Question-Answer (In)Congruence in the Acquisition of *Only*

Martin Hackl, Ayaka Sugawara, and Ken Wexler

1. Introduction – Crain's puzzle

   Crain et al. (1994) reported a curious asymmetry in children's ability to interpret sentences with *only* in an adult-like manner; their participants, whose age range included 6 years of age, had a pronounced tendency to interpret sentences with pre-subject *only* (hereafter Subj-*only*) such as in (1a) in a non-adult-like way, but no such tendency was observed for sentences with VP-*only* such as (1b). Interestingly, the reasons children provided to justify the non-adult-like interpretations indicated that they assigned Subj-*only* sentences the same meaning that they would assign to the corresponding VP-*only* sentences.

(1) a.   Only the cat is holding a flag.
     b.   The cat is only holding a flag.

   Subsequent research has shown that this asymmetry between Subj-*only* and VP-*only* is robust across a number of dimensions: it is robust across truth-values meaning that children would accept Subj-*only* sentences when adults wouldn't and vice versa, it has been found across wide age range going from 2.6 yrs of age to 6+ yrs of age, and it has been replicated across variety of languages (cf. Philip & Lynch 2000, a.o. for English; Notley et al. 2009, Zhou & Crain 2009 a.o. for Mandarin; Müller et al. 2011 for German; Endo 2004, Sano 2011 for Japanese) leading researchers to conclude that VP-*only* is acquired before Subj-*only* and therefore might be more basic than Subj-*only*.
   In this paper we argue for a more nuanced picture. We first point out that in previous studies the test sentences with *only,* which were presented as answers to a question like *What happened?,* were not directly congruent to that question

---

and as such not completely felicitous. This raises the possibility that previous studies underestimated children's competence regarding *only*. Next, we show experimentally, using subject- and object-questions rather than *What happened?* as prompts*,* that children are very sensitive to question-answer congruence. More specifically, we show that when question-answer congruence is satisfied their understanding of *only* is essentially adult like for both Subj-*only* and VP-*only* and that when question-answer congruence is violated their response strategy is to ignore the syntactic position of *only* irrespective of whether it is Subj-*only* or VP-*only.* Instead they associate *only* with the constituent that corresponds to the *wh*-phrase in the question as question-answer congruence demands. Taken together these data argue that previous literature both underestimated and overestimated children's command of *only*. Finally, we argue that *only*-sentences that are intended to answer *What happened?* require accommodation of a suitable sub-question (e.g., a subject- or an object-question), and that Crain's puzzle, presented in the literature as a difference between Subj-*only* and VP-*only,* can be insightfully restated as a difference in how easy it is to accommodate a subject- or object-question.

2. Question-Answer Congruence (QAC)

It is well-known that answers to *wh*-questions are subject to a principle known as Question-Answer Congruence (QAC) which requires, roughly speaking, that the constituent in the answer that corresponds to the *wh*-phrase in the question has to be focused, (Paul 1880, Rooth 1985, 1992, etc.) This is illustrated in (2), where we see that the answer in A1 to the question in (2a) satisfies QAC and thus is well-formed (**congruent**) because (2a) is a subject-question and the subject DP in A1 is indeed focused.[1] By contrast, A2 does not satisfy QAC and so is **incongruent** since it is the object DP that is focused in A2. Exactly the opposite is the case in (2b).[2]

(2) a. Q: Who is holding a flag?
    A1: THE CAT$_F$ is holding a flag.
    A2: *The cat is holding A FLAG$_F$.
  b. Q: What is the cat holding?
    A1: *THE CAT$_F$ is holding a flag.
    A2: The cat is holding A FLAG$_F$.

Let us, next, turn to answers that include the focus-sensitive operator *only,*

---

[1] In the examples in (2) and hereafter, we use F-subscripts to indicate the presence of focus marking and SMALL CAPS to represent prosodic prominence, which typically constitutes the phonetic exponent of focus-marking.

[2] We restrict our discussion to sentences with one focus which are intended as answers to a single *wh*-question. More complex cases involving answers to multiple *wh*-questions, multiple foci, or Second Occurrence of Focus are beyond the scope of this paper.

which requires an F-marked constituent in the sentence it occurs. The position of the focus associate of *only* is subject to a syntactic constraint whose effect can be summarized in a purely descriptive form as in (3) and exemplified in (4).[3]

(3) **Focus Co-occurrence Constraint** for *only* (FCC-*only*)
    A.   When *only* occurs to the immediate left of the subject its associated F has to be on or inside the subject.
    B.   When *only* occurs immediately to the left of the VP its associated F has to be on or inside the VP.

