The role of Question-Answer Congruence (QAC) in child language and adult sentence processing

by

Ayaka Sugawara

B.A., University of Tokyo (2008)
M.A., University of Tokyo (2010)

Submitted to the Department of Linguistics and Philosophy
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

September 2016

Signature redacted

Signature of author: .....
Department of Linguistics and Philosophy
September 9th, 2016

Signature redacted

Certified by: .....
Martin Hackl
Associate Professor of Linguistics
Thesis Supervisor

Signature redacted

Certified by: .....
Ken Wexler
Professor Emeritus of Psychology and Linguistics
Thesis Supervisor

Signature redacted

Accepted by: .....
David Pesetsky
Professor of Linguistics
Head, Department of Linguistics and Philosophy

The role of Question-Answer Congruence (QAC)
in child language and adult sentence processing

by

Ayaka Sugawara

Submitted to the Department of Linguistics and Philosophy
on September 9th, 2016 in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

This dissertation investigates experimentally the role of Question-Answer Congruence (QAC, von Stechow 1990, Rooth 1985, 1992) in child language acquisition and adult sentence processing. Specifically, I present two case studies: sentences with "*only*" and sentences with the Rise-Fall-Rise contour (RFR, Jackendoff 1972).

Case study 1 investigates a long-standing puzzle concerning the acquisition of "*only*." Since Crain et al. (1994), many studies report children display difficulties understanding sentences with pre-subject "*only*" while having no difficulty with pre-verbal "*only*." The current study manipulates the levels of QAC to study the puzzle. The results show that QAC plays a significant role in the asymmetry between Subject-*only* vs. VP-*only*, and more generally, both children and adults are sensitive to QAC, with congruence facilitating processing, and incongruence interfering with processing. We conclude that the cost of the accommodation of a sub-question to obey QAC is different for Subject-*only* and VP-*only*. I propose that the way information is packaged makes it easier to accommodate a set of questions asking about the object of the sentence, which makes it easier to process VP-*only*.

We also showed that the extent to which QAC influences comprehension is different for children and adults. For children, QAC has a stronger influence than the syntactic condition on *only*. For adults, the syntactic condition is not violated. We suggest that the status of the syntactic condition regarding *only* is a factor that distinguishes children from adults in processing *only*.

Case study 2 investigates whether children can construe the inverse scope interpretation, which children are reported to disfavor (Musolino 1998), with the RFR prosody, which is standardly taken to require inverse scope. The results show that both children and adults are sensitive to RFR and able to compute the implicature associated with RFR to disambiguate the inverse scope interpretation.

As in the "*only*" study, QAC levels varied in the RFR experiment. However, the results revealed no effect of question type. This leaves an important open question: why QAC seems more active in children in determining the associate of *only* than in the determination of quantifier scope when the prosody makes the scope clear.

Thesis Supervisor: Martin Hackl
Title: Associate Professor of Linguistics

Thesis Supervisor: Ken Wexler
Title: Professor Emeritus of Psychology and Linguistics

Acknowledgments

First and foremost, I owe countless thanks to the members of my dissertation committee: Martin Hackl, Ken Wexler, and Danny Fox. Their unwavering help, dedication, detailed and thoughtful comments, sharing insightful ideas, and encouragements were extraordinary supportive. It goes without saying that this dissertation would not have been possible without them.

Martin has been my advisor since I attended MIT. He was always patient, calm, supportive, generous with time, and always keen in detecting a crucial point of view with my research. He kept pushing me to understand materials better and not to jump to a conclusion, when I was sloppy. He listened to and encouraged me when I was frustrated or needed a boost of motivation. He was a great mentor throughout the years and I enjoyed our conversations on linguistics and life. The ideas and principles in this dissertation I owe to him; especially related to the various projects on "*only*", which he initiated, and I hope we can continue our lengthy discussions about this topic moving into the future. I am extremely privileged to work with and be advised by someone I truly trust.

Ken has always been like the sun in my research universe. He is irreplaceable; especially with his enthusiasm and encouragement. He is so eager to learn new ideas and implements them tirelessly. This goes hand in hand with his amazing ability to remember so well. His insatiable desire for more knowledge humbles my efforts. He is also very good at making me realize the big picture and how independent-looking phenomena could be connected. His attitude towards research is something I would always strive for. I am honored that I had many opportunities to work with him.

Danny was so inspiring as a teacher and mentor for me. I was always astonished by how quickly he sees through the core elements of issues in question. His advice has always helped me with solidifying ideas, which otherwise would have been vague, and also prevented me from making embarrassing mistakes. Due to my initial lack of organization, he had to familiarize himself with the contents of my dissertation in a short period time but nevertheless helped me tremendously; I am indebted to him for his detailed and crucial comments that were so helpful in improving the contents.

In my first year, Martin and Danny co-taught the Introduction to Semantics class, where I for the first time encountered the world of formal semantics. I was scared of semantics before entering MIT and more terrified after the first few classes... I didn't know what I was doing with those lambdas! At that point I was devastated, but our TA, Sam Alxatib, and the only semanticist in our year, Wataru Uegaki, helped me overcome my fears. I cannot thank them enough, for their time and putting up with my bizarre questions. Martin and Danny were also the victims of those types of questions, but I hope they didn't suffer as much. It was an exciting moment when the course materials started becoming clear a few weeks later. I am grateful to Martin, Danny, Sam, and Wataru for their support, patience, and encouragements, which really helped me succeed.

I was very fortunate to work with Ken Wexler and Irene Heim for Independent Study the following spring semester. Discussions with Ken triggered me to become very interested in the study of language acquisition. During another meeting, Irene referred me to Noah Constant's work on Rise-Fall-Rise, which developed into some results reported in this thesis. I learned a lot from a number of meetings with Ken and Irene. The experience I obtained that first year meant so much to me, and I cannot imagine finishing the program without those critical interactions and discussions.

It feels strange that I have not mentioned David Pesetsky up to this point. I joined the New Philharmonia Orchestra of Massachusetts, where David was second violin principal. Occasionally we played as stand partners, and I enjoyed every moment at New Phil with him, including viola jokes. David has been a great teacher and friend! I was very lucky to get to know such a positive, outgoing and fun person who I can talk with about everything linguistics, music, and playing the violin. David, thank you so much!

I would also like to thank some other MIT professors who gave me lots of advice, suggestions and encouragements, and made my days at MIT so valuable and enjoyable. Their insightful comments at meetings always made me realize new things, and I enjoyed the many brief chats with them in the elevator or kitchen area. I would like to thank Shigeru Miyagawa, Kai von Fintel, Irene Heim, Sabine Iatridou, Roger Schwarzschild, Donca Steriade, Michael Kenstowicz, Edward Flemming, Adam Albright, Michel DeGraff, and Norvin Richards. As for the post-doctoral teaching associates, I am especially grateful to Omer Preminger, Rick Nouwen, and Noah Constant.

My life at MIT was enriched by many wonderful fellow students and friends. First of all, many thanks go to my classmates of ling-10, which consisted of 10 people and 1 four-legged friend: Michelle Fullwood, Isaac Gould, Gretchen Kern, Ted Levin, Ryo Masuda, Sam Steddy, Wataru Uegaki, Coppe van Urk, Suyeon Yun, and Sally the dog. Their friendship and support made it possible for me to survive the five years at MIT, and keep my spirit for the year in Japan. I thank my cool officemates, some for the short period and others for the longer: Yasutada Sudo, Sasha Podobryaev, Igor Yanovich, Iain Giblin, Edwin Howard, Sam Steddy, Rafael Nonato, and for the time Rafa was away, Gary Thoms. I liked the office a lot, primarily because of you guys (another reason would be that we always kept a box of snacks). There was a lot of great times that we spent playing basketball, slacklining, hanging out for drinks and coffee, and/or grabbing ice cream with my close friends, Yasutada Sudo, Sam Alxatib, Sasha Podobryaev, Natasha Ivlieva, Liuda Nikolaeva, Rafael Nonato, YoungAh Do, Guillaume Thomas, Micha Breakstone, Hadas Kotek, Michael Yoshitaka Erlewine, Yusuke Imanishi, Junya Nomura, Iain Giblin, Michelle Fullwood, Isaac Gould, Suyeon Yun, Mia Nussbaum, Aron Hirsch, Erin Olson, and Athulya Aravind. I love you all! I would also like to thank my collaborators for the various kinds of joint projects over the years: Miwako Hisagi, Dimitrios Pantazis, Sachiko Kato, Masatoshi Koizumi, Daichi Yasunaga, Negin Ilkhanipour, Cory Bill, Jacopo Romoli, Yasutada Sudo, Chris O'Brien, Shintaro Hayashi, Satoshi Ito, Hisao Kurokami, and Koji Shimamura. Some of the research has not been published, but I am very grateful to them for broadening my horizons by teaching me new concepts and exchanging ideas.

As an acquisitionist, I cannot express my thanks enough to the people who took part in my experiments, whatever stage their commitments were. First of all, my research was absolutely not possible without the help of all the children who participated in the experiments, along with their supportive parents and teachers at daycares. I am very grateful for all their generosity. I would also like to thank the participants in the Amazon Mechanical Turk experiments, for their time and commitment. I also thank the undergraduate research assistants over the years for their efforts and help with creating experimental stimuli, contacting daycares, getting consent from parents, and conducting experiments. With that being said, there is a special thanks to Irina Onoprienko, Su Lin Blodgett, and Erika Trent, who engaged themselves in early-stage discussions of designing the experiments. I am indebted to Erin Olson for her excellent skills in programming and statistics, as well as in teaching me from scratch how to run statistics. I am

fortunate to be involved in establishing the MIT Language Acquisition Lab, and hope the lab will continue to thrive long into the future. Many thanks to our lab manager, Leo Rosenstein, for putting up with lots of irrational requests and for making the life of an experimentalist smooth.

I am grateful for the relationships I had with all the wonderful people outside MIT linguistics. I thank Chris Tancredi for continuously listening to my ideas and giving me advice on various projects I had. Many thanks to friends at Harvard linguistics: Yimei Xiang, Edwin Tsai, Louis Liu, Dorothy Ahn, Pooja Paul, Andreea Nicolae, Greg Scontras, Jacopo Romoli, and Sverre Stausland. I thank Utako Minai, Takuya Goro, Hajime Ono, Koji Sugisaki, Masaya Yoshida, Jeff Lidz, Kristen Syrett, Tom Roeper, Jeremy Hartman, Katalin É Kiss, Loes Koring, Stephen Crain, and many more for helpful and critical comments and suggestions at conferences and workshops. Meeting and caching up with Lyn Tieu, Emma Nguyen, Morgan Moyer, Jinsun Choe, Lilla Pintér, Susi Lima, Alex Nazarov, Lisa Bylinina, Naho Orita, Atsushi Oho, and Yu Tanaka at conferences has always energized me. I would also like to thank the members of the Japanese Association of MIT (JAM), Sokki Chen, Rafa's fun non-linguist friends, Livia Souza, Ning Tang, Pamela, Aidan, Masako, Yuika, Julie, Alistair, and Greg, for their warm friendship. And thank you, Elina Itelman, for being a fun and always encouraging roommate. I am deeply appreciative to the company of the New Philharmonia Orchestra members. Many players are professionals in their own field, and at the same time, all are very serious about the music. It was an incomparable memory to have played in New Phil with the extraordinarily talented people, especially in the last concert with our late maestro, Ron Knudsen.

I spent my last dissertating year teaching at Mie University in Japan. I am grateful to the colleagues and friends at Mie University, for solid discussions, as well as their support and friendship. I especially want to thank Koji Sugisaki, Seiki Ayano, Osamu Sawada, Shintaro Hayashi, Satoshi Ito, Sachiko Kiyama, Hisao Kurokami, Junya Nomura, Masanobu Sorida, Koji Shimamura, and Hideharu Tanaka. I also thank my students for being such a wonderful audience. My thanks extend to Katsuo Tamaoka, Ikumi Imani, Koji Kawahara, Chris Tancredi, Makoto Kanazawa, Kyoko Yamakoshi, and Shigeto Kawahara, for fruitful discussions and technical support on various projects.

I owe a deep gratitude to my teachers in Japan: Special thanks go to Yoshitaka Nishikawa, who introduced me to the exciting field of linguistics. I am deeply thankful to Akira Watanabe and Noriko Imanishi at the University of Tokyo (UT), for training me as a linguist and inspiring me to further explore the field. It was a great fortune to have done my bachelor's and master's degrees under Akira's supervision. I learned a lot from his careful and critical insights, and I always feel humbled by his vast knowledge. I would also like to thank my friends and teachers at UT and beyond, Sakumi Inokuma, Shun'ichiro Inada, Hiroyuki Shimada, Terue Nakato, Yuki Ishihara, Chizuru Nakao, Tomoe Arii, Mioko Miyama, Yuki Ito (whose memories are always with us), Hiromune Oda, Tetsuya Sano, Miwa Isobe, Reiko Okabe, Takuya Goro, Masaya Yoshida, Hajime Ono, Ken Hiraiwa, and Asako Uchibori. I am sure there are people I am forgetting, and I am sorry; but thank you to all!

Last but not least, I would like to thank my family, Michiko and Tadahiro Sugawara, Yukiko Tamaki, and Mio Sugawara, for all the opportunities I was fortunate to have, their endless support and love throughout my life.

Table of Contents

Abbreviations

CT	contrastive topic
df	degrees of freedom
D-tree	Discourse-tree
F, $_F$	focus, focus-marked
FCC-*only*	focus co-occurrence constraint of *only*
GEN	genitive marker
GLMEM	generalized linear mixed effects model
L1	first language
L2	second language
LF	logical form
LMEM	linear mixed effects model
M	mean age
ms	millisecond
n	number in sample
NEG	negation
NOM	nominative case marker
Obj-*only*	pre-object *only*
p	*p*-value
PAST	past tense
PRT	particle (specifically sentence-final particle)
QAC	question-answer congruence
QAR	question-answer requirement
QUD	question under discussion
r	observed *r* value, correlation
RFR	rise-fall-rise
RT	response time
Subj-*only*	pre-subject *only*
t	*t*-value
TOP	topic marker
TT	thematic topic
TVJT	truth-value judgment task
UG	Universal Grammar
VP-*only*	pre-verb phrase *only*
#;#	year;month when describing children's age (such as 5;2)

Chapter 1

Introduction and background

How are children different from adults in their linguistic ability and what are those differences? Is it grammar or some extra-grammatical factor that makes children appear to be different from adults? Furthermore, are there any aspects that are not different between children and adults? These are the questions that this thesis started with and have produced extensive research to answer these questions. The current research tries to give some answers to these questions by experimentally investigating two case studies with different populations in semantics and pragmatics.

The first case study is the comprehension of the sentences with the word *only*. As we will review in §1.1, it has been pointed out since Crain et al. (1994) that children are often found not adult-like in their understanding of sentences with *only* when it is attached to the subject of a sentence. Children interpret those sentences with *only* in the subject position (hereafter Subj-*only*) differently from adults, while exhibiting adult-like performance for sentences with *only* attached to the VP level (hereafter VP-*only*). This has led researchers to conclude that children acquire the grammar of Subj-*only* later than that of VP-*only*. We tested English-speaking children and adults, whose results led us to conclude that Question-Answer Congruence (QAC, Rooth 1985, 1992, etc.) plays an important role for focus association of *only*, both in child and adult grammar.

The second case study investigates the effect of prosody on scope relations between a universal quantifier and negation. Since Musolino's (1998) research, many studies have reported that English-speaking children strongly prefer to assign the surface scope interpretation to a sentence with a universal subject and negation. We tested English-speaking children and adults (and also Japanese adult learners of English, which is reported in Appendix 7), to investigate whether they are sensitive to prosodic cues for scope assignment. The results

from English-speaking populations show that the effect of prosody is robust and the different question types do not seem to be crucial in comprehension, while the results from the L2 population suggest that question types affect processing of marked sentences. We will discuss the apparent lack of the effect in the question type with the English-speaking population in the discussion section.

The two case studies might sound independent of each other, especially since they are concerned with two different phenomena. However, the experimental results from both case studies suggest that (i) children are more competent than previous relevant literature had assumed, and that (ii) what they are not adult-like about is the ability to accommodate an appropriate sub-question required under the QAC; in other words, their pragmatic skills are not adult-like.

The remainder of this chapter provides background for the experiments on the "only" case study and follow-up discussions. Background for the experiments on the "RFR and scope" case study will be provided in the Chapter 4. The word "*we*" will be used when discussing the details of the research, since the experiments discussed in this thesis were carried out with a tremendous amount of help from my collaborators: Martin Hackl (experiments in Ch. 2-5), Ken Wexler (Ch. 2, 5), Erin Olson (Ch. 3), and Irina Onoprienko (Ch. 5). All errors are mine.

1.1 Crain's puzzle

Children's acquisition of sentences with *only* has attracted much attention since Crain et al.'s (1994) finding that English-speaking children seem to have a tendency of not interpreting sentences as in (1a) in an adult-like way; while at the same time they exhibit adult-like competence in interpreting sentences as in (1b). To better understand, imagine a picture of a cat holding a flag, a goose holding a flag and a balloon, and a frog holding a balloon. If an adult speaker is asked if (1a) is true relative to the picture, they would reject the sentence; and if they are asked if (1b) is true, they would accept the sentence. However, what Crain et al. found is that children tend to incorrectly accept (1a) and that they correctly accepted (1b).

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

This observation is not because of a child's "yes-bias". It was revealed that children do reject sentences, as we will see shortly. The interesting point is that the participant's reasons for rejection are not adult-like and this is apparent based on their justifications to reject a sentence. Suppose a picture contains two animals, a dinosaur and an elephant; and the dinosaur is paining a house and a chair while flying a kite, and the elephant is painting a car and holding a balloon. If an adult speaker is asked if (2a) is true, they would respond true; while a child would reject the sentence in (2a). One reason that a child provided the experimenter for rejecting the sentence in (2a) was "the dinosaur was flying a kite and painting a chair, as well as painting a house" (Crain et al. 1994: 461).

(2) a. Only the dinosaur is painting a house.
    b. The dinosaur is only painting a house.

Children seem to interpret sentences with Subj-*only* to mean what a corresponding VP-*only* would mean. Furthermore, children do not have trouble interpreting VP-*only* in an adult-like way. Subsequent research on children's understanding of sentences with *only* has shown that this curious asymmetry in children's understanding, between sentences with Subj-*only* and sentences with VP-*only*, is robust across numerous dimensions. As demonstrated from the results of Crain et al., it is also robust across truth-values. This is when a child accepts Subj-*only* sentences and an adult rejects them; or, a child rejects Subj-*only* sentences and an adult accepts them.[1] The asymmetry has been found across a wide age range going from 2;6 years to over 6 years of age. Furthermore, it has been replicated across a variety of languages (cf. Philip and Lynch 2000, Paterson et al. 2006, a.o. for English;[2] Notley et al. 2009, Zhou and Crain 2009, 2010 a.o. for Mandarin; Müller et al. 2011 for German; Endo 2004, Sano 2011 for Japanese).[3] The same trend is replicated in different experimental methodologies, such as

[1] Here it is assumed that the truth-value of Subj-*only* sentence and the hypothetical truth value of VP-*only* of the corresponding sentence are made counter-balanced.

[2] There are other interesting studies such as by Gualmini et al. (2003), who studied children's comprehension of the sentences with VP-*only* in the double-object construction, and by Paterson et al. (2003), who studied Subj-*only* and VP-*only* to conclude that children ignore *only*, rather than that there is an asymmetry between comprehension of Subj-*only* and VP-*only*.

[3] Interestingly, a study in Hungarian by Pintér (2015) indicates that Hungarian-speaking children tend to associate *only* with the subject more often than with the object. She suggests that this might be due to the language-specific features regarding focus constructions and the word order. Further studies on *only* cross-linguistically might reveal more facts of our interest.

picture-sentence verification task, sentence-picture verification task, and Truth-value judgment task (TVJT, Crain and Thornton 1998). Those results have led researchers to conclude that the grammar of VP-*only* is acquired before that of Subj-*only*, and therefore VP-*only* might be more basic than Subj-*only*.

In the first case study of this thesis, I discuss the need for a more nuanced view versus the acquisition field's conventional understanding and explanation of the results. First, Crain et al.'s hypothesis to account for the Crain's puzzle is reviewed (§1.2), and then in §1.3, it is pointed out that in previous literature of TVJT, the question prompting puppet's stimulus sentence was infelicitous or not ideal for incorporating a sentence with *only*. This idea is based on a notion of Question-Answer Congruence (QAC), which in turn gives us several pairs of questions and answers that allow for the differentiation of the congruence aspect. It is pointed out that the previous studies failed to pay attention to the role of QAC, and the motivation for the experiments reported in Chapter 2 and §3 is to examine to what extent the factor of QAC might contribute to Crain's puzzle. The new way to state the question regarding Crain's puzzle is proposed; it is evaluated and compared with Crain's hypothesis in §1.4.

To evaluate the contribution of QAC, the experiments with children reported in Chapter 2 test a hypothesis that congruent Q-A pairs are understood in an adult-like way, more often than incongruent Q-A pairs, and irrespective of attachment-site of *only*. The hypothesis has been widely evaluated, and we show that the accuracy rates for congruent Subj-*only* answers were above 70%; which provides strong evidence that children's difficulty with Subj-*only*, as presented in the literature, should not be considered that they have not acquired Subj-*only*. Furthermore, the results from studies with adults reported in Chapter 3 conform closely to those with children; in that we observed the asymmetry between Subj-*only* and VP-*only* in an experimental setting which does not take QAC into account, and our adult participants showed high sensitivity to QAC in an experimental setting where the congruence is manipulated.

While the results clearly show that QAC plays an important role in the comprehension of the sentences with *only*, the asymmetry between Subj-*only* and VP-*only* is yet to be accounted for. I will propose an idea linking the observed asymmetry and the way of information packaging, referring to the semantics of superiority obeying/violating multiple *wh*-questions, in the discussion section of Chapter 2, and I will come back to the discussion in Chapter 3.

1.2 Hypothesis proposed for the Subj-*only* vs. VP-*only* asymmetry

To account for the asymmetry in children's comprehension with Sub-*only* and VP-*only*, Crain et al. (1994) hypothesized the following;[4] children initially misanalyse Subj-*only*, which is in fact an attributive *only*, as an adverbial *only*, which attaches to the sentential-level, and thus they tend to associate *only* with VP, wherever it attaches in a given sentence. This hypothesis can be paraphrased as: the grammar of Subj-*only* (= attributive *only*) is not in place in child grammar.[5] Zhou and Crain (2010) claim that the hypothesis is motivated by the fact that many adverbs (such as *sometimes*, and *usually*) tend to take sentential scope and are not associated with the subject, and that a sentential adverb tend to be associated with/modifying the VP across a variety of human languages.

Crain's hypothesis predicts a couple of things; first, children will not misanalyse VP-*only* sentences. Secondly, assuming that the hypothesis is reduced to a statement that children misanalyse attributive *only* as adverbial *only*, it predicts that children would have difficulty with *only* that is adjacent to an object NP/DP, since Object-*only* (Obj-*only*) is also an attributive *only*. Finally, since the hypothesis assumes that children lacks the grammar of Subj-*only* (or attributive *only*), it predicts that children do not produce Subj-*only*. Another possible prediction is that children misuse apparent Subj-*only* to mean VP-*only*; If children do have the competence of deriving the word order of Subj-*only* (as a sentential adverb) but have preference of assigning the associate to the object in production as well, then it is predicted that children use Subj-*only* to mean VP-*only*. All in all, production errors would be expected. We will evaluate how our proposal will differ in prediction from Crain's hypothesis in §1.4.

Crain's hypothesis answers the question of what is the distinction between adults and children, in a simple manner; namely, the grammar is different. Child grammar lacks the mechanism to interpret Subj-*only*/attributive *only*, and only after they have acquired attributive *only*, are they adult-like. This statement highlights the learning question: What does it take to

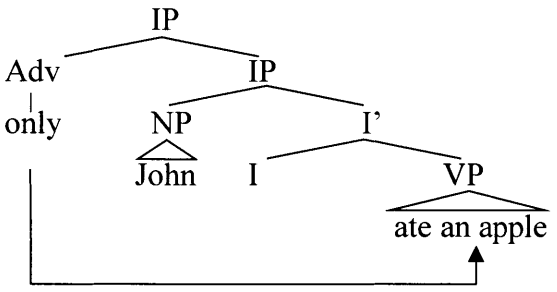
[4] Crain et al. (1994) do not clearly state the hypothesis, in the way it is articulated here; but instead, they argue that "children initially hypothesize just one of the interpretative [options] from UG (Crain et al. 1994: fn. 7)", which is the grammar of VP-*only*. They further argue that "[t]he interpretations that are lacking in early child grammars are simply added on the basis of positive evidence (ibid.)." The way this hypothesis is articulated here is largely based on the description of Crain et al.'s (1994) hypothesis by Zhou and Crain (2010).

[5] Paterson et al. (2003) hypothesized that children ignore *only* when they interpret sentences with *only*. That is, the children interpreted a sentence with *only* as what its prejacent means. Paterson et al. (2006) reached a different conclusion, which is similar to what Crain et al.'s (1994) view.

learn Subj-*only*? Crain et al. argues in their fn. 7 (see our fn. 4) that, it is done through positive evidence.

Let me further briefly note that the following configuration is hypothesized for the Subj-*only* sentences, according to Zhou and Crain (2010).

(3) "Only John ate an apple"



(Zhou and Crain 2010: 981)

Z&C further hypothesized that the presence of negation would block the association between *only* and the element in the VP, based on the assumption of Relativized Minimality (RM, Rizzi 1990, 2001) that the association between two elements is blocked when a third element, with a potential to participate in the relevant relation, intervenes. Details aside, they predict that negation in the sentence such as "Only John did not eat an apple" will prevent the association between *only* and the element in the VP, and thus children would produce more adult-like behavior with such sentences. Their results, tested with Mandarin-speaking children, confirmed the expectation. However, questions remain; how is the assumption of the association between *only* and its associate constituting a legitimate configuration for RM be motivated (it does not constitute a Head Chain, A-Chain, or A-bar Chain, which are listed as exhibiting typical RM effects by Rizzi)?[6] Why isn't the same type of blocking phenomenon found with the other adverbs that Z&C consider to be similar to *only* (i.e. negation does not block the association between *sometimes* and VP, for example, in a sentence "*Sometimes John doesn't eat spinach*")?

In the next subsection, I will review the principle of QAC, specifically in connection with sentences with *only*. It is pointed out that the importance of QAC has been ignored in the

[6] It is possible that they assume that it is a case of focus intervention (Beck 2006, a.o.), though it is still not clear regarding the second question: how it could be comparable with non-focal sentential adverbs such as *sometimes*. More importantly, the intervention by negation only explains why speakers associate pre-subject *only* with the subject in the presence of negation. It does not explain why in the adult grammar speakers always associate it with the subject regardless of the environment. Therefore the RM account is not an explanation of the learning that must take place. I owe this discussion to Martin Hackl.

literature of the acquisition of *only*, but should have been paid attention to. It is important to note that since QAC is a general principle, it should be taken into account in other phenomena outside of *only*.

1.3 Question-Answer Congruence (QAC)

It is well-known that answers to *wh*-questions are subject to a principle known as Question-Answer Congruence (QAC) (Paul 1880 via Krifka 2006, von Stechow 1990, Rooth 1985, 1992, etc.). QAC requires the constituent in the answer that corresponds to the *wh*-phrase in the question be focused (or "F"-marked). This is illustrated in various Q-A pairs in (4). First, we see that the answer in A1 to the question in (4a) is well-formed and is considered congruent. This is because the Q-A pair of (4a) and A1 satisfies QAC; (4a) is a question seeking information about the subject (subject-question) and the subject DP in A1 is the focus. By contrast, A2 does not satisfy QAC and is considered incongruent. This is because the object DP is the focus in A2 even though the question (4a) is a subject-question. Exactly the opposite of that is presented in (4b).

(4) a. Q: Who is holding a flag?
A1: THE CAT_F is holding a flag.[7]
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 $FLAG_F$.

The constraint of QAC is summarized as follows: following Rooth (1992), the focus, or "F," is always interpreted as introducing an anaphor of the type of a question (a set of propositions;

[7] The words in capital letters are intended to bear pitch accent in its stressed syllable to indicate focus of the answer, and the subscript "F" stands for focus. Consequently, the words in lower letters do not bear pitch accent. Putting stress on the answer term for incongruent answers (e.g. the cat in A2 in (4a)) would make the answer sound less unfavorable, but that manipulation is not considered here. In any case, the "less unfavorable" example is not saved from incongruence. Also note that there are many other answers possible, other than an answer that is introduced here as a congruent one, such as a fragment answer (e.g. "Who is holding a flag?" – "The cat.").

Hamblin 1973), and the choice of antecedent is constrained to be a subset of a set of propositions denoted by the answer (as described in (5a) in terms of analogy to presupposition).[8]

(5) a. $\varphi \sim \Gamma$ presupposes that Γ is a subset of the focus semantic value for φ and contains both the ordinary semantic value of φ and an element distinct from the ordinary semantic value of φ. (Rooth 1992: 93)

b. where defined, $[\![\varphi \sim \Gamma]\!]^o = [\![\varphi]\!]^o$

The ordinary semantic value of (4a-Q) is a set of propositions as in (6a), and the focus semantic value of the answer (4a-A1) is a set of propositions as in (6c). The focus semantic value of the answer and the ordinary semantic value of the question is identical; i.e., the ordinary semantic value of the question is a subset of the focus semantic value of the answer. This satisfies the constraint of QAC, which is why this is a congruent Q-A pair.

(6) a. $[\![(4a\text{-}Q)]\!]^o = \{\lambda w.\ x \text{ is holding a flag in } w: x \in D_e\}$

b. $[\![(4a\text{-}A1) \sim Q]\!]^o = [\![[\text{the cat}]_F \text{ is holding a flag} \sim (4a\text{-}Q)]\!]^o$

c. $[\![(4a\text{-}A1)]\!]^f = \{\lambda w.\ x \text{ is holding a flag in } w: x \in D_e\}$

On the other hand, when we look at the incongruent pair with the answer (4a-A2), then the focus semantic value of the answer is the one in (7c). This time, the ordinary semantic value of the question is not a subset of the focus semantic value of the answer; e.g., an example proposition in the set $[\![(4a\text{-}Q)]\!]^o$ could be "*The goose is holding a flag*", but this is not contained by the focus semantic value of (4a-A2).

