to a sentence uttered by the experimenter and repeating it. A background story was created to make the game more enjoyable. Participants were introduced to a puppet who likes to listen to stories, but to those told by children rather than adults. This way, children were encouraged to interact with the puppet and to
repeat the sentence for it. For each trial, whether correctly repeated or not, the child was allowed to pick a construction block out of a bag and build something (e.g., a tower). Children’s answers were recorded. The task lasted about 15 minutes per child, without a break. Generally, children were involved and found the task fun.

3. Analysis and results

We analyzed the data with generalized linear mixed-effects models with a logit link function (Baayen et al., 2008; Jaeger, 2008), using the `lme4` package (Bates et al., 2015) in R (R Core Team, 2017). The response sentence the participants produced in each trial was coded as a binary dependent variable, assigned 1 when it was a correct repetition or 0 otherwise. The factor SENTENCE TYPE was coded as 0.5 (Relative Clause) and -0.5 (Indirect Question), and similarly the factor GAP TYPE was coded as 0.5 (Object Gap) and -0.5 (Subject Gap). The covariate AGE was inserted into the model as a centered continuous variable. We thus included all the participants in a single group and checked the effect of age on their performance. In the fixed-effects part, we estimated the main effects and all possible interactions of SENTENCE TYPE, GAP TYPE and AGE. In the random-effects part, we included only an intercept for subjects and one for items, in order to avoid overparametrization and, as a consequence, model results which might be uninterpretable (Bates et al., submitted; Matuschek et al., 2017).

Children’s proportion of correct repetitions across conditions and in relation to age is shown in Figure 1. As can be seen, there is an overall increase in correct repetitions as children get older, a pattern that is reflected in the significant main effect of AGE (coef=2.46, SE=0.59, z=4.12, p<.001). There was also a significant main effect of SENTENCE TYPE, whereby children had more correct repetitions of object relatives than of indirect questions (coef=0.86, SE=0.32, z=2.74, p=0.006). There was no difference between children’s performance on subject and object gap sentences, as evinced by the non-significant effect of GAP TYPE (coef=0.15, SE=0.31, z=0.47, p=0.63). Moreover, none of the interactions in the model was significant (all p-values > 0.05).
3. Error analysis

Next, we looked at the trials in which children answered incorrectly, in order to assess the various types of errors. Overall, children were incorrect on 54% of the trials, producing a wide range of errors that we classified into various categories. Most of the errors concerned the production of the *ce que* sequence. It is important to point out that children were highly accurate on the syntactically complex practice trials. Thus, it is not likely that their errors are due to weak cognitive skills like memory or attention; rather, they are likely to derive from difficulties with the *ce que* sequence.

The most prominent type of error (31.4% of the trials) was the omission of the determiner *ce* (see example 20 and Figure 2). Such sentences, which are ungrammatical in French, were produced more in indirect question trials than in relative clause ones, although this effect was weak (coef=-0.65, SE=0.39, z=-1.67, p=0.096). There was no difference in the omission of the determiner between subject and object gap trials or in relation to age (all p-values > 0.05). The second frequent type of error (9.7% of the trials) was the use of a complex wh-form *qu’est-ce que*, which we will write *kesk*, based on its pronunciation (see example 21 and Figure 3). This wh-form is used in French root questions (22), but not in sentences like the ones we tested (at least in standard French; in colloquial French,
this wh-form is frequently used in subordinate clauses too). Children produced such sentences more in indirect questions than in relative clauses, but the effect did not reach statistical significance (coef=-0.64, SE=0.35, z=-1.83, p=0.067). Also concerning the production of *kesk*, there was no difference between sentences with subject and object gap, or in relation to age (all p-values > 0.05).

(20) Le maître demande que la fille a dessiné. [Carla, 5;3]  
   the schoolteacher asks QUE the girl has drawn

(21) L’élève demande qu’est-ce que la maîtresse a lancé. [Sarah, 5;6]  
   the-pupil asks kẽsk the schoolmistress has thrown

(22) Qu’est-ce que le maître a lancé? kẽsk the teacher has thrown
   ‘What did the teacher throw?’