(4) a.   Only THE CAT$_F$ is holding a flag.
    b. *THE CAT$_F$ is only holding a flag.
    c. *Only the cat is holding A FLAG$_F$.
    d. The cat is only holding A FLAG$_F$.

Given these facts, we can now ask how association with focus for *only* interacts with QAC. We see, as illustrated in (5), that in order to satisfy QAC, the associate of *only* has to be the constituent that corresponds to the *wh*-phrase in the preceding question: the pairs of Q-A1 in (5a) and Q-A2 in (5b) are congruent because focus marking as dictated by QAC is on the associate of *only*, while the pairs of Q-A2 in (5a) and Q-A1 in (5b) are incongruent since the focus associate of *only* does not correspond to the *wh*-phrase in the question.

(5) a. Q: Who is holding a flag?
       A1: Only THE CAT$_F$ is holding a flag.
       A2: *The cat is only holding A FLAG$_F$.
    b. Q: What is the cat holding?
       A1: *Only THE CAT$_F$ is holding a flag.
       A2: The cat is only holding A FLAG$_F$.

Finally, let's consider cases where a sentence with *only* is intended as answer to a question like *What happened?,* i.e. the type of question that is typically used in acquisition experiments that employ the TVJ tasks to prompt the target sentence (which is normally uttered by a puppet), (6).[4]

(6) (Scene at the end of the story shows a cat holding a flag, a goose holding a flag and a balloon, and a frog holding a balloon.)

---

[3] (3) is canonically thought to be a consequence of *only* being a scope taking operator. E.g. on the assumption that Subj-*only* forms a DP constituent with the subject (similar to determiners) while VP-*only* is adverbial in nature we can state the relevant principle in terms of scope: the associate of *only* needs to be in the surface c-command domain of *only*. See Jackendoff (1972) and much subsequent work for discussion.
[4] We are aware of only one exception to this in the literature on the acquisition regarding *only* – Experiment 3 of Notley et al. (2009), which we discuss in section 5.

Q: Kermit, can you tell us what is happening in the picture?
A1: [(?)]Only THE CAT$_F$ is holding a flag.
A2: [(?)]The cat is only holding A FLAG$_F$.

Note that the Q-A pairs in (6) do not satisfy QAC as we have characterized it above since the constituent in the answer that corresponds to the *wh*-phrase here would have to be clausal in nature and express a proposition. However, neither *the cat* in A1 nor *a flag* (or *holding a flag* assuming focus projection) in A2 are constituents of the sort that could provide the information sought in the question. QAC is therefore not satisfied and we would expect these dialogues to be ill formed and thus infelicitous. Interestingly, they are not judged to be infelicitous, although native speakers typically do find them a bit less crisp than fully congruent cases like (5a)-A1 or (5b)-A2 and somehow more difficult to process (See Appendix).

Following Roberts (1996/2012), Büring (2003), Beaver & Clark (2008) etc., we can characterize what might be going on in (6) as a case of question accommodation triggered by a need to avoid a violation of QAC. More concretely, we suggest that questions can be broken down into sub-questions and that QAC can be satisfied **indirectly** if the answer is congruent with a contextually relevant sub-question that is accommodated by the comprehender. For instance, given the scene in (6), contextually relevant sub-questions of Q might be *Who is holding a flag?, Who is holding a balloon?, What is the cat holding?, What is the goose holding?,* etc. Answers to these sub-questions are relevant because they provide part of the information sought by *What is happening in the picture?* and a speaker answering with either A1 or A2 may be judged to be cooperative and their answers judged felicitous if the comprehender accommodates a suitable sub-question – *Who is holding a flag?* for A1 and *What is the cat holding?* for A2.