(7) a. $[\![(4a\text{-}Q)]\!]^o = \{\lambda w.\ x \text{ is holding a flag in } w: x \in D_e\}$

b. $[\![(4a\text{-}A2) \sim Q]\!]^o = [\![\text{the cat is holding } [\text{a flag}]_F \sim (4a\text{-}Q)]\!]^o$

c. $[\![(4a\text{-}A2)]\!]^f = \{\lambda w.\ \text{the cat is holding } x \text{ in } w: x \in D_e\}$

Next, let us turn to answers that include the focus-sensitive operator *only*, which requires an F-marked constituent in the sentence it occurs. First let us review the definition of *only* in (8); the

[8] The squiggle (~) operator introduces presuppositions, which imposes a condition on its sister represented on the right in (5b).

sentence with *only* presupposes the ordinary semantic value of its prejacent, and asserts that all of the elements of the focus semantic value of its prejacent are false, except the one that is the ordinary semantic value of the prejacent.

(8) $[\![\text{only S}]\!]^o = \lambda w: [\![S]\!]^o(w) = 1.\ \forall S' \in [\![S]\!]^f (S' \neq [\![S]\!]^o \rightarrow S'(w) = 0)$.[9]

There is still something missing in (8) – how can we know the member of the focus semantic value of the prejacent? The set of alternative propositions is generated by replacing the focus with its alternatives, and the focus is the associate of *only*. The position of the focus associate of *only* is subject to a syntactic constraint whose effect in simple sentences like these can be summarized in a purely descriptive form presented in (9) and exemplified in (10).

(9) Focus Co-occurrence Constraint for *only* (FCC-*only*)[10]
   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.

(10) 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. The Q-A pairs in (11) illustrate that the associate of *only* has to be the constituent that corresponds

[9] This definition plainly assumes that the sentence with *only* presupposes its prejacent, and does not take into account the scalar nature of the presupposition; namely, the condition "the prejacent is ranked relatively low among its alternatives" is also presupposed (following Bonomi and Casalegno 1993, Guerzoni 2003, Klinedinst 2005, Beaver and Clark 2008, Alxatib 2013, a.o.). Though I do consider that the scalar presupposition is a part of the semantics of *only*, the component is not crucial in the context of the interaction between QAC and *only*, and thus I will keep the definition here simpler.

[10] This is an extension of a constraint considered in Rooth (1985), which is cited below:

(i) In LF, *only* must be the sister of a phrase bearing the feature F. (Rooth 1985: 40)

This is canonically thought to be a consequence of *only* being a scope taking operator. For example, *only* and the subject NP forms a DP constituent, and that *only* and a VP forms a VP constituent. Then we can state the relevant constraint, in terms of scope, is that the associate of *only* has to be in the surface c-command domain of *only* (it does not have to be the sister of *only*, as long as the sister of *only* contains it). See Jackendoff (1972) and more subsequent research related to this topic for discussion.

to the *wh*-phrase in the preceding question. The pairs of Q-A1 in (11a-b) and Q-A2 in (12a-c) are congruent while the other pairs are incongruent.

(11) a. Q: Who is holding a flag?
b. A1: Only THE CAT_F is holding a flag.
c. A2: *The cat is only holding A $FLAG_F$.

(12) a. Q: What is the cat holding?
b. A1: *Only THE CAT_F is holding a flag.
c. A2: The cat is only holding A $FLAG_F$.

Let us assume that a focus sensitive expression, here using *only*, is anaphoric on the question under discussion (QUD), [11] to be regarded as analogous to the focus in an answer term being anaphoric to the question. It then follows that the focus, in our example the associate of *only*, has to be on an expression that gives a focus semantic value that is a superset of the question to observe QAC. In other words, Rooth's theory on QAC correctly allows that an intonational focus in the above examples could serve as both an answer to the question and an associate of *only*.

Furthermore, Rooth's theory correctly predicts which pairs are congruent and which are not. This is because the prejacent of *only*, for example in (11b), has focus on *the cat*, which corresponds to *who* in the question; the focus semantic value of the prejacent of (11b) is identical to the ordinary semantic value of the question (11a). This is why the Q-A pair is congruent. The reason for the incongruence for the Q-A-pairs with an asterisk above is the same as presented in example (4).

Finally, let us consider cases where a sentence with *only* is intended as an answer to a question such as "*What happened?*"; i.e. the type of question that is typically used in Truth-value judgment tasks in acquisition experiments to prompt the target sentence, as in (13).[12]

(13) (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.)
Q: Kermit, can you tell us what is happening in the picture?

[11] Following Roberts (1996/2012), von Fintel (1994), and Beaver and Clark (2008).
[12] We are aware of only one exception to this, in the literature on the acquisition of *only* – Experiment 3 of Notley et al. (2009), which will be presented in the discussion section of Chapter 2.

A1: (?)Only THE CAT_F is holding a flag.

A2: (?)The cat is only holding A $FLAG_F$.

Note that neither of the answers in (13) is congruent with the question asked. The ordinary semantic value for the "broad" question in (13-Q) would consist of, for example, {the cat is holding a flag, there are three animals in the picture, the goose got two things, the frog is very big, ...} and so on. The set of alternatives could contain anything happening in the picture. Since it is a broad focus question that guides the addressee to describe what is happening in general as long as it is relevant to the context, the ordinary semantic value of (13-Q) is not a subset of the focus semantic value of the prejacent of either (13-A1) or (13-A2). QAC is therefore not satisfied and we would expect that the Q-A pairs would become infelicitous. Interestingly enough, they are perceived to be acceptable unlike the incongruent pairs in (11-12), although adult native speakers typically do find them a bit less crisp than fully congruent cases.[13]

Why doesn't the Q-A pairs in (13) end up being infelicitous? Following e.g. Roberts (1996/2012), we propose that this is because there emerges a need to avoid violation of QAC in individual's mind when they try to comprehend the sentence. As a result, they can accommodate an appropriate sub-question that would make a congruent Q-A pair.[14] Roberts proposes a model of discourse structure, inspired by Stalnaker's (1978) view that conversation is a set of possible worlds we might be in that are being narrowed down as the conversation goes on. She argues that discourse is compared to a game where speakers are engaged in figuring out a "big question" – which can be paraphrased as "What is the way things are?" The way this is done is by resolving sub-questions of the big question. Suppose the big question that we would like to figure out is "*What is happening in the picture?*" This question can be broken down into a number of sub-questions; for example, "*What animals are in the picture?*", "*Who is holding a flag?*", "*What is the cat holding?*" and so on. An example of what a QUD stack looks like in our case is presented below:

[13] The experiment measuring response times carried out with adult native speakers to indicate that this intuition is on the right track is reported in Chapter 3 and confirms that there are different levels of congruence.

[14] To account for how the accommodation of a sub-question works, let us assume the question-under-discussion stack (QUD stack), or its visualized counterpart, Discourse-tree (D-tree), following Roberts (1996/2012), Büring (2003), and Beaver and Clark (2008), in conjunction with the notion of discourse move in Carlson (1982). Cf: the work of van Kuppevelt (1995, 1996).

(14) What is happening in the picture?

a. Who is holding a flag?

i. Is the cat holding a flag?

ii. Is the goose holding a flag?

iii. Is the frog holding a flag?

b. What is the cat holding?

i. Is the cat holding a flag?

ii. Is the cat holding a balloon?

Answers to these sub-questions are relevant.[15,16] That is because the answers provide part of the information sought by "*What is happening in the picture?*", and a speaker answering with either (13-A1) or (13-A2) are judged to be cooperative and their answers judged felicitous once the comprehending individual accommodates an appropriate sub-question modulo focus – "*Who is holding a flag?*" for (13-A1) and "*What is the cat holding?*" for (13-A2). As long as the comprehender accommodates an appropriate sub-question upon hearing an answer that is not totally congruent, QAC is indirectly satisfied. We will call the Q-A pair of a broad question and a sentence with *only* "indirectly congruent Q-A pairs." On the other hand, QAC is directly satisfied when the congruent Q-A pairs do not require accommodation of a sub-question. We call such a Q-A pair a "directly congruent Q-A pair."

The idea that the Q-A pairs in (13), which use a broad question and require the mechanism of the accommodation of a sub-question raises the possibility that previous studies on the acquisition of *only* have underestimated children's competence regarding *only*. That is, it is possible that children have difficulty accommodating a suitable sub-question, and that contributes to at least some parts of Crain's puzzle. It might be that Crain's puzzle is more

---

[15] An answer needs to be relevant to the question (Roberts 1996/2012, van Rooy 2003). Roberts characterizes the notion of Relevance as follows:

(i) A move *m* is *Relevant* to the question under discussion q, [...], *iff m* either introduces a partial answer to *q* ([i.e.,] *m* is an assertion) or is part of a strategy to answer *q* ([i.e.,] *m* is a question).

(Roberts 1996/2012: (15))

[16] The notion of Relevance could be defined in other terms as well; assuming the notion of a question as providing a partition of a set of possible worlds into subsets of possible worlds (Groenendijk and Stokhof 1984, and subsequent work), the partition can be considered to describe what is relevant given the question (Potts 2006).

As suggested by Danny Fox (p.c.), it might be possible to explain the interaction of QAC and sentences with *only* in the context of Relevance and the notion of maximizing informativity, which are more basic factors of pragmatics. However, I will leave the further examination of different approaches for future research.

aligned to the ease and difficulty of accommodation types, rather than the grammar of Subj-*only* and VP-*only*. In other words, the hypothesis we propose to account for the asymmetry between Subj-*only* and VP-*only* is that the accommodation of object-questions requires less processing load than that of subject-questions, and therefore VP-*only* sentences are easier to comprehend than Subj-*only* sentences, as an answer to a broad question. If the hypothesis is confirmed, such a discovery would suggest that Crain's puzzle should be restated as a question about "Why are object-questions easier to accommodate than subject-questions", which is concerned with discourse strategies, and a question about "Why does child grammar allow the association of the pre-subject *only* with the object DP, disregarding the FCC-*only*?", which is a topic about what constitutes child grammar. A proposal responding to the first question and the discussion regarding the second question are found in the discussion section of Chapter 2, §2.4.

1.4 Comparison between hypotheses

Crain's hypothesis, reviewed in §1.2, predicts the following: (i) children will not misanalyse VP-*only* sentences, (and will often misanalyse Subj-*only*), (ii) children would have difficulty with *only* that is adjacent to an object DP, and (iii) children do not produce Subj-*only*, and/or they misuse Subj-*only* to mean VP-*only*. What differentiates children from adults is the state of their grammar; child grammar lacks the grammar of Subj-*only*/attributive *only*.

What does the hypothesis of accommodation predict for the first point? According to the hypothesis, the difficulty with Subj-*only* is due to the difficulty with the accommodation of a suitable subject-question when a broad question is presented. This means that when one does not have to accommodate a sub-question, i.e., when a directly congruent question precedes a sentence with *only*, the adult-like performance is expected to increase. To be more specific, if the difficulty in accommodation is the whole story, when the accommodation cost is set to zero, the perfectly adult-like performance is expected. Conversely, when the accommodation cost is somehow manipulated to be large, the performance is expected to be sensitive to the manipulation. In order to make concrete predictions as to whether the adult-like performance will decrease or whether the chance-level performance is expected, we need to discuss in more detail what kind of strategy a comprehender might take when they encounter an incongruent Q-A pair (§2.3.1). However, what we can safely say now is that the accommodation hypothesis

predicts that the performance will be modulated according to the levels of congruence, and it should hold true for either type of sentences with *only*. This prediction is not as simple as to claim that VP-*only* is not misanalysed while Subj-*only* is often misanalysed.

Regarding the comprehension of Obj-*only*, the accommodation hypothesis predicts that it will pattern with VP-*only*, where the plausible associate of VP-*only* is located on the object of the sentence.[17] We have seen in §1.3 that the immediately congruent question for VP-*only* is an object-question, because the focus associate of VP-*only* is on the object DP. Note that the focus associate of Obj-*only* is located on or inside of the object DP; which means that the immediately congruent question for Obj-*only* is an object-question. Taken together, the accommodation hypothesis predicts that the comprehension of Obj-*only* will pattern with that of VP-*only*. In fact, this hypothesis is not directly tested in the experiments reported in this thesis, though Experiment 1-prime in Chapter 3 (§3.2.3), which was tested with adults, will be suggestive regarding this point.

Does the accommodation hypothesis predict anything about the production of sentences with *only*? Basically, the hypothesis is only concerned with the comprehension aspect; it is about processing cost of accommodation of different sub-questions, and it is hearers who accommodate a sub-question when necessary, in conversation exchange. Therefore, the hypothesis predicts either of the following: children do not make mistakes in production of sentences with *only*, wherever the attachment site is, or there will be no difference in error rates between the production of Subj-*only* and VP-*only*, if they make mistakes.

The accommodation hypothesis assumes that children have the grammar of Subj-*only*, but the accommodation cost prevents them to interpret Subj-*only* in an adult-like way. What differentiates children from adults under this hypothesis is the ability to accommodate a suitable subject-question for Subj-*only*, when a broad question, rather than a directly congruent question, is presented.

The predictions described so far are idealized; it is possible that a combination of multiple factors constitutes Crain's puzzle. For example, it is possible that resolving Crain's puzzle might call for both explanations after all; in other words, the accommodation of a subject-question is more difficult than that of an object-question, and on top of that, the grammar of

[17] This is partly because of the nature of the situation; in our cases of animals holding items, all the animals are engaged in the same action, so there is no reason to put focus on the verb itself or the VP as a whole. In addition, there seems a tendency to associate VP-*only* with the object DP (as suggested the results of Experiment 2 from adult participants in Crain et al. 1994), even when different actions are presented in a given situation.

Subj-*only* is indeed more complex than that of VP-*only*. The hybrid hypothesis – the combination of the accommodation difficulty and the grammar difficulty [18] – predicts somewhere in the middle of the Crain's hypothesis and the accommodation hypothesis, but is distinct from Crain's hypothesis in important aspects, to be shown in a comparison table shortly.

In the hybrid hypothesis, what differentiates children from adults? First, children need to be able to accommodate a suitable subject-question for Subj-*only* when a broad question is presented. It is also expected that children need to be able to overcome the difficulty with associating Subj-*only* with the subject DP, to show adult-like behavior; in other words, children need to know FCC-*only*, and understand that association between *only* and its associate is not determined arbitrarily.

The hybrid hypothesis asks to what extent QAC (or, different levels of accommodation) contributes to Crain's puzzle. This hypothesis will be discussed as the most plausible account for Crain's puzzle. The predictions of each hypothesis towards child performance are summarized below.

[18] By the difficulty in grammar, I mean that the grammar to interpret sentences with the focus associate of *only* on the subject position is somehow difficult to process. I do not mean that children do not *have* the grammar of Subj-*only*, and that I do not particularly commit to discuss that attributive *only* and adverbial *only* are distinct.

Table 1: Comparison between different hypotheses on Crain's puzzle

<table>
<tr><th></th><th colspan="3">Comprehension</th><th rowspan="2">Production of only-sentences</th></tr>
<tr><th></th><th>Subj-only</th><th>VP-only</th><th>Obj-only</th></tr>
<tr><td rowspan="2">No grammar of Subj-only</td><td>Not good</td><td>good</td><td>= Subj-only,
Not good</td><td>Similar asymmetry to comprehension, S worse than V
(meaning S-only is produced less/Subj-only is misused to mean VP-only, etc.)</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to acquire the grammar of Subj-only/ attributive only.</td></tr>
<tr><td rowspan="2">Difficult accommodation of subject-Q</td><td colspan="2">Regardless of the attachment site;
QAC satisfied = good,
QAC unsatisfied = not good</td><td>= VP-only,
good</td><td>Does not particularly predict production errors</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to be able to accommodate a suitable sub-question when it is implicit in discourse.</td></tr>
<tr><td rowspan="2">Difficult accommodation of subject-Q
+
difficult grammar of Subj-only</td><td>Performance modulated by QAC, but not good</td><td>Performance modulated by QAC, but good</td><td>Better than Subj-only (b/c the required accommoda-tion is object-Q)</td><td>Difference in errors between S and V is smaller than comprehension</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to be able to accommodate a suitable sub-question when it is implicit in discourse, and to know FCC-only.</td></tr>
</table>

The goal of Chapter 2 is to evaluate which hypotheses serve better than others, and Chapter 3 provides supporting evidence from experiments with adults. To test whether QAC plays a role in processing of sentences with *only*, we use directly congruent Q-A pairs, as well as irrecoverably incongruent Q-A pairs, and ask whether the rate of adult-like responses increases for the former and decreases for the latter compared to the indirectly congruent Q-A pairs 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*. A conjecture on what

distinguishes children from adults is presented by closely examining the data on incongruent pairs, as well as comparing them with the results on the other conditions.

Chapter 2

Question-Answer Congruence and incongruence in the acquisition of *only*[*]

2.1 Hypothesis and predictions

As we have reviewed in §1.1, acquisition of sentences with *only* has observed Crain's puzzle for long – English-speaking children seem to exhibit an asymmetric capability in comprehending sentences with *only*, as in (1).

(1) a. Only the cat is holding a flag. → Not understood well
b. The cat is only holding a flag. → Understood well

Specifically, children do not interpret sentences like in (1a) in an adult-like way, but they tend to interpret the sentence in (1a) to mean what (1b) would mean. Similar results have been replicated in different languages and with children in a wide age range, which led researchers to conclude that the grammar of Subj-*only* is somehow more complex and acquired later than that grammar of VP-*only*.

We question that claim by pointing out that the previous studies did not use directly congruent Q-A pairs. Given the principles of QAC discussed in detail in §1.3, it was revealed that the Q-A pairs such as in (2a) are in fact not directly congruent, since comprehenders must

[*] This chapter reports the results of our experiments reported in Hackl et al. (2015) and discusses its results in a more extensive way. For the experiments conducted for this chapter, I would like to thank Rachel Magid, Sammy Floyd and Laura Schulz at the MIT Early Childhood Cognition Lab at Boston Children's Museum for their courtesy. I would also like to thank our MIT Undergrad Research Assistants, Amaya Arcelus, Su Lin Blodgett, Jiapei Chen, Sebastian Garza, Lucie Lozinski, Irina Onoprienko and Laya Rajan for their help conducting experiments.

accommodate an appropriate sub-question so that the answer that has a focus-sensitive particle makes a congruent Q-A pair. We labeled such Q-A pairs (i.e., a broad focus question and an answer sentence with *only*) as indirectly congruent Q-A pairs. By contrast, the Q-A pairs in (2b-c) are congruent ones. We labeled such Q-A pairs as directly congruent Q-A pairs.

(2) a. 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 $FLAG_F$.
b. Q: Kermit, who is holding a flag?
A1: ^{ok}Only THE CAT_F is holding a flag.
c. Q: Kermit, what is the cat holding?
A2: ^{ok}The cat is only holding A $FLAG_F$.

Notice that a standard protocol of TVJT typically uses a broad question to prompt puppet's target sentence. There are good things about this, of course – it is a broad question and therefore it does not restrict the relevant scope of an answer, which is convenient because it can be used in almost all kinds of experiments. We point out that given the theory of D-tree, using a broad question to prompt the target sentence basically always requires accommodation of a sub-question. [19] This raises the possibility that previous studies, which forced children to accommodate an appropriate sub-question and thus presumably required more processing load, underestimated children's competence regarding *only*. The hypothesis is the following; if children are sensitive to QAC, we will observe different rates of adult-like responses among indirectly congruent Q-A pairs (*What happened?* – target), directly congruent Q-A pairs, and incongruent Q-A pairs. More specifically, if accommodation of a suitable sub-question actually involves processing load, directly congruent Q-A pairs yields higher rates of adult-like responses than those of indirectly congruent Q-A pairs, which might in turn yield higher rates than those of incongruent Q-A pairs.[20] We will test this hypothesis in the experiments reported in this chapter. We also analyze children's responses to incongruent Q-A pairs, to assess

[19] As pointed out in fn. 16, by relying on the notion of Relevance and information maximization, it might be possible to reduce the intuition for the Q-A pairs such as in (2a) of being less crisp than directly congruent Q-A pairs to the combination of more basic components in pragmatics.
[20] See §2.3.1 for predictions for different strategies with incongruence.

whether they rely more on QAC as guide for determining the associate of *only*, or on the syntactic position of *only* (which we dubbed as "FCC-*only*" in §1.3), as adults would do.

The plan is as follows. First, we will replicate the findings from previous studies by using a broad question (an indirectly congruent Q-A pair) to prompt puppet's utterance. That way we will set the baseline of the adult-like responses for comparison with directly congruent Q-A pairs and with incongruent Q-A pairs. Next, we show experimentally, using subject- and object-questions as prompts, that children are in fact very sensitive to QAC. More specifically, we show that the hypothesis is borne out; when QAC is directly satisfied, their comprehension of *only* is essentially adult-like for both Subj-*only* and VP-*only*. Furthermore when QAC is violated, children's response strategy is to ignore the syntactic position of *only* for both Subj-*only* and VP-*only*. Instead, they associate *only* with the constituent that corresponds to the *wh*-phrase in the preceding question – this is predicted if one conforms to QAC. Although the results clearly show that the directly congruent Subj-*only* is understood in an adult-like way to a great extent, with the levels of the congruence being equal, there still seemed to be an asymmetry between Subj-*only* and VP-*only*. Thoughts on what might be going on are presented as a proposal in the discussion section, §2.4.2.

The discussion section also compares the results of the experiments in this chapter and the predictions for different hypotheses, discussed in §1.4. We argue that the "hybrid" hypothesis serves as the best bet to account for the results. We also argue that, since the sentences with *only* that are intended to answer a broad question require accommodation of an appropriate sub-question, Crain's puzzle can be insightfully restated in terms of a difference in how easy it is to accommodate a subject- or object-question. The proposal in §2.4.2 is proposed to account for why object-questions are easier to accommodate than subject-questions.

2.2 Experiment 1 – Baseline

2.2.1 Methods and materials

Before we embark on investigating whether children are sensitive to QAC, we would like to establish that we can replicate the results to show Crain's puzzle with our own methods and materials. Experiment 1 serves this purpose and we call this the baseline experiment. In this

experiment, children were told simple stories supported by PowerPoint slides which showed animations of animal characters getting food items. Figure 1 shows the procedure and typical experimenter's narration during the presentation of the stories.

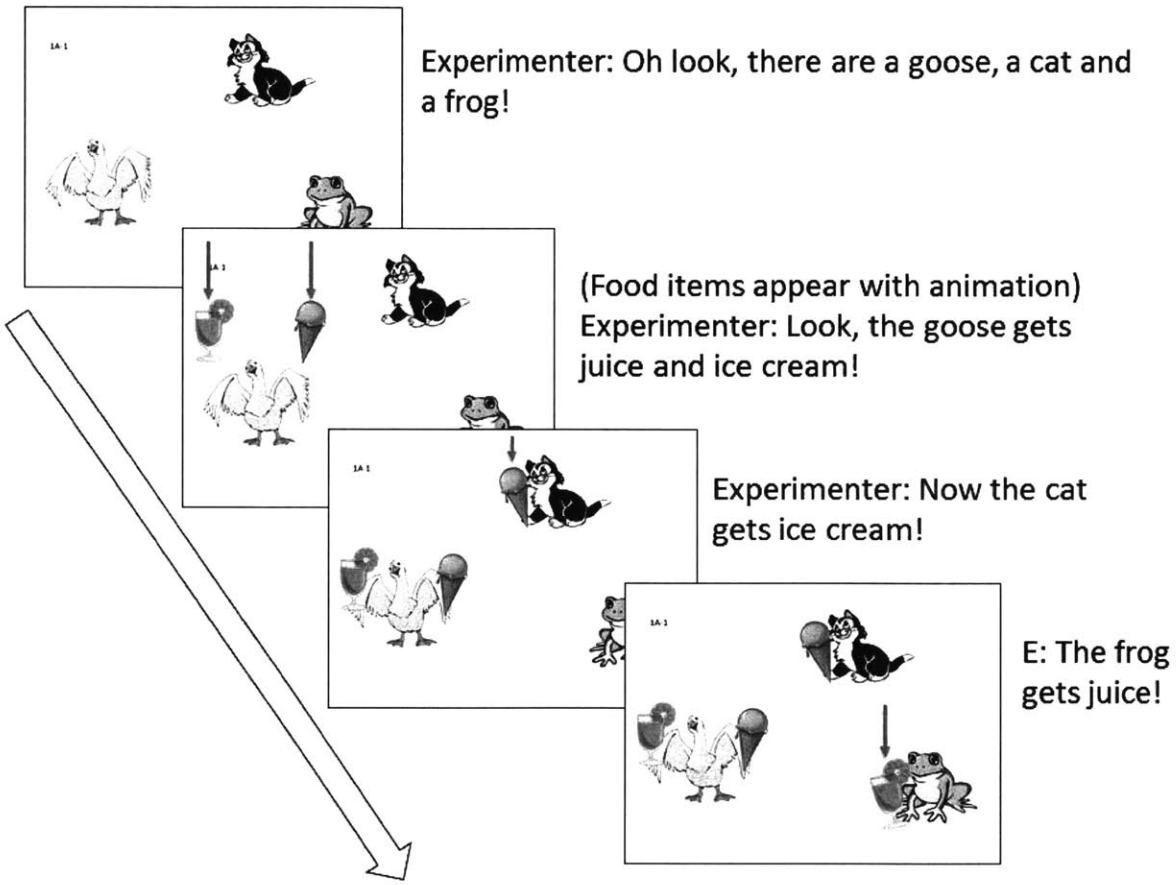


Figure 1: Illustration of experimental procedure

The experimenter, who was a native speaker of English, narrated the story and advanced the slides to show what was happening in the story as illustrated above. At the end of the story, the experimenter asked a puppet who was watching the story with them, "Kermit, 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*. The list of the target sentences and filler items, accompanied with the rates of adult-like responses per each item are in Appendix 1. Child participants were asked to judge if the puppet was right or wrong, and if they thought he was wrong, they were encouraged to give a justification. 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 some others have conducted. Children were encouraged and praised equally for Yes and No answers but encouraged to provide justifications only in case of the No answer, for it is less natural to ask for justifications for Yes answers.

The experiment consisted of 8 target items (4 with Subj-*only* and 4 with VP-*only*) as well as 4 filler items within one session. The presentation order was pseudo-randomized, and we prepared two different lists of items randomly assigned to each participant. Half of the items in each condition were designed to be true on an adult-like interpretation relative to the story, and the other half to be false. Importantly, stories were designed so that if a target sentence was true on the adult-like interpretation, it would be false on a non-adult-like interpretation with the attachment of *only* being on the other condition. Namely, if a given Subj-*only* sentence was designed to be true, the hypothetical VP-*only* interpretation of the sentence would be false, and vice versa. One example from each attachment-site and the story with which it was presented is illustrated below. There were the same number of True items as well.

(3) **Situation:** A goose gets juice and ice cream, a cat gets ice cream, and a frog gets juice.
Sentence: Only the cat got ice cream. [False, but on the VP-*only* interpretation, True]

(4) **Situation:** An elephant gets cake and jello, a horse gets jello and candy, and a dog gets cake.
Sentence: The horse only got candy. [False, but on the Subj-*only* interpretation, True]

Test sessions took about ten minutes per participants.[21] Some sessions were conducted at local daycares in a relatively quiet room or space in a classroom, and other sessions at the Boston Children's Museum in a quiet room with caretakers present throughout the test session. The caretakers sat behind the participants so the children were not able to see their reactions and complexions during the session, and the caretakers were asked to not interrupt children's games.

[21] We were quite limited in terms of the time the study could take, for recruitment of some participants took place at a museum. We did not want to interrupt their visit to the museum for unnecessarily long.

2.2.2 Results

We recruited 40 English-speaking children from Boston area daycares and the Boston Children's Museum across all socioeconomic and ethnic backgrounds. Children who incorrectly answered two or more fillers out of four were excluded (n = 2). Data from 38 participants, ranged from 4;0 to 6;9 (M = 5;2) are included in the analysis.

Figure 2 summarized the results of Experiment 1 plotting adult-like responses.[22] The error bars in the graph indicate 95% confidence interval. The rates of adult-like responses for the Subj-*only* when the designated answer is false was 22%, when the answer is true was 37%, the rates of VP-*only* when the answer is false was 75%, when the answer is true was 84%. In the analysis reported here, we combine responses from different items and from all the participants without dividing them into groups. See Appendix 1 for adult-like rates for each item and for two age groups (younger vs. older).

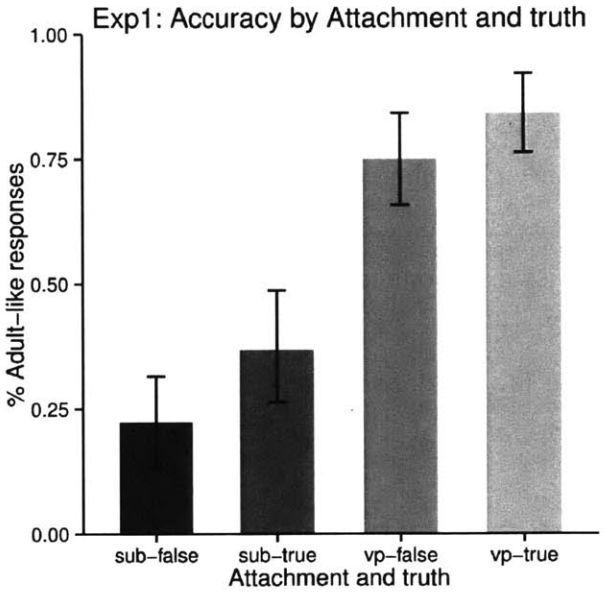


Figure 2: Results of Experiment 1

[22] Since we were not aware of the effect of truth in Hackl et al. (2015), the rates by the truth are not reported there. Subsequent analyses might look different from what is reported there because the truth is now considered as a potential main effect, but the body of the results which analyses are run on is the same.

Table 2: Statistical analysis of the results of Experiment 1

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-7.990	2.905	-2.751	0.00594 **
attachmentVp	11.419	4.353	2.624	0.00870 **
truthTrue	3.079	1.081	2.847	0.00441 **
attachmentVp:truthTrue	-1.788	1.269	-1.409	0.15887

Statistical analysis of the response rates was conducted in R[23] using linear mixed effects model with logistic regression (generalized linear mixed effects model, hereafter GLMEM). The resulting output is presented in Table 1. Since a maximally specified model did not converge, the order of presentation was not considered as a potential main effect. The analysis reveals main effects of the attachment site ($p < .009$) and of truth ($p < .004$). There was no interaction detected. The effect of truth is naturally interpreted as revealing children's yes-bias and the difficulty of correctly rejecting a false sentence in general. The fact that the truth effect was detected strongly indicates, as has been maintained as a merit of TVJT, the importance of using both true-items and false-items in a TVJT whenever possible, for it may well be possible that adult-like rates from true-items inflate the assessment of children's competence in grammar. Even VP-*only* items, which yielded higher adult-like rates, exhibited lower accuracy on the false items. In order to assess children's competence with fairness, we need to also test false-items. We get back to this issue in connection with the results from Experiment 2 of the RFR study reported in §5.3 and discuss in §5.4.