---

![Figure 2](image)

**Figure 2.** Proportion of omissions of the determiner *ce* (with ±1 standard error). On the x-axis, age is represented in years. The left panel shows the performance on Indirect Questions and the right panel on Relative Clauses. The dark bars show the performance on Object Gap and the gray bars on Subject Gap sentences.
Figure 3. Proportion of utterances with the complex wh-form *qu’est-ce que* (*kesk*). On the x-axis, age is represented in years. The left panel shows the performance on Indirect Questions and the right panel on Relative Clauses. The dark bars show the performance on Object Gap and the gray bars on Subject Gap sentences. The error bars represent ±1 standard error.

4. Discussion

The results of our experiment show that French children have difficulties in repeating *ce+que* sequences, particularly at a younger age. This difficulty is however modulated by the syntactic structure: children were overall more accurate in the relative clause condition than in the indirect question condition. The nature of the dependency, namely whether it involves a subject or an object gap, seems to play no significant role. In what remains we will briefly discuss these two main findings.

4.1. No reanalysis

The first finding, and in particular the fact that children seem to have more trouble with interrogatives than with relative clauses, appears to disconfirm the reanalysis hypothesis of *ce+que* as the wh-element *səkə* that we entertained as plausible in Section 1. As we argued in details in Section 2.1 such reanalysis would predict either no difference between the two conditions (if children are aware that wh-structures can be both a question or a relative clause) or a higher accuracy in the indirect question condition, where a wh-element is obligatory and fully interpreted. The higher accuracy with RCs that we have found in our experiment
suggests that children analyze this complex morpheme as a D+C sequence, which they find less appropriate in an indirect question, where a wh-element is required.

The errors analysis goes in the same direction: the omission of ce that is observed with indirect questions more than with relative clauses, illustrated in (20) can be interpreted as a substitution error: children replace the disturbing ce que with the wh-que. Notice that the use of wh-que is not acceptable in this context in adults’ French (see 23b.), because its clitic nature requires it to be adjacent with an inflected element in C, as in 23a., and inversion in C is not possible in embedded context, as shown in (24). Children are ready to break this rule (or ignore it) in order to eliminate the disturbing nominal element ce.

(23) a. Que veux-tu?
   What want.2s - you
   ‘What do you want?’

   b. *Que tu veux?
      What you want.2s

(24) a. *Je me demande que veux-tu
     I me wonder what want.2s-you

   b. *Je me demande que tu veux
      I me wonder what you want.2s

The other strategy that we observe in children’s behavior is the replacement of the ce+que sequence with the wh-element kesk, as in (20). In adults’ grammar, this wh-element is supposed to be ungrammatical in embedded contexts, at least in Standard French, but many adult speakers use it frequently in indirect questions as well as in root questions. Again, the fact that children opt for this option instead of the ce que element, more so in wh-questions, is coherent with the hypothesis that they perceive ce que as being something else than a simple wh-element, something that is less natural in an interrogative context: a DP structure.

These results overall can lead us to discard the hypothesis that ce+que has undergone a reanalysis as a wh-element in French, as happened in many other Romance varieties. Once we discard this possibility, we are left with the question of what the syntactic function of this determiner is in the interrogative context, where no such determiner is neither selected nor has any other clear role to play. We leave this question to further research.

4.2. No subject-object asymmetry

As for the absence of any subject-object asymmetry in children’s behavior, this might be surprising prima facie at the light of such overwhelming evidence that object dependencies are more difficult, and acquired later, than subject dependencies in French and in most languages (Fragman, 1998 for French; Goodluck and Tavakolian, 1982 for English; Costa, Lobo and Silva, 2011 for
Portuguese; Friedmann and Novogrodsky, 2004 for Hebrew; Adani, 2011 for Italian; Adani, Sehm and Zukowski, 2012 for German). It has also been shown, however, that the difficulty of object dependencies is modulated by the featural endowment of the subject and of the moved element (Friedmann et al. 2009 and much subsequent work): in a nutshell, it has been shown that object dependencies are difficult only when the moved element and the subject are featurally similar, and the latter thus acts as an intervenor disrupting the dependency. When they are not featurally similar, the asymmetry between subject and object dependencies disappears, or is at least substantially reduced, and object dependencies are not more difficult than subject ones.