Our hypothesis about the mechanics of comprehending the answers in (6) relative to a question like *What is happening in the picture?* raises the possibility that previous studies have underestimated children's competence regarding *only*. It could be, after all, that children have difficulty accommodating a suitable sub-question and that this contributes to Crain's puzzle. Of course, showing that comprehending (6)-A1 and (6)-A2 is more complex than meets the eye and that children have difficulty with this complexity is not yet an explanation of Crain's puzzle. It might, however, suggest that Crain's puzzle may be more a puzzle about question accommodation than a puzzle about Subj-*only* and VP-*only* per se. In other words, such a discovery would suggest that Crain's puzzle should be restated as a question why, in situations like those used in Crain et al.'s experiments, object-questions are easier to accommodate than subject-questions and why children's grammar allows them to associate Subj-*only* with the object DP (disregarding the FCC-*only*) as would be appropriate for obeying QAC with an accommodated object-question.

The goal of our paper is to evaluate this idea. To do this, we use directly congruent question-answer pairs such as (5a)-A1 and (5b)-A2 as well as

irrecoverably incongruent question-answer pairs like (5a)-A2 and (5b)-A1 and ask whether the rate of adult-like responses increases for the former and decreases for the latter across the board for both Subj-*only* and VP-*only*. We also analyze children's response strategies for irrecoverably incongruent pairs to assess whether they rely more on QAC as guide for determining the associate of *only* or on the syntactic position of *only*.

3. Experiment 1 – Baseline
3.1. Methods and materials

Before we embark on studying children's sensitivity to QAC as outlined above we want to establish that we can replicate Crain et al.'s original observation with our own methods and materials. Experiment 1 serves this purpose. Children were told simple stories supported by PowerPoint slides which showed simple animations of animal characters getting food items. The experimenter, who was a native speaker of English, narrated the story and advanced the slides to show what was happening in the story. At the end of the story, the experimenter asked the puppet *Can you tell me what happened?* The puppet responded with a target sentence (either with Subj-*only* or VP-*only*) or a filler sentence that did not contain *only*. Child participants were then asked to judge if the puppet was right or wrong, and if they thought he was wrong, they were asked to give a justification. See (7) for an example.

(7) Experimenter: In this story we have a goose, a cat, and a frog. <click> Oh, the goose gets ice cream and cheese. <click> Oh, the cat gets ice cream. <click> Oh, the frog gets cheese and juice. Now Kermit, can you tell me what happened?
Kermit: Oh, I know what happened. (Only) the frog (only) got juice.

While being asked for their judgments and justifications, the child could see the completed events as a still image on the computer screen, which would be comparable to the picture verification task that Crain et al. and others have conducted. Children were encouraged and praised equally for *Yes* and *No* answers but encouraged to provide justifications only in case of the latter.

The experiment consisted of 8 target trials (4 with Subj-*only* and 4 with VP-*only*) as well as 4 filler trials within one session. The presentation order was pseudo-randomized, and we had two different sets of orders randomly assigned to each participant. Half of the items in each condition were designed to be true on an adult interpretation relative to the story and the other half to be false. Importantly, stories were designed so that if a target sentence in a given condition, say Subj-*only*, was true on the adult Subj-*only* interpretation it would be false on a non-adult, VP-*only* interpretation and vice versa. Test sessions took about ten minutes per participant and were conducted at local daycares in a relatively quiet room or space in a classroom, and at the Boston Children's Museum in a quiet room with caretakers present throughout the test session.

3.2 Results

40 English-speaking children from Boston area daycares and the Boston Children's Museum across all socioeconomic and ethnic backgrounds were recruited for this experiment. Children who incorrectly answered 2 or more filler items were excluded (N=2). Data from 38 participants (4;0 – 6;9, M=5;2) were included in the analysis. Figure 1 summarizes the results of Experiment 1 plotting adult-like responses for Subj-*only* and VP-*only*. We observed a significant effect of Condition (Subj-*only* vs. VP-*only*) on accuracy rates with the rate of adult-like responses being significantly lower for Subj-*only* (29.6%) than for VP-*only* (79.6%). [5] This replicates Crain et al.'s original discovery and means that we can confidently use our methods and materials to study children's sensitivity to QAC.
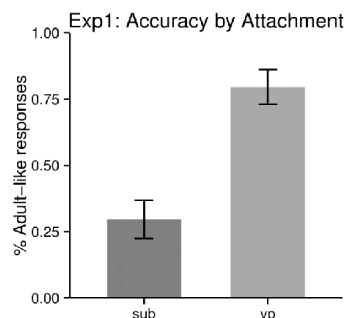

Figure 1: Results of Experiment 1

4. Experiment 2A & 2B
4.1 Design and predictions

Experiments 2A and 2B investigate children's sensitivity to QAC. We used the same target sentences as in Experiment 1 but changed the question prompting the puppet's answer. Specifically, we used subject- and object-questions in place of *What happened?* and crossed question type and attachment site of *only* (Subj-*only* vs. VP-*only*). This gives rise to two congruent and two incongruent conditions as illustrated in Table 1.