To take a closer look at the children's responses, the situation in a picture and the target example for the "True" Subj-*only* trial is presented below:

(5) **Experimenter:** Look, a goose gets ice cream and cheese. And a cat gets ice cream. And a frog gets cheese and juice. Kermit, tell me what happened!
Puppet: Only the frog got juice. [True]

[23] R project for statistical computing by CRAN. R version 3.1.3 was used through the platform of RStudio version 0.99.893.

The expected answer is "correct/right", but the accuracy rate was below 40%. Common justifications referred to the other item that the animal in question is holding. For example, with the situation depicted in (5), "Because he has cheese too."

All in all, the accuracy of Subj-*only* was lower (Average: 29.6%) than that of VP-*only* (Average: 79.6%). This replicates Crain et al.'s and subsequent works' discovery – whatever the cause of the asymmetry is – and this means that we can confidently use our methods and materials to further study children's sensitivity to QAC.

2.3 Experiments 2A and 2B

2.3.1 Design and predictions

We would like to test the hypothesis that the broad Q-A pairs as used in Experiment 1 are not directly congruent and thus processing the conversation requires more processing load than when directly congruent Q-A pairs are presented. In Experiments 2A and 2B, when we prompt puppet's answer we use questions that have different levels of congruence to investigate children's sensitivity to QAC. 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 described in the table below.

Table 3: 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.

Truth was defined in terms of adult responses to the target sentence, and it was counter-balanced across the attachment site of *only* in the same way it was in Experiment 1. It might be a little difficult to imagine what an incongruent item would look like; the example in (6) illustrates an incongruent trial which is coded as false, because the adult response to the target sentence relative to the situation would be false. Notice that the hypothetical Subj-*only* interpretation of the target sentence will return true.

(6) **Situation:** A goose gets ice cream and cheese, a cat gets ice cream, a frog gets cheese and juice.
Experimenter: 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 direct congruence in the broad question-answer pairs in Experiment 1 contributed to the results, we expect children to exhibit higher accuracy rates on the congruent conditions for both types of sentences with *only*. We also predict possible outcomes as to how children react to the incongruent conditions; for example, we could expect them to exhibit lower accuracy rates for the incongruent conditions for both types as well, for the following reason. If the notion based on the D-tree is on the right track, incongruent Q-A pairs require restructuring a totally different D-tree, while the accommodation required by indirectly congruent Q-A pairs is not as extensive. If children are sensitive to QAC, we will expect the difference between indirectly congruent pairs and incongruent pairs, as well as indirectly congruent pairs and directly congruent pairs. If, on the other hand, children are insensitive to QAC and Crain's puzzle is indeed a product of the grammar of Subj-*only* being acquired later than that of VP-*only*, our manipulation should not qualitatively change the accuracy rates that we have seen in Experiment 1.

Let us take a closer look at the incongruent conditions. If children are sensitive to QAC, they are confronted with a very broken – almost ungrammatical – Q-A pair, and might adopt a variety of strategies which could be revealing in their own right in response: (i) they might simply guess if they cannot make 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 sentences as if Subj-*only* were VP-*only* and vice versa, or (iv) they might ignore the question they heard and not even try to accommodate an appropriate sub-question and therefore interpret

the target sentence according to the syntactic position of *only* (possibly with the same preference for a VP-*only* interpretation as observed in Experiment 1).

What kind of reactions would we expect for the possible strategies? If they adopt strategy (i), we expect the accuracy rates for the incongruent conditions to be around 50% across the board.[24] If they adopt strategy (ii), we also expect 50% of accuracy – not because they use "yes" and "no" randomly but because they would say "yes" indiscriminately.[25] If one answers with "yes" all the time, they would be accidentally correct for the half of the time and wrong for the rest of the time, given the truth of our items being counter-balanced. If they adopt strategy (iii), namely, if their responses are guided by a desire to respect QAC at the expense of the FCC-*only*, we expect them to interpret the target sentences as if the position of *only* was swapped. On this strategy, the QAC cue for the location of F in the answer is taken to be decisive. For example, Subj-*only* preceded by an object-question should be interpreted as if it were VP-*only*, and similarly VP-*only* sentences preceded by a subject-question should be interpreted as if it were Subj-*only*.[26] As a result, we expect the accuracy rates for the incongruent items to be close to 0%, on this strategy. Finally, if children adopt strategy (iv) – that is, if they ignore the question and follow the FCC-*only*, we expect them to interpret the target sentences as adults do. On this strategy, the position of *only* in the target sentence is taken to be decisive for determining the location of F. This should produce accuracy rates close to 100% for the incongruent items.[27]

These predicted accuracy rates are, of course, idealized. It could very well be that children's response pattern is the product of a mix of these strategies. Among these four strategies, we are particularly interested in the last two strategies. We would like to find a way of assessing whether children pay attention to QAC more than to the syntactic position of *only* to determine the associate of *only*. In order to do that, we implemented the basic design in Table 2 in two

---

[24] One might suspect that participants' preferred strategy would be to reject the sentence plainly out of confusion (caused by encountering an incongruent Q-A pair). Given that in our design truth was counter-balanced, we would expect the "chance-level" performance with that "all-reject" strategy as well. That is, if one rejects all the incongruent items, (since half of the trials are coded as true and the other half as false) they would be accidentally correct for the half of the time and wrong for the rest half of the time.

[25] To avoid presupposition failure, the prejacent of the *only*-sentences are always made true in the experiment.

[26] 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.

[27] However, we all by now know that children have a bias towards a VP-*only* interpretation, so the odds for taking this strategy are high that the accuracy rates look very similar to those in Experiment 1.

ways. In Experiment 2A, the attachment site of *only* is treated as a within-subjects factor while the question type is 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* is a between-subjects factor and the question type is 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. That way, the constant cue for focus comes from the attachment site of *only* in Experiment 2B. The design is summarized in Table 4.

Table 4: Design of Experiment 2A and 2B

	Subj-Q	**Obj-Q**
Subj-*only*	Congruent	Incongruent
VP-*only*	Incongruent	Congruent

Exp. 2A

	Subj-Q	**Obj-Q**
Subj-*only*	Congruent	Incongruent
VP-*only*	Incongruent	Congruent

Exp. 2B

The idea behind this manipulation is the following; the levels of the between factors will stay constant for participants, and so we hope that it will provide a stronger cue for determining the location of F in the target sentences for both experiments than the levels of the within factors, which vary within each participant. For example, in Experiment 2A, the QAC cue is expected to provide a stronger cue for determining the location of F than the cue of the syntactic position of *only*. In Experiment 2B, the syntactic cue is expected to provide a stronger cue for determining the location of F than the QAC cue.

A comparison of Experiment 2A and 2B would provide us with an indication whether the factors are equally important to them. To see this more concretely, consider what would be expected if the QAC cue is highlighted for the child as strongly in Experiment 2A as the attachment site cue is highlighted in Experiment 2B. 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); we expect close to 100% accuracy rates for incongruent items. The graphs in Figure 3 illustrates the predictions.[28]

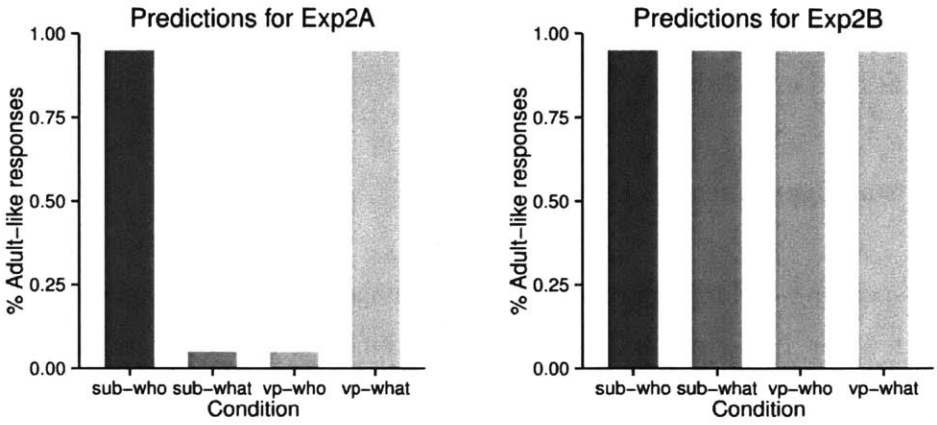


Figure 3: Predictions of Experiments 2A and 2B

If 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*, then the actual results of Experiment 2A should mimic its idealized predictions in Figure 3 more closely than the results of Experiment 2B would do, and if the attachment site of *only* provides a stronger cue the opposite should be true.

2.3.2 Procedure

The presentation of the task and 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 object-question. Then Kermit utters the target sentences, for the participants to judge. We also encouraged them to provide justification for negative answers. Materials were prepared based on the ones in Experiment 1 with some addition of characters that are popular among children. The list of the target items can be found in Appendix 2.

[28] 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.

There were 8 target items (4 congruent Q-A pairs and 4 incongruent Q-A pairs; see Table 3) as well as 4 filler items without *only* within one session. The truth was counter-balanced as discussed in §2.3.1. The presentation order was pseudo-randomized, and we prepared two different lists of the items for each sub-experiments randomly assigned to each participant. The whole session took about 10 minutes per participant. Experiments took place either at local daycares or at Boston Children's Museum.

2.3.3 Results of Experiments 2A and 2B

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

Figure 4 summarizes the results of Experiment 2A, in which question type was the between-subjects factor. Error bars indicate 95% confidence interval (they do so in Figure 5 as well). 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. In the former condition, Subj-*only* sentences constitute congruent Q-A pairs, and VP-*only* sentences constitute incongruent Q-A pairs. This group yielded accuracy rates for Subj-*only* sentences of 72.9% and for VP-*only* sentences of 31.3%. On the other hand, in the object-question condition Subj-*only* sentences make incongruent Q-A pairs and VP-*only* sentences make congruent Q-A pairs. The accuracy rates for this group were 6.3% for Subj-*only* and 95.8% for VP-*only*. The accuracy rates for each test item can be found in Appendix 2.

Statistical analysis of these results (using maximally specified GLMEM) reveals main effects of attachment site (p = .009) and of the question type (p = .001), and importantly, an interaction (p < .001).[29] The summary of the analysis is presented in Table 5. These results combined indicate that congruent items have higher accuracy rates than incongruent items, and this is true irrespective of question type and attachment site of *only*.

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

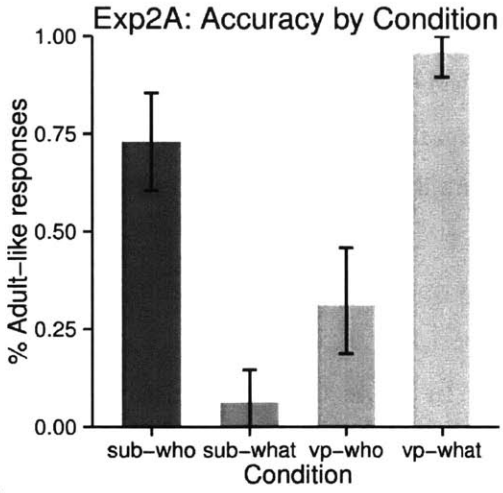


Figure 4: Results of Experiment 2A

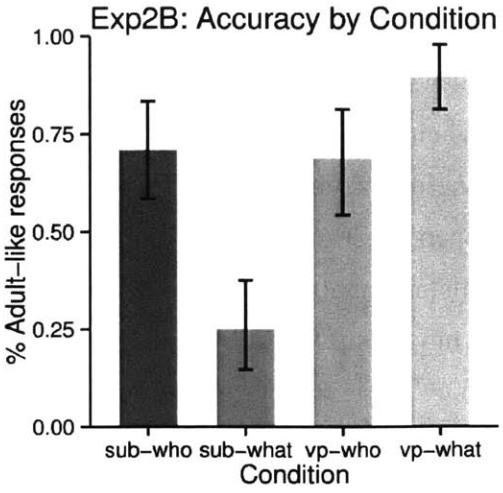


Figure 5: Results of Experiment 2B

Table 5: Statistical analysis of the results of Experiment 2A

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-11.7469	5.4265	-2.165	0.030408 *
attachmentVp	21.6620	8.3379	2.598	0.009377 **
qTypeWho	18.3089	5.7522	3.183	0.001458 **
order2	-1.1454	0.8691	-1.318	0.187545
attachmentVp:qTypeWho	-30.5889	8.9078	-3.434	0.000595 ***

Table 6: Excerpt of the statistical analysis of the results of Experiment 2B

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.1318	0.6419	-1.763	0.077872 .
attachmentVp	3.5516	1.0658	3.332	0.000861 ***
qTypeWho	1.9927	0.7085	2.812	0.004916 **
order2	-0.4896	1.0126	-0.483	0.628771
attachmentVp:qTypeWho	-4.0246	1.1401	-3.530	0.000416 ***

Figure 5 summarizes the results of Experiment 2B, in which attachment site 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. In the Subj-*only* group, when a target sentence is preceded by a subject-question, it makes a congruent pair, while it is

preceded by an object-question, it makes an incongruent pair. This group yielded accuracy rates for subject-question of 70.8%, and for object-question of 25%. By contrast, in the VP-*only* group, when a sentence is preceded by a subject-question it constitutes an incongruent Q-A pair while when a sentence is preceded by an object-question it becomes a congruent Q-A pair. This group yielded accuracy rates for subject-question of 68.8% and for object-question of 89.6%. See Appendix 2 for accuracy rates for each test item.

Statistical analysis of these results using maximally specified GLMEM reveals main effects of attachment site ($p < .001$) and of the question type ($p = .005$) and again an interaction ($p < .001$).[30] The excerpts of the analysis is presented in Table 6. These results indicate that congruent items have higher accuracy rates than incongruent items, and that is true irrespective of question type and attachment site of *only*.

2.3.4 Discussion of Experiments 2A and 2B

The clearest result of Experiments 2A and 2B is that congruent answers are understood significantly more often in an adult-like way than incongruent answers, furthermore it was so irrespective of question type and attachment site of *only*. Specifically, in Experiment 2A, congruent Subj-*only* had a higher accuracy rate (73%) than the incongruent VP-*only* condition (31%). In Experiment 2B, congruent Subj-*only* had again a higher accuracy rate (70%) than the incongruent Subj-*only* condition (25%). And of course, the following was observed as well; congruent VP-*only* was more accurately answered than incongruent Subj-*only* (Exp2A), and congruent VP-*only* was more accurately answered than incongruent VP-*only* (Exp2B). This shows, first of all, that our participants were sensitive to QAC. This is an observation that, to our knowledge, has not been made previously in the acquisition literature. Our data, in

[30] In the analysis of Experiment 2B, the maximally specified GLMEM converged. However, for a better comparison to Experiment 2A, the summary of the experiment in Table 6 only shows excerpts of the whole output of the analysis. The whole output of the analysis is as follows:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.1318	0.6419	-1.763	0.077872 .
attachmentVp	3.5516	1.0658	3.332	0.000861 ***
qTypeWho	1.9927	0.7085	2.812	0.004916 **
order2	-0.4896	1.0126	-0.483	0.628771
attachmentVp:qTypeWho	-4.0246	1.1401	-3.530	0.000416 ***
attachmentVp:order2	1.0682	1.6607	0.643	0.520084
qTypeWho:order2	0.8226	1.1297	0.728	0.466541
attachmentVp:qTypeWho:order2	-0.2659	1.7988	-0.148	0.882493

particular the fact that the accuracy rates for congruent Subj-*only* yielded above 70% of accuracy, 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.

We have observed in Experiment 1 that children's behavior on VP-*only* is essentially adult-like (80% when averaged across truth). But now that we saw in Experiments 2A and 2B that adult-like response rate for VP-*only* could be as low as 31% (Exp 2A), when the Q-A pairs are incongruent. By parallel reasoning as above, the claim that children's competence with VP-*only* is truly adult-like has to be questioned. If they really have an adult-like command of VP-*only*, they should have ignored the preceding question and compute the meaning of the target sentence as conforming to the FCC-*only* when they encountered incongruent Q-A pairs. Our participants, however, interpreted in such cases 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* (= strategy (iii)). In other words, children seem to ignore the position of *only* in both Subj-*only* as well as VP-*only* when they search for an associate and so interpret the incongruent sentences as if the position of *only* was swapped. This entails that children's command of VP-*only* is not fully adult-like either. An example of children's justifications below for the incongruent VP-*only* sentences indicate that it is indeed the case that children were taking the strategy (iii).

(7) **Situation:** A goose gets juice and ice cream, a cat gets ice cream, a frog gets juice.
Experimenter: Now Mr. Cat, can you tell me who got ice cream?
Mr. Cat: Oh, I know who got ice cream. The cat only got ice cream.
Child: Wrong… Because goo got ice cream too.

In Chapter 3, we show that adults employ the response strategy to incongruent Q-A pairs that they ignore QAC and instead identify the associate of *only* according to the FCC-*only* (= strategy (iv)).

Note that the results of Experiment 2B seem to conform less well to our idealized predictions. Specifically, although we see higher accuracy rates for incongruent items in Experiment 2B than in Experiment 2A, they are strikingly symmetric and the accuracy with incongruent items are far from 100%; also Subj-*only* generates fewer adult-like responses than

VP-*only*. What do the results mean? We divided the experiment in two ways so that Experiment 2A would have the QAC cue consistently (i.e., the QAC cue is designed to be stronger than the syntactic cue) and Experiment 2B would have the syntactic cue consistently (i.e., the syntactic cue is designed to be stronger than the QAC cue). We hypothesized that, if QAC provides a stronger cue in processing than the attachment of *only*, the actual results of Experiment 2A should look like its idealized predictions more closely than the results of Experiment 2B would do, and that if the attachment site of *only* provides a stronger cue the opposite would be true. If, on the other hand, both of the cues are equally influential, then the actual results of Experiment 2A and those of Experiment 2B would look like their predictions. Now that we observe the results where the outcome of Experiment 2A looks like its predicted figure, and that of Experiment 2B does not seem to conform well to its predicted figure. This suggests that while children take the QAC cue to be a reliably decisive factor in determining the associate of *only*, they pay less attention to the syntactic cue when they have to identify the associate of *only*. Instead, they seem to adopt a similar strategy to the one they use in Experiment 1 when the QAC cue varies within participants.

Finally, comparing Subj-*only* and VP-*only* at the same levels of the congruence, one notices that there is still an asymmetry between Subj-*only* and VP-*only*. This is apparent in the significant effects of attachment site in both experiments. Which of the hypotheses do the results conform to the better? The comparison between different hypotheses, just with the part of the prediction on the comprehension of Subj-*only* and VP-*only*, is repeated from Table 1 in §1.4 below.

Table 7: Comparison of predictions on Subj-only vs. VP-only comprehension by different hypotheses

Hypotheses	**Comprehension of Subj-*only***	**Comprehension of VP-*only***
No grammar of Subj-*only*	Not good	Good
Difficult accommodation of subject-Q	Regardless of the attachment site; QAC satisfied = good, QAC unsatisfied = not good	
Difficult accommodation of subject-Q + difficult grammar of Subj-*only*	Performance modulated by QAC, but not good	Performance modulated by QAC, but good

The results of the current set of experiments conform well with the predictions of the third, hybrid hypothesis. It consists of two components; the difficulty of accommodation of subject-questions, and the difficulty of the grammar of Subj-*only*. What are their components? Thoughts on what might be going on are presented in the following discussion section.

2.4 General discussion and conclusion

2.4.1 Summary

Previous literature on the acquisition of *only* observed an asymmetry between the comprehension of Subj-*only* and VP-*only* (Crain's puzzle) and concluded that the grammar of Subj-*only* is more difficult for children to acquire than that of VP-*only*. However, we argued that the experimental strategy employed in that literature such as Picture-verification tasks and TVJTs overlooked a complication in their materials that might have led to underestimating children's competence of *only*. Specifically, comprehending sentences with *only* as an answer to a broad question (whether it is explicit as in a TVJT or implicit as in a Picture-verification task) requires accommodation of an appropriate sub-question to satisfy QAC, because of the nature of *only* being a focus-sensitive particle. We hypothesized that the difficulty with accommodation of a suitable sub-question 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-*only* and VP-*only* sentences in a fully crossed design.

The experiments revealed that children's responses for directly congruent pairs were 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 in certain circumstances (e.g., when they have to accommodate a suitable sub-question to satisfy QAC), that they seem to misanalyse Subj-*only* for VP-*only*. However, the results also show that the accommodation is not the whole story; even directly congruent Subj-*only* sentences were understood less well compared to the directly congruent VP-*only* sentences, and similarly, incongruent Subj-*only* sentences were almost always misanalysed while incongruent VP-*only* sentences were misanalysed less often. This seems to be a residue of the same asymmetry that Crain et al. initially identified. We also have seen the asymmetry favoring VP/Obj association

under the indirectly congruent conditions of our experiments. To account for why this residue is brought about, I propose that a set of object-questions is by default generated when a broad question is presented in such cases as our experimental stories. This can be achieved, assuming that Topic of the sentence (often the case, the subject) serves as a sorting key for generating D-tree of a broad question. The proposal and its supporting evidence that suggests generating a set of subject-questions are less well supported are based on a few assumptions in the literature of QUD stack and D-tree, as well as the semantics of the multiple *wh*-questions. The next subsection will spell out the proposal in more detail.

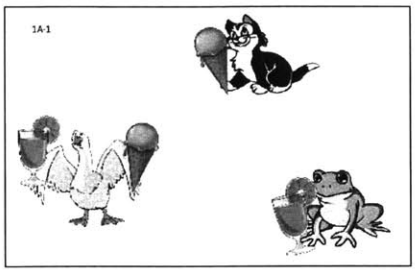
The structure of the rest of this section is the following. The discussion in §2.4.3 provides a piece of supporting evidence from an experimental study that QAC plays a significant role in children's comprehension of sentences with *only*. The subsection §2.4.4 discusses what we can say from the results of incongruent sentences. The subsection §2.4.5 adds some new data from child corpus, to evaluate the prediction regarding the production of sentences with *only*. By comparing different hypotheses we entertained in §1.4, it is made clear what to investigate with the adult population to confirm the hybrid hypothesis, and I will conclude the chapter.

2.4.2 Proposal: Topic as a sorting key for generating D-tree of a broad question

What makes the accommodation of subject-questions more difficult than that of object-questions, when a broad question such as "*What happened?*" is presented? Assuming Roberts' (1996/2012) QUD stack and Büring's (2003) D-tree, a broad question is broken down into a set of sub-questions; and there are multiple ways the sub-questions are constructed. For example, one can break down a broad question into a set of object-questions or a set of subject-questions, illustrated respectively in (8-9).[31]

[31] There are other ways to list a set of sub-questions, such as "How many animals are there?", "Did each animal receive something?", and so on. The proposed D-trees in (8) and (9) could be understood to be under a sub-question that has multiple *wh*-phrases; "Who is holding what?" – I point out the similarity in D-tree configuration between a broad question and a multiple *wh*-question shortly.

(8) *Broad question broken down into a set of object-questions* (Subject as a sorting key)[32]



What is happening in the picture?

[Question about the cat] — *[Question about the goose]*

What is the cat holding? — What is the goose holding? …

Is the cat holding ice cream? Is the cat holding juice? …

(9) *Broad question broken down into a set of subject-questions* (Object as a sorting key)

What is happening in the picture?

[Question about ice cream] — *[Question about juice]*

Who is holding ice cream? — Who is holding juice? …

Is the cat holding ice cream? Is the goose holding ice cream? …

Either of the configuration works equally fine as a D-tree for a broad question, in that the interlocutors will eventually figure out what is happening in the picture by answering each of the sub-questions. However, I propose that the D-tree configuration in (8) is generated with less processing cost than (9); and therefore is the default strategy to configure a D-tree of sub-questions under a broad question.[33] One support to this claim can be found in the theory of Information Structure.

It is important to understand what defines "Topic" with the ability to be used naturally as a sorting key, and that the subject of a sentence tends to be a Topic of the sentence (Kuno 1972, Reinhart 1981, Lambrecht 1996 Vallduví and Engdahl 1996, Erteschik-Shir 1997, 2007, and

[32] In the sense of a "sortal key" in Kuno (1982).

[33] At least, in such cases where the subject does some action towards the object.

much subsequent research). Reinhart (1981) observes two major approaches to the definition of Topic: (i) a Topic is an element that the sentence is about, and (ii) a Topic is an element that is old information (i.e., being *given*).[34] In our case, both the animals and the items of food are present in the picture, so both can be considered old information. However, if one is asked to describe the picture in (8), they would use sentences that are sorted by the animal; e.g., "In the picture, the goose is holding juice and ice cream, the cat is holding ice cream, and the frog is holding juice" In our experimental examples, and in general, sentential subjects tend to be Topics, and the rest of the sentence tends to be the Topic's comment. Configuring sub-questions which are sorted according to Topics in a given situation seems natural enough to be the default strategy when breaking down a broad question into sub-questions.

The idea that the D-tree configuration in (8) is less costly than (9) can be independently motivated by seeing the D-tree configurations as an analogy to a superiority-obeying multiple *wh*-question and a superiority-violating multiple *wh*-questions, respectively.[35] For each multiple *wh*-question that has a pair-list reading, a set of hierarchically nested alternatives (a set of sets of propositions) has been proposed for the denotation of the question (referred to as "a family of questions"; Roberts 1996, Hagstrom 1998, Krifka 2001, Büring 2003, Fox 2012, Nicolae 2013, Kotek 2014, a.o.). The following pair of denotations of multiple *wh*-questions are cited from Kotek (2014). Consider there are three persons, John, Mary, and Bill and three books, Moby Dick, War and Peace, and Oliver Twist as relevant alternatives.

(10) A family of questions denotation for "*Which* student read *which* book?"

$$\left\{ \left\{ \begin{array}{l} \text{John read Moby Dick} \\ \text{John read War and Peace} \\ \text{John read Oliver Twist} \end{array} \right\}, \left\{ \begin{array}{l} \text{Mary read MD} \\ \text{Mary read WP} \\ \text{Mary read OT} \end{array} \right\}, \left\{ \begin{array}{l} \text{Bill read MD} \\ \text{Bill read WP} \\ \text{Bill read OT} \end{array} \right\} \right\}$$

(11) A family of questions denotation for "*Which* book did *which* student read?"

$$\left\{ \left\{ \begin{array}{l} \text{John read Moby Dick} \\ \text{Mary read MD} \\ \text{Bill read MD} \end{array} \right\}, \left\{ \begin{array}{l} \text{John read War and Peace} \\ \text{Mary read WP} \\ \text{Bill read WP} \end{array} \right\}, \left\{ \begin{array}{l} \text{John read Oliver Twist} \\ \text{Mary read OT} \\ \text{Bill read OT} \end{array} \right\} \right\}$$

via Kotek (2014: 59)

[34] See Tancredi's (2016) research for discussion on the necessity of distinguishing the phenomena of Topic, Focus and Givenness. One is not reduced to another notion.

[35] When they have the pair-list reading, rather than a single answer reading.

When a multiple *wh*-question such as in (10) is presented, a set of sets of propositions, bundled by the subject, is created. In order to entertain the set of sets of propositions, bundled by the object and presented in (11), one has to assume a superiority-violating multiple *wh*-questions.

Let us turn to how to break down a broad question, specifically in our case, where different subjects take actions toward different objects, i.e., a situation where two arguments could be potential Topics or Foci. In such cases, since each of the arguments could be understood as variables, a broad question can be paraphrased as if it were a multiple *wh*-question, i.e., "*Who got what?*"[36] Assuming the semantics of multiple *wh*-questions as above, it is possible that the configuration with the subject as a sorting key is given preference, because subjects are good Topics and making other elements as Topic would violate superiority.

This assumption leads to a consequence that subject-questions are easier to accommodate than object-questions, when a broad question is presented. To see this, we will first review how the principle of QAC satisfies the congruent Q-A pair, and requires accommodation of sub-questions in indirectly congruent and incongruent Q-A pairs. Consider there are three animals: a goose, a cat, a frog, and three items of food: ice cream, juice and cheese.

(12) Congruent case: "Who got ice cream?" – "Only the cat_F got ice cream."

i) $[\![Q]\!]^o$ = {the cat got ice cream, the goose got ice cream, the frog got ice cream}

ii) $[\![A]\!]^f$ = {the cat got ice cream, the goose got ice cream, the frog got ice cream}[37]

$\therefore [\![Q]\!]^o \subseteq [\![A]\!]^f$, QAC is satisfied.

(13) Incongruent case: "What did the cat get?" – "Only the cat_F got ice cream."

i) $[\![Q]\!]^o$ = {the cat got ice cream, the cat got juice, the cat got cheese}

ii) $[\![A]\!]^f$ = {the cat got ice cream, the goose got ice cream, the frog got ice cream}

$\therefore [\![Q]\!]^o \not\subset [\![A]\!]^f$, QAC is not satisfied.

The following illustrations show how to accommodate an indirectly congruent question (i.e., a broad question "*What happened?*") for VP-*only* sentences, with an assumption that the sub-

[36] It is expected that the number of the arguments and/or adjuncts that could be topics or foci in a given situation should correspond to the number of the *wh*-phrase. Confirming this issue is left for future research.

[37] To be exact, this is the focus semantic value of the prejacent of the answer. As discussed in §1.3, I assume that the focus sensitive operator is anaphoric to the QUD. To obey QAC, the assumption requires the associate of *only* be on the expression that gives a focus semantic value that is a superset of the question, i.e., the focus semantic value of the prejacent of the sentence with *only* has to be the superset of the question.

questions are sorted by Topic of the sentence – the subject. First, the broad question in our case is understood as a multiple *wh*-question, "*Who got what?*", and the family of questions is generated, using the Topic as a sorting key. The focus semantic value of the prejacent of the answer is obtained as a form of a set of propositions, and the principle of QAC tells us that the ordinary semantic value of the question is not contained by the focus semantic value of the answer. To observe QAC, one tries to accommodate a question so that it is contained by the focus semantic value of the answer. By "pruning" the two branches, a congruent question can be obtained.

(14) Indirectly congruent case: "What happened?" – "The cat only got ice cream$_F$."

i) $[\![Q]\!]^o =$

$$\left\{ \begin{array}{lll} \textit{Sub-Q about the cat} & \textit{Sub-Q about the goose} & \textit{Sub-Q about the frog} \\ \left\{ \begin{array}{l} \text{The cat got ice cream} \\ \text{The cat got juice} \\ \text{The cat got cheese} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The goose got IC} \\ \text{The goose got juice} \\ \text{The goose got cheese} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The frog got IC} \\ \text{The frog got juice} \\ \text{The frog got cheese} \end{array} \right\} \end{array} \right\}$$

ii) $[\![A]\!]^f =$ {the cat got ice cream, the cat got juice, the cat got cheese}

$\therefore [\![Q]\!]^o \not\subset [\![A]\!]^f$, QAC not satisfied, so accommodate a sub-question!

iii) [accommodated Q]$^o =$

$$\left\{ \begin{array}{lll} \textit{Sub-Q about the cat} & \sout{\textit{Sub-Q about the goose}} & \sout{\textit{Sub-Q about the frog}} \\ \left\{ \begin{array}{l} \text{The cat got ice cream} \\ \text{The cat got juice} \\ \text{The cat got cheese} \end{array} \right\}, & \left\{ \begin{array}{l} \sout{\text{The goose got IC}} \\ \sout{\text{The goose got juice}} \\ \sout{\text{The goose got cheese}} \end{array} \right\}, & \left\{ \begin{array}{l} \sout{\text{The frog got IC}} \\ \sout{\text{The frog got juice}} \\ \sout{\text{The frog got cheese}} \end{array} \right\} \end{array} \right\}$$

= {the cat got ice cream, the cat got juice, the cat got cheese}

$\therefore$ [accommodated Q]$^o \subseteq [\![A]\!]^f$

Assuming that the broad question is treated as a set of questions sorted by the Topic, we see that one only needs to modify two branches of sub-questions in (14), when a VP-*only* sentence follows the broad question.

