When *Must* Children Acquire Long Distance Wh-Extraction?

Jessica Kotfila and Jill de Villiers

1. Introduction

The young child’s response to wh-questions that require long distance movement has been a topic of interest in acquisition work for some time. Since it was established that children were capable of long distance movement (de Villiers, Roeper, & Vainikka, 1990), various studies have asked to what extent the child’s grammar compares to adult grammars in this respect. One fascinating finding was that the child often answers argument questions like 1. unlike adults do.

(1) a. What did mom say she bought?

For 1, imagine that mom bought *oranges* but made a mistake and said she bought *apples*. An adult provides the answer *apples*, but the young child aged 3 or 4 years says *oranges*. This failure to integrate the matrix and embedded verbs, resulting in a “reality” type error accounting for only the lower verb, applies to both argument and adjunct questions with verb complements under *say*, *tell*, and *think* (de Villiers, Kotfila, & Klein, in press)). Here is a sample story, 2 a., from that study, which had illustrations to accompany each stage of the story:

(2) a. Mom needed some rice for supper so she gave the money to her daughter and asked her to go to the store for rice. The girl ran as fast as she could but when she got to the store the money was not in her hand anymore. Look, it fell out when she opened the door to the store. She couldn't buy the rice. She went home and told her Mom in the kitchen that she dropped the money when she was running.

b. Test question: Where did the girl say she dropped the money?

Two possible adult answers to 2 b. are: a) where she *said* it: in the kitchen at her house, or b) where *she said she dropped it*: when she was running. The immature answer is where she *really* dropped it, at the door of the store. There is considerable debate about why the child makes such an error. Is it the result of cognitive development, pragmatics, or the grammar? If it is the immature grammar, what is it about the child’s grammar that differs from that of the adult?

* Jessica Kotfila, Georgetown University, jak328@georgetown.edu. Jill de Villiers, Smith College.

Interestingly, manipulating the linguistic structure and semantic mood of the question appears to change the child’s answer. Children are less likely to make such an error when confronted with irrealis infinitival complements as in 3.a. than they are with realis tensed complements in 3.b. (de Villiers, Harrington, Gadilauskas & Roeper, 2012).

(3) a. What did Mom say to buy?
   b. What did Mom say she bought?

That is, if Mom said to buy *apples* but *oranges* were really bought, the child will answer like an adult with *apples* in 3.a., but with *oranges* in 3.b. when Mom misspeaks.

In keeping with these findings concerning irrealis infinitival complements, yes/no questions using *think… should*, with a modal in the lower clause were easier for children to answer like adults than tensed complements such as *think… is* constructions (de Villiers, 2005). Here we ask whether these findings generalize to *wh*-questions containing modal complements, that is, we investigate a new structure with an irrealis mood.

As mood and structure appear to influence the child’s behavior we hypothesized that questions containing modal complements would be easier for children. Furthermore, we hypothesized that modal flavor, epistemic versus deontic, could influence child answers, as various arguments suggest these flavors differ in terms of their structure and moods.

First we will explain why questions containing modal complements should be easier for the child because of their structure and mood. Roeper and de Villiers (2011) argued that these so called “reality” answers arise because the child answers the wh-trace in the lower clause at the first phase of interpretation. A mature grammar instead waits until the wh-word moves up to the top spec-CP before interpretation occurs at the final phase, allowing adults to integrate across both the matrix and embedded verbs. The mechanism through which the appropriate embedding occurs is hypothesized to be covert tense movement, through which the point of view on the lower clause subordinates to that of the matrix clause. If this does not happen, the wh is interpreted only with respect to the last clause, and the Point of View of the matrix subject - e.g. what Mom said, or thinks, is neglected. Root modals scope below tense and aspect, located lower in the syntactic structure (Hacquard, 2017), and may not entail this covert tense movement (de Villiers, et al., 2012). The child’s immature grammar may be equipped to deal more readily with wh-movement from a deontic modal complement. This prediction fits with the previous findings that for irrealis complements the child seems to resist reality type answers.

Consider further the question of modal flavor, as theorized structural and semantic differences exist between deontic and epistemic modals. Deontic modals would potentially be easier for children to provide verb-integrated answers for two possible reasons. It has been argued that deontic modals have a different structural position on the tree than epistemic modals (Hacquard, 2017).
In contrast to them, epistemic modals are located in a higher syntactic position scoping over tense and aspect (Hacquard, 2017). They may also be considered closer to a more realis mood, exemplified in 4b.