Table 1: 2x2 design for Experiment 2

	Subject-question	Object-question
Subj-*only*	*Congruent* Kermit, can you tell me who got juice? Only the frog got juice.	*Incongruent* Kermit, can you tell me what the frog got? Only the frog got juice.
VP-*only*	*Incongruent* Kermit, can you tell me who got juice? The frog only got juice.	*Congruent* Kermit, can you tell me what the frog got? The frog only got juice.

---

[5] Statistical analysis of the response rates was conducted in R using linear mixed effects model with logistic regression. Since a maximally specified model did not converge, the order of presentation was investigated only for a potential main effect, which was not detected (p = .93).

Truth was again defined in terms of adult responses to the target sentences, and counter-balanced across the attachment site of *only* in the same way it was in Experiment 1. The example in (8) is an instance of an incongruent trial which is coded as false because the adult response to the target sentence relative to the situation would be false.

(8) Experimenter: In this story we have a goose, a cat, and a frog. <click> Oh, the goose gets ice cream and cheese. <click> Oh, the cat gets ice cream. <click> Oh, the frog gets cheese and juice. Now Kermit, can you tell me who got juice?
Kermit: Oh, I know who got juice. The frog only got juice.

If children are sensitive to QAC and the lack of congruence in Experiment 1 contributed to the results because accommodation of a suitable sub-question is difficult, we expect children to exhibit higher accuracy rates on the congruent conditions and lower accuracy rates for the incongruent conditions for both types of *only*. If, on the other hand, children are insensitive to QAC and Crain's puzzle is indeed a simply product of Subj-*only* being acquired later than VP-*only*, our manipulation should not qualitatively change the accuracy rates we have seen in Experiment 1.

Our incongruent conditions deserve a closer look. If children are sensitive to QAC, they are confronted with an ungrammatical question-answer pair and might, in response, adopt a variety of strategies which could be revealing in their own right: (i) they might simply guess if they cannot makes sense of the items at all; (ii) they might ignore *only;* (iii) they might identify the associate of *only* according to QAC and so, in effect, interpret the target sentence as if Subj-*only* were VP-*only* and vice versa; or (iv) they might ignore the question and not even try to accommodate an appropriate alternative question and therefore interpret the target sentence according to the syntactic position of *only* (possibly with the same bias towards a VP-*only* interpretation as observed in Experiment 1). If they adopt strategy (i), we expect the accuracy rates for the incongruent conditions to be around 50%.[6] If they adopt strategy (ii) we also expect 50% accuracy – not because they use "yes" and "no" randomly as in (i) but because they would say "yes" indiscriminately (and, given the design of our items, half of those would be accidentally correct). If their responses are guided by a desire to respect QAC at the expense of the FCC-*only* – i.e. if the QAC cue for the location of F in the answer is decisive – we expect them to interpret the target sentences as if the position of *only* was swapped. I.e., sentences with Subj-*only* should be interpreted as adults interpret sentences with VP-*only* but also sentences with

---

[6] One might suspect that participants' preferred strategy would be to reject the sentence plainly out of confusion (with encountering an incongruent Q-A pair). Given that in our design truth was counter-balanced (i.e., half of the trials are coded as true and the other half as false), we would expect the "chance-level" performance with that strategy as well.

VP-*only* should be interpreted as adults interpret sentences with Subj-*only*.[7] As a result, we expect the accuracy rates for the incongruent items to be close to 0%. Finally, if children adopt strategy (iv) and ignore the question and follow, instead, the FCC-*only* – i.e. they take the position of *only* in the target sentence as decisive for determining the location of F – we expect them to interpret the target sentences as adults do. This should produce accuracy rates close to 100% for the incongruent items.