Next, the following shows how to accommodate an indirectly congruent question for Subj-*only* sentences, with the assumption that the sub-questions are sorted by Topic of the sentence, just as before. This time, the accommodation involves reconfiguration of the way the set of questions is sorted and bundled, as illustrated in (15). The step (i) is the same as the case in (14).

The focus semantic value of the prejacent of the answer is something in (15-ii), since this is a Subj-*only* sentence. The principle of QAC tells us that the ordinary semantic value of the question is not contained by the focus semantic value of the answer. To observe QAC, one tries to accommodate a question so that it is contained by the focus semantic value of the answer, but none of the members of the family of questions is qualified. Therefore one reconfigures the broad question as a family of subject-questions, as in (15-iii). The broad question itself does not prohibit a family of subject-questions from being built for it, but it amounts to a superiority-violating question. Finally one can accommodate a question that is contained by the focus semantic value of the answer, by pruning the two branches, to satisfy QAC. Under this strategy, the accommodation involves more steps than (14).

(15) *Indirectly congruent case:* "What happened?" – "Only the cat_F got ice cream."

i) $[\![Q]\!]^o =$

$$\left\{ \begin{array}{lll} \textit{Sub-Q about the cat} & \textit{Sub-Q about the goose} & \textit{Sub-Q about the frog} \\ \left\{ \begin{array}{l} \text{The cat got ice cream} \\ \text{The cat got juice} \\ \text{The cat got cheese} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The goose got IC} \\ \text{The goose got juice} \\ \text{The goose got cheese} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The frog got IC} \\ \text{The frog got juice} \\ \text{The frog got cheese} \end{array} \right\} \end{array} \right\}$$

ii) $[\![A]\!]^f = \{$the cat got ice cream, the goose got IC, the frog got IC$\}$

$\therefore [\![Q]\!]^o \not\subset [\![A]\!]^f$, reconfigure the D-tree to accommodate a sub-question!

iii) [reconfigured $Q]\!]^o =$

$$\left\{ \begin{array}{lll} \textit{Sub-Q about ice cream} & \textit{Sub-Q about juice} & \textit{Sub-Q about cheese} \\ \left\{ \begin{array}{l} \text{The cat got ice cream} \\ \text{The goose got IC} \\ \text{The frog got IC} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The cat got juice} \\ \text{The goose got juice} \\ \text{The frog got juice} \end{array} \right\}, & \left\{ \begin{array}{l} \text{The cat got cheese} \\ \text{The goose got cheese} \\ \text{The frog got cheese} \end{array} \right\} \end{array} \right\}$$

iv) [accommodated $Q]\!]^o =$

$$= \left\{ \begin{array}{lll} \textit{Sub-Q about ice cream} & \textit{\sout{Sub-Q about juice}} & \textit{\sout{Sub-Q about cheese}} \\ \left\{ \begin{array}{l} \text{The cat got ice cream} \\ \text{The goose got IC} \\ \text{The frog got IC} \end{array} \right\}, & \left\{ \begin{array}{l} \sout{\text{The cat got juice}} \\ \sout{\text{The goose got juice}} \\ \sout{\text{The frog got juice}} \end{array} \right\}, & \left\{ \begin{array}{l} \sout{\text{The cat got cheese}} \\ \sout{\text{The goose got cheese}} \\ \sout{\text{The frog got cheese}} \end{array} \right\} \end{array} \right\}$$

$=$ {the cat got ice cream, the goose got IC, the frog got IC}

$\therefore [\![$accommodated $Q]\!]^o \subseteq [\![A]\!]^f$

Two ideas are presented to support the proposal that a broad question is denoted (in our case, by default) as a set of object-questions (or questions *about* Topic): Topic as a sorting key is a general phenomenon in the theory of Information Structure, and the family of questions using the subject as a sorting key correspond to the semantics of a superiority-obeying multiple *wh*-question.

If the proposal is on the right track, it explains why the accommodation of subject-questions is more difficult than that of object-questions, when a broad question is presented. It might also explain the residue of Crain's puzzle in the directly congruent (and possibly incongruent) cases. This is a speculation that needs further careful thoughts, but if, in the first place, placing focus in the subject is not desirable (because subjects tend to be topic, or because it violates superiority), it might follow that computing the semantics of subject-questions is relatively difficult than object-questions. It might also follow that computing the focus semantic value of Subj-*only* (i.e., focus on the subject) is relatively difficult than VP-*only* (i.e., focus on object in our cases).

It is important to note that the asymmetry in difficulty between Subj-*only* and VP-*only* is now the issue of processing and discourse strategies interlocutors take. It should not only affect children. It is expected that adults also exhibit the asymmetric behavior in processing Subj-*only* and VP-*only* as a response to a broad question.

2.4.3 Manipulation of preceding question

We are aware of one other experiment in the literature, Experiment 3 of Notley et al. (2009), that yielded similar results to ours; children exhibited (unexpectedly) high rates of adult-like responses to Subj-*only* sentences. Notley et al. investigated whether Mandarin-speaking four-year-old children interpret sentences with a focus construction, which is claimed to be similar to sentences with *only*, in an adult-like way. The participants in Experiment 3 had participated in their Experiment 2. To understand the context behind the experiments better, let us first briefly summarize Experiment 2; the target sentences were Subj-*only* sentences such as in (16). The word *zhiyou* is translated as *only*. In the story, Mr. Pig gets a silver and a gold coin, while Mr. Horse only got a gold coin. Adults responded "yes" 100% of the time, while children (n = 20, M = 4;7) accepted the sentence 10% of the time.

(16) Example target sentence in Experiment 2:

Zhiyou	zhu-xiansheng	nadao-le	yinse	yinbi.
only	pig-sir	get-ASP	silver	coin

'Only Mr. Pig got a silver coin.'

In Experiment 3, they tested the children who exhibited very strong bias toward VP-*only* interpretations in Experiment 2. They tested as stimuli the subject focus construction (*she...de*), which is claimed to be similar to *zhiyou* (only), as in (17). In the story, Winnie the Pooh jumps over the fence, and Tigger jumps over the house and the fence. According to them, Mandarin *shi...de* constructions behave like English *only*- and Mandarin *zhiyou*-sentences, rather than English cleft sentences as some other literature claims. Under the situation described above, Mandarin-speaking adults would respond "no" to the sentence, on the grounds that Winnie the Pooh also jumped over the fence.

(17) Example target sentence in Experiment 3:

Shi	Tiaotiaohu	tiaoguo-le	liba.
FOC	Tigger	jump-over-ASP	fence

'It was Tigger who jumped over the fence.'

Their experimental manipulation was to put the sentence in "subject-biased stories," which, interestingly, ended with a subject-question prompt rather than a broad question. The children (n = 14, *M* = 4;6) responded "no" 100% of the time to the sentence as in (17) and justifications that they provided suggest that they rejected the sentence for an adult-like reasoning (e.g., because Pooh also jumped over the fence) 61.9% of the time, while they gave an VP-*only*-biased reasoning (e.g., because Tigger also jumped over the house) 38.1% of the time.[38] It is important to note that the same participants who once exhibited a very strong bias towards VP-*only* interpretation now provide adult-like, appropriately Subj-*only*-biased reasons over 60% of the time. This clearly shows that the children in fact have competence to interpret Subj-*only* sentences in an adult-like way, but something about experimental setting – in our words, incongruence in QAC – hinders them from exhibiting the competence. Even though Notley et

[38] They tested sentences which would yield "yes" responses according to adult-like interpretation, and the "Adult-True" condition yielded 76% of "yes" rate.

al.' experimental design was not intended to test QAC itself, we think that their results provide support for our claim that sensitivity to QAC and the need for accommodation of a sub-question play an important role in Crain's original observation.

2.4.4 What incongruent Q-A pairs tell us about Crain's puzzle

Our data from incongruent pairs are equally significant and informative with regard to the proper characterization of Crain's puzzle. We saw that children's primary strategy for determining the associate of *only* was based on QAC. Our results show that the children identified the associate of *only* with the constituent that corresponds to the *wh*-phrase in the preceding question – even when the FCC-*only* would not permit such association according to the grammar adults hold. This fact reinforces our observation that children are quite sensitive to QAC, or dependent on it, more than they are on the syntactic position of *only*. Importantly, this tendency 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 of *only* – that is, even for VP-*only*, children do not exercise their command of *only* independently of the QAC cue when some distracting QAC cue exists.

As has been noted, the results showed that incongruent Subj-*only* sentences were almost always misanalysed, while incongruent VP-*only* sentences were misanalysed less often. This residue of Crain's puzzle is not incompatible with the proposal that generating a set of object-questions is less costly in such cases as our experimental stories. Future research will have to clarify this issue, and other factors such as scalar presupposition of *only* will have to be integrated to further investigate the semantics of *only*.[39]

Taken together, our data on incongruent Q-A pairs suggests that what distinguishes children from adults in their command of *only* is the status of the FCC-*only*. For adults, it is a principle when computing the meaning of sentences with *only*. By contrast, for children it seems quite possible to ignore it when other factors such as QAC are in conflict with it. It is possible that for children the sentence with *only* is ambiguous with *only* adjoined at IP,[40] and since they do not know the FCC-*only* yet, they assign F on the constituent according to the QAC cue.

---

[39] Hackl et al. (2014, paper in progress) show the evidence that the scalar presupposition of *only* (Bonomi and Casalegno 1993, Guerzoni 2003, Klinedinst 2005, Beaver and Clark 2008, Alxatib 2013, a.o.) is an important factor in determining the difficulty of comprehending a sentence with *only*, with experimental data acquired from adult participants.

[40] Similarly to Crain et al.'s hypothesis of *only* always being adverbial.

2.4.5 Comparison between hypotheses, some production data, and conclusion

We will review how different hypotheses serve on the predictions discussed in §1.4. The prediction table is repeated below. The results from the experiments seem to conform pretty well to the predictions of the hybrid hypothesis regarding the comprehension of Subj-*only* and VP-*only*.

Table 8: Comparison between different hypotheses on Crain's puzzle

<table>
<tr><th></th><th colspan="3">Comprehension</th><th rowspan="2">Production of only-sentences</th></tr>
<tr><th></th><th>Subj-only</th><th>VP-only</th><th>Obj-only</th></tr>
<tr><td rowspan="2">No grammar of Subj-only</td><td>Not good</td><td>good</td><td>= Subj-only,
Not good</td><td>Similar asymmetry to comprehension, S worse than V
(meaning S-only is produced less/Subj-only is misused to mean VP-only, etc.)</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to acquire the grammar of Subj-only/ attributive only.</td></tr>
<tr><td rowspan="2">Difficult accommodation of subject-Q</td><td colspan="2">Regardless of the attachment site;
QAC satisfied = good,
QAC unsatisfied = not good</td><td>= VP-only,
good</td><td>Does not particularly predict production errors</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to be able to accommodate a suitable sub-question when it is implicit in discourse.</td></tr>
<tr><td rowspan="2">Difficult accommodation of subject-Q
+
difficult grammar of Subj-only</td><td>Performance modulated by QAC, but not good</td><td>Performance modulated by QAC, but good</td><td>Better than Subj-only (b/c the required accommoda-tion is object-Q)</td><td>Difference in errors between S and V is smaller than comprehension</td></tr>
<tr><td colspan="4">In order to become adult-like, one needs to be able to accommodate a suitable sub-question when it is implicit in discourse, and to know FCC-only.</td></tr>
</table>

Let us turn to the question regarding the production of *only*. Our preliminary results from the search for the utterances with *only* in the CHILDES corpus (MacWhinney 2000) show intriguing facts. The age range in the scope of search is from 2;0 to 6;6, and children's utterances

that contain the word *only* were extracted with some context before and after the utterance. The total of 311 utterances were found. Each utterance was first coded as "attributive *only* (attaches to DP)" or "adverbial *only* (VP-level)", if the coding was possible. Then the utterances of attributive *only* were further coded as "Subject", "Object", "Predicate nominal", or "Answer (to a question; most of them are fragment answers and require further considerations as to how to treat them)", if the coding was possible. Keeping in mind that these results are still preliminary, a quick summary is presented below.

Table 9: Preliminary results of distribution of only of different types in CHILDES

Age	Total number of utterances with *only*	Attributive *only*	Adverbial *only*	Not categorized
2;0-2;6	19	10 (53%)	7 (37%)	2
2;6-3;0	39	19 (49%)	14 (36%)	6
3;0-3;6	32	18 (56%)	8 (25%)	6
3;6-4;0	42	26 (62%)	12 (29%)	4
4;0-4;6	62	31 (50%)	22 (36%)	9
4;6-5;0	33	21 (64%)	9 (27%)	3
5;0-5;6	33	16 (49%)	12 (36%)	5
5;6-6;0	45	22 (49%)	16 (36%)	7
6;0-6;6	6	3 (50%)	3 (50%)	0

Table 10: Preliminary results of distribution among attributive only

Age	Number of Attributive *only*	Subject	Object	Predicate nominal	Answer	Not categorized
2;0-2;6	10	5 (50%)	1 (10%)	0 (0%)	4 (40%)	0
2;6-3;0	19	4 (21%)	8 (42%)	3 (16%)	4 (21%)	0
3;0-3;6	18	2 (11%)	5 (28%)	4 (22%)	6 (33%)	1
3;6-4;0	26	4 (15%)	4 (15%)	5 (19%)	7 (27%)	6
4;0-4;6	31	8 (26%)	2 (7%)	5 (16%)	9 (29%)	7
4;6-5;0	21	3 (14%)	1 (5%)	6 (29%)	5 (24%)	6
5;0-5;6	16	5 (31%)	1 (6%)	4 (25%)	1 (6%)	5
5;6-6;0	22	2 (9%)	1 (5%)	10 (46%)	7 (32%)	2
6;0-6;6	3	0 (0%)	0 (0%)	1 (33%)	1 (33%)	1

The results are suggestive in that (i) utterances containing *only* can be found as early as two years old, (ii) the proportion of attributive *only* and adverbial *only* are about the same, or the proportion of attributive *only* tends to be slightly larger, (iii) among the attributive uses of *only*,

there are more uses of *only* on the subject than those on the object on average (except for the age 2;6-3;6). If children only had the grammar of VP-*only* and VP-*only* is ambiguous between Subj-*only* and VP-*only* for them, it was expected that children produce no or little Subj-*only* sentences. However, the results show that does not seem right. Do the children, similarly to the comprehension of Subj-*only*, make mistakes with Subj-*only*? Specifically, do they use Subj-*only* to mean VP-*only* (which is expected if children do have the competence of deriving the word order of Subj-*only* but have preference of assigning the associate of the pre-subject *only* to the object)? An examination on the utterances reveals that children do not seem to misuse the sentences with *only*. Some examples of Subj-*only* from relatively young children are presented below, with some context. It is clear that the children mean what the Subj-*only* sentence would mean by using Subj-*only*.

(18) Child: These blocks are part of the sofa, Sleepy.
Mother: Oh can I jump on them?
Child: No.
Mother: Oh okay. Sleepy you better come here.
Child: Only I get to touch them.[41]
Mother: Umm. (age 2;11, Providence corpus, Demuth et al. 2006)

(19) Investigator: How about sweet potetoes?
Child: White potetoes!
Invstgtr: White potetoes?
Child: I'm not allergic to anything.
Invstgtr: Oh, good!
Child: I can eat anything I want.
Invstgtr: Who's allergic to white potetoes?
Child: No one.
Invstgtr: No one?
Child: Only the dog is allergic to that.[42]
(age 4;4, Weist corpus, Weist et al. 2009; Weist & Zevenbergen 2008)

[41] The video and audio are available for the Providence corpus. Stress on "I" was observed in this utterance.
[42] The audio is available for the Weist corpus. Vowel lengthening and stress on "dog" was observed in the utterance.

Crain's hypothesis, which assumes the grammar of Subj-*only* is added to the child grammar later in the development, would not predict there to be a number of correct Subj-*only* utterances at an early stage of the development. By contrast, the accommodation hypothesis does not particularly predict production errors, and the hybrid hypothesis predicts few errors, with possibly more errors with Subj-*only* than VP-*only*. Conducting further analysis on the production data is under way, which surely will benefit the acquisition literature on *only*.

The results presented so far, comprehension and production together, are suggesting that Crain's puzzle can be reduced to the two components: the asymmetry in processing load regarding discourse strategies between comprehension of Subj-*only* and VP-*only*, and the different status of FCC-*only* that children have, compared to adults. Notice that the first component concerns a general issue in pragmatics, and therefore it should not only affect children. It is expected that adults also exhibit the asymmetric behavior in processing Subj-*only* and VP-*only* as a response to a broad question. Furthermore, if QAC plays an important role, as seen with child participants, then adults are also expected to exhibit better performance with the congruent Q-A pairs than with the indirectly congruent Q-A pairs and the incongruent Q-A pairs. We have seen in this chapter that children's primary strategy to approach the incongruent Q-A pairs was to respect QAC at the expense of observing FCC-*only*; what is the primary strategy for adults? Also, among the columns in the prediction table above, the column on the comprehension of Obj-*only* has not been examined yet. The experiments in the next chapter confirm the expectations and provide data on the processing of Obj-*only*.

Chapter 3

Processing of QAC in sentences with *only*[*]

3.1 Hypothesis and predictions

Chapter 2 presented the following: (i) children have competence in interpreting Subj-*only* sentences when the QAC is satisfied, (ii) children's non-adult-like behavior increases when they encounter incongruent Q-A pairs, and (iii) Crain's puzzle can be restated from "Why is VP-*only* easier than Subj-*only* for children?" to "Why is accommodation of an object-question easier than that of a subject-question for children?" In this Chapter 3, we ask if Crain's puzzle applies to adults' processing of language. The question becomes more general: Is the accommodation of an object-question easier than that of a subject-question in adult grammar?

As an adult, one might not find a difference in comprehending Subj-*only* sentences and VP-*only* sentences; and hence, they may ask what the point of the question is in the first place. The answer to the question whether the accommodation of an object-question is indeed easier than that of a subject-question will benefit the understanding of children's acquisition of *only*. If we observe adults exhibiting a greater processing difficulty with Subj-*only* than with VP-*only*, as seen in children, then that would suggest the reason children do not exhibit adult-like behavior is due to the processing load being too large in accommodation of sub-questions and not related to their lack of necessary grammar. The rationale behind the claim is the following. If children lack the necessary grammar at first and acquire it as they grow up, that means that the parameter of having the grammar about Subj-*only* is categorical; and adults, who are obviously in the

[*] For the experiments reported in this chapter, I would like to express many thanks to Erin Olson for setting up experiments on Ibex and AMTurk, fixing problems promptly, running and teaching me statistical analyses.

adult-like stage, are expected to exhibit good performance across the board. By contrast, if children's difficulty with Subj-*only* is due to the profile of Subj-*only* itself (i.e., being difficult for the comprehending individual to accommodate a suitable sub-question), that means listeners have continuous difficulty comprehending sentences with *only*. In addition, Subj-*only* is expected to be difficult compared to VP-*only*; were as, adults are expected to exhibit different levels of performance in comprehension/processing depending on the ease of accommodation.

3.2 Experiments

3.2.1 Method

We conducted series of timed inference tasks (Experiment 1, 1-prime, 2A and 2B) to investigate comprehension of sentences with *only* by adult English speakers. The task was more challenging, since we anticipated ceiling effects for accuracy when we tested adults with the same TVJTs that were run with children. The participants only received the story content by reading written sentences without pictures on a computer screen; and once they understood the situation, the descriptive sentences disappeared. At the same time, a target sentence appeared on the screen, and the participants would evaluate the truth-value of that target sentence. The experimental step-by-step procedures are presented later in the subsections of each experiment by using figures as a means of illustration.

The task is presented on the Internet using the platform of Ibex Farm (http://spellout.net/ibexfarm/), and participants are recruited through Amazon Mechanical Turk (https://www.mturk.com/). Experiment 1 with adults is parallel to Experiment 1 with children reported in Chapter 2, because it uses a broad question to prompt the target sentence, and crosses the attachment site of *only* (Subj-*only* and VP-*only*). In the experiments with children, we used sentences such as "*The cat only got ice cream*" for a VP-*only* scenario. Although it was clear by the prosody that the experimenter indicated the preverbal *only* attaches to the predicate, it is possible for some English speakers to associate *only* to the subject, with great support of prosody, such as, "*The cat ONLY,... got ice cream*" by putting the pitch peak on *only* and more importantly, making sure that there is no prosodic boundary between the subject DP and *only*, and that prosodic boundary before the predicate indicates that the subject DP and *only* are in

the same prosodic domain. Since there is no phonological cue to clearly indicate that pre-verbal *only* attaches to the predicate if we were to employ the same sentences for the experiments with adults, we used an auxiliary to separate *only* from the subject; e.g., "*The cat has only gotten ice cream.*" Experiment 1-prime is minimally different from Experiment 1. It uses a broad question, but the attachment site of *only* was manipulated between Subj-*only* vs. Object-*only* (such as "*The cat has gotten only ice cream*", Obj-*only*). The point of conducting Experiment 1-prime is to ask whether the difficulty observed for Subj-*only* was in fact a special case of a more general difficulty with attributive *only*. Since both Experiments 1 and 1-prime use a broad question to prompt the target sentence, they both serve as a baseline for the experiment on QAC. We investigated whether this baseline exhibits the asymmetry between Subj-*only* and VP/Obj-*only*, which was observed with children (with the measurements of Subj-*only* vs. VP-*only*). Experiments 2A and 2B jointly are designed to test the hypothesis that the QAC affects adult language processing in similar way to children, by using directly congruent Q-A pairs and incongruent Q-A pairs; which are analogous to sub-experiments of Experiment 2A, discussed in Chapter 2.

3.2.2 Experiment 1 – Baseline (Subj-*only* vs. VP-*only*)

3.2.2.1 Design

In Experiment 1, the participants first read a set of three sentences on a computer screen; which jointly describe a situation that would be equivalent to a picture shown in a child experiment. Below these "premise" sentences, participants saw a question that says "What happened?" with a link. Upon clicking the question, the premise disappeared and a target sentence, either a Subj-*only* or a VP-*only* sentence, appeared alongside buttons that said "True" and "False". The participants' task was to evaluate the truth-value of the target sentence relative to the situation described by the premise they previously read. The accuracy of the responses and the response times (RTs) from the click on the question till they make a judgment were recorded. Participants received feedback to their selection; so when they were wrong, the screen shows the following message for two seconds until the next item was displayed: "Oops! That was the wrong answer. Please wait for the next sentence.". When they were correct, the screen shows the following message for two seconds until the next item was displayed: "Please wait for the next sentence."

The feedback was meant to encourage the participants; especially when they were wrong it was meant to encourage them to pay more attention to the task. An example of the experimental procedure is shown in Figure 1.[43]

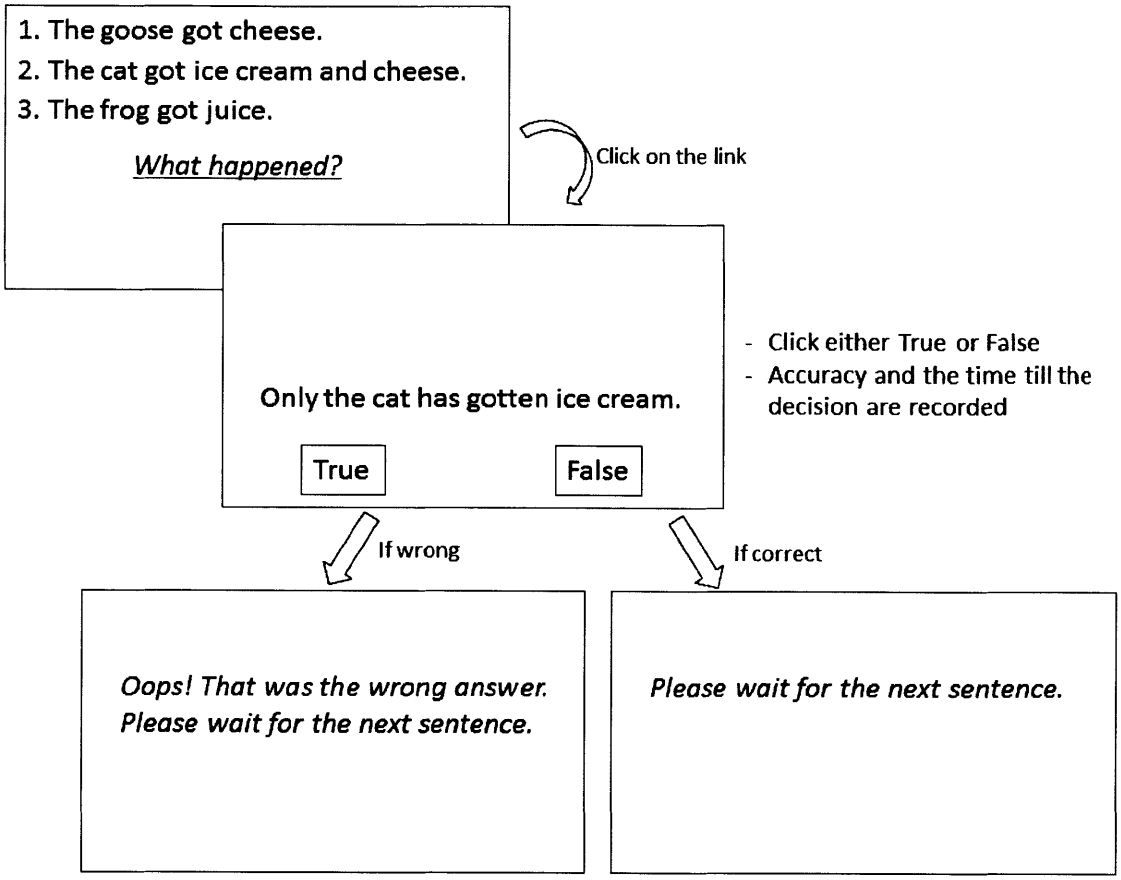


Figure 1: Illustration of experimental procedure

An experiment consists of 24 target items (12 Subj-*only* and 12 VP-*only* sentences) as well as 34 filler items. The truth is counter-balanced with 6 to be true and 6 to be false in each condition.

[43] We are aware that the way the information is packaged is an important factor and might have an effect on the processing of the sentences of the type that we are interested in. A possible critique here is that the way these pieces of information are introduced (as the sentence form of the premises) might be biasing toward generating a family of object-questions (i.e., sorted by the subject). In fact, it was one of the experimental manipulations to have the information packaging sorted by the subject and by the object. The pilot results suggested issues were not that simple. To put it plainly, the information packaging sorted by the object does not seem to very strongly bias the comprehenders to package information with the object as a sorting key. In a nutshell, the experimental procedure employed in Figure 1 might be biasing the way to package the information to some extent, but this will not distort the story of Topic as a sorting key.

Similar to the children experiments, each item is designed to have the opposite truth-value if it were interpreted to be sentences which has *only* attached to the other condition. In other words, if a target sentence in the Subj-*only* condition is made true, the hypothetical interpretation of the VP-*only* counterpart of the same sentence is false; and vice versa. The target items are distributed across lists in a Latin Square design, and the order of the items in each list was randomized. The list of target and filler sentences can be found in Appendix 3.

If VP-*only* is processed easier than Subj-*only* by adult speakers, we will see differences between the results of VP-*only* and Subj-*only*; so that VP-*only* has higher accuracy and/or shorter RTs.

3.2.2.2 Results

For Experiment 1, 64 English-speaking adults who reside in the U.S. participated.[44] Average time spent on an experiment is 18 minutes 23 seconds.[45] The exclusion criteria are as follows: We excluded participants with less than 75% accuracy across all items, because those participants can be treated as not paying enough attention to the experiment. Secondly, trials that were answered with RTs that are longer than 15 seconds are excluded from the analysis; because that would invite in unnecessary noises to the results, and such long RTs would indicate that the participants were distracted during the trial. Thirdly, RTs for incorrectly answered items were excluded from RT analysis; however, they were included in the accuracy analysis. Accuracy was analyzed using R software with generalized linear mixed effects model with logistic regression (GLMEM). The mode is maximally specified wherever possible. RT analysis used maximally specified LMEM for raw RTs and logged RTs. The summary of the analysis of logged RTs are reported. Figure 2, Tables 1 and 2 summarize the results. The error bars in graphs indicate 95% confidence intervals.

[44] A filter was put in place to isolate U.S.-based IP addresses only. The potential participants were posed a question about being a native speaker or not. All participants were confirmed to be native speakers.

[45] The average time was calculated on the Amazon Mechanical Turk website. This experiment was run in September 2013, with the reward per assignment being 30 cents. The results of this experiment were reported elsewhere (Hackl et al. 2014).

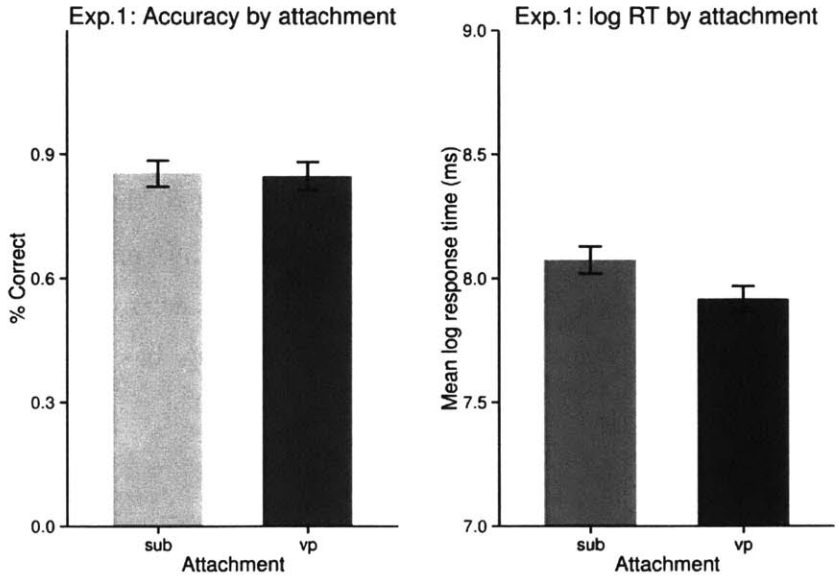


Figure 2: Results of Experiment 1

Table 1: Statistical analysis of the accuracy of Experiment 1

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.88560	0.17285	10.909	<2e-16 ***
attachmentVp	0.04937	0.21883	0.226	0.821

Table 2: Statistical analysis of the log RTs of Experiment 1

	Estimate	Std. Error	*t* value
(Intercept)	8.08376	0.07015	115.23
attachmentVp	-0.14569	0.03359	-4.34 *

The accuracy of VP-*only* (84.7%) was not statistically different from the accuracy of Subj-*only* (85.5%).[46] The raw RTs of VP-*only* was 3183 ms, and those of Subj-*only* was 3747 ms. The analysis on logged RT showed the main effect of attachment (*t* value = -4.34). The results from RTs replicated the asymmetry observed in child language acquisition with the accuracy measure. The results indicate that processing sentences with VP-*only* is easier than that with Subj-*only* in the adult grammar, when the sentence is in response to the broad question "*What happened?*"

[46] Since the maximally specified model did not converge, slopes for the random factors (items and subjects) did not include estimates for truth; only attachment site was estimated.