(4) a. Mom must wash the car
   b. Mom must be washing the car
   c. Mom is washing the car

Epistemic and Deontic modals are truth evaluable in a different way compared to one another and compared to tensed complements. Suppose 4.a. carries a deontic flavor such that a. is true if in all possible worlds where it is an obligation for Mom to wash the car, Mom washes the car. $A$ will refer to the obligation of washing a car, and the act of washing the car will be referred to as $B$. If in all possible worlds $B$ is contained in $A$ then 4.a. is true. However, $A$ can still be true in this world independent of whether or not Mom is currently washing the car. Mom *must* eventually wash the car, but 4.a. is not false at the time of the utterance if she is not currently washing the car. The deontic statement 4.a. is not anchored to the time of the utterance in the sense that the event execution of the car washing need not occur for 4.a. to hold true.

However, 4.b., where b. is meant to carry an epistemic flavor, behaves differently than 4.a. For example, “Mom must be washing the car” can be a valid statement under the following conditions:

I. there are possible worlds where given what is known about Mom, washing cars and other pertinent information, it is valid to make an assumption about Mom washing cars
II. there are possible worlds where Mom is actually washing the car
III. there is overlap between the possible worlds in I. and II.

The event execution of Mom washing the car during or before the time of the utterance is essential to the truth of the statements in terms of their deniability. This is also true for the ordinary tensed sentence 4.c., where the event execution of the car washing is essential for the statement to be true. The truth of the epistemic case 4.b. can be denied in two ways, that is the statement can be false if either the reasons to believe Mom is washing the car are incorrect or the event execution of mom washing the car is not occurring. In this way epistemic modals are more realis than deontic modals: their truth is tense-dependent. If so, children might treat epistemic modals more like tensed complements, and make reality-type errors.

We hypothesized first, that children overall will provide fewer reality type answers for questions with modal complements than for non-modal complements, subsequently resulting in more adult-like long distance responses.

---

1 Throughout this paper we often use *must* + *have* or *be* to express epistemic modality. Epistemic *must* occurs most frequently in this construction but need not exclusively use this construction to be epistemic.
Second, children were expected to provide more reality type errors for questions with an epistemic modal (reals) than a deontic modal (irreals).

2. Design

The present study was designed using the same methodology and experimental framework as de Villiers, Kotfila, and Klein (in press) in order to provide comparable non-modal and modal conditions. Test questions were constructed using the same nine possible adjunct and matrix verb combinations as de Villiers et al (in press). Of the nine test questions for each of the three modal conditions, three were *when* questions, three were *where* questions, and three were *why* questions. For each of the different adjunct questions, the wh-word was paired with a communication or mental state verb: *say*, *tell X*, or *think*. The second verb was selected based on its ability to carry a semantic sense of obligation or necessity as well as an epistemic flavor. The verbs were the same in every condition: *get*, *show*, *wash*, *buy*, *make*, *clean*, *feed*, and *clean up*. These verbs are different from de Villiers et al. (in press) as they needed to pair easily and sensibly with necessity and/or obligation in the deontic condition and belief in the epistemic condition.

Every test question was paired with a contextual story and picture set as in de Villiers et al (in press). These stories contained information that could provide a “reality” type answer to the question, contrasting what was said or thought with the reality of the actualization of the second verb. In this way the stories contained all the information necessary to allow three possible answers, the short distance, long distance, or reality answer to each question. In order to rule out a recency effect, the order in which these possible answers occurred within the stories varied across the nine different stories. In addition, the nine question and story scenarios were presented to each individual child in a quasi-random order to insure that they would not answer three *where*, *when*, or *why* questions sequentially.

The important distinction the present study makes from de Villiers et al. (in press), is that the complement in each question contains a modal. Following the adjunct and communication or mental state verb, the complement of each question contained the modal auxiliary verb *must* before one of the final verbs listed above. See Figure 1 for an example. In a between-subjects design, this study contained three different conditions depending on modal flavor. The questions in the two Epistemic conditions were marked by the phrase *must be* while the Deontic condition contained *must* independent of the auxiliary *be*. The same images were used to accompany the stories in Epistemic 1 and Deontic condition. The stories used in these two questions were as identical in content and language as possible with minor differences to clearly distinguish the two modal uses. Epistemic 2 only differed from Epistemic 1 in terms of one image. In Epistemic 1 the image that depicted the reality answer clearly confirmed that the event described in the modal complement was not actualized in the same manner that the event was *said*, *told*, or *thought* to have been actualized. In the drawing for Epistemic 2, reality is not depicted directly. These images required
epistemic reasoning to infer a different actualization of the event described in the modal complement, prompting additional reasoning to understand what was real. This was done to see whether children treat epistemic modals differently on the basis of their deniability. It is possible to deny the assertion in 5 in two different ways (explained in Figure 1).

(5) a. Mom must be making soup

That is to say, it might be evident that she is not currently doing so, or, the inference itself is mistaken. The actualization of the event in the modal complement can be denied or the basis for the epistemic reasoning to arrive at this conclusion can be denied.