These predicted accuracy rates are, of course, idealized. It could very well be, for instance, that children's response pattern is the product of a mix of these strategies. Since we are particularly interested in the last two strategies, we would like to find a way of assessing whether children pay more attention to QAC or to the syntactic position of *only* as they determine the associate of *only*. In order to do that we implemented the basic design in Table 1 in two ways. In Experiment 2A the attachment site of *only* was a within-subjects factor while the question-type was a between-subjects factor. I.e. each participant heard 4 subj-*only* sentences, 4 VP-*only* sentences as well as 4 fillers and, depending on the question-type they were randomly assigned, all of the sentences were either preceded by a subject-question or by an object-question. In Experiment 2B, by contrast, the attachment site of *only* was a between-subjects factor and the question type a within-subjects factor. I.e. participants either heard 8 subj-*only* sentences preceded by 4 *who*-questions and 4 *what*-questions as well as 4 fillers, or they heard 8 VP-*only* sentences preceded by 4 *who*-questions and 4 *what*-questions as well as 4 fillers. Thus the constant cue for focus comes from the attachment site of *only* in Experiment 2B. This is summarized in Table 2.

Table 2: Design for Experiment 2A and 2B

	Subj-Q	Obj-Q		Subj-Q	Obj-Q
Subj-only	Congruent	Incongruent	**Subj-only**	Congruent	Incongruent
VP-only	Incongruent	Congruent	**VP-only**	Incongruent	Congruent
	Exp. 2A			Exp. 2B	

The idea behind this manipulation is that the levels of the between factors will stay constant for participants and so possibly provide a stronger cue for the location of F in the *only* sentences than the levels of the within factors, which

---

[7] There are at least two ways in which such a strategy could be implemented. One is for the parser to literally relocate *only* from its pre-subject position to the VP-adjoined position and vice versa. Another possibility is to assume that the parser locates F, which identifies the associate of *only,* in accordance with QAC, disregarding the FCC-*only*. Choosing between these two options is not crucial for the present paper.

vary for each participant. A comparison of Experiment 2A and 2B might then provide us with an indication whether the factors are equally important.

To see this more concretely, consider what would be expected if the two types of cues are equally strong and, in fact, decisive for both sub-experiments. For Experiment 2A it would mean that participants would behave according to strategy (iii) – incongruent items would yield close to 0% accuracy. For Experiment 2B, by contrast, it would mean that participants would behave according to strategy (iv) and we expect close to 100% accuracy rates for incongruent items, Figure 2.[8] If, on the other hand, the two factors are not equally strong and children pay more attention to one of the two, we would not see a perfect flip in accuracy. Instead, if QAC provides a stronger cue than the attachment site of *only* the results of Experiment 2A should mimic its idealized predictions in Figure 2 more closely than the results of Experiment 2B and if the attachment site of *only* provides a strong cue the opposite should be true.
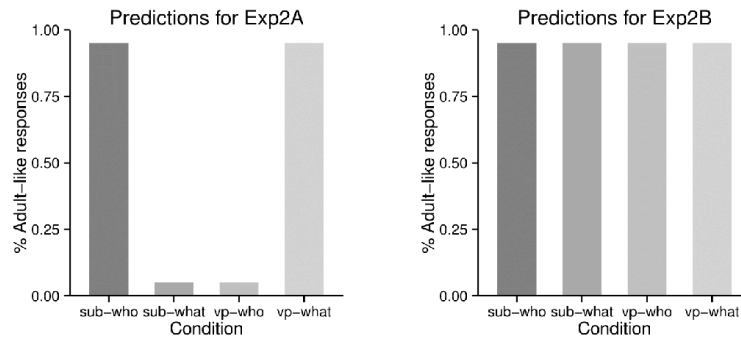

Figure 2: Predictions for Experiment 2A and 2B

4.2 Methods

The presentation of the task and the testing environment were parallel to the ones in Experiment 1; the experimenter uses PowerPoint slides as narrating which animal gets what food item(s), and asks Kermit a question with either of the subject- or the object-question, and then Kermit utters the target sentence, for the participants to judge. We also asked for justification for negative answers. Materials were prepared based on the ones in Experiment 1 with some addition of characters that are popular among children. There were 8 target trials (4 congruent pairs and 4 incongruent ones; see Table 2) as well as 4 filler trials without *only* within one session. The truth was counter-balanced as discussed in 4.1. The presentation order was pseudo-randomized, and we had two different

---

[8] If the two factors are equally strong but not decisive we would expect the increase in accuracy rates for incongruent items in Experiment 2A to be as large as the decrease in accuracy for incongruent items in Experiment 2B.

sets of orders randomly assigned to each participant. The whole session took about ten minutes per participant.