With this approach in mind, let us now go back to our findings. In all the object gap cases, the object and the subject display a double featural mismatch: the object is inanimate (ce) and the subject is animate; the object is a bare determiner ([-NP]) and the subject is a lexically specified NP [+NP]). This configuration of double mismatch is illustrated in (25).

(25) \(ce\) [\(\text{CP que}\) \(\text{la fille}\) \(a\) \(\text{dessiné t}\)]  
\([-\text{animate}]\)  
\([+\text{animate}]\)  
\([-\text{NP}]\)  
\([+\text{NP}]\)

(25) is exactly the kind of configuration where object asymmetries are claimed to disappear under the subject intervention approach just outlined. The absence of such an effect, far from being surprising in our data set, confirms the importance of the featural composition of the subject intervenor in object dependencies.

5. Conclusion

In this paper we have tried to test a hypothesis concerning the grammar of French by looking at how it is acquired by children. The hypothesis concerns the sequence \(ce\) que that can introduce either an indirect question or a light headed relative clause in French, depending on the verb selecting the entire construction. A possible explanation for this double nature of \(ce+que\) sequences is that \(ce+que\) has undergone the same reanalysis into a wh-element that is observed in many Romance varieties, including various Italian dialects and Portuguese. If this is the case, the ambiguity of \(ce+que\) sequences is explained as yet another case of the systematic ambiguity that wh-structures display in many languages, allowing for an indirect question and a free relative interpretation/distribution.

The acquisition data that we obtained from our repetition test clearly go against such a hypothesis: children do not behave as if \(ce+que\) were a wh-element, since they tend to be less accurate in repeating this very sequence in indirect questions, where such a wh-element would be expected, being obligatory and interpreted. The conclusion that we can draw from this result is that children do analyze \(ce+que\) as a complex sequence, possibly nominal hence more acceptable in a relative clause environment. This is of course only a conclusion about children: it could still be the case that French learners go from a phase in which
they analyze *ce+que* as a complex sequence to a phase where they reanalyze it as a simple wh-element. This is however not very likely since it would involve going from a more complex construction to a simpler one. Besides, this conclusion about the nature of *ce que* in children grammar converges with other findings concerning the grammar of Standard French (Konrad 2018), going in the same direction: in French, *ce que* does not undergo any reanalysis process, and still keeps its nominal component. Acquisition data can hence be seen here as an important complement of information, providing a strong argument against a syntactic analysis that has been proposed for Adult grammar.

It would be interesting then to inquire what happens in a language where a similar reanalysis has indeed taken place for sure, such as the Italian dialects discussed in (10-11), or Portuguese (12). We plan to run the same experiment with Portuguese children, and see how they perform in an elicitation task including an ambiguous *o que* sequence. The prediction is that they will outperform French learners of the same age range, and that no clause type effect should arise. To be continued...

**Appendix: List of the experimental items.**

<table>
<thead>
<tr>
<th>Subject gap</th>
<th>Relative clauses</th>
<th>Indirect questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le papa mange ce qui est tombé du placard.</td>
<td>Papa sait pas ce qui est tombé du placard.</td>
<td></td>
</tr>
<tr>
<td>La directrice efface ce qui est au tableau.</td>
<td>La directrice demande ce qui est au tableau.</td>
<td></td>
</tr>
<tr>
<td>L’enfant attrape ce qui est sorti de la boîte.</td>
<td>L’enfant devine ce qui est sorti de la boîte.</td>
<td></td>
</tr>
<tr>
<td>Le monsieur attrape ce qui a roulé par terre.</td>
<td>Le monsieur demande ce qui a roulé par terre.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Object gap</th>
<th>Relative clauses</th>
<th>Indirect questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Le maître efface ce que la fille a dessiné.</td>
<td>Le maître demande ce que la fille a dessiné.</td>
<td></td>
</tr>
<tr>
<td>L’élève attrape ce que la maîtresse a lancé.</td>
<td>L’élève demande ce que la maîtresse a lancé.</td>
<td></td>
</tr>
<tr>
<td>La dame mange ce que la maman a cuisiné.</td>
<td>La dame sait pas ce que maman a cuisiné.</td>
<td></td>
</tr>
<tr>
<td>La nounou attrape ce que bébé a caché.</td>
<td>La nounou devine ce que bébé a caché.</td>
<td></td>
</tr>
</tbody>
</table>

**References**