<table>
<thead>
<tr>
<th>Figure 1: Example Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>*the same images were used for both the epistemic and deontic conditions. Image 3.b. was used in place of image 3.a. for only the epistemic 2 condition.</td>
</tr>
<tr>
<td>Epistemic Story (used for epistemic conditions 1 and 2)</td>
</tr>
</tbody>
</table>
Every time Bella gets sick Mom makes her soup so she feels better. Dad heard Bella sneeze and smelled soup cooking so he told Grandma Mom must be making Bella soup because she is sick. But really Bella sneezed because she opened a really dusty book so Mom must be making soup because it’s almost dinnertime.

<table>
<thead>
<tr>
<th>Test Question</th>
<th>Test Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why did Dad tell Grandma Mom must be making soup?</td>
<td>Why did Dad tell Grandma Mom must make soup?</td>
</tr>
<tr>
<td>1. Short Distance- Bella sneezed/smells soup</td>
<td>1. Short Distance- Bella sneezed</td>
</tr>
<tr>
<td>2. Reality- almost dinner time</td>
<td>2. Reality- because it is what they are having for dinner</td>
</tr>
<tr>
<td>3. Long Distance- Bella is sick</td>
<td>3. Long Distance- Bella is sick</td>
</tr>
</tbody>
</table>

### 3. Method
#### 3.1. Participants

36 English-speaking children (11 boys and 25 girls) aged 3;2 to 5;6 were recruited from private daycare centers in Northampton Massachusetts. The mean age of the children was 4 years 4 months. There were 12 children in each condition, 4 three-year-olds, 4 four-year-olds, 4 five-year-olds. These children comprise the remaining population of participants after 4 children were excluded due to their inability to complete the task. In addition 35 native English-speaking adults were tested, 12 in the deontic condition, 10 in Epistemic 1, and 9 in Epistemic 2. All adults received the stimuli through a group presentation of the stories, isolated by condition, and received course credit for their participation.

#### 3.2. Procedure

##### 3.2.1. Children

The children were tested individually in one session. Children were randomly assigned to one of the three conditions based on their age and availability to ensure that the conditions were age balanced. The stories were read to the children while the picture stimuli were presented as the events occurred within the story. The pictures were left in front of the child as the story continued so that they could refer back to the events that occurred previously. Once the story finished the experimenter allowed for a brief pause before asking the target question. If the child did not answer, the target question was repeated.
If the child still did not answer, the experimenter asked if they wanted to hear it again, and on a few occasions read the story again. This step proved important as the children were tested onsite at their school or daycare and the test area was occasionally noisy and distracting. If the child merely pointed to a picture rather than giving a verbal answer, the experimenter prompted them by asking “can you tell me with words?”. Every child received the nine stories in a random order to control for order effects. The child’s answers were written down word for word and coded later as short distance answer, reality answer, or long distance answer.

3.2.2. Adults

Adult participants were recruited through courses being taught at Smith College in Northampton Massachusetts and received course credit for their participation. The student participants received the picture and oral stimuli in a similar manner to the children, but rather than being tested individually they were tested as a group. The experimenter read the stories and the questions out loud while projecting the corresponding pictures as a PowerPoint presentation. The adults were asked to write their answers down rather than produce them verbally. This procedure was followed on three separate occasions to accommodate the three conditions. Adults wrote down their own answers that were later coded for answer type just as the child answers were.

4. Results

For each participant a percentage of long distance, short distance, and reality type answers was calculated based on how many of each answer type they provided out of the total number of answers they provided. One-way ANOVAS were used to compare these answer percentages across conditions to evaluate the hypotheses. As expected, the adult controls showed an overwhelming long distance answer preference and little to no effect of condition on answer preferences. Following the results of previous studies, with the children there was a small but significant effect of age group on reality-type answers, as five-year-old children were less likely to produce reality type answers (F(2,27)=2.7, p=.048) shown in figure 2.
Interestingly, there was no significant effect of modal flavor on answer type on the percentage of reality-type answers (F(2,33)=.241, p=.7), unlike originally hypothesized. In fact, modal flavor did not significantly influence answer type in terms of the percentage of short or long distance answers either.