3.2.3 Experiment 1-prime – Another baseline (Subj-*only* vs. Obj-*only*)

3.2.3.1 Motivation of the experiment and design

A number of studies considered that there are two kinds of *only*'s: Adverbial *only*, which behaves like an adverb and attaches to a phrasal level, and attributive *only*, which behaves like an adjective and attaches to a DP (von Stechow 1991, Bayer 1996 and subsequent research, cf: Büring and Hartmann 2001 for a unified approach). If the asymmetry between Subj-*only* and VP-*only* that we observe is due to this characteristics of *only* (i.e. simply because they are two different *only*'s), then we need to revise our assumptions so that the difference is taken into account to explain the asymmetric behavior of Subj-*only* and VP-*only*; such as, adverbial *only* is easier/less costly than attributive *only*, as proposed by Crain et al. This concern led us to examine whether the results stay the same when we use attributive *only* in both conditions.

We used the same methodology to run the experiments; which were hosted on the Ibex Farm platform, and participant recruitment was conducted through Amazon Mechanical Turk. The experimental materials were identical to those in Experiment 1, except that the VP-*only* sentences in Experiment 1 were replaced with Obj-*only* sentences.[47]

If the asymmetry we observed in the experiments with children and Experiment 1 is caused because adverbial *only* is easier than attributive *only*, we will not see the difference between Subj-*only* and Obj-*only* conditions. However, if the asymmetry is caused because the accommodation of an object-question is easier/less costly than that of a subject-question, we will still see the asymmetric results between Subj-*only* and Obj-*only*.

3.2.3.2 Results

For Experiment 1-prime, 64 English-speaking adults who reside in the U.S. participated.[48] Average time spent on an experiment is 24 minutes 40 seconds.[49] We do not exactly know why

[47] Some of the premise sentences for filler items were replaced by new sets of sentences, to amend some confound. However, no change except for the experimental manipulation was made to the target items.
[48] Participant selection filters can be found in fn 44.
[49] The average time was calculated on the Amazon Mechanical Turk website. This experiment was run in November 2014, with the reward per assignment being $2.75, which is higher than that for Experiment 1, given an inflation in the standard of reward amount since 2013. It is possible that with a larger amount of reward the

Experiment 1-prime overall took so much longer than Experiment 1, but the raw response times of target items were statistically not longer in Experiment 1-prime than in Experiment 1. This indicates that the participants in Experiment 1-prime took longer time with reading instructions, the premises, and/or possibly filler items, but not with answering the target trials. The exclusion criteria and the models to use for analyses are the same as for Experiment 1. Figure 3 presents the results.

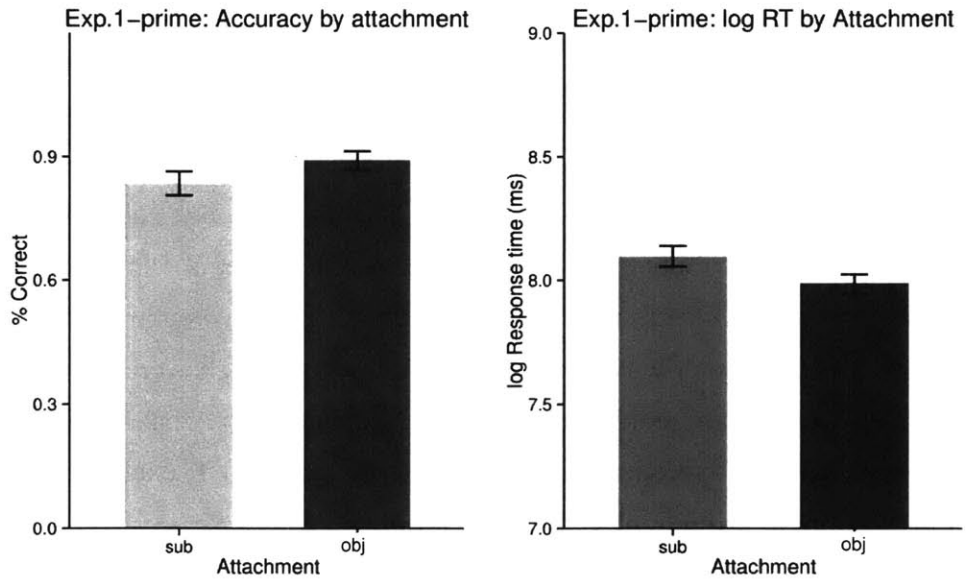


Figure 3: Results of Experiment1-prime

Table 3: Statistical analysis of the accuracy of Experiment 1-prime

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.7508	0.1522	11.503	<2e-16 ***
attachmentVp	0.4650	0.1854	2.509	0.0121 *

Table 4: Statistical analysis of the log RTs of Experiment 1-prime

	Estimate	Std. Error	*t* value
(Intercept)	8.09830	0.05099	158.81
attachmentVp	-0.11050	0.02238	-4.94 *

participants took the experiment more seriously, and paid more attention, which might have led to the longer time spent on the experiment. However, as I noted above, the RTs of target trials did not differ between Experiment 1 and Experiment 1-prime.

The accuracy of Obj-*only* (89.2%) was significantly higher than that of Subj-*only* (83.5%), (p = .012).[50] The response time of Obj-*only* was 3340 ms and that of Subj-*only* was 3734 ms. The statistical analysis on RTs showed that there was a significant main effect of attachment (the raw RTs (t value = -4.15), and log RTs (t value = -4.94)).

The results of RTs, again, replicated the asymmetry that was found in the experiments with children. Furthermore, the results of accuracy analysis revealed that Subj-*only* was less understood compared to Obj-*only*. The overall results from Experiment 1-prime reinforce the observation that the asymmetry between Subj-*only* and VP/Obj-*only* can be found in adult grammar as well; which in turn supports the claim that children's difficulty with Subj-*only* is due to the profile of Subj-*only* being difficult for a listener to accommodate an appropriate sub-question, and not due to the lack of the necessary grammar. The results of Experiment 1 and 1-prime are not explained if the grammar about Subj-*only* was treated as categorical. The results suggest that the ease and difficulty of accommodation in order to respect QAC play an important role in comprehending sentences with *only*. The following subsections report a set of experiments that are parallel to Experiment 2A with children, in that QAC is manipulated. We expect adults to exhibit different levels of performance depending on the levels of accommodation of an appropriate sub-question.

3.2.4 Experiments 2A & 2B

3.2.4.1 Design and methods

In Experiments 2A and 2B, we aim to investigate whether the difference in congruence (indirectly congruent, directly congruent, and incongruent) affects the processing of the sentences with *only*. As we did in Experiment 1 and 1-prime, we used Ibex Farm platform and recruited participants via Amazon Mechanical Turk. The way the test items were presented to the participants was modified after Experiment 1, to make sure the participants read the question sentence; which is a crucial manipulation in Experiments 2A and 2B. The "premise" sentences appeared line by line, prompted by clicking a button on a keyboard in a self-paced-reading

[50] Slopes for the random factor of items did not include estimates for truth and attachment site. Slopes for the random factor of subjects did not include estimates for truth, but included those for attachment site.

fashion. After reading three premise sentences, the premise disappeared and a question appears that either says "*What happened?*" or a more specific question, e.g., "*Who got [food item]?*" or "*What did [the animal] get?*" The question sentences were hyperlinked, and the participants clicked on the link to advance the procedure. By making participants click on the link that shows either type of question, and letting the question sentence appear by itself on the screen, we tried to make sure that the participants read the question sentence before clicking on it to further advance.[51] When the participant clicked on the question, the question disappeared and a target sentence appeared, alongside buttons that say "True" and "False." The participant's task was to evaluate the truth value of the target sentence relative to the situation described by the three-sentence premise. The dependent measures were accuracy and RTs, from the time they clicked the question line till they judged the truth-value of the sentence. After answering each question, the screen showed the following message regardless of the accuracy for two seconds until the next item was displayed: "Please wait for the next story." The participants did not receive any feedback for their choice in this experiment. We did not provide the participants with feedback because part of the goals of this study is to examine adults' strategies to deal with incongruence. From the experiments with children, we observed that children's primary strategy to "fix" the incongruence was to respect QAC and ignore the FCC-*only*. That means, one would interpret a Subj-*only* sentence in response to an object question as if it were a VP-*only* sentence, and a VP-*only* sentence in response to a subject-question as if it were a Subj-*only* sentence. Their judgements based on such interpretations would be coded as incorrect. We did not affect the participants' way of thinking so they would adjust their strategy after they received a negative feedback. That is why we decided not to give them feedback during the session. An example of the experimental procedure is illustrated in Figure 4.

[51] We had to conceal the premises while displaying the question sentence in any case. If a participant saw a question while they can still read the premise to check the answer, the task would have become too easy for an adult participant.

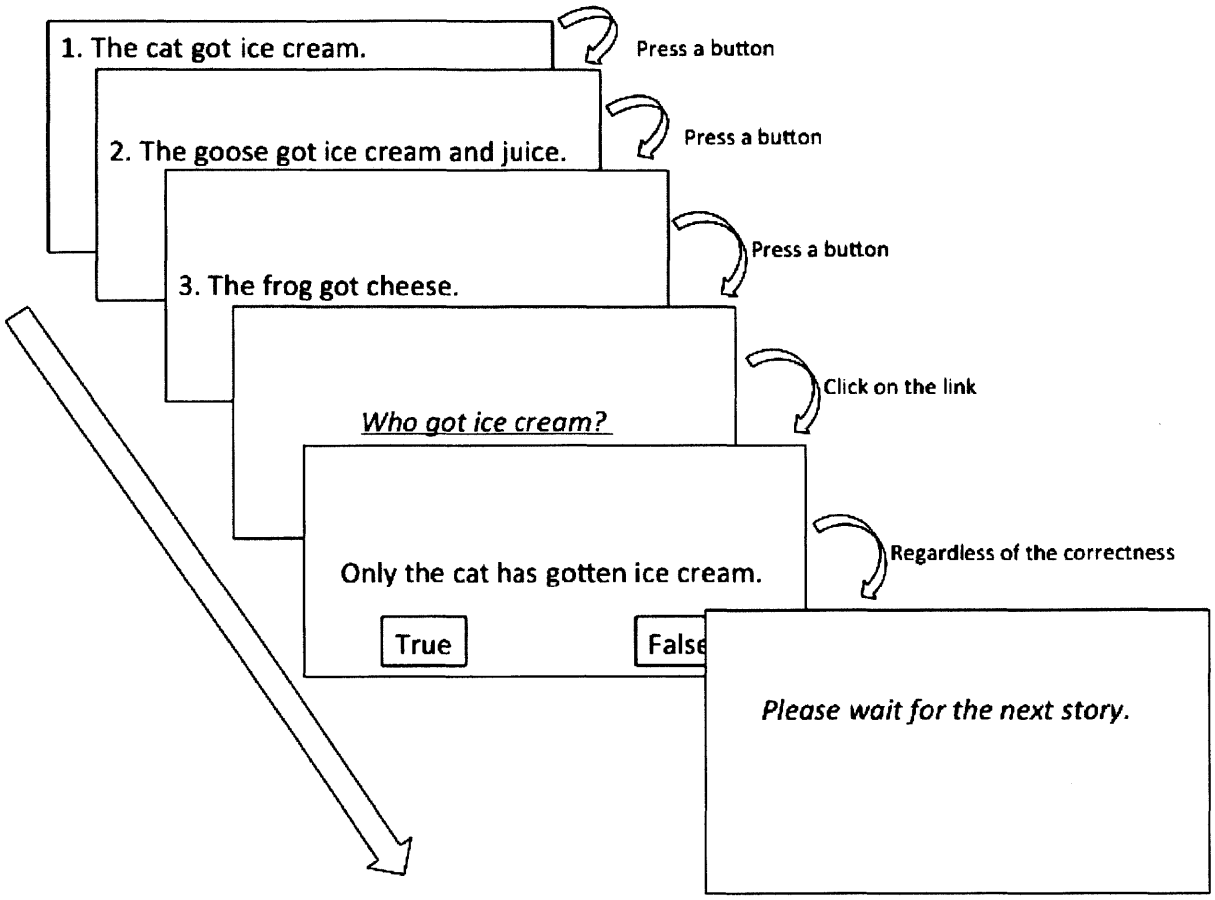


Figure 4: Illustration of experimental procedure

In order to study the effect of QAC, we varied the types of questions in one session. In the experiment 2A, participants read either "*What happened?*" or a subject-question such as "*Who got ice cream?*" On the other hand, in the experiment 2B, participants read either "*What happened?*" or an object-question such as "*What did the cat get?*" The participants were randomly assigned to one of the experiments and no participant took both experiments. In the children experiments, the broad Q-A pairs were tested in a separate experiment (Experiment 1), and we also tested the broad Q-A pairs within one session to study if a directly-congruent or incongruent pairs are deviant from the baseline.[52] The attachment site of *only* was a within-

[52] We have run Experiment 1 and 1-prime, whose experimental design is parallel to Experiment 1 with children (i.e., All the questions are "*What happened?*" and the attachment site of *only* is a within-subject factor). Nevertheless we tested the baseline condition in Experiments 2A and 2B, because we wanted to determine the baseline for each sub-experiment, and to see whether they are different from each other and from Experiment 1 and 1-prime.

subjects factor. The experimental design gives rise to two indirectly congruent conditions, one directly congruent condition, and one incongruent condition for each sub-experiment. The experimental design is presented in Table 5.

Table 5: 2x3 design for Experiment 2A and 2B

	Subject-question	**Baseline question**	**Object-question**
Subj-*only*	***Congruent*** Who got juice? Only the frog has gotten juice.	***Indir. congruent*** What happened? Only the frog has gotten juice.	***Incongruent*** What did the frog get? Only the frog has gotten juice.
VP-*only*	***Incongruent*** Who got juice? The frog has only gotten juice.	***Indir. congruent*** What happened? The frog has only gotten juice.	***Congruent*** What did the frog get? The frog has only gotten juice.

Exp. 2A

Exp. 2B

Each experiment consists of 24 target items (6 items in each condition) and 34 filler items. Truth of the items are counter-balanced so that half of the items are designed to be true and the other half to be false, and the truth of a given sentence would have the opposite value if the sentence was interpreted to have *only* at the other attachment site. The target items are distributed to create lists in a Latin Square design, and the presentation order of items are randomized. The target items are the same as the ones in Experiment 1. The only manipulation that was made for Experiment 2A and 2B was to use subject-questions and object-questions in addition to baseline questions, to prompt the target sentence.

3.2.4.2 Predictions

The discussion of QAC in §1.3 argues that a Q-A pair proceeded by a broad question "*What happened?*" is indirectly congruent because the Q-A pair becomes congruent only via accommodation of an appropriate sub-question. If the processing system of adults is affected by the accommodation of a sub-question, the processing load is predicted to be the following: Directly congruent Q-A < Indirectly congruent Q-A < Incongruent Q-A. However, if the

accommodation of a sub-question is done automatically and trivial for adults, it is possible that we see the processing load appearing as follows: Directly congruent Q-A = Indirectly congruent Q-A < Incongruent Q-A.

In the section §2.3.1, we discussed four strategies as to how to deal with incongruent Q-A pairs. The strategies are: (i) guessing, (ii) to ignore *only*, (iii) to identify the associate of *only* according to QAC and in effect interpret the target sentence as if Subj-*only* were VP-*only* and vice versa, and (iv) to ignore the preceding question and interpret the target sentence according to the syntactic position of *only*. We concluded that children's primary strategy seems to be (iii), on the grounds that the adult-like responses to incongruent Q-A pairs are very low (6.3% for Subj-*only* and 31.3% for VP-*only* in Experiment 2A) and that the results of Experiment 2B; which is designed to guide the participants to take the strategy (iv), did not conform well with the prediction.

We can make predictions for adult participants in Experiment 2A and 2B when they encounter incongruent Q-A pairs. If they take the strategy (i) for incongruent Q-A pairs, we will expect about 50% of accuracy from incongruent conditions. If they take the strategy (ii), we will again see about 50% of accuracy from incongruent conditions. If they ignore the word *only*, they would indiscriminately choose True; because the prejacent of the target sentence is always made true (i.e. the presupposition is always met) in the experiment. Since half of the items are true and the other half are false, the accuracy rates would be predicted to be around 50%. If they take the strategy (iii), we expect the accuracy to be close to 0%. That is because the strategy (iii) specifies that the QAC cue is the decisive factor to attract focus in the target sentence, in which case a participant interprets a target sentence as if the sentence had *only* at the swapped position. Thus we expect participants would always make mistakes if on the strategy (iii). Finally if they take the strategy (iv), we expect the accuracy to be close to 100%.

3.2.4.3 Results of Experiment 2A

For experiments 2A, 64 participants who reside in the U.S. are recruited. Average time spent on a session for Experiment 2A was 19 minutes 14 seconds.[53] The exclusion criteria and the models to use for analyses are the same as for Experiment 1; we excluded participants with less

[53] These times reported are obtained from the Amazon Mechanical Turk website. The experiments were conducted in November 2013, with the reward per assignment being 50 cents.

than 75% accuracy across all items, and individual responses with longer RTs than 15 seconds from the analysis, and finally, RTs for incorrectly answered items were excluded from the RT analysis, but included in the accuracy analysis. The results were analyzed using R software and the maximally specified GLMEM for the accuracy analysis, the maximally specified LMEM for the RT analysis were used. The results of Experiment 2A (n = 64) are shown in the following tables and the figure. The error bars in the figures 5-6 indicate 95% confidence intervals.

Table 6: Average values of each condition

	Subject-question	**Baseline question**
Subj-*only*	***Congruent*** Accuracy: 94% RT: 2775 ms	***Indirectly congruent*** Accuracy: 81% RT: 3871 ms
VP-*only*	***Incongruent*** Accuracy: 66% RT: 4098 ms	***Indirectly congruent*** Accuracy: 80% RT: 3208 ms

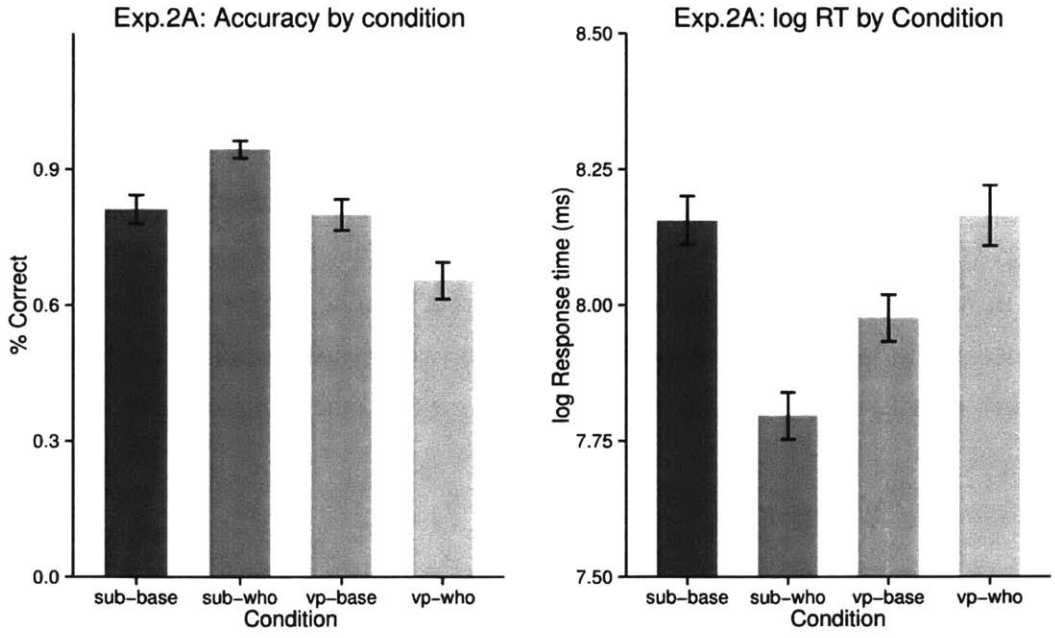


Figure 5: Results of Experiment 2A

Table 7: Statistical analysis of the accuracy of Experiment 2A

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.8155	0.2684	6.765	1.34e-11 ***
attachmentVp	0.5340	0.5064	1.054	0.291736
qTypeWho	7.1201	2.0755	3.430	0.000603 ***
attachmentVp: qTypeWho	-8.5858	2.1053	-4.078	4.54e-05 ***

Table 8: Statistical analysis of the log RTs of Experiment 2A

	Estimate	Std. Error	*t* value
(Intercept)	8.17855	0.06986	117.06
attachmentVp	-0.18193	0.06152	-2.96 *
qTypeWho	-0.38570	0.06510	-5.92 *
attachmentVp: qTypeWho	0.56370	0.11708	4.81 *

The analysis on accuracy reveals that there was a significant main effect of question type ($p < .001$) and more importantly, there was a significant interaction of the attachment of *only* and the question type ($p < .001$).[54] This indicates that when a subject-question precedes a VP-*only* (i.e., the Q-A pair is incongruent), more errors were observed. The analysis on log RT revealed two main effects and an interaction; the effect of the attachment site was observed (*t* value = -2.96), the effect of question type was observed (*t* value = -5.92), and the interaction of them was observed (*t* value = 4.81).[55] The values of the two main effect might look conflicting with the observed *t* values, but they only appear so because of the very large interaction.

3.2.4.4 Discussion on Experiment 2A

The results from Baseline-Q conditions did not show a difference in accuracy of Subj-*only* (81%) and VP-*only* (80%); while the Baseline-Q results show that VP-*only* was faster than Subj-*only*, which we can tell from the significant effect of the attachment site of *only*. This replicates the results of Experiment 1, and we can conclude that the asymmetric behavior is intact when the Baseline-Q conditions are intermingled with other types of questions. The

[54] Slopes for the random factor of items did not include estimates for attachment site and question type. Slopes for the random factor of subjects included estimates for both of the factors and their interaction.
[55] The analysis run on the raw RTs and z-Scored RTs also had two main effects and an interaction.

results from subject-questions confirm two points. Firstly, when the Q-A pair is congruent, the answer sentence is processed faster (2775 ms) than the baseline (3871 ms); and when the Q-A pair is incongruent, the answer sentence is processed slower (4098 ms) than its baseline (3208 ms). There is no VP-*only* privilege (i.e., across-the-board ease with VP-*only*) found here. The results strongly suggest that adults are also sensitive to QAC, and the processing load is directly affected by the accommodation of an appropriate sub-question. The results are suggestive in that the cost of the accommodation that was saved on the Subj-*only* condition, and the cost that was added on the VP-*only* condition, are about the same. As for the Subj-*only* condition, 28% of the raw RT on the Baseline-Q condition was saved on the subject-question condition ((3871 – 2775)/3871 = 0.28); and on the VP-*only* condition, 28% of the raw RT on the Baseline-Q condition was added on the subject-question condition ((4098 – 3208)/3208 = 0.28).

Secondly, the analysis on accuracy reveals that congruent Q-A pairs are answered more correctly (94%) than the baseline (81%); and incongruent Q-A pairs are answered less accurately (66%) than the baseline (80%). The results tell us two things; (a) adults have difficulty with comprehending incongruent answer sentences with *only*, and (b) the strategy to deal with the incongruence seems to ignore the question and observe the FCC-*only*. This strategy was dubbed as the forth (iv) strategy in Chapter 2. Recall that children's primary strategy with incongruence was to respect QAC and ignore the FCC-*only*, which was strategy (iii). We will come back to this point in the discussion of Experiment 2B, and general discussion.

3.2.4.5 Results of Experiment 2B

Let us turn to the results of Experiment 2B. We recruited 64participants through Amazon Mechanical Turk and tested using Ibex Farm. The average time per assignment for Experiment 2B was 18 minutes 39 seconds. The accuracy and average raw RTs are summarized in Table 9, and the graphs to illustrate the results of accuracy and log RTs are shown in Figure 6.

Table 9: Average values of each condition

	Baseline question	**Object-question**
Subj-*only*	***Indirectly congruent*** Accuracy: 86% RT: 3523 ms	***Incongruent*** Accuracy: 73% RT: 3402 ms
VP-*only*	***Indirectly congruent*** Accuracy: 82% RT: 3320 ms	***Congruent*** Accuracy: 90% RT: 2458 ms

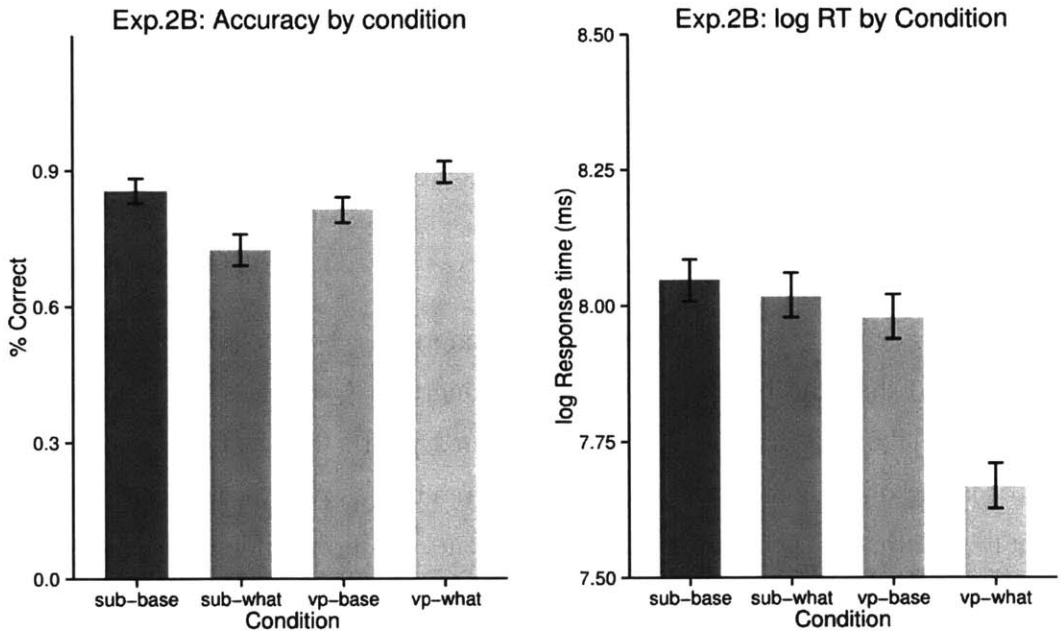


Figure 6: Results of Experiment 2B

Table 10: Statistical analysis of the accuracy of Experiment 2B

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.7904	0.4217	6.616	3.68e-11 ***
attachmentVp	-0.6318	0.4789	-1.319	0.18703
qTypeWhat	-1.0371	0.3301	-3.141	0.00168 **
attachmentVp: qTypeWhat	1.9800	0.6247	3.170	0.00153 **

Table 11: Statistical analysis of the log RTs of Experiment 2B

	Estimate	Std. Error	*t* value
(Intercept)	8.05742	0.06579	122.48
attachmentVp	-0.07947	0.05840	-1.36
qTypeWhat	-0.05977	0.06106	-0.98
attachmentVp: qTypeWhat	-0.25618	0.08645	-2.96 *

The analysis on accuracy reveals a significant main effect (with an object-question, lower accuracy was observed, p = .002) and more importantly, a significant interaction of the attachment site of *only* and the question type (p = .002).[56] The results indicate that when an object-question precedes VP-*only* (i.e., the Q-A pair is congruent), less errors were observed. The analysis on log RTs revealed a significant interaction of the attachment site of *only* and the question type (t value = -2.96). No main effects were observed from the RT analysis.

3.2.4.6 Discussion on Experiment 2B

The results on accuracy is straightforward; incongruent Q-A pairs were understood less accurately (73%) than the baseline (86%), and congruent Q-A pairs were understood more accurately (90%) than its baseline (82%). This again supports that the QAC plays an important role in processing of sentences with *only* in adult grammar as well as children's grammar. It is important to note that the accuracy of the incongruent condition being 73% suggests again that adults' primary strategy to respond to incongruent Q-A pairs is to ignore the question and stick to the FCC-*only* (strategy (iv)).

The analysis of RTs revealed two interesting points. In the discussion of Experiment 2A, it was suggested that the cost of the accommodation that was saved for the directly congruent answer was about 28% of the RT on the baseline condition. In Experiment 2B, we see a similar phenomenon; comparing the baseline and the object-question conditions with VP-*only* sentences, 26% of the raw RT on the Baseline-Q condition was saved on the object-question condition ((3320 – 2458)/3320 = 0.26). This can be interpreted as the cost of accommodating a suitable sub-question.

[56] Slopes for the random factor of items did not include estimates for attachment site and the question type. Slopes for the random factor of subjects included estimates for both of the factors and their interaction.

Another interesting point is that the difference between the Baseline-Q condition and the object-question condition with the Subj-*only* sentences was not observed. The directly congruent question for Subj-*only* is a subject-question. If it is easier to accommodate a subject-question when a broad question is presented than when an object-question is presented, as it was with VP-*only* sentences, we would have seen an increase of RTs on the object-question condition with Subj-*only*. However, we did not observe the increase on the object-question condition. The proposal, which was made in §2.4.2, that the broad question is analyzed as a set of object-questions by default seems to acquire support. If the broad question is essentially understood as a set of object-questions, then it is possible that a broad question and an object-question are quite similar to each other.

3.3 General discussion

The first thing to note with the adult experiments, is the asymmetry between Subj-*only* and VP-*only* was observed in the baseline conditions. The RTs across experiments, i.e., in Experiment 1 and in the baseline conditions of Experiment 2A and 2B, are not statistically different from one another. The interesting observation is that the asymmetry was only found in RTs and not in accuracy (except for Experiment 1-prime). It is possible that the processing cost of accommodating subject-questions is not substantial enough to affect the comprehension by adults, but since the cost exists, the RTs are affected. The accuracy rates were affected in the experiments with children, which suggests that the processing cost of accommodating subject-question had a substantial effect on the comprehension of the child population, whose cognitive resources (or, cognitive ability concerning linguistic processing) is presumably smaller than that of adults.

