1. Introduction

Complementation structures consist of a complex clause with a main verb that selects an embedded clause as its direct object. The structural and semantic properties of complementation are topic of many syntactic and semantic theories. Studies address a wide variety of syntactic and semantic constructions: long-distance movement and barrierhood, point-of-view phenomena such as direct versus indirect speech and sequence of tense, factivity, opacity, true versus false complements under verbs of saying and mental verbs.

There is also a large body of research on the acquisition of complementation in the generative tradition. It goes back to the 80’s when researchers investigated long-distance wh-movement in children (De Villiers, Roeper & Vainikka, 1990; van Kampen, 1997; Oiry, 2008; Thornton, 1990). Other acquisition of complementation phenomena include: direct versus indirect speech (Hollebrandse, 2007), sequence of tense (Hollebrandse, 2000; Lungu, 2012), factivity (Schulz, 2003); referential opacity (De Villiers & Fitneva, 1996; De Villiers, 2001), and
true versus false complements (De Villiers & Pyers, 2002). The tradition continues up until today, witness recent work on double embedding (De Villiers, Hobbs & Hollebrandse, in press; Hollebrandse, Hobbs, De Villiers & Roeper, 2008; Hollebrandse & Roeper, under review).

The core question in this research tradition is: When do children acquire the properties of complementation? Our study presents a new angle on long-distance wh-movement by testing wh-questions with a super long distance between wh-phrase (filler) and trace (gap).

2. Background

Wh-extraction is only possible out of complement clauses, not out of adjunct clauses, noun phrases or wh-clauses (Rizzi, 1982, 1990; Ross, 1967). Locality effects arise in complex wh-questions with an embedded wh-clause (with an upstairs wh and a medial wh) such as (1). The medial wh-phrase functions as a relativized minimality barrier—blocking a long-distance relation between the upstairs wh-word and a trace in the embedded question—when the upstairs wh-word and the intervener are of the same type. Thus, how is an intervener in (1b), because when and how are both adjuncts, but not in (1a) because who is an argument and how an adjunct.

(1) a. Who did Big Bird ask how to paint?
   b. When did the boy say how he hurt his arm?

The first acquisition study to investigate children’s sensitivity to barrierhood and long-distance (LD) movement was De Villiers et al. (1990). Children made a crucial distinction between argument and adjunct extraction. For wh-arguments such as who in (1a), the children correctly associated the wh-phrase to a gap in the lower clause (answering who he painted) or the higher clause (answering who he asked). This is a target-like pattern, since the question is ambiguous. For wh-adjuncts such as when in (1b) on the other hand, which are not ambiguous, children properly restricted the wh-phrase to short-distance readings (answering when he said it), and never gave LD readings (when he hurt his arm). De Villiers and colleagues argue that these results show that the children have acquired barrierhood for LD movement. They point out furthermore that the children did not seem to have parsing problems relating the filler to the gap in the lower clause in (1a) (see Roeper & De Villiers, 2011, for a recent overview of subsequent studies on this topic).
Yet, the younger children (3-year-olds) sometimes produced a particular type of error for (1b): they answered the medial wh-word, saying how he hurt his arm (instead of when he said it). De Villiers et al. (1990) explain these so-called medial answers as a consequence of a different parameter setting for wh-questions. They argue that in the grammar of these 3-year-olds, the upper wh-word is merely a copy of the medial one, just like the grammar of some German dialects. The medial wh-phrase contains the actual question operator, while the upstairs wh-phrase functions as some kind of scope marker, marking the question as a whole as a wh-question. For an account of this developmental pattern couched in more recent syntactic terminology, see De Villiers, De Villiers & Roeper (2011). The essence of their analyses is supported by the fact that children regularly double the wh-word in their own production of complex questions (e.g., *What do you think what’s in the box?*) (for English, Thornton, 1990; for Dutch, van Kampen, 1996). De Villiers and colleagues conclude that, even though children have firmly acquired barrierhood, and thus complementation, they have not fully acquired all aspects of the English grammar of LD wh-constructions.