4.3 Results of Experiment 2A and 2B

53 English-speaking children from Boston area daycares and the Boston Children's Museum were recruited for Experiment 2A & 2B combined. None of them had participated in Experiment 1. Children who incorrectly answered 2 or more filler items were excluded (N=5). 48 children (4;0 – 6;11, M=5;2) are included in the analysis (24 on Experiment 2A and 24 on Experiment 2B).

Figure 3 summarizes the results of Experiment 2A, in which question type was the between-subjects factor. 12 children (M=5;1) were tested in the subject-question condition and 12 different children (M=5;0) were tested in the object-question condition. The former group exhibits accuracy rates for Subj-*only* sentences of 72.9% and for VP-*only* sentences of 31.3%. By contrast, the latter group exhibits accuracy rates for Subj-*only* sentences of 6.3% and for VP-*only* sentences of 95.8%. Statistical analysis of these results (using maxiamally specified logit-LMEM) reveals main effects of attachment site (p=.009) and of the question-type (p=.001), and importantly, an interaction (p<.001) indicating that congruent items have higher accuracy rates than incongruent ones irrespective of question-type and attachment site of *only*.[9]
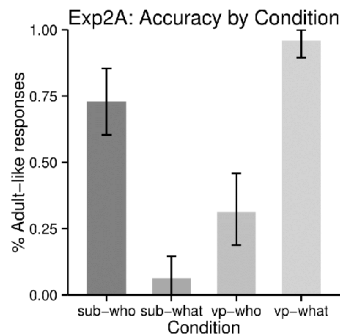

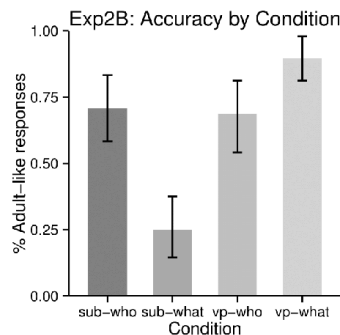**Figure 3: Results of Experiment 2A**   **Figure 4: Results of Experiment 2B**

Figure 4 summarizes the results for Experiment 2B, in which attachment of *only* was the between-subjects factor. 12 children (M=5;3) were tested in the Sub-*only* condition and 12 different children (M=5;3) were tested in the VP-*only* condition. The Subj-*only* group exhibits accuracy rates for sentences preceded by a subject-question of 70.8% and for sentences preceded by an object-question of 25%. By contrast, the VP-*only* group exhibits accuracy rates for sentences preceded by a subject-question of 68.8% and for sentences preceded by an

---

[9] The order of presentation was investigated only for a potential main effect, which was not detected (p=.19).

object-question of 89.6%. Statistical analysis of these results (using maxiamally specified logit-LMEM) reveals main effects of attachment site (p=.001) and of the question-type (p=.005), and, again, an interaction (p<.001) indicating that congruent items have higher accuracy rates than incongruent ones irrespective of question-type and attachment site of *only*.

4.4 Discussion of Experiments 2A and 2B

The most important result of Experiments 2A and 2B is that congruent answers are understood significantly more often in an adult-like way than incongruent answers irrespective of question-type and attachment site of *only*. This shows, first of all, that our participants were sensitive to QAC, an observation that, to our knowledge has not been made previously in the acquisition literature. Our data, in particular the fact that the accuracy rates for congruent Subj-*only* answers are above 70%, also provide strong evidence that children's difficulty with Subj-*only* in Experiment 1 should not be taken to mean that they have not acquired Subj-*only* at all. Clearly, such a stance is incompatible with above chance behavior when QAC is (directly) satisfied.

By parallel reasoning, the fact that children's behavior on VP-*only* in Experiment 1 is essentially adult-like can no longer be taken to show that their command of VP-*only* is truly adult-like. Such a stance is incompatible with our observations from the incongruent pairs in Experiment 2A. We saw there that children's behavior was systematic rather than arbitrary but also clearly non-adult-like for both Subj-*only* and VP-*only*. Specifically, our participants interpreted incongruent Subj-*only* sentences to mean what the corresponding VP-*only* sentences mean at a rate of 94% and they interpreted incongruent VP-*only* sentences to mean what the corresponding Subj-*only* sentences mean at a rate of 69%. This result conforms pretty well to the predictions we derived from the assumption that children pay attention to the QAC cue at the expense of violating the FCC-*only*. In other words, children seem to ignore the position of *only* in both Subj-*only* as well as VP-*only* sentences when they search for an associate and so interpret the incongruent sentences as if the position of *only* was swapped. This entails, of course, that children's command of VP-*only* is not fully adult-like either.[10]