Note that King and Just (1991) observed that some amount of processing cost affects both comprehension and reading time with a smaller working memory population; but the same task only affects reading time, not the accuracy, with a larger working memory population. Their experiments tested adult participants with processing of subject relative clauses and object relative clauses in an elaborated self-paced reading task; the participants read three sentences word by word with the method of self-paced reading, and at the end of the sentence, one has to recall the last word of the sentence. After reading the second sentence, one has to recall the last

word from the first sentence and the last word from the second sentence. After reading the third sentence, one has to recall the last word from the first, second, and the third sentences. A comprehension question follows the last sentence they read. The reading time by the word of the target sentence (subject relatives and object relatives) and the participants' accuracy on the comprehension questions were recorded. This experimental manipulation pressures the participants' working memory, since the participants had to preserve a word or two when they read and comprehended other sentences. According to the number of words that the participants were able to recall, K&J grouped them into two groups: large working memory group and a small working memory group. The results reveal that both groups showed slower reading times with subject relative clauses than with object relative clauses. However, the results of accuracy were different; as for the small working memory group, the accuracy on the subject relative clause condition was significantly lower than that on the object relative clause condition, but on the other hand, the larger working memory group did not show the difference between the two conditions, and both subject and object relative clause conditions were answered overall more correctly than with the small working memory group.

The results from K&J's experiment suggest that time-sensitive measures can manifest the subtle differences in processing load regardless of the individual differences in cognitive resources, while the measures such as accuracy will demonstrate the difference in processing load, only when the cost of the computation exceeds the person's cognitive resources. Although the type of phenomenon tested and the experimental methods are different, the message obtained from K&J's experiment can be seen as an analogy to what we observed in the baseline conditions; RTs manifests the difference between Subj-*only* and VP-*only* regardless of the size of the population's cognitive resources (i.e., adults and children). On the other hand, the accuracy measures exhibit the difference between Subj-*only* and VP-*only* only with children, where their cognitive resources are limited; but with adults, their cognitive resources are presumably not affected by the extra processing load that is required for Subj-*only*.

The results from Experiments 2A and 2B combined show that adults are also sensitive to QAC, and that the RTs are shorter and accuracy is higher with the congruent Q-A pairs than their baseline counterparts. It was also pointed out that the results seem to support the proposal of a broad question being understood as a set of sets of questions that are sorted by Topic of the sentence. In addition, the way a family of questions with a set of object-questions is formed corresponds to a superiority-obeying multiple *wh*-questions. This proposal successfully accounts for the asymmetric processing loads between Subj-*only* and VP-*only*, and for the fact

that there was no difference in RTs between the BaselineQ-Subj-*only* condition and the ObjectQ-Subj-*only* condition in Experiment 2B.

The results of Experiments 2A and 2B show that adult speakers' primary strategy to resolve the incongruence is to ignore the preceding question and respect the FCC-*only* (strategy (iv)). This strategy is different from children's primary strategy (strategy (iii)). Why does this occur? Is there a difference in processing load between the two strategies?

To examine these questions, we compared RTs of correct answers (produced by following strategy (iv)), and RTs of incorrect answers (produced by following strategy (iii)), on the incongruent conditions (i.e., object-Q and Subj-*only*, and subj-Q and VP-*only*). A statistical analysis reveals that correct answers took significantly longer than incorrect answers (*t* value for log RT: 3.69). A main effect on the *only* attachment site (the pair of subj-Q and VP-*only* was read longer than the pair of obj-Q and Subj-*only*) was also observed (*t* value for log RT: 2.50). See Figure 7 for an illustration. The average raw RTs of incorrect Subj-*only* was 2592 ms, incorrect VP-*only* was 3242 ms, correct Subj-*only* was 3402 ms, and correct VP-*only* was 4098 ms.

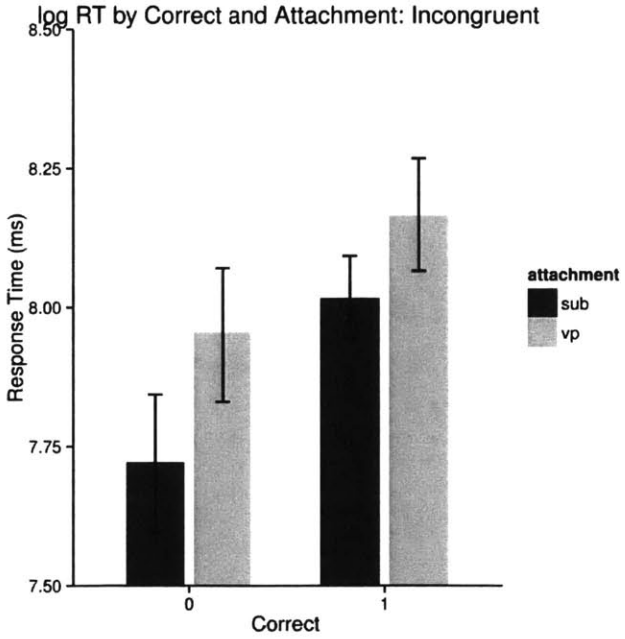


Figure 7: log RT by correct and attachment

The results suggest that adopting strategy (iv) might be costlier than adopting strategy (iii). Recall that children's primary strategy is to take strategy (iii). It is possible that children initially adopt the less costly strategy when they encounter incongruent items. However, since adults know that the FCC-*only* is a principle to obey under the grammar of English, adults adopt the strategy (iv), despite the high processing load; which is caused because they try to reconfigure the question so that they can observe the FCC-*only*.

Chapter 4

Background for English Rise-Fall-Rise

4.1 Prosody and interpretation

This chapter provides background for the second case study on the effect of the "Rise-Fall-Rise" prosody and scope relations. Specifically I focus on the scope relations between a universally quantified subject and negation.

When an English sentence contains multiple quantifiers (e.g., *some conductor studied every music in the library*) or quantifier(s) with a logical operator (e.g., *every violin player didn't show up at the rehearsal*), the sentence has multiple meanings. Speakers, however, generally should be as clear as possible and avoid ambiguity to observe Grice's maxim of manner (Grice 1975); hence, only one of the possible meanings should be intended. In turn, listeners assume that the speaker obeys the cooperative principles and thus believe that the sentence would intend only one of the meanings. Interlocutors usually know which of the possible meanings is more plausible in a given situation. This is possible because even though multiple LF representations are equally obtainable from what they can infer from the PF representation of the sentence, there are ways to disambiguate the possible meanings in real-life conversation. Contexts prior to the utterance are most likely to serve as a disambiguation. Follow-up comments can disambiguate a previous utterance if the speaker thinks they made an unclear statement, and phonological information such as special pitch contour or stress on certain elements also serves as a good indicator in many languages. Among ways to disambiguate possible meanings, it has been noticed that prosodic information can play a significant role since Jespersen's (1933) observation of (1), where sentences with negation and a *because*-clause can have different scope

relations depending on the prosody; *call* in (1a) has a rising tone and *call* in (1b) has a falling tone.[57]

(1) a. I didn't call because I wanted to see her (but I called her for some other reason).
b. I didn't call because I wanted to avoid her. (Jespersen 1933: 299)

Aside from intonational information such as high tone and low tone, other prosodic aspects such as prosodic prominence (higher pitch, longer duration, larger volume) and prosodic boundary (indicated by a halt and/or a lengthened segment before a boundary) are known to play a role in structural ambiguity resolution (Lehiste 1973, Tanenhaus et al. 1995, Kjelgaard and Speer 1999, Snedeker and Trueswell 2003, etc.), indication of importance, new and given (Bolinger 1972, Halliday 1967, Selkirk 1995, Schwarzschild 1999, Dahan et al. 2002, etc.), contrast (Spivey-Knowlton 1996, Ito and Speer 2008, Ito et al. 2012, etc.), pronoun reference (Akmajian and Jackendoff 1970, Hirschberg and Avesani 2000, Baauw et al. 2004, etc.), and association of focus sensitive operator (Hirschberg and Avesani 2000, Beaver and Clark 2008, etc.). The experimental approaches to the effect of prosody and scope relation between a universal quantifier and negation are reviewed in §1.4 (Ianucci and Dodd 1980, Syrett et al. 2014a, b). For extensive review of the literature on experimental approach to comprehension of prosody, see Wagner and Watson (2010).

4.2 Rise-Fall-Rise (RFR) to mark Contrastive Topic (CT)

For such sentences where a quantifier and negation interact, prosody is found to play a role affecting interpretations in languages such as English and German (Jackendoff 1972, Liberman and Sag 1974, Ladd 1980, Büring 1997, 2003, Krifka 1998, Jackson 2007, Constant 2012, 2014 a.o., Cf: Gussenhoven 1983, Ward and Hirschberg 1985). Let us now closely look at what Jackendoff (1972) proposes, in order to understand what it takes to prosodically indicate a certain meaning. He proposes that the pitch accent on the universal quantifier and falling at the

[57] For recent experimental literature related to disambiguation of sentences with negation and a *because*-clause, see Hirschberg and Avesani (2000) and Koizumi (2009).

end of a sentence such as in (2), correspond to the "All > Not" reading; and that the pitch accent on the universal quantifier and rising at the end correspond to the "Not > All" reading.

(2) a. ALL the men didn't go.$_{L\%}$ ($\forall > \neg$)[58]

b. ALL the men didn't go…$_{L\text{-}H\%}$ ($\neg > \forall$)

The contour that is associated with (2b) and Figure 1 has been variously referred to as 'B-accent', 'Fall-Rise' etc. under different pieces of literature. However, following Constant, I would like to call it a Rise-Fall-Rise (RFR) contour, since the name is theory-neutral and it transparently describes the pitch tracks of the contour (in order to fall after *all*, there must be a rise beforehand). Figure 1 illustrates what a pitch track typically looks like with the RFR contour.[59] As we can see in the pitch track, there is a sharp rise on *all* and a rising boundary tone at the end of the (intonational) phrase (IntP), which always happens to be sentence-final in our examples. As we will see shortly, the part of "rise and fall with stress" (to be more exact, the part with the tone of L*+H or L+H*) corresponds to a Contrastive Topic (CT) of the sentence.[60] RFR occurs entirely (stretched, if necessary) within one intonational phrase.

[58] Intonation of some sentences are indicated with low (L) and high (H) tones within the convention of ToBI. Core elements of tones are:

(i) Pitch accents: H*, L*, L+H*, L*H, etc.
(ii) Phrase tones: H-, L-
(iii) Boundary tones: H%, L%

[59] The audio recordings of the examples in Constant (2012) are available at http://semanticsarchive.net/Archive/jhmYTI5M/.

[60] How to define Contrastive Topic has mixed views in the literature. There is a disagreement even on the issue as to whether CT is essentially given or new (See discussions and references in Büring 2003, Constant 2014). However, for our convenience, CT can be descriptively understood as "[t]he phrase denoting what the question being addressed is about. Implies other questions about different topics (Constant 2014: 17).", or "shifting the topic of discourse from one item to another, narrowing down the referent of a topic or simply implicating the existence of another salient item (Vermeulen 2009: 362)".

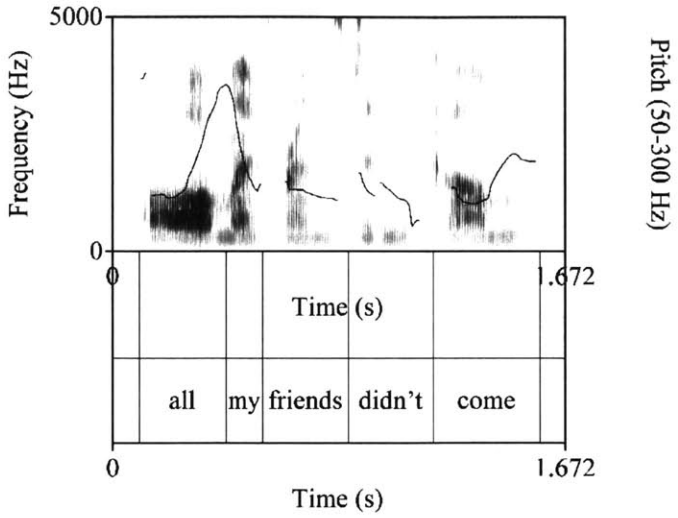


Figure 1: Pitch track generated by Praat (Constant 2012: (46))

Note that the RFR is not peculiar to such sentences with a universal quantifier and negation. See some more examples from Constant (2012) below. Take a conversation between person A and person B, where person B's utterance is pronounced with RFR. If you were person A, who wanted to know whether person B's friends liked the movie or not, you would get an impression upon hearing person B's response, either that the other friends of B's did not like the movie, or that B does not know if the others liked it and thus B is not entitled to make a claim about the other friends. Now consider the conversation between person A and person C. RFR is compatible with the response in (3C) as well, in which case what person A would perceive as an implication is that either there are some friends of C's who did not like the movie but it was not as many as half of the group, or there are some friends of C's whose thoughts toward the movie person C does not know about. On the other hand, the answer in (3D) is not viable with the RFR prosody. Why is that? We will have to review Constant (2012, 2014) to understand this mechanism.

(3) A: Did your friends like the movie?
B: JOHN liked it…
C: MOST of my friends liked it… (Constant 2012: (23), (33a))
D: #ALL my friends liked it…

Büring (1997, 2003) discusses the effect of the CT contour on scope relations in German and English. Based on his analysis, Constant (2012, 2014) proposes the possibility that the RFR contour is treated as a "lone" CT, i.e., Contrastic Topic not accompanied by a Focus.[61] The following conversations illustrate some examples of how CT and/or Focus are present in English sentences (examples are from Constant 2012: 39-42). Suppose the conversation takes place at a potluck party, and interlocutors try to figure out who brought what.

(4) A: What about Fred? What did he bring?
B: [Fred]$_{CT}$ brought the [beans]$_{F}$. CT + F
L+H* L-H% H* L-L%

(5) A: What about beans? Who brought them?
B: [Fred]$_{F}$ brought the [beans]$_{CT}$. F + CT
H*L- L+H* L-H%

(6) A: Which people brought something?
B: [Fred]$_{CT}$ brought something... lone CT
L+H* L-H%

In the examples in (4-5), what is contextually given and contrasted ("*Fred*" in (4) and "*beans*" in (5)) is read with the CT prosody, i.e., RFR. On the other hand, the answer of the question ("*beans*" in (4) and "*Fred*" in (5)) is read with the falling prosody. The example of "lone CT" has a pitch accent on the CT element, and the rise is found at the boundary of an intonational phrase (IntP) that contains CT.

RFR has two important properties; it is not compatible with (i) thoroughly exhaustive answers and (ii) alternative dispelling elements. These concepts are explained below, by taking a look at the unlicensed examples violating these constraints. The examples in (7-8) show that the responses that exhaustively answer the question are not compatible with RFR (exemplifying the feature of (i)). The example in (9) shows that the response is a maximal element that logically entails all the other alternatives and thus RFR is not compatible (exemplifying the feature of (ii)).

[61] In Constant (2012), RFR and CT are treated separately though he suggests that proposing a unified account of RFR and CT is a possible path. Constant (2014) does not particularly use the term RFR and instead he calls it lone CT. In the current study, the term RFR will be used to cover the union of the phenomena.

(7) Q: Did Carol win the first place in the race?

A: I think so... / #Yes... / #No...

L+H* L-H% L+H* L-H% L+H* L-H%

(8) Q: Who won the first place in the race?

A: Debora didn't... / #Carol did...

L+H* L-H% L+H* L-H%

(9) Q: Did everyone arrive yet?

A: Most of them did... /#All of them did...

L+H* L-H% L+H* L-H%

These properties of RFR tell us that it has a requirement that there should be at least one alternative proposition remaining unresolved after the utterance. That is, the mechanism of RFR can be broken down into these steps – (i) The L+H* on the F-marked item (Contrastive Topic) activates its relevant alternatives, (ii) Alternative propositions are generated based on the alternatives activated in (i), (iii) Multiple possible LFs if any are computed, and (iv) Identify which LFs satisfy condition that there is at least one proposition expressed by a relevant LF that is not yet resolved/dispelled after assertion of the sentence.

The unlicensed examples in (7-9) show why the answers under the RFR-contour are infelicitous: when pronounced with RFR, the alternative sets in (10-12) are generated respectively. RFR further requires that there should be unresolved or not-dispelled LFs remaining post assertion. However, the assertion of the unlicensed sentences entails the truth/falsity of all the other LFs in the alternative set. Hence, no alternative proposition remains open for discussion and this violates the condition imposed by RFR.

(10) {Yes, she did, No she didn't}

(11) {Carol won the race, Debora won the race, Elaine won the race, ...}

(12) {Someone arrived, most of them arrived, all of them arrived, ...}

The licensed utterances, by contrast, leave at least some LFs in the relevant alternative set remaining unresolved after assertion. As a result, the utterances are compatible with RFR and they implicate (based on Gricean quantity reasoning) that the speaker does not know the information about the truth of the other alternatives or that they deliberately avoid mentioning the information that dispels the other alternatives at this moment.

4.3 RFR and scope interaction between a universal quantifier and negation

Finally, let us turn to the case of the scope interactions between the universal quantifier and negation. The mechanism at work is exactly the same as the previous subsection. The sentence "All my friends didn't come" is potentially ambiguous between the "All > Not" reading and the "Not > All" reading (without any prosodic information). If one entertains the interpretation of "All > Not," the set of alternatives looks like the one in (14). Once the sentence is asserted, the other LF(s) are dispelled since they are entailed by the assertion. This does not meet the condition of RFR that at least one LF must remain unresolved after assertion. On the other hand, the set of alternatives for the "Not > All" interpretation of the sentence will be presented in (15). The other LF(s) are not entailed by the assertion, and therefore they remain unresolved after assertion. This satisfies the condition of RFR, and that is why this interpretation of the sentence survives when pronounced with RFR.

(13) A: Did all the viola players show up at the rehearsal last night?
B: ALL the violi didn't show up...
L+H* L-H%

(14) <u>All > Not LF:</u>
Assertion: For all viola players x, x did not show up.
Alternatives: { For all viola players x, x did not show up,
For some viola player x, x did not show up, ...}

(15) <u>Not > All LF:</u>
Assertion: It is not the case that all the viola players showed up.
Alternatives: { It is not the case that all the viola players showed up,
It is not the case that some viola player did not show up, ...}

There have been proposed counterarguments to the thesis that the RFR contour disambiguates interpretations of a sentence (Gussenhoven 1983, Ward and Hirschberg 1985). Based on the examples such as in (16), they claim that it is context, not intonation, which disambiguates the interpretations.

(16) A: The foreman wants to know which union meeting some of the men missed.
B: ALL of the men didn't go to the last one...
L+H* L-H%

(Ward and Hirschberg 1985: 771)

The most natural interpretation of person B's utterance in (16) is the "All > Not" meaning; no one went to the last meeting. The sentence, however, can be naturally accompanied with the prosody of RFR. Isn't this a problem with proposed RFR mechanism?

To respond to the apparent problem, Constant points out that speakers confirmed that *last* in person B's utterance needs stress as well as *all*. That is, person B's utterance in fact has two F-marked elements, and thus the set of alternatives to the utterance would be something similar to (17b).

(17) a. {all, some, ...} of the men didn't go to the {first, second, ... last} one.
b. {all of the men x, x did not go to the last one,
all of the men x, x did not go to the second to the last one,
...
some of the men x, x did not go to the last one.
some of the men x, x did not go to the second to the last one, ...}

Notice that even when the "All > Not" LF (the first line of (17b)) is asserted, it only partially resolve other alternatives (e.g., the second and last lines of (17b) are not resolved). Since there remains at least one LF that is not resolved after assertion, person B's utterance is compatible with RFR and still interpreted as conveying the "All > Not" meaning. The example in Ward and Hirschberg is not a counterexample to the theory of RFR, but it is congruent to it. The important conclusion from this discussion is that Constant's (and Büring's) mechanism does not claim that RFR is a *disambiguating* contour. Rather than that, the claim is that RFR is conditioned so that there is at least one alternative proposition among relevant alternatives that is not resolved/dispelled after assertion of the sentence. If disambiguation happens as a result of conforming to the condition, then the meaning of a sentence is disambiguated; but if the result from conforming to the condition leaves ambiguity, then there is no wonder that the sentence remains ambiguous or preferably conveys the logically stronger meaning. The

existence of a sentence, which puts on the RFR prosody and has the "All > Not" interpretation at the same time, is not controversial; and Constant's mechanism correctly predicts so. However, as readers notice by now, that is the case when there is another variable to generate alternatives (in the example above, a scalar term, *last*) present in a sentence. In the current work, we are only concerned with sentences with minimal number of alternative-generating elements; hence, in our examples the RFR prosody will always "disambiguate" as a result of obeying the condition of RFR.

4.4 Question-Answer Congruence (QAC) again, and RFR

While the Rise-Fall-Rise contour and its scopal interaction attract much attention, the discussion on under what circumstances the contour is felicitously used has attracted less attention. Büring (2003) and Constant (2012, 2014) discuss what makes a congruent Q-A pair when an answer contains a Contrastive Topic (CT) such as in (4-6), repeated below as (18-20), without explicitly mentioning the possibility that RFR with universal quantifier and negation would be subject to the same condition. Büring calls the condition CT congruence, which is referred to in (21), and summarized in (22) using informal terms.

(18) A: What about Fred? What did he bring?
B: $[\text{Fred}]_{CT}$ brought the $[\text{beans}]_{F}$. CT + F

(19) A: What about beans? Who brought them?
B: $[\text{Fred}]_{F}$ brought the $[\text{beans}]_{CT}$. F + CT

(20) A: Which people brought something?
B: $[\text{Fred}]_{CT}$ brought something… lone CT

(21) CT congruence[62]
An utterance U containing a contrastive topic can map onto a move M_U within a d-tree D only if U indicates a strategy around M_U in D.

[62] D-tree (Büring 2003) is a visualized version of the QUD stack (Roberts 1996/2012); which we have reviewed with some examples in §1.3 (for the QUD stack), and in §2.4.2 (for D-tree). More examples follow on the next page. Each node in a D-tree is called a *move* (Carlson 1982). A *move* represents a declarative or interrogative sentence, without CT/F-marking. The other notions should be self-explanatory in the definition, but readers are also referred to the useful list of formal definitions for more concreteness in the Appendix of Büring (2003).

U indicates a strategy around M_U in D iff there is a non-singleton set Q' of questions such that for each $Q \in Q'$, (i) Q is identical to or a sister of the question that immediately dominates M_U, and (ii) $[\![Q]\!]^o \in [\![U]\!]^{ct}$. (Büring 2003: 528)

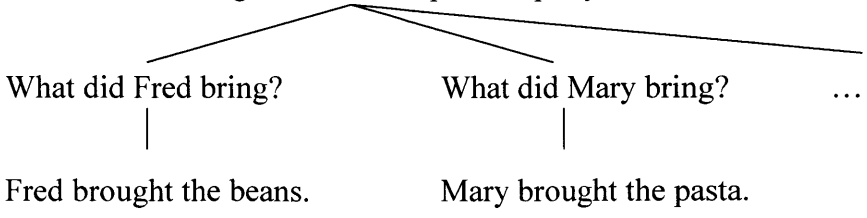
(22) An utterance (U) with a CT marked element answers a sub-question in a D-tree of a larger question, but there is at least one other alternative sub-question within the same D-tree which is not immediately answered by the U.[63] Furthermore, the alternatives are generated according to the placement of CT (and F).

Let us first look at how CT congruence predicts what kind of question is congruent for the sentences in (18-20) (D-tree configurations based on what is described in Constant 2014: 39-42). Crucially the focus (the answer) in the utterance will create the set of alternatives first, and then the CT creates its alternative set. Note the difference between D-trees in (23) and (24); because the subject is CT in (23), the CT semantic value of the utterance is a set of questions about the object (i.e., F in the sentence).[64] The congruent sub-question for the utterance in (23) is something like "*What did Fred bring?*", and the utterance also implicates that other sub-questions such as "*What did Mary bring?*" is not resolved yet. On the other hand, since the object is CT in (24), the CT semantic value of the utterance is a set of questions about the subject (i.e., F in the sentence). The congruent sub-question for the utterance in (24) is "*Who brought the beans?*", and furthermore the utterance implicates that another sub-question such as "*Who brought the pasta?*" is not resolved.

(23) $[\![[\text{Fred}]_{CT} \text{ brought the } [\text{beans}]_F]\!]^{ct}$ CT + F

a. { [[What did Fred bring?]], [[What did Mary bring?]], … }
= { Fred brought beans, Fred brought gazpacho, Fred brought pasta,
Mary brought beans, Mary brought gazpacho, Mary brought pasta, …}

b. Who brought what at the potluck party?



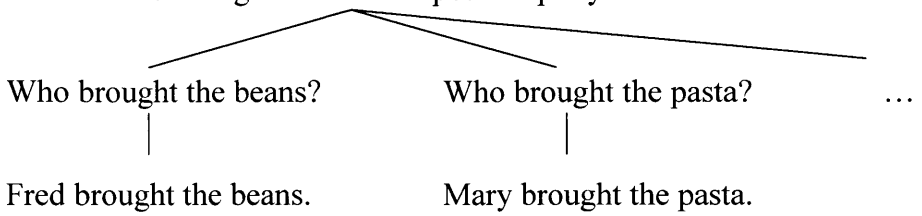
[63] In Büring's terms, it is specified as a sister of the immediate question, but the condition paraphrased here does not necessarily require it be so, following Constant (2014).
[64] CT semantic value is obtained by replacing CT-marked element in focus semantic values with variables.

(24) ⟦ [Fred]$_F$ brought the [beans]$_{CT}$⟧^{ct} F + CT

a. { ⟦Who brought the beans?⟧, ⟦Who brought pasta?⟧, … }

= { Fred brought beans, Mary brought beans, Sarah brought beans,
Fred brought pasta, Mary brought pasta, Sarah brought pasta, …}

b. Who brought what at the potluck party?



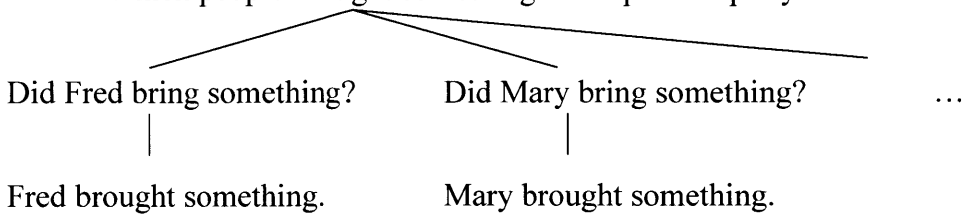
Then what about the case of lone CT? The procedure to arrive at the congruent sub-question is not different from previous examples. The subject *Fred* is CT in (25), and the CT semantic value of the utterance is a set of questions as in (25a). Note that *Fred* is not the focus (i.e., answer) of a question, so the CT semantic value of (25) will not be as simple as {Fred brought something, Mary brought something, …}; but it will be a set of sets of alternatives. The congruent sub-question for the utterance (25) is "*Did Fred bring something?*", with an implication that other sub-questions such as "*Did Mary bring something?*" is not answered yet.

(25) ⟦ [Fred]$_{CT}$ brought something ⟧^{ct} lone CT

a. { ⟦Did Fred bring something? ⟧, ⟦Did Mary bring something?⟧, …}

= { Fred brought something, Fred did not bring anything,
Mary brought something, Mary did not bring anything, …}

b. Which people brought something at the potluck party?



If the utterances with CT was made without an explicit congruent sub-question, a listener accommodates an appropriate sub-question, just as we have seen in §1.3. Consider the conversation in (26) for instance. Person B's utterance is not preceded by an immediately dominating question (i.e., directly congruent question), which would have been "*Did Fred bring*

something?", but instead it is preceded by a broad question about how things are.[65] Though the Q-A pair is not directly congruent, the conversation is viable because the interlocutors (specifically, person A in this situation) accommodate the appropriate sub-question and can think of other sub-questions under the same D-tree that are not answered by person B.

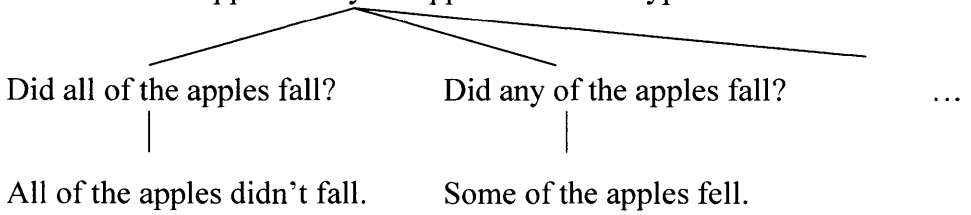
(26) A: Hey, the number of dishes is much fewer than the people who said would cook! Among them, which people brought something?
B: Hm, [Fred]$_{\text{L+H*}}$ brought something...$_{\text{L-H\%}}$

RFR in a sentence with a universal quantifier and negation, should be discussed along the lines of D-tree configuration. The CT semantic value of the utterance in (27), with CT on *all*, will be similar to (27a). Even when the immediately dominating sub-question (i.e., directly congruent question) does not precede the utterance, the listener would accommodate it and receive an implication that there is at least one other relevant sub-question under the same D-tree, and that the speaker does not answer it yet.

(27) ⟦ [All]$_{\text{CT}}$ of the apples didn't fall ⟧^{ct} RFR, a kind of lone CT

a. { ⟦Did all of the apples fall? ⟧, ⟦Did any of the apples fall?⟧, ...}
= { all of the apples fell, it is not the case that all of them fell,
some of the apples fell, it is not the case that some of them fell, ...}

b. What happened to your apple tree after a typhoon?



That is, the use of RFR is most naturally uttered as an answer to a question raised by an interlocutor specifically about a CT-marked element. Furthermore, the RFR condition, understood along with the CT congruence, requires that there should be another sub-question that is unanswered under the same D-tree. This means that RFR is not felicitous in most

[65] The term "immediately dominating sub-question" is used to mean a "directly congruent question" that we used in the chapters 1-3. Since the definition of CT congruence contains the notion of "immediately dominate", and the notion of dominance is intuitively easily captured by the illustration of D-trees, I will use the term in this chapter.

monologue contexts;[66] that is because if there is only one person involved in describing something, then they would have knowledge about what they are going to talk about. In other words, they know about the theme of their monologue, which will not leave unidentified answers of other sub-questions under the same D-tree. This felicity condition is expected to hold true with all the CT instances, since CT congruence requires an unanswered sub-question remaining under the same D-tree. The way it works with the RFR prosody with a universally quantified subject and negation is exemplified in the conversation in (28), where the immediate (i.e., directly congruent) question is not given in the preceding context, and the conversation is less natural versus, in (29), where the conversation is perfectly natural. The example in (30) illustrates a monologue that does not license RFR.