In her later work, however, Jill de Villiers argues that complementation may not be fully acquired early after all; certain aspects of complementation structures seem to be acquired late. De Villiers (1999) and De Villiers and Pyers (2002) developed a complementation test with true and false propositions embedded under verbs of saying, using complement clauses describing mistakes such as (2). The main verb embeds a false proposition (it is not true that there is a bug in the girl’s hair), and the question asks for this description, thus probing children’s memory of complements.

(2) This teacher said there is a bug in the girl’s hair.
But it was really just a leaf.
What did the teacher say?

The target answer is the teacher’s mistaken claim (*bug*); many 3-year-old children answered with a true description (*leaf*), even though this is not what the teacher had said. Note that this answer does not reflect any failure of false-belief reasoning, because the verb in the main clause is a verb of communication, not mental representation. Rather, the children had trouble reporting what was actually said when that claim was false (a mistake). De Villiers and Pyers conclude that certain features of syntactic complementation are not yet in place; children “fail to incorporate the complement under the scope of the top verb” (de Villiers & Pyers,
De Villiers (1999) speculates that children lack a truth-value feature in their complementation structure; this leads to a more restricted grammar which allows only true complements.

Summarizing so far, children can deal with LD wh-questions (supplying LD answers). Furthermore, they have acquired barrierhood of embedded questions (no island violations in LD wh-questions). However, they have trouble with double wh-questions (medial answers, wh-word doubling in production). Moreover, they cannot deal with embedded false complements. So, certain structural aspects of complementation seem to be acquired, whereas other syntactic and semantic properties are not.

Our goal is to determine whether or not children have acquired the structural properties of complementation, specifically, syntactic embedding. We approach this question by investigating recursive syntactic embedding. For sentences with one level of embedding, it is possible that the construction is formed with an actual recursive rule that combines—in technical terms merges—main verb with embedded clause, or, alternatively, in some non-target-like way, e.g., as two coordinated main clauses or with some form of adjunction structure. In fact, even in the adult grammar, there are alternative ways of expressing embedded propositions with verbs of thinking and saying; the content of one proposition can be ascribed to someone using a sequence of two main clauses, (3a), effectively making the same claim as a single-embedded clause, (3b). Roeper (2007) observes, however, that this is impossible for double-embedded constructions: (4a) does not mean the same as (4b).

(3) a. It is raining. John thinks that.
     b. John thinks that it is raining.

(4) a. It is raining. John thinks that. Mary knows that.
     b. # Mary knows that John thinks that it is raining.

It is possible to express the semantics of belief ascription without syntactic complementation, but only at first-order level. In order to report a second-order belief, i.e., a belief about a belief, a double-embedded structure is necessary. The same applies for verbs of saying.

The reason why we extend the domain of investigation to recursive structures is that single-embedded structures, such as (1) and (3b), do not necessarily involve syntactic complementation. Embedding structures with two or more levels on the other hand constitute a case of recursive, syntactic complementation, since recursive constructions reveal the true
nature of a rule (Hollebrandse & Roeper, in press). If children can do extraction across one level of embedding, can they also do it recursively?

In order to find out when recursive complementation is acquired, we compare the acquisition of embedded structures in single and double-embedded structures. We tested two types of recursive embedding: double-embedded declarative clauses with a production task and LD wh-questions with a comprehension task. Our specific research questions are: i) Can children produce double-embedded clauses? ii) Can children do super-long-distance argument extraction in double-embedded questions with false complements? If recursive, syntactic complementation is in place, we expect children to be able to produce double embedded clauses. Moreover, if there are no parsing problems relating the filler to the gap in the lower clause (De Villiers et al., 1990), we expect children to be able to give super-long-distance answers.

3. Method

Twenty two Dutch 3rd and 4th grade children (8;2–9;11, mean age: 8;10) and a control group of adults participated. Two tasks were administered to each participant.