As there was no effect of modal flavor on answer type, the epistemic and deontic modal conditions were collapsed into a single modal condition to compare to the data on the group of children who were tested on non-modal complements in de Villiers et al., (in press). A between-subjects ANOVA compared the results across the two studies, each of which had similar samples of children participating. The children in the current study with modals were much less likely to give reality answers than those who had received non-modal complements (Figure 3) (F(1,80)=12.39, p=.001). The group receiving modals gave more verb-integrated long distance answers than the group receiving non-modals, as shown in Figure 4.
5. Discussion

As predicted, questions containing modal complements proved easier for children than questions containing tensed complements in that the children were able to provide more adult like verb-integrated answers. Here we demonstrate how, again, semantic mood and grammatical structure matter for these complex questions. In contrast, a subtle difference in the mood and grammatical structure of epistemic and deontic modals did not significantly influence the child’s
answers to these wh-questions. Why is it that modality matters but modal flavor does not? The answer to this question may stem from the child’s grammatical development.

The irrealis low modal structure of deontic modals allows for the child to resist reality answers more readily than with a realis tensed complement. The explanation we offer follows nicely from the analysis by Roeper and de Villiers (2011). They argue that reality answers may arise when the child fails to activate the covert tense movement that allows the Point of View of the lower verb to subordinate to that of the matrix clause. Recall that in a deontic modal clause, tense is irrelevant to whether or not the action has been or is being completed at the time of utterance:

(5) a. Dad must clean the house
   b. Dad must be cleaning the house
   c. Dad must have cleaned the house

5. a. is only deniable based on Dad’s obligation to clean the house, not whether or not he is cleaning the house at the time of the utterance, while 5 b. and c. are deniable based on the evidence that Dad is cleaning the house/has cleaned the house at the time of the utterance. Therefore there may be no separate Assertion with a different Point of View for the lower clause in these deontic modal complements, making correct long distance answers easier for the child.

The child’s performance with epistemic modal questions matched that with deontic questions, in that the two conditions did not differ significantly from one another in terms of frequency of short distance, reality, and long distance type answers. We predicted epistemic modals would yield more reality type answers compared to deontic modals on the basis of their more realis mood and higher syntactic position. Perhaps this divergence of the results from our hypothesis is best explained by the epistemic gap.

The Epistemic Gap refers to a stage children undergo, around the age of three, in which they fail to produce and comprehend epistemic modals (Cournane, 2015). Interestingly, during this time children easily produce and comprehend lexical epistemics like maybe (Papafragou & Ozturk, 2007). It appears that this delay only applies to functional modals, suggesting grammatical development could be the key to understanding epistemics and not a conceptual delay. Cournane (2015) argues that the functional modals differ in syntactic complexity. Modals have an event variable (e) which must be locally bound at LF. Must in the low position has an event variable bound by Aspect, and is deontic. In the high position, the event variable of must is bound by the speech act event, and is epistemic, linked to the speaker’s knowledge. In root interpretations the modal scopes over a predicate, but epistemic interpretations require that the modal scope over a proposition. For this reason, Cournane predicts, and finds in spontaneous data, that mastering epistemic modals requires the ability to embed propositions.

Two possible explanations for our data arise, considering this account of the Epistemic Gap. In the first, the epistemic modal is misanalyzed as a root modal,
in the lower position, and hence does not scope over the proposition. Two problems occur with this account. One is that the children seem to understand the epistemic meaning, so we would have to argue that it takes on that meaning by analogy to a lexical epistemic, such as probably, given the data from Cournane (2015). But even lexical epistemics must eventually be acknowledged as having a speaker’s perspective. Cournane also finds that children produce epistemic modals like must at a relatively young age, once they have sentential embeddings. Given that the age of our participants is 3 to 5 years, this account in terms of a low analysis of epistemics contains a lot of assumptions for which we do not yet have evidence.

The second theory follows from the analysis of an epistemic modal containing a hidden embedding (Cournane, 2015). That is, the epistemic reading of 6 a. covertly manifests as 6 b. in Logical Form.

(6) a. Mom said that Dad must have bought apples
   b. Mom said (it must be) that Dad bought apples

Now in our questions, this implied embedding creates a three clause recursive structure as in 7 a.

(7) a. Why did Mom say (it must be) that Bella must be washing the dog?

In our recent work on three clause recursive wh-adjunct questions, we found that children were significantly less likely to give reality type answers for these complex structures than they were from tensed complements with single embedding. Surprisingly, children are biased to the correct long distance interpretation under these recursive questions (de Villiers, Kotfila, & Roeper, in press). We argue that the three clause structure removes ambiguity and thus triggers the cyclic movement of the wh-question. The natural possibility arises in our current study that children are succeeding with the epistemic modals not because they are analyzed low as root modals, but precisely because they are analyzed as requiring further embedding. Thus cyclic wh-movement is an immediately available option, and the premature misinterpretation of the wh question in the last clause is bypassed. Epistemic modal complements are thus also easier than tensed complements, but the explanation is different across the two types of modal.

It should be evident that much further careful work is required to arbitrate between these alternatives, as this is the only study to date on the acquisition of wh-questions from modal complements of distinct flavors.

References