(28) A: What happened at the party?
B: # $\text{ALL}_{\text{L+H*}}$ my friends didn't come...$_{\text{L-H\%}}$ (with RFR)

(29) A: Hey, did all your friends come to the party?
B: $\text{ALL}_{\text{L+H*}}$ my friends didn't come...$_{\text{L-H\%}}$ (with RFR)

(30) I'm remembering how the party last night went.
#$\text{ALL}_{\text{L+H*}}$ my friends didn't come...$_{\text{L-H\%}}$

To sum up this section, it was discussed that RFR is a general phenomenon in English that is not limited to quantified expressions; despite the fact that the interactions between RFR and scope relations attract much attention. The mechanism of RFR, along with the principle of CT congruence, requires that there should be at least one alternative proposition remaining unresolved after assertion, and the unresolved alternative proposition is a relevant one in the context. In order for a sentence to felicitously have the RFR prosody, an immediate sub-question in D-tree should prompt the utterance; but the listener can most of the time accommodate an appropriate sub-question and think of another follow-up sub-question to figure out the broad question, even if the RFR utterance is not preceded by an immediate sub-question.

[66] Monologues are not felicitous for RFR unless one asks themselves a question, whose relevant sub-question they do not know the answer of. Or one can imagine that a speaker, in front of an audience, could felicitously use RFR; but in that situation the audience is involved and therefore one can regard it as a kind of conversation.

4.5 Previous experimental studies – universal quantifier and negation

4.5.1 Children's preference for surface scope interpretation

In the next two subsections, I will review another longstanding puzzle in children's language acquisition; which is the tendency of isomorphism found in the literature of acquisition of quantifier scope relations. As we have seen in §4.1, some English sentences are ambiguous but we do not often encounter a situation in real life where one cannot decide which of the possible interpretations is being conveyed by the sentence. That is because there are ways to disambiguate the meanings such as contexts prior to the utterance, follow-up comments, and prosodic information. However, what if you were *tested* in an experimental setting; where the sentence is deliberately left potentially ambiguous, and you were asked to judge if the sentence fits a situation which favors one of the meanings? Adult speakers obey the Principle of Charity (Davidson 1984, cf: Grice 1975) and accept the sentence, under the interpretation that is true given the situation; which is shown in the results of adult controls from experiments, which will be discussed shortly. This in turn leads to a question about what is happening with children.

For about twenty years since Musolino's (1998) observation of isomorphism, it has been reported in the acquisition literature that English-speaking children have a strong preference for the surface scope reading or the isomorphic interpretation. Consider the sentence in (31), which contain a universal quantifier in the subject position and negation.

(31) Every horse didn't jump over the fence.

a. No horse jumped over the fence. ($\forall > \neg$, surface scope reading)

b. It is not the case that all horses jumped over the fence. ($\neg > \forall$, inverse scope)

Many experiments have been conducted to evaluate children's comprehension of the sentence similar to (31), using the procedure of a Truth-value judgment task (TVJT, Crain and Thornton 1998). For an extensive review of the literature, see Musolino (2011). A typical procedure would be similar to the following: an experimenter tells a participant a story, using experimental materials such as props and pictures. Then a puppet, manipulated typically by another experimenter, tells the child what he thought happened in the story. The child is asked to judge if the puppet was right or wrong about the story.

In such experiments which test sentences as in (31), a participant is presented a story where two out of three horses jumped over a fence, and a puppet utters the target sentence as in (31), which participants are asked to indicate whether it fits the story or not. While English-speaking adults do not have difficulty accepting the sentence in (31) in such a case, meaning they interpret (31) to mean (31b), English-speaking children tend to reject the sentence on the grounds that they interpret (31) to mean (31a).

The preference is robustly replicated in a number of studies; children reject the sentence (31) as a description of a situation where a proper subset of the horses (e.g. two out of three) jumped over the fence 75%-93% of the time (Musolino 1998, see also comparable conditions in Musolino et al. 2000, Musolino and Lidz 2006, Conroy 2008, and in Viau et al. 2010, a.o.). It has also been noted that children at the same age are competent in accessing the interpretation of "Not > All" per se in sentences such as (32). They successfully accept the sentence as a description of a situation where the Smurf bought a proper subset of the oranges (e.g. one out of three), 85% of the time (Musolino et al. 2000).

(32) The Smurf didn't buy every orange.

This fact clearly indicates that children do not have difficulty with the interpretation of "Not > All" itself, but some mechanism to derive the interpretation from a sentence such as in (31). What would a common explanation to access the inverse scope interpretation be? In English, a standard analysis to account for the "Not > All" LF for (31) is that the universal quantifier in subject position can reconstruct/undergo Quantifier Lowering (QL) to a position lower than the negation at LF, which is generally assumed to be [Spec, vP], so that the negation c-commands the universal quantifier at LF. In cases such as in (32), the "Not > All" interpretation is less complex to derive; negation c-commands the object universal quantifier in the overt syntax. Thus there is no need for reconstruction/QL to access the "Not > All" interpretation, on the assumption that the lowest interpretable position for a quantifier in object position is below sentential negation.

Musolino and colleagues argue that the phenomenon that children have a strong preference for the surface scope interpretation can be characterized as the observation of Isomorphism-by-default (IBD); children prefer interpretations which are isomorphic to the surface scope relations between quantifiers and/or operators.

Though children do have the preference of being isomorphic, research has revealed that they are able to access the inverse scope reading to a reliably higher extent than the chance-level under certain circumstances. Specifically, when children's processing load is alleviated by experimental manipulation, they seem to be capable of accessing the "Not > All" LF (Musolino and Lidz 2006, Viau et al. 2010) to be reviewed shortly. In the current research, I take a different approach to children's competence with quantified sentences; we studied the effect of prosody on scope interactions. However, let us first review what the previous studies have found regarding children's capability to access the inverse scope interpretation.

4.5.2 Children can access the inverse scope reading

Let us review two pieces of literature that evidently show that children do access the non-isomorphic reading under certain circumstances. Musolino and Lidz (2006) hypothesize that children do not interpret the sentence in (31) to mean "Not > All" because children are not ready for processing negative sentences under the "Not > All" reading out with no context. Experiment 1 tested the set of conditions exemplified in (33). Condition 1 replicates the results reported in the previous literature and serves as the baseline for Condition 2. In Condition 2, an affirmative sentence precedes the target sentence, "in order to familiarize children with the intended domain of quantification (ibid: 825)."

(33) a. **Condition1:** Every horse didn't jump over the fence.

b. **Condition2:** Every horse jumped over the log, but every horse didn't jump over the fence.[67]

M&L tested 10 five-year-old children with Condition 1 (mean age 5;7) and another 10 five-year-old children with Condition 2 (mean 5;2) on a TVJT. They also tested 20 adult controls (10 for each condition). All of the target sentences in the experiment had a universally quantified subject (*every NP*) and negation. The stories were constructed so that a proper subset of the objects completed the action in question, i.e., "$\neg > \forall \wedge \exists$" is made true in the story. The authors observed children's "yes" responses to mean that they access the "Not > All"

[67] They also tested the sentences that employed *and* instead of *but*. The results did not differ from each other.

interpretation, and their "no" responses to mean that they interpret the sentence to have the "All > Not" meaning.

Results are the following: In Condition 1, children's "yes" response rate was 15%; whereas in Condition 2, children's "yes" response rate was 60%, which was significantly higher than in Condition 1. The adult control on Condition 1 yielded 92.5% "yes" response rate and on Condition 2 100%. These results from children show that the children are in fact able to access the "Not > All" interpretation even when it is not the isomorphic interpretation. They argue that, together with the results from their Experiment 2,[68] what is lacking in children's linguistic system regarding this phenomenon is pragmatic abilities; and that under certain contextual manipulation, such as using an affirmative sentence that would make the following negated sentence more readily comprehensible,[69] children's ability to deploy linguistic knowledge can be boosted.

Viau et al.'s (2010) Experiment 3 tests whether children would be "primed" by the interpretation (LF) of "Not > All". The two conditions in the experiment are illustrated in (34). Condition 1 is the baseline condition, in which six target sentences in the session have the same structure – a universally quantified subject precedes negation. On the other hand, Condition 2 has two phases. In the first half of the session, *Not* precedes a universal quantifier, and the latter half of the session uses the same sentences as the baseline condition.

(34) a. **Condition 1:** [trials 1 through 6] Every horse didn't jump over the pig.

b. **Condition 2:** [trials 1 through 3] Not every horse jumped over the pig.
[trials 4 through 6] Every horse didn't jump over the pig.[70]

[68] Experiment 2 tests whether negative sentences which have a universally quantified NP in the object position are interpreted differently depending on the mode (prediction mode vs. description mode).

[69] The idea is that "not p" is felicitous in a context where the possibility of p has been explicitly raised in the previous context.

[70] In this paper the authors do not discuss the issue of prosody. One wonders whether it is possible that in certain conditions the experimenter used something like RFR, which is natural for a "Not > All" interpretation. Were the equivalent sentences in Condition 1 and Condition 2 (e.g. Every horse didn't jump over the pig) presented as exactly the same auditory taper, for example splicing in the tape from Condition 1 into Condition 2, to insure that prosody remained constant? Is it possible that some of the experimenters unconsciously used more RFR prosody in trials 4 through 6 in Condition 2 than in the trials in Condition 1? This would be natural given that the first 3 trials of Condition 2 involved *not every*, and it may be natural to read with special prosody related to RFR. It would be good to control prosodic effects in these studies, as we do in our experiment.

In Condition 2, what is manipulated is as follows: The isomorphic interpretation of the sentence in the first half is "Not > All," and thus the first interpretation people will get with such a sentence is the "Not > All" reading.[71] That is, the participants will be forced to access the "Not > All" LF in the first half of the session. The prediction Viau et al. have is that if the participants are primed by the LF of "Not > All," then it will be easier to access the same "Not > All" LF in the latter half of the session even though the interpretation is the non-isomorphic, relative to the structure of the actual sentences given. The relevant stories support the "$\neg > \forall \wedge \exists$" situation just like the experiment in Musolino and Lidz (2006). Therefore, children's "yes" responses indicate that they access the "Not > All" interpretation and "no" responses indicate that they interpret the sentence to mean "All > Not".

Viau et al. tested 12 four-year-old children with Condition 1 and another 12 four-year-olds with Condition 2 on a TVJT (mean age of the 24 children = 4;6). In the baseline condition, the "yes" response rate for the first half is about 20%, and for the last half around 40%. On the other hand, the primed condition shows higher rate of "yes" responses overall. Both the first and last halves yielded 80% of "yes" response rate. The last halves of the two conditions were compared and they found that there was a significant difference between them. Based on the results, the primed children exhibited the higher rates of a "yes" response and the authors conclude that children do access the inverse scope "Not > All" interpretation; but accessing such an interpretation seems to involve processing difficulty, which can be alleviated by a structural priming of the relevant LF.

In sum, Viau et al. (2010) argue that children can access the non-isomorphic interpretation when the processing load is lessened. It seems reasonable that the results of Musolino and Lidz (2006) and those of Viau et al. (2010) can be narrated along similar lines, provided that the motivation of having an affirmative sentence prior to the target sentence in the M&L study, was to introduce the relevant domain of the quantification to the children so they can easily compute the LF that allegedly involves some processing load. That is, the experimental manipulation in Experiment 1 of M&L can be seen as removing some of the processing load from children. I would summarize that the two studies employed different methods that both found ways of making inverse scope LF more readily available by means of lowering the processing load.

[71] The "All > Not" interpretation is not contradictory to the sentence, although Viau et al. refers to this type of sentence as an unambiguous test sentence. The "All > Not" interpretation is generally excluded because of the implicature computations (often called "indirect" scalar implicature).

Gualmini et al. (2008) present an important set of results to show that contextual manipulation to support the QUD for the inverse scope interpretation leads children to access the inverse scope interpretation more often. They propose that the principle of Question-Answer Requirement (QAR) is observed in the comprehension of both children and adults, but children are more restricted in terms of the ability to come up with a relevant question. The idea of the principle of QAR is that a sentence is understood as an answer to a question, and the answer has to be relevant and informative for the question. The basic idea of every sentence being an answer to a relevant question has been fairly widely accepted as a standard assumption in the literature of pragmatics (Groenendijk and Stokhof 1984, von Fintel 1994). Gualmini et al. claim that in the cases where children prefer the surface scope reading, the inverse scope reading is not informative in terms of QUD. This leads to a prediction that if the inverse scope reading would give a relevant and informative answer with regard to the QUD, the inverse scope reading is preferred to the surface scope reading. This was confirmed with their stimulus sentences concerning the interaction of an existential quantifier and negation, such as in (36) and (38).

In their experiments, a situation of a story is as described in (35). Since this is a pizza-delivering context, a question under discussion is whether the Troll successfully delivered all of the pizzas. With the sentence as in (36), both the surface and inverse scope readings answer the question with relevance and informativity. The QAR view predicts that, children would mostly accept the sentence, based on the Principle of Charity. By contrast, with the sentence as in (38), only the inverse scope interpretation (39a) answers the question with relevance and informativity. The surface scope interpretation (39b) does not answer the question, since it just conveys that some pizzas were delivered. The QAR view predicts that children would reject the sentence more often, based on the inverse scope interpretation, and the Principle of Charity is overridden by QAR.

(35) **Situation:** Troll delivers 4 pizzas as quickly as possible, but on his way he loses 2 pizzas because he rushes too much.

(36) [QUD upon expectation: Did the Troll deliver all of the pizzas?]
Condition 1: Some pizzas were not delivered.

(37) a. It's not the case that some pizzas were delivered. (inverse, FALSE)
= No pizzas were delivered.

b. There are some pizzas that were not delivered. (surface, TRUE)

(38) [QUD upon expectation: Did the Troll deliver all the pizzas?]
Condition 2: Some pizzas were not lost.

(39) a. It's not the case that some pizzas were lost. (inverse, FALSE)
= No pizzas were lost.
b. There are some pizzas that were not lost. (surface, TRUE)

The results are the following: child participants (2;6-5;6, M = 4;6) accepted the sentences on Condition 1, 88% of the time, while child participants (2;10-6;1, M = 4;9) accepted the sentences on Condition 2, 43% of the time. Their adult controls accepted the sentences 100% and 93% on Condition 1 and 2, respectively. While Gualmini et al.'s experiments investigated the scope interaction between an existential quantifier and negation, it is quite possible that the idea can be extended to the case of the universal quantifier and negation.

To sum up this subsection, we have reviewed that in the literature on the interaction of universal quantifier and negation, children have a strong preference for the surface scope reading; however, they are capable of accessing the inverse scope reading to a good extent when processing load is lowered by setting up some context that lets the children be ready for the complex sentence, or by priming children with a certain interpretation prior to a target sentence. Our experiments, on the other hand, tackled the topic from a different perspective. We tested whether prosodic manipulation discussed in §4.2 and §4.3 affect children's (and adults') interpretation of the sentences in question; and more importantly, we ask whether children can sensibly reach the desired interpretation solely by the prosodic information, not relying on any of the explicit contextual manipulation.

4.5.3 Does prosody have an effect on scope-taking? – Adults

As we have seen in detail in §4.3, sentences with RFR which have a universal quantifier in the subject position and negation will most probably have the interpretation of "Not > All". Although this phenomenon has been discussed in a number of previous works based on native speakers' intuitions (Jackendoff 1972 and references in §4.2), there is little experimental work investigating (i) under what conditions speakers use RFR and (ii) what effects it has on listeners' comprehension. To my knowledge there are two exceptions provided by McMahon et al.'s (2004) and Syrett et al.'s (2014a, b) studies.

McMahon et al. (2004) study child-directed speech of parents and reported that parents do not differentiate prosodic patterns on scopally ambiguous sentences, as reported by Jackendoff, when they read to children. They also report that their adult participants showed no sensitivity to distinct prosodic cues in disambiguating a seemingly scopally ambiguous sentences in their comprehension study. While the participants did not show the effect with the scope disambiguation tasks, they seemed to show sensitivity to prosodic cues on another task, the pronoun referencing task; they tested whether parents use certain prosodic cues to disambiguate referents of pronouns (as in *Eddie rammed Mark into a haystack, and then HE rammed HIM right out of the yard.* (McMahon et al.'s (2))). They found that 64% of the parents utilized a pitch accent to indicate the reverse order in the pronoun tasks. Their comprehension task on pronoun disambiguation run with adult participants showed a significant effect of intended interpretation. The study showed that parents do not use prosodic cues to convey an intended interpretation in a scopally ambiguous sentence and therefore they concluded that children's failure to compute the inverse scope should not be due to their lack of attention to the prosodic cues.

Syrett et al. (2014a, b) conducted a production experiment and a set of comprehension experiments. In the production experiment in Syrett et al. (2014a), participants read stories in which potentially ambiguous target sentences appear. The context prior to the target sentence with a continuation make it clear which of the potential meanings the target sentence should convey. An example which they predicted to be accompanied with the RFR contour is cited below.

(40) A few years ago, the township decided to plant magnolia saplings to line a path through the park. The saplings on the north side were planted mainly in sand, and haven't been getting nearly enough nutrients. However, the soil near the south side is rich, and the magnolias are thriving there. All the magnolias won't bloom. But I bet the ones on the south side will.

The results, however, showed that the target sentence was read with *falling* contour 89% of the time. It follows that their prediction was not borne out, which lead them to conclude that the sentence-final contour, whether falling or non-falling, is not a good indicator of scope for the sentences with universal subject and negation. In fact, this statement is misleading; the results only suggest that *a speaker*, who knows which of the possible interpretations should be encoded

in the target sentence, does not employ the RFR contour to read this sentence. This by no means follows that the contour itself does not indicate one reading over the other for *a listener* to conjecture.

Syrett et al. (2014b), in turn, report an adult comprehension study. It shows that the participants were able to access different scope interpretations depending on what they heard. Let us review their Experiment 1, in which participants first viewed the target sentence on a computer screen, which comes with an audio stimulus of the sentence. An example of the target sentence is: *All the moms didn't allow eyeliner.* The audio stimuli were taken from those elicited in the production experiment. Then the second screen shows two kinds of continuations, one of which indicates the "All > Not" interpretation and the other the "Not > All" interpretation (e.g., A: *They were all in agreement.* B: *Only the moms of the older girls let their daughters wear it*). Participant's task is to choose the most natural continuation.

Their experimental items include 48 sentence. Among them, 24 are target items and 24 are control items (such as sentences with negation and *because*).[72] The two conditions that are relevant to our interest are (i) the prosodic contour that was produced to convey the "All > Not" meaning in the production experiment and (ii) the one that was produced to convey the "Not > All" meaning. When the sentence was read with the prosody (i), which was originally uttered to convey the "All > Not" reading, the rate of choosing the intended continuation was 64%. On the other hand, when the sentence was read with the prosody (ii), which was intended to convey the "Not > All" reading, the rate of choosing the desired continuation was 66%.[73] These correct response rates were significantly above chance, though clearly did not indicate full disambiguation by contour. Based on these results, the study argues that comprehenders can take advantage of the difference in contour that was produced with only one meaning in mind to arrive at the appropriate interpretation successfully.

---

[72] The target items consist of two kinds of sentences with *all ... not* read with two different prosodic contours (2*2 = 4 items) read by four different speakers (4*4 = 16 items), and of one kind of sentence with *many/most ... not* read with two different prosodic contours (1*2 = 2 items) read again by four different speakers (2*4 = 8 items). This adds up to (16 + 8 =) 24 items. The auditory stimulus was presented three times to the participants. My understanding is that it means one participant hears the same sentence 24 times in one session, three repeated by four different speakers and two different prosodic patterns for each speaker. I would question naturalness of this experimental setting, though it might make sense if a large effect size of a within-subject factor is observed.

[73] There was a relatively large variation in the rates of participants' accuracy between different speakers that elicited the target sentences. For example, speaker 1 utters a sentence with the prosody to convey the "Not > All" reading, and the participants access the correct interpretation 83% of the time. By contrast, speaker 4 utters a sentence with the prosody to convey the "Not > All" interpretation, and the participants access the correct interpretation 52% of the time.

This is a promising result in that it suggests that naïve adult English speakers could reach an interpretation specified solely by prosodic manipulation, as the theoretical literature argues. However, given that the average correct response rates were barely above 60%, the experimental effect might seem to some people to be weak, and one might wonder if experiment-specific and/or task-related factors (e.g., the auditory stimuli were originally recorded as a part of an elicitation task), might have prevented the participants from exhibiting the full competence to map the specific prosodic information onto a specific interpretation. As we will see in Chapter 5, our experimental results show that the prosody-meaning relationship is indeed stronger.

Let us discuss what we can learn from these experiments. First, I suspect that we can focus on consequences of a particular prosodic pattern in a clearer way if an experiment uses stimuli produced by a trained speaker. The implicature computation regarding RFR is imposed on *listeners*; that is, the mechanism does not particularly predict *speakers* to produce RFR when, for example, "Not > All" is expressed. The auditory stimuli in Syrett et al. (2014b) were recorded as a part of a production experiment. This means that the speakers were not expected to produce those sentences for others to judge. The participants in their experiment "were encouraged to read the items as naturally as possible, as though they were recording them for an audiobook or reading to children (Syrett et al. 2014b: 467)." An example of the scenario that they used in the production experiment is cited below:

(41) *An example story of their experimental item for "Not > All"*
Several moms were helping their daughters get ready for the upcoming school dance. This is a progressive school, and moms are usually lenient about certain things, so even the younger girls thought their moms would approve of eyeliner. But at the dance only the older girls were wearing it. All the moms didn't allow eyeliner. Only the moms of the older girls let their daughters wear it.

I suspect that "reading something as if reading to children" would lead to "being clear and lively," which does not exactly fit the circumstances where RFR is naturally licensed. Using RFR is going in the opposite direction of being clear; rather, the speaker is "being mysterious", which with more formal terms would be explained as "implicating that there are other relevant questions that the speaker has not answered." To me it seems that the speaker almost challenges the listener to figure out what is not said but is implicated; which is formally defined as the

principle of CT congruence. Speaking of speakers and listeners, it is also important to note that RFR is most naturally produced in dialogue where questions and answers are exchanged. In any case, in the experiment in question, intended meaning (e.g., whether "All > Not" or "Not > All") was clearly cued by context, and therefore I consider that there is no need for the speaker to be mysterious and employ RFR; in addition, there is no guarantee that the speaker was making a coherent speech regarding prosody within each trial. The participants were not expected to read the sentences aloud to let others who are blind to the interpretations to conjecture the intended interpretations experimentally. In our experiment, by using a trained speaker, we tried focusing on the effect of the prosody on interpretations listeners get, manipulated by the speaker.[74]

Secondly, it is possible that an experiment will yield clearer results if the environment that RFR appears in conforms to the principle of QAC. This is related to what I just pointed out in the previous paragraph; some context licenses RFR better than others. As the discussion in §4.4 suggests, RFR – more specifically, a sentence with a universally quantified subject and negation read with RFR, in our case – is most felicitous in a context where the broad question is to figure out the number of individuals who *VP*-ed. As discussed in Büring (1997, 2003), Constant (2012, 2014), and also suggested in examples in Jackendoff (1972), the "sharp rise and fall" part of RFR is an indicator of/analyzed as Contrastive Topic (CT). The CT congruence ((21), repeated below) requires that there is at least one alternative sub-question within the same D-tree which is not immediately answered by the utterance. The alternative sub-questions are generated by replacing the CT-marked element with its contextually relevant alternatives.

(42) CT congruence (=21)

An utterance U containing a contrastive topic can map onto a move M_U within a d-tree D only if U indicates a strategy around M_U in D.

U indicates a strategy around M_U in D iff there is a non-singleton set Q' of questions such that for each $Q \in Q'$, (i) Q is identical to or a sister of the question that immediately dominates M_U, and (ii) $[\![Q]\!]^o \in [\![U]\!]^{ct}$. (Büring 2003: 528)

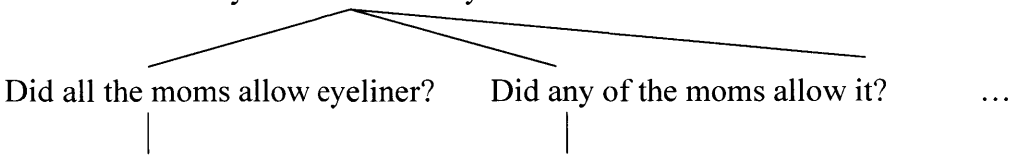
[74] There are two points of departure regarding this issue: trained vs. naïve speakers, and instruction to pronounce the sentence with the intent to disambiguate via contour vs. no such instruction. Specifically, there is a question whether naïve speakers will produce RFR when instructed to disambiguate meanings by prosody, as they are blind to the literature on RFR. An important question is which factors are key at this point; but this will remain unanswered for the time being.

Consider the story in (41). It seems that using RFR in this context is not felicitous because the target sentence describes an entailment of the previous sentences and is not easily understood as contrasting with different options. In other words, there is no other sub-question that is left open. Since the option of "All > Not" reading is already ruled out, this does not fit the felicitous circumstances where RFR is licensed. Then, what would the felicitous circumstances be? Since *all* is a CT in the sentence in question, the alternative sub-questions are about quantity of the moms who allowed eyeliner; and the immediate sub-question that the utterance in question answers is "Did all the moms allow eyeliner?" This is illustrated by using D-tree below.

(43) ⟦ [All]$_{CT}$ the moms didn't allow eyeliner ⟧^{ct}

a. { ⟦Did all the moms allow it? ⟧, ⟦Did any of the moms allow it?⟧, …}
   = { all the moms allowed it, it is not the case that all of them allowed,
   some moms allowed it, it is not the case that some of them allowed, …}

b. How many moms allowed eyeliner?



If the immediate sub-question is not present before the RFR utterance, a comprehender would accommodate an appropriate sub-question that the utterance answers (as well as a relevant sub-question for the sake of contrastiveness). If we set the context in the form of dialogue, and the immediate sub-question explicitly precedes RFR, a comprehender does not have to perform an accommodation and it might lead to results that are crisper. The dialogue that I would propose is the following:

(44) Q: So, what happened to the girls at the school dance party after all? Did their moms allow them to wear eyeliner?
A: Well, ALL the moms didn't allow eyeliner… Only the moms of the older girls did.

In our experiment in Chapter 5, we consider that if we employ dialogue and provide an appropriate QUD to obey the CT congruence as in (44), it might reveal a stronger effect and that we could extend their discussion. This is a crucial difference in our experiment compared to the experiments presented in the literature.

4.5.4 Testing whether children know prosody affects scope-taking

In an older study with rich data, Ianucci and Dodd (1980) carry out a picture-selection experiment with children ranged from K, Grades 2, 4, to 7 and with adults, to test whether differences in prosody lead to different interpretations.[75] Their experimental stimuli include five sentences, each demonstrating different kinds of sentence types with quantifiers. Among them, three conditions are relevant for our purpose and are cited below.

(45) **Condition 1:** All the rabbits aren't in the cages. (Stress on *all*, rise at the end – RFR)
Condition 2: All the rabbits aren't in the cages. (Fall at the end)
Condition 3: Not all the rabbits are in the cages.

In an experiment, participants were told to help the experimenter to create a picture book, by selecting a picture that would match the stories they hear. Their choice was between two pictures, one of that illustrates the "All > Not" situation, and the other picture that illustrates the "Not > All" situation.[76] They tested between 15 and 22 subjects per group and the results are summarized below.

[75] Many thanks to Thomas Hun-tak Lee for referring me to this very relevant paper.
[76] By the "Not > All" situation, I mean the situation where the implicature is computed; a "Not all, but some" situation. As discussed shortly, the "none" interpretation is not contradictory to the "Not > All" interpretation. However, the "none" interpretation is generally excluded through the computation of Scalar Implicature. In Ianucci and Dodd's study, the "Not > All" picture showed two out of five rabbits being captured in a cage.

Table 1: Rates of choosing "Not > All" pictures in Ianucci and Dodd (1980)

Groups	Condition 1	Condition 2	Condition 3
K	38%	27%	55%
Grade 2	47%	19%	79%
Grade 4	46%	17%	83%
Grade 7	58%	18%	100%
Adults	96%	18%	100%

Children (and of course, adults) behave differently to Condition 1 (RFR) and Condition 2 (Falling). On Condition 2, the participants chose the "All > Not" pictures most of the time. By contrast, on Condition 1 adults chose the "Not > All" pictures most of the time; while the "Not > All" rates by children were between 38% and 58%. Given that the choices were among two options, the children chose the "All > Not" pictures 62-42% of the time. Though they do not provide statistical analyses, the rates of choosing the "Not > All" pictures on Condition 1 seem consistently higher than those of Condition 2. If this actually holds statistically, the results suggest that the children interpreted the two kinds of prosody differently, although the connection between RFR and the "Not > All" interpretation does not seem to be as strong as for adult speakers.

It is also worth noting that in their experiment the rates of choosing the "Not > All" picture on Condition 3 for preschoolers were as low as 55%. The results follow if children do not compute, or at least have difficulty with, Scalar Implicatures (SIs). Generally, a sentence such as in Condition 3 of (45) has a scalar implicature that there exist some rabbits in the cages (such SIs found in downward entailing environments are often called "indirect" scalar implicatures, but derived via the same mechanism as for direct SIs).

(46) a. Not all the rabbits aren't in the cages.

b. $[\![(46a)]\!]^f = \{$ it is not the case that all the rabbits are in the cages,
it is not the case that some of the rabbits are in the cages, …$\}$

c. Logical strength: $\neg\exists$ entails $\neg\forall$

d. Scalar implicature of (41a): $\neg(\neg\exists)$ $\therefore\exists$

If children do not compute SIs, then the interpretations of "none" is not disfavored relative to the "Not all but some" interpretation, upon listening to (46a). What would happen if they had to choose among the two pictures in question? If one picks randomly since both are fine to them, then would account for 55%, which would be regarded as chance. It is well-known that children perform at lower rates than adults on computing SIs, at least through Grade 2, often later (Noveck 2001, Papafragou and Musolino 2003, Guasti et al. 2005, Huang and Snedeker 2009, Barner et al. 2011, a.o., and for comparison of direct SI and indirect SI, see Shetreet et al. 2012, Bill et al. 2014). Not computing implicatures is likely to explain why the kindergartners (and to some extent the second graders) did not perform adult-like on Condition 3.

Another possible, more trivial explanation on the low rates of choosing the "Not > All" picture on Condition 3 by children could be that the preschoolers in this experiment, plainly did not particularly prefer to choose "Not > All" pictures even when the isomorphic interpretation is "Not > All." The children, 4-years old, in Viau et al.'s (2010) experiment accepted the "Not > All" interpretation in a TVJT for sentences like those in Condition 3 of (45) more than 80% of the time. In comparison, it is possible that the children in Ianucci and Dodd's study had a stronger preference for the "All > Not" interpretation overall; or that some experimental confound had prevented the children from accessing the "Not > All" interpretation when adults preferred to do so.