For the production task (Hollebrandse et al., 2008), the experimenter told short stories about a boy called Billy and his family, illustrated with pictures; see Figure 1 and (5) for a sample item. In the stories one family member (here, Billy) talks to another member (the sister), who reports to a third member (the mother) what the first person (Billy) said. It is crucial in this task that the proposition (“sisters are stupid”) be linked to the proper person, keeping straight who says what. So, the experimenter asks what the first person said (targeting a single embedding: “Billy said that ...”) or

---

1 The data presented in this paper are part of a larger project that furthermore included two second-order false-belief tasks. We collected data from 6-7-year-olds and 8-9-year-olds on the production task presented below and both false-belief tasks. After testing the older children on the comprehension task presented below and seeing that they did not do well, we decided to drop this task with the younger children, especially given that our task battery was already fairly large. Here we focus on the 8-9-year-olds, comparing their production and comprehension of double embedding. See Hollebrandse, van Hout & Hendriks (2014) for a comparison of the younger and older children on the theory-of-mind tasks.
what the second person said about the first person (targeting a double embedding: “The sister said that Billy said that ...”). Note that the natural way of answering these questions is to start with a complementizer, thus dropping the main clause subject and verb, which was in fact what most participants did. The task targeted six single-embedded clauses, (5a), and six double-embedded clauses, (5b), making a total of 12 stories.

Figure 1: Sample elicitation item in production task, see protocol in (5)

(5) a. Exp: Wat zei Billy?
‘What did Billy say?’
Target: Dat alle zusjes stom zijn.
‘That all sisters are stupid.’

b. Exp: Wat vertelde Jane aan haar moeder?
‘What did Jane tell her mom?’
Target: Dat Billy zei dat alle zusjes stom zijn.
‘That Billy said that all sisters are stupid.’

The comprehension task involved questions after complex stories, modeled after the De Villiers and Pyers’ (2002) test with false complements. We created an additional level: someone makes a claim about the claim of another; neither claim is true, but the complex clause as a whole (about someone making a certain claim) is true. In the stories a police officer (or a dad) receives a phone call in which someone reports a problem and then tells his colleague (or the son) about it. The essence of the paradigm is that (i) the officer mishears and says something else than the caller had actually said, and (ii) both caller and officer are wrong, because the reality turns out to be different. Figure 2 and (6) illustrate a sample item. The first character (a woman) calls a police officer and makes a certain claim (“there is a cat in my bag”). The officer reports a
different claim to his colleague (“a woman says there is a rat in her bag”). He adds his own opinion (“there must be a wallet in her bag”). The final picture shows that in reality something different was the case (there is a stuffed animal in her bag).

De vrouw belde de politie en zei dat er een kat in haar tas zat.
“Wat” zei de agent, “een vrouw zegt dat er een rat in haar tas zat. Maar dat is raar. Er zit vast een portemonnee in.”
Maar kijk, het is eigenlijk een knuffelhondje.

‘The woman called the police and said that there was a cat in her bag. “What!” said the officer, “a woman said that there was a rat in her bag. That’s weird. There must be a wallet in it”. But look, it really is a stuffed dog.’

Figure 2: Sample comprehension item in comprehension task, see protocol in (6)

(6) a. Wat zei de agent dat de vrouw zei dat er in haar tas zat?
‘What did the officer say that the woman said there was in her bag?’
Target: Een rat ‘a rat’

b. Wat zei de vrouw dat er in haar tas zat?
‘What did the woman say she there was in her bag?’
Target: Een kat ‘a cat’

c. Wat zei de agent dat er in haar tas zat?
‘What did the officer say there was in her bag?’
Target: Een portemonnee ‘a wallet’

The stories were illustrated with a series of three pictures, shown one by one on a laptop screen; all three pictures remained visible when the experimenter asked what the caller and the officer said. After each story two questions were asked: first, a double-embedded question about what
the officer said about the caller, (6a); next, a single-embedded question about either the caller (6b) or the officer, (6c). Participants had to interpret the wh-phrase by linking it to the proper gap position. The task consisted of ten stories, yielding ten double-embedded questions, five single-embedded questions about the first character and five about the second character.

4. Results

The adults performed at ceiling in both tasks. Figure 3 shows the results for the school children on the production task (Billy): the children performed at ceiling for single-embedded clauses, and they were also good at producing double-embedded clauses: 80% target answers for (5b). A paired sample t-test shows no significant difference in performance on single and double embedding.

![Figure 3: Mean correct answers in production task (Billy)](image)

Non-target answers sometimes involved two coordinated clauses with an anaphor in the second clause ("That girls are stupid and that Billy said that.") (compare 3a).