Finally, note that the results of Experiment 2B seem to conform less well with our idealized predictions. Specifically, although we do see higher accuracy rates for incongruent items in Experiment 2B than in Experiment 2A, they are strikingly asymmetric; Subj-*only* generates fewer adult-like interpretations than VP-*only*. This suggests that children pay less attention to the syntactic cue (the attachment site of *only*) when they have to identify the associate of *only*. Instead,

---

[10] In related work in progress we show that adults' response strategy to incongruent question-answer pairs is to ignore QAC and instead identify the associate of *only* according to the FCC-*only* (Sugawara, in progress; see Appendix for a summary).

they seem to adopt a similar strategy to the one they use in Experiment 1 when the QAC cue varies within participants.

5. General discussion and conclusion

Previous literature on the acquisition of *only* observed an asymmetry between Subj-*only* and VP-*only* (Crain's puzzle) and concluded that Sub-*only* is more difficult for children to acquire than VP-*only*. However, we argued that the experimental strategy employed in that literature overlooked a complication in their materials that might have lead to underestimating children's competence of *only*. Specifically*,* we argued that comprehending *only*-sentences which are intended to be answers to *What happened?* requires accommodation of a sub-question to satisfy QAC and that difficulty with question-accommodation could be a contributing factor to the results reported in the literature. To investigate this possibility we used subject- and object-questions as prompts for Subj- and VP-*only* sentences in a fully crossed design.

For congruent pairs we saw that children's interpretation was close to adult-like for both types of *only*. This shows that children do not have difficulty understanding Subj-*only* in an adult-like way across the board. Rather, it is only in certain circumstances, e.g. when they have to accommodate a suitable question to satisfy QAC, that they seem to misanalyze Subj-*only* for VP-*only*. We are aware of one other experiment in the literature, Experiment 3 of Notley et al. (2009), that yielded similar results to ours. Notley et al. reported that four-year old Mandarin-speaking children tested on the subject focus construction (*shi…de*), which is claimed to be similar to *zhiyou* (*only*), exhibited adult-like responses in "subject-biased stories", which, interestingly, ended with a subject-question prompt rather than a broad question like *What happened?*. Even though Notely et al.'s design was not intended to test QAC, we think that their results provide support for our claim that sensitivity to QAC and the need for question-accommodation play an important role in Crain's original observation.

Our data from incongruent pairs are equally significant and informative wrt. the proper characterization of Crain's puzzle. We saw that children's primary strategy for determining the associate of *only* in these cases was based on QAC since the associate of *only* was identified with the constituent that corresponds to the *wh*-phrase in the preceding question even when the FCC-*only* would not permit such association. This reinforces our observation that children are quite sensitive to QAC, more so, in fact, than they are to the syntactic position of *only*. Importantly, this was true for both Subj-*only* and VP-*only* at least when the QAC cue was sufficiently strong (Experiment 2A). This shows that children's command of *only* is in fact not fully adult-like for both types and not just for Subj-*only* as was previously believed.

Taken together, our data suggest that Crain's discovery can be restated as follows: What distinguishes children from adults in their command of *only* is the status of the FCC-*only*. For adults, it is an inviolable constraint while for children it seems quite possible to ignore it when other factors such as QAC are

in conflict with it. Moreover, Crain's original observation of an asymmetry between Subj-*only* and VP-*only* can now be stated as an asymmetry between accommodating two types of sub-questions. In other words, our results pose new research questions: What makes object-questions easier to accommodate than subject-questions in contexts such as those canonically used to test *only* and, more generally, what factors govern the ease with which sub-questions are accommodated?