The comprehension results are illustrated in Figure 4. In this task, the children had no trouble understanding long-distance dependencies in wh-questions with one level of embedding: 92% correct answers for (6b). In contrast, they had severe problems with double-embedded wh-clauses: only 43% target answers for (6a). The errors show that they interpreted
the double-embedded questions sometimes as single-embedded ones (40%). For example, when asked the double-embedded question (6a), they answered *wallet*, which was the target for the single-embedded question in (6c).

![Figure 4: Mean correct answers in comprehension task (Officer)](image)

A repeated-measures ANOVA shows a significant effect of question type (F=29.16; p< 0.001). Paired-sample t-tests show that the difference between the single-embedded (woman) and the double-embedded condition is significant (t(21)=-9.08; p<0.001), as well as the difference between the two single-embedded conditions (woman vs. officer) (t(21)=7.63; p< 0.001).

Surprisingly, the two types of single-embedded questions turned out to be different for the children: they performed at ceiling for the questions about the person who called the police, the woman in (6b), but they performed much worse (only 41% correct) on the question about what the police officer told his colleague about the bag, (6c). We had expected both types of questions to be equally easy (or difficult), because they both involve one level of embedding (“The woman said there was a rat in her bag” and “The officer said there must be a wallet in her bag”). Moreover, the correct answer is equally non-factive in both cases: at the end of the story it turns out that that there is something else in the bag than what both the woman and the officer said there was. We will return to this difference in the Discussion.

So, we found an asymmetry between the two tasks: in the production task, the school children reveal knowledge about double embedding,
whereas in the comprehension task they often fail to perform well on double-embedded questions.

5. Discussion and conclusions

We investigated when children acquire the recursive structure of clausal embedding, asking i) Can children produce double-embedded clauses? ii) Can children do super-long-distance movement? The answer is: i) yes, Dutch 8 and 9-year-old school children produce double-embedded structures. But ii) no, they do not do recursive long-distance movement. Both types of structures involve double embedded complements. Why, then, do they not manage to employ this same structure in both tasks alike?

The literature shows that children can deal with extraction across one level of embedding. (De Villiers et al., 1990). Our results on the single-embedded condition support this, which was to be expected given that our children are much older than the children in De Villiers et al.’s study. But our children could not deal with recursive wh-extraction, across two embeddings.

Do school children still have trouble dealing with the truth-value feature on complement clauses? This seems highly unlikely, given that De Villiers and Pyers (2003) established that children perform well on a memory of complementation task with false propositions by the age of 5, at least for single-embedded structures, like (2). In our study too, the children performed well on the single-embedded wh-questions such as (6b), which involved a false proposition. So, they are able to remember and assess false complements.

Don’t they have a recursive structure for complement clauses? Maybe they have a single level of complementation structure, but cannot construe double-embedded structures? Our findings on the elicitation task show that they can: they have no problems producing double-embedded clauses. One could raise an objection here, given that the children provided elliptical answers, starting their answers for the double-embedded condition with the complementizer and effectively leaving out the subject and main verb, (5): Dat Billy zei dat alle zusjes stom zijn. ‘That Billy said that all sisters are stupid.’ With no embedding main verb present, what is the that-clause embedded in? The presence of the complementizer dat ‘that’ and the verb-final word order are overt indications of syntactic embedding. Moreover, children link the right complement to the right person (i.e., they know who

---

2 As did an anonymous reviewer.
said what), thus showing knowledge of semantic embedding, in a recursive way. Thus, the children have acquired double-embedded complement structures both syntactically and semantically. And they have the proper truth-value features on those clauses. We conclude that complementation is fully and recursively acquired. Why, then, can’t they answer super-long distance argument wh-questions?

We argue that the asymmetry between the production and comprehension of double-embedded clauses reveals a novel instance of the kindergardenpath effect (Trueswell et al., 1999). Our proposal is that the parser fails, when it must wait too long for the gap in the second embedded clause. There are earlier, temporarily potential gap positions to which the parser can link the wh-word. These create a garden path from which the children are not able to backtrack and revise. Thus, in a complex wh-question such as (6a), the parser posits the gap at the first possible filler position, i.e., after the first subject NP the officer, (7). It must subsequently revise this parse to incorporate the remaining sentence, ultimately positing the gap at the super-long-distance position, (8). A child parser, however, does not manage this revision, hence, the short-distance answers.

(6) a. Wat zei de agent dat de vrouw zei dat er in haar tas zat?
   ‘What did the officer say that the woman said there was in her bag?’

(7) Wat zei de agent gap ...
what said the officer gap ...

(8) Wat zei de agent dat de vrouw zei dat
what said the officer that the woman said that

er gap in haar tas zat?
there gap in her bag sat

This proposal makes a straightforward prediction. Children should be able to deal with double-embedded clauses as long as they do not involve a filler-gap dependency. Suppose the task is changed to a truth-value judgment task: children have to judge declarative sentences that vary the most deeply embedded complement: “The officer said that the woman said that there is a rat / wallet / stuffed dog in her bag. Is that right?”. If recursive complementation is acquired by age 8, children are expected to judge such double-embedded sentences correctly for the different
complements (i.e., they should only accept “wallet” and reject “rat” and “stuffed dog”). We are presently investigating this in a follow-up study.

This leaves us to discuss the surprising finding that the children did not perform equally well on the two types of single-embedded questions, (6b) about the woman’s claim and (6c) about the officer’s claim.

(6) b. Wat zei de vrouw dat er in haar tas zat?
‘What did the woman say she there was in her bag?’
Target: Een kat ‘a cat’

c. Wat zei de agent dat er in haar tas zat?
‘What did the officer say there was in her bag?’
Target: Een portemonnee ‘a wallet’

We suspect that the cause for the lower performance on the latter question lies in the set-up of the task, specifically, in the presentation of the questions. For each story, we asked two questions, first the double-embedded question (6a) and then a single-embedded question, either (6b) or (6c). However, when participants responded incorrectly to (6a), they typically gave a single-embedded answer, effectively providing an answer to (6c). When they were subsequently asked (6c), they were surprised to get the “same” question again. Some even made their confusion explicit, saying “But I just told you”. This confusion may have caused the low performance on this condition. If so, these errors should disappear in a set-up where only one question is asked per story, taking away this potential confusion.

The incorrect answers to (6c) may also indicate confusion about which complement that the officer used, is being targeted, as he actually used two different ones: he (mis)-quotes the woman’s claim (rat) and he provides his own claim (wallet). And so one might parse (6c) as a question about both things the officer mentioned—rat and wallet. Note that this problem does not arise with (6b), because the woman makes only one claim (cat). This explanation raises the question why the children gave two possible answers to (6c), whereas the adults did not.

A different type of explanation for the difference between the two types of single-embedded questions refers to the “modality” of the officer’s statement in the story, signaled by the modal adverb vast:3 Er zit vast een portemonnee in ‘There must be a wallet in it’. If modality is somehow difficult for children, we expect that they will perform better on the single-embedded question about the officer if we slightly modify the

---

3 An anonymous reviewer offered this suggestion.
story removing the suggestion of modality. In fact, another aspect on which the story is not quite balanced is the way in which the two claims are presented: the woman’s claim is given in indirect speech and the officer’s claim in direct speech. One could make both of these aspects more equal in the following way, (9).

(9) ‘The woman called the police and said: “There is a cat in my bag”. “What!” said the officer, “a woman said that there was a rat in her bag. That’s weird. There is a wallet in her bag.” But look, it really is a stuffed dog.’

To conclude, our kindergardenpath explanation for these school-aged children is in line with recent work that also establishes parsing difficulties in similar-aged children with object wh-questions that are disambiguating by number agreement (Metz et al., 2010; Schouwenaars et al., 2014; Strangmann et al., 2014). Taken together, these studies suggest that parsing in children is not yet adult-like up to quite an advanced age. This calls for more research into the developmental track of parsing.

Acknowledgements

We thank the Acquisition Lab participants at the University of Groningen and the audience at GALA 2013 at the University of Oldenburg for their sharp questions and lively discussion. We are looking forward to write up our next paper comparing the present study to a follow-up study by our student Floor Banning where we used a truth value judgment paradigm. We gratefully acknowledge ESF (grant no. 028395, PI M. Krifka and U. Sauerland).

References

http://irs.ub.rug.nl/dbi/4ef4a1ef1c357