Before closing, we would like to point out that an asymmetry in question-accommodation is likely not the whole story. One indication that there might be additional factors at play comes from our directly congruent question-answer pairs where we see that congruent VP-*only* is responded to at the higher adult-like rate than congruent Subj-*only*. This may be a residue of the same asymmetry that Crain et al. initially identified or it may be due to subject-questions being less well supported in the context and thus less well attended to/ more easily ignored by children than object-questions. Future research will have to clarify this issue.[11]

Appendix – Adult's reaction to incongruent question-answer pairs

In related work using a timed inference task we investigate how adults process *only* in the environments we discussed here. Experiment 1 in Hackl et al. (to appear) is comparable to our Experiment 1 here and shows that adults exhibit a parallel processing asymmetry between Subj-*only* and VP-*only* in case *What happened?* was used as a prompt. Specifically, response time for VP-*only* sentences was significantly faster than for Subj-*only* sentences (no difference in accuracy was observed here). Furthermore, parallel experiments to our Experiment 2A (Sugawara, in progress) reveal that the accuracy rate by adult participants on congruent Subj-*only* sentences was 94%, on incongruent VP-*only* sentences 66%, on incongruent Subj-*only* 73%, and on congruent VP-*only* 90%. These results suggest that adults also have difficulty comprehending incongruent *only*-sentences, but that the strategy they primarily adopt is to ignore the question and take the position of *only* to be the decisive factor for identifying the associate of *only* (strategy (iv)). This is different from children's primary strategy with incongruent items, which is to respect QAC and disregard the FCC-*only* (strategy (iii)). Moreover, a closer look at the adult data for incongruent items, comparing response times of correct answers (produced by following strategy (iv)) with those for incorrect answers (produced by following strategy (iii)) reveals that correct answers took significantly longer than incorrect answers. This suggests that adopting strategy (iv) might be costlier than strategy (iii) and that children adopt the less costly strategy with incongruent items.

---

[11] See Hackl et al. (to appear) for evidence that the scalar presupposition of *only* is an important factor in determining the difficulty of comprehending a sentence with *only*.

References

Beaver, David and Brady Clark (2008) *Sense and sensitivity: How focus determines meaning*, Blackwell, Oxford.

Büring, Daniel (2003) "On D-trees, beans, and B-accents," *Linguistics and Philosophy* 26, 511-545.

Crain, Stephen, Weijia Ni, and Laura Conway (1994) "Learning, parsing, and modularity," in *Perspectives on sentence processing*, ed. by Clifton, Charles, Jr., Lyn Frazier, and Keith Rayner, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 443-467.

Endo, Mika. (2004) "Developmental issues on the interpretation of focus particles by Japanese children," in *BUCLD 28 Proceedings*, Cascadilla Press, Somerville, MA, 141-152.

Hackl, Martin, Erin Olson, and Ayaka Sugawara (to appear) "Processing *only*: Scalar presupposition and the structure of ALT(S)," in *Proceedings of Sinn und Bedeutung 19 (SuB 19)*.

Jackendoff, Ray (1972) *Semantic interpretation in generative grammar*, MIT Press, Cambridge, MA.

Müller, Anja, Petra Schulz, and Barbara Höhle (2011) "How the understanding of focus particles develops: Evidence from child German," in *Proceedings of the 4th conference on Generative Approaches to Language Acquisition North America (GALANA 2010)*, Cascadilla Proceedings Project, Somerville, MA, 163-171.

Notley, Anna, Peng Zhou, Stephen Crain, and Rosalind Thornton (2009) "Children's interpretation of focus expressions in English and Mandarin," *Language Acquisition* 16:4, 240-282.

Paul, Hermann (1880) *Prinzipien der Sprachgeschichte* (8th edition, 1970, Niemeyer, Tübingen)

Philip, William, and Emily Lynch (2000) "Felicity, relevance, and acquisition of the grammar of every and only," in *BUCLD 24 Proceedings*, Cascadilla Press, Somerville, MA, 583-596.

Roberts, Craige (1996/2012) "Information structure in discourse: Towards an integrated formal theory of pragmatics," *Semantics and Pragmatics* Volume 5, 1-69.

Rooth, Mats (1985) *Association with focus*, PhD Dissertation, University of Massachusetts, Amherst.

Rooth, Mats (1992) "A theory of focus interpretation," *Natural Language Semantics*, 75-116.

Sano, Tetsuya (2011) "Abstract CP/IP configuration in child Japanese," in *BUCLD 36 Proceedings*, Cascadilla Press, Somerville, MA, 525-535.

Zhou, Peng, Stephen Crain (2009) "Focus in child language: Evidence from the acquisition of Chinese," in *Proceedings of the 3rd conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*, Cascadilla Proceedings Project, Somerville, MA, 336-346.