The study by Ianucci and Dodd is suggestive in that it shows that children regard the two types of prosody (RFR and falling) differently, and the preferences in interpretation are in the same direction as adults. Namely, with RFR, more "Not > All" reading is elicited. Moreover, the study obtained excellent results with adults, strongly confirming linguists' intuitions regarding the effects of prosody. Our current study reported in Chapter 5 aims, to follow-up on their study, with more standardized procedure such as wider variety of items, using recorded sound files by a trained speaker, and employing a dialogue which makes the context felicitous for RFR. These items also improve on the experiment of Syrett et al.

Chapter 5

Acquisition of Quantifier Scope: Evidence from English Rise-Fall-Rise*

5.1 Hypothesis and predictions

There are two points of departure for this study: the rich theoretical discussion on the RFR contour and associated semantic consequences; and the acquisition literature that consistently finds the isomorphic tendency of children. The current study aims to experimentally confirm the prosody-semantic link and examine whether children know the effect of RFR on the scope assignment, despite their isomorphic tendency. As we have seen in Chapter 4, a sentence with a universally quantified subject and negation[77] read with the RFR prosody has the interpretation of "Not > All." This is possible because RFR is conditioned to maintain at least one LF among relevant alternatives that is not resolved after assertion. Furthermore, in order for a sentence to felicitously license RFR, an immediate sub-question in a D-tree should serve as the current QUD, rather than a broad question. As stated in the acquisition literature, and following the review of studies in §4.5.1 and §4.5.2, children can access the inverse scope interpretation to a good extent when processing load is lowered by setting up certain contexts or by priming children with the target interpretation with another grammatical structure. Could RFR be yet

* Some parts of the contents reported in this chapter are based on the results of our experiments reported in Sugawara et al. (to appear). For the experiments conducted in this chapter, I would like to thank Sammy Floyd and Laura Schulz at the MIT Early Childhood Cognition Lab at Boston Children's Museum for their courtesy. I would also like to thank our MIT Undergrad Research Assistants, Sebastian Garza and Irina Onoprienko for their help with recording sound files and conducting experiments. Also many thanks go to Athulya Aravind, who helped me with running some of the experiments.

[77] Without other alternative-generating elements – in order to avoid counterarguments raised by Ward and Hirschberg (1985).

another manipulation that makes children access the inverse scope reading as it does with adults? This potential manipulation is different from the experimental manipulations that we have already observed; we do not manipulate contexts/stories before children encounter the target sentence. In other words, the story does not bias either of the interpretations; and thus, prosody is the only cue that the participants' interpretations are based on. The following will demonstrate the functionality of this process.

The question that this study asks is whether children know the effect of RFR. This question is decomposed into two hypotheses, and the discussion on QAC raises a third hypothesis to test.

(1) **Hypothesis 1:** Children are sensitive to the difference in contour, and therefore we would expect different reactions to different contours (RFR contour vs. a neutral contour).

Hypothesis 2: If H1 is confirmed; Children compute the effect of RFR in the same way adults do, and therefore the rates of the "Not > All" reading will increase with the RFR contour.

Hypothesis 3: If H2 is confirmed; Children are sensitive to QAC (more specifically, CT-congruence), and the rates of the "Not > All" reading will increase with the immediate sub-question preceding RFR.

The first two hypotheses might sound redundant, but they are worth presenting separately. The acquisition literature on contrastive stress has seen mixed results and it remains controversial whether children are sensitive to the difference in prosodic information, such as being suprasegmental or not.[78] The studies that report children's sensitivity to the prosodic cues (Arnold 2008, Ito et al. 2012, etc.), investigated contrastive stress on a noun phrase or an

[78] Pre-test experiments in McDaniel and Maxfield's (1992) study revealed that children did not behave adult-like in an act-out task where they were expected to pet Grover upon hearing trigger sentences such as "Grover is patting Bert. Now YOU pat HIM." We will take a closer look at their results, in a developmental context, in the discussion section in §5.2.4.2. Baauw et al. (2004) study Spanish-speaking children and found that the rate of children's adult-like performance on pronominal referent in sentences like "First, the woman$_i$ kissed the girl$_k$, and then *ELLA$_k$ besó al niño* (SHE kissed the boy)" was around 44%, which was significantly lower than non-stressed sentences. Wells et al. (2004) report asymmetry between the production and the comprehension of the focus-marking prosody, with the comprehension developing slowly compared to production. On the other hand, using an eye-tracking method, Arnold (2008) report that 4- and 5-year-old English-speaking children have the bias toward *given* objects with deaccenturation, and Ito et al. (2012) report that 6-year-old Japanese-speaking children are sensitive to prosodic cues to mark contrast. It seems that results from studies using offline measures support insensitivity, while the results from studies using online measures tend to reveal sensitivity.

attributive adjective. It is important to investigate whether RFR, which is stretched over the whole sentence and not just about one word, is perceived differently compared to a neutral prosody by children. Hypothesis 2 mentions adult reactions to RFR; however, there is little experimental support for the argument that naïve adult speakers will compute the effect of RFR and generate the "Not > All" reading. To the best of our knowledge, there are two exceptions for this; the study by Syrett et al. (2014b) and the study by Ianucci and Dodd (1980), which were reviewed in §4.5.3 and §4.5.4. It is important to note that the current study expands in novel ways on their studies.

5.2 Experiment 1 – Picture-selection task

5.2.1 Method and design

We conducted an elaborated version of the picture-selection task, where the picture-selecting phase is preceded by a short story involving two people. We employed dialogue in the stories, since RFR is most naturally elicited as an answer to a question. Since the RFR sentences that we use contain a universally quantified subject and negation, a question that asks about the quantity of the subject NP licenses the RFR utterance, as discussed in §4.4. Figure 1 illustrates the procedure of one of the experimental items. The experiment was conducted using a PowerPoint slides on a laptop computer. The sound files were pre-recorded and played to the subjects as the experimenter clicks a button to advance the story.[79]

[79] The female voice was recorded by Irina Onoprienko, who is an undergraduate research assistant and is familiar with the literature. The male voice was recorded by another undergraduate research assistant. Both of the speakers are native speakers of English. The recording took place in a sound proof booth. After recording, we normalized the amplitude of all the sound files.

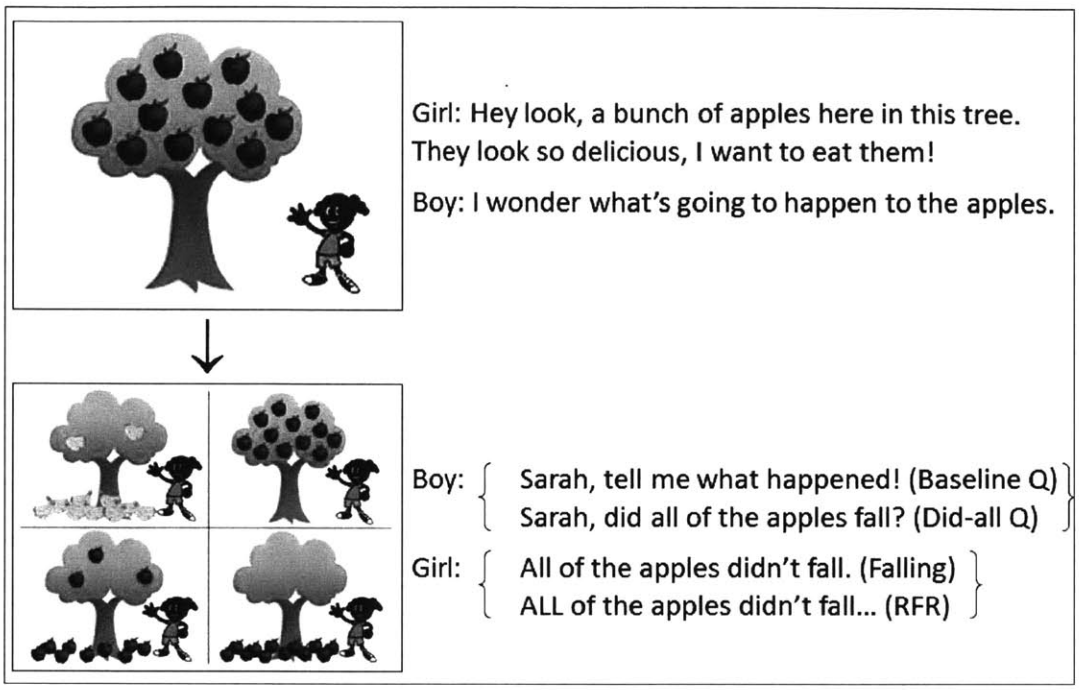


Figure 1: Experimental procedure

The first phase of a story is an introduction. The girl mentions the objects in the story, and the boy says he is wondering what is going to happen to the objects. The second phase of the story shows four pictures on the computer screen. The four pictures describe situations where (i) All > Not (none of the objects VP-ed), (ii) Not > All (Some but not all of the objects VP-ed), (iii) All did (All of the objects VP-ed), and (iv) irrelevant situation, are depicted. The position of the pictures were pseudo-randomized. Shortly after the second phase picture is shown, the boy asks the girl a question to prompt the answer. In response, the girl utters a target sentence. A subject was asked to point to the picture that they thought the girl was in or talking about, and their responses were recorded by an experimenter.

We cross-examined the question type (Baseline Question vs. Did-All Question) and contour (Falling vs. RFR), which was established as a between-subjects factor.[80] Figure 2-3 show a pair of the pitch tracks from one of the experimental items.

[80] We made contour a between-subjects factor because we ran a pilot experiment that employed the contour as a within-subjects factor, and we observed a seeming order effect during the pilot experiment, which was hard to interpret if the sentences were treated independently by the children. It is also possible that the effect of prosody carried over between sentences with different contours.

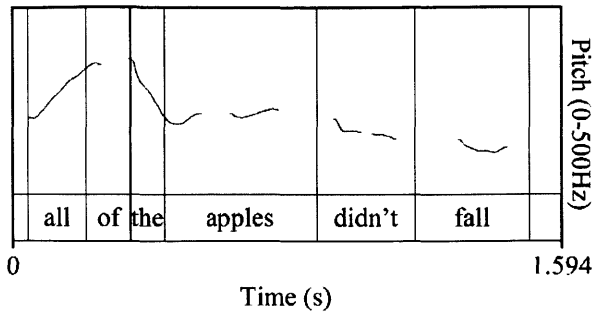


Figure 2: Pitch track of an item on the Falling condition

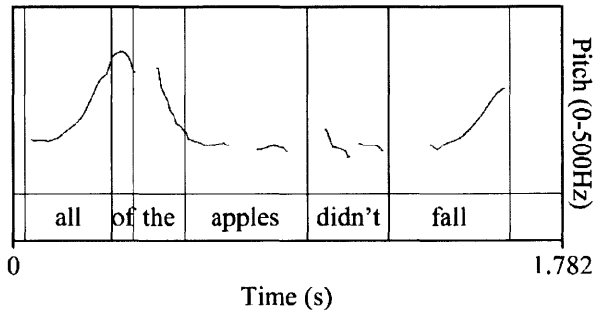


Figure 3: Pitch track of an item on the RFR condition

An experiment consists of 8 target items (4 with Baseline-Q and 4 with All-Q) as well as 6 filler items, pseudo-randomized. We prepared two lists with items in different order, and each participant was randomly assigned to one of the lists. The sentences in the filler items were all non-negated sentences. An example of a target sentence and two examples of the filler trials, as well as the list of the verbs used are found below in (2). See Appendix 4 and 5 for the list of all the target items.

(2) An example of the target sentence

a. All of the apples didn't fall.

Two examples of the filler sentence

b. All of the glasses broke. (Basic simple past)

c. All of the helicopters did fly. (Emphatic past)

List of the verbs

d. *break, dry, fall, fly, grow, open* (in addition, *close* as a verb was used in a filler)

Since the filler items are affirmative sentences, the participants had to pay attention to positive pictures as well as the "Not > All" and/or the "All > Not" pictures. Moreover, half of the filler items had emphatic "did" as in (2c).[81] Because of this manipulation, the participants could not jump to an early conclusion that the sentence should mean either of the negated situations upon hearing *did*. Instead they had to pay attention to what follows, to see whether the negation is present or not. When a child wanted to listen to the sentence again, the question and answer pair were played together. A session for each participant typically took about ten minutes to complete. Some sessions took place at local daycares in a relatively quiet room or space within the classroom environment; and other sessions took place at Boston Children's Museum in a quiet room with caretakers present throughout the session. The caretakers sat behind the participants and were asked not to interrupt children's games. Children were praised and encouraged regardless of the choice during the session. For adult controls, an answer sheet with sets of four cells for each item was handed to them, and they were asked to mark the appropriate answer.

5.2.2 Results – Adults

We recruited 24 adult volunteer participants (12 on the Falling condition and 12 on the RFR condition) with no or little linguistics background. The participants were mostly undergraduate students either at MIT or at Wellesley College.

Of 192 relevant data points (8 items * 24 subjects), 2 data points were excluded from the analysis since they were responses that chose a positive "All Did" picture. That is, 190 data points were analyzed as binomial parameters. Figure 4 shows the rates of choosing "Not > All" pictures by condition. The error bars indicate 95% confidence intervals.

[81] In order for the emphatic sentence to be naturally uttered, the preceding prompt for the sentence type in (2c) was always a "Did-All" Question. E.g., Q: Sarah, did all of the helicopters fly? A: All of the helicopters *did* fly.

To make the number of the different question types equal, the filler sentences with basic simple past was precede by a broad question. E.g., Q: Sarah, tell me what happened! A: All of the glasses broke.

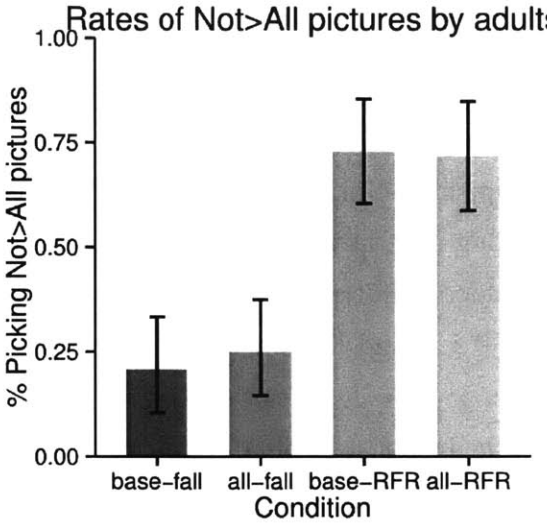


Figure 4: Results from adults

Table 1: Statistical analysis of the results from adults

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.4466	2.0767	-1.660	0.0970 .
contourRFR	7.7889	3.7209	2.093	0.0363 *
qTypeBase	0.9689	1.5518	0.624	0.5324
contourRFR: qTypeBase	-2.3461	2.8214	-0.832	0.4057

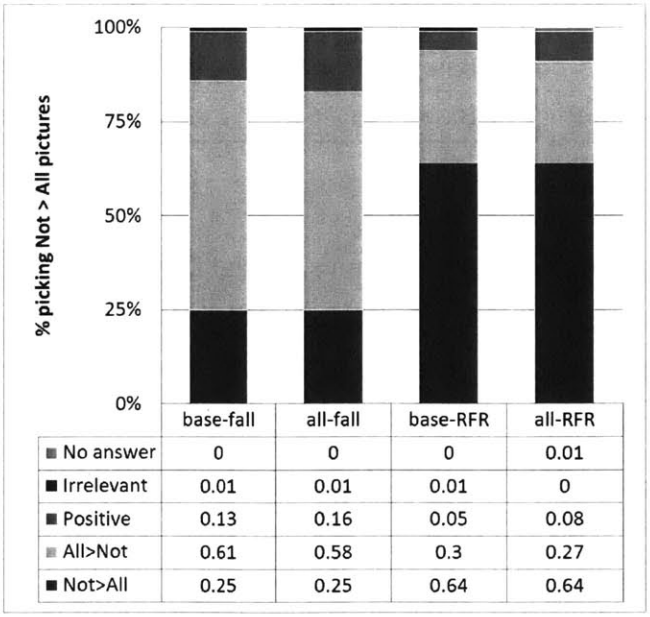
The rates of choosing the "Not > All" pictures for the BaselineQ-Falling condition was 21%, for the AllQ-Falling condition was 25%, for the BaselineQ-RFR condition was 73%, and for the AllQ-RFR condition was 72%. See Appendix 4 for the list of the sentences and the rates of the "Not > All" pictures for each item. Using GLMEM, statistical analysis of these rates revealed a main effect of contour (p = .036).[82]

[82] Since the maximally specified model did not converge, the order of the presentation was not considered as a possible factor.

5.2.3 Results – Children

We recruited 36 child participants from Boston area daycares and at the Boston Children's Museum across all socioeconomic and ethnic backgrounds. All the children are native speaker of English. Children who skipped two or more items (n = 3) and children who answered three or more fillers out of six wrong (n = 1) were excluded. A total of 32 children are included in the analysis. Among the included children participants, 16 were tested on the Falling condition (ranging from 4;4 to 6;10, M = 5;3) and 16 were tested on the RFR condition(ranging from 4;5 to 6;7, M = 5;2).

Figure 5 summarizes all the responses from the child participants, including the errors (i.e., choosing positive "All Did" pictures or irrelevant pictures). The errors account for 30 data points out of 256 data points (8 items * 32 subjects), and the popular errors were to choose the positive pictures as presented in Figure 5 below in the graph. The error rates were statistically not different across the four conditions. The 30 error responses are excluded from the analysis hereafter. This allows us to treat the responses as binomial answers since the remaining data points are either responses for the "Not > All" or "All > Not" pictures.



	base-fall	all-fall	base-RFR	all-RFR
No answer	0	0	0	0.01
Irrelevant	0.01	0.01	0.01	0
Positive	0.13	0.16	0.05	0.08
All>Not	0.61	0.58	0.3	0.27
Not>All	0.25	0.25	0.64	0.64

Figure 5: Grand results from children

The normalized results are summarized in Figure 6. The error bars indicate 95% confidence intervals. It shows the rates of choosing the "Not > All" pictures by condition. The rates for the BaselineQ-Falling condition was 29%, for the AllQ-Falling condition was 30%, for the BaselineQ-RFR condition was 68%, and for he AllQ-RFR condition was 71%. See Appendix 5 for the rates of choosing "Not > All" for each item in the experiment. Statistical analysis of these rates using GLMEM reveals a main effect of contour ($p = .038$).[83] The effect of question type was not observed.

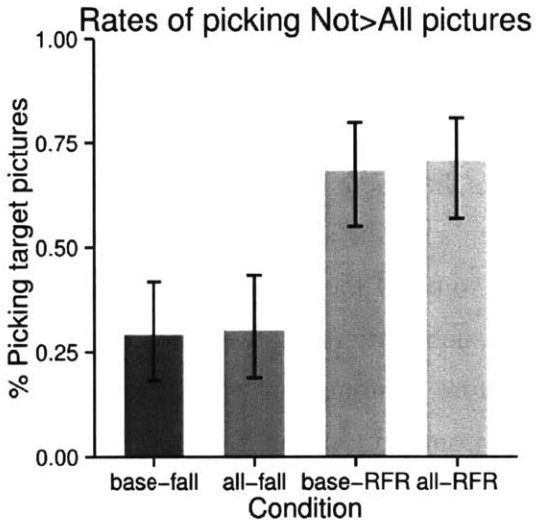


Figure 6: Normalized results from children

Table 2: Statistical analysis of the results from children

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.021	1.232	-1.640	0.1009
contourRFR	3.480	1.678	2.074	0.0381 *
qTypeBase	-1.814	1.622	-1.119	0.2633
contourRFR: qTypeBase	2.432	2.195	1.108	0.2679

[83] Again, since the maximally specified model did not converge, the order of the presentation was not considered as a possible factor.

5.2.4 Discussion

5.2.4.1 Discussion - Adults

The results from adults show that naïve speakers chose the "Not > All" interpretation significantly more often when the sentence is read with the RFR contour. The finding supports the hypothesis that the "Not > All" interpretation is strongly preferred with RFR. Another important thing to note is that in our experimental setting, the preceding context (i.e., Phase 1 in the story) was always the same across and neutral for both of the Falling and the RFR conditions. Participants were not able to tell which picture to choose until they heard the target sentence. The participants could arrive at their respective interpretations using *only* prosodic cues.

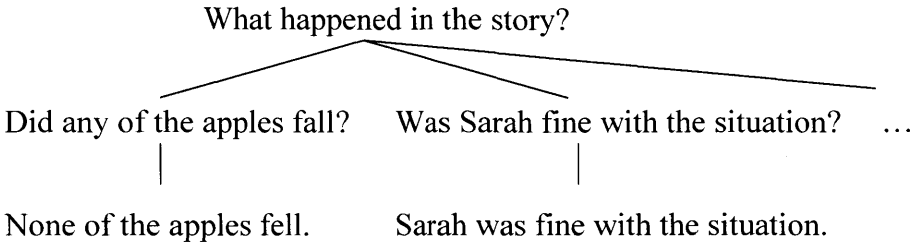
A curious thing to note is that there was no effect of the question type detected. This was also the case with the children's results and we will discuss this point in §5.4.

One might wonder why some of the adults did not choose the "Not > All" pictures even with RFR. It is possible that such participants were not aware of the specific contour and failed to compute relevant implicature computation with RFR. This may have occurred because the experiment was done with contour type as a between-subjects factor, to create a parallel experiment to the child experiment. An individual subject analysis shows that the "All > Not" responses under the RFR condition came from a small subset of the participants; one subject consistently (8 out of 8 items) picked "All > Not" pictures, and three subjects strongly preferred "All > Not" reading (6 or 7 out of 8 items). The other 8 subjects always (100%) picked the "Not > All" pictures. In other words, the choices did not follow a normal distribution, but was strongly bimodal, with modes occurring at the extreme values. The fact that some subjects kept choosing "All > Not" pictures even when the sentence was read with RFR, suggests that those subjects have a strong preference when it comes to the sentences with a universally quantified subject and negation; it is possible they did not pay much attention to or ignored the prosody due to this strong preference.

Alternatively, it is possible to yield the "All > Not" reading even when the implicature is correctly computed, and this possibility might account for the rationale for the few isomorphic subjects. As Constant (2012, 2014) states, and as we discussed in §4.3, RFR is *not* a disambiguating contour; RFR is conditioned so that there remains at least one alternative

proposition among relevant alternatives that are not resolved/dispelled by the assertion. Disambiguation is often observed as a result of computing the effect of RFR; however, if at least one contextually relevant alternative proposition remains unresolved after asserting, such as with the "All > Not" meaning, it is technically possible to maintain the "All > Not" interpretation for a sentence with RFR. This can be considered with a D-tree presented in (3).

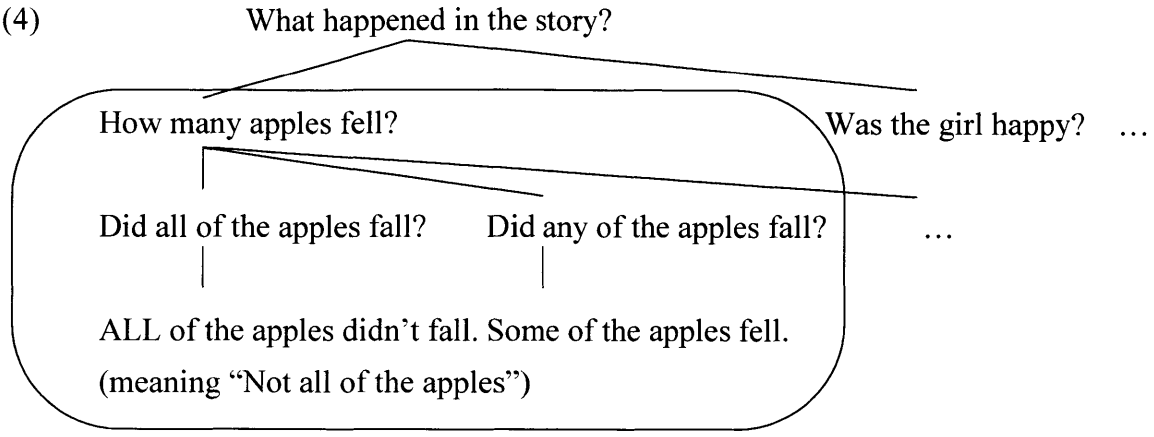
(3)



In the D-tree in (3), with the question at the top being a broad question "*What happened?*", one can imagine a wide variety of sub-questions that are relevant to figure out what is happening in the story. For example, "*Was Sarah fine with the situation?*" is a possible sub-question to ask to figure out what is happening in the story. Under such a circumstance, the "All > Not" meaning, i.e., "*None of the apples fell*", does not resolve all the other sub-questions. In other words, if one entertains a D-tree similar to the one illustrated in (3), it is technically possible to interpret the sentence with RFR to mean "All > Not", while respecting the condition regarding RFR.

In the current study, we have tried to narrow down the range of the alternative questions by having the boy explicitly raise the question about the object (e.g., "*I wonder what's going to happen to the apples.*"); and, for half of the trials, he asks about the quantity of the objects (e.g., "*Sarah, did all of the apples fall?*"). Due to this manipulation, the relevant D-tree is deliberately prepared based on the circled portion presented in (4), where the only possible interpretation is "Not > All", since the "All > Not" interpretation would resolve all the other sub-questions under the relevant D-tree.

(4)



However, one can become very creative so that it might be possible for them to assume a D-tree such as in (3), possibly driven by an idiosyncratic preference for the surface scope interpretation. We entertained two possible accounts for the "All > Not" interpretations by some adult speakers. However, the plausibility of the accounts will not be discussed at this point.

In conclusion, the results from adult controls in our experiment provide strong support for the idea that the RFR prosody leads a listener to arrive at the "Not > All" interpretation in the sentences. Our study, with a new methodology, highlighted the differences between a neutral and an RFR prosody. Even though these difference were already identified in Ianucci and Dodd's study and in Syrett et al.'s study, our study produced cleaner results to provide clarity to these differences. In addition, explanations involving context could be more easily ruled out in our study, since everything right before the target sentence was exactly the same across the Falling and the RFR conditions. This allows us confidently argue that adults have knowledge of the prosody-semantic link, which they are capable of exercising independent of the context.

5.2.4.2 Discussion - Children

Results from child participants conforms closely with those of adults. We aimed to test the following three hypotheses, repeated in (5).

(5) **Hypothesis 1:** Children are sensitive to the difference in contour, and therefore we expect different kinds of reactions to different contours (RFR contour vs. a neutral contour).

Hypothesis 2: If H1 is confirmed; Children compute the effect of RFR in the same way as adults do, and therefore the rates of the "Not > All" reading will increase with the RFR contour.

Hypothesis 3: If H2 is confirmed; Children are sensitive to QAC (more specifically, CT-congruence), and therefore the rates of the "Not > All" reading will increase with the immediate sub-question preceding RFR.

The result of this study proved the first two hypotheses. We observed the main effect of contour, and suggest that children are indeed sensitive to the difference in prosody. Since the literature is controversial as to whether children around the age of 4 and 6 are sensitive to the suprasegmental information, and studies report that adults do not reliably produce the distinctive contours to indicate distinct scope relations in child-directed speech (McMahon et al. 2004), our results might come as a surprise. However, our experimental materials do not allow other possible accounts for the significant main effect of contour. As we discussed and specifically illustrated in Figure 1, the preceding stories continued until target sentences were exactly the same (same sound files) across the testing conditions. The preceding stories did not favor one interpretation to another (See Appendix 4-5), and the only difference was the prosody of the target sentences.

Furthermore, the direction of the main effect of contour was similar as the results for adults; meaning that children compute the effect of RFR in an expected fashion. As we have seen in Chapter 4, the computation involved in RFR seems complex. It involves alternative generating and implicature computation. In spite of the complexity, children do get the effect. We have reviewed studies in §4.5 that show children access the non-isomorphic interpretation when the processing load is alleviated. Our results add to the literature that children do get a substantial amount of non-isomorphic reading in our experimental settings. It is also important to note that our results contrast with previous studies in that children's ability to access the non-isomorphic interpretation increases without a manipulation to lessen the processing load; computing the effect of RFR might put more processing load on children, since RFR involves alternative generation and implicature computation. [84] The possible reason why the children were successful in computing the effects of RFR, is that the pressure to make a sentence favorable to

[84] Specifically, it is possible that RFR is a conventional implicature, as Ward and Hirschberg (1985) and Constant (2012) maintain.

the readings is so strong that it works as a driving force for the child accessing the "Not > All" reading. One way to think of this result is that RFR is a visible, kind of "lexical" signature forcing a calculation of a particular kind over alternatives. This calculation can be done with the RFR prosody.

The following is a closer examination of adult and child RFR results. At first glance, the figures look very similar; 73% and 72% (mean by collapsing contours = 72%) with adults and 68% and 71% (mean = 70%) with children. As we discussed in §5.2.4.1, all of the "All > Not" responses from adults come from a small subset of the participants. That is, there are several "isomorphic" people whose responses contribute all of the "All > Not" responses in the RFR condition. What about children? It turns out that the "All > Not" responses, which account for 30% of the answers in the RFR condition, are also coming from a subset of children and not sparse across individuals. What is interesting is that these "isomorphic" children are found virtually only in the population younger than 5 and a half of age. Figure 7 shows the correlation between the rates of choosing "Not > All" pictures (out of 8 trials) on the RFR condition and children's age in days. Indicators of the age of 4;5, 5;0 and 6;0 as a guide are also shown. The correlation was statistically just near significant, but we observe that there is a trend that the age and the "Not > All" rates on the RFR condition are likely to be positively correlated with a medium-to-large effect size ($t = 1.78$, $df = 14$, $p = .097$, $r = .429$).[85]

As presented in Figure 7, children can be separated into two groups: the group who almost always understands the effect of RFR, and the group who are almost always isomorphic. Just as for adults, the distribution of choices for children is not a normal distribution, but rather bimodal, with the modes at the extreme values. The way different kinds of population distribute in the results of children look very similar to those of adults; whereas, the percentages of isomorphic participants is quite close for the child and adult populations. The correlation graph shows that children who almost always get the effect of RFR range from four years old to six, but the children who are almost always isomorphic can only be found in the population younger than 5 and a half years old. One possibility is that the isomorphic group might undergo some kind of maturation in sensitivity to the prosodic contour that calls for computation over scopal alternatives. The other possibility is that the maturation involves the ability to compute scalar implicatures, or possibly to learn/construct the relevant set of alternatives for the scalar alternative computation. Five and a half years old appears to be an age that sometimes divides

[85] Unfortunately, the sample size is too small for proper analysis, which likely resulted in near significant results.

certain levels of a good performance on scalar implicatures with other types of items such as *some/all* pairs.

An analysis was run with participants tested on the Falling condition, but no correlation between the age and the rates of choosing "Not > All" pictures was found ($t = 0.599$, $df = 14$, $p = .559$ $r = .158$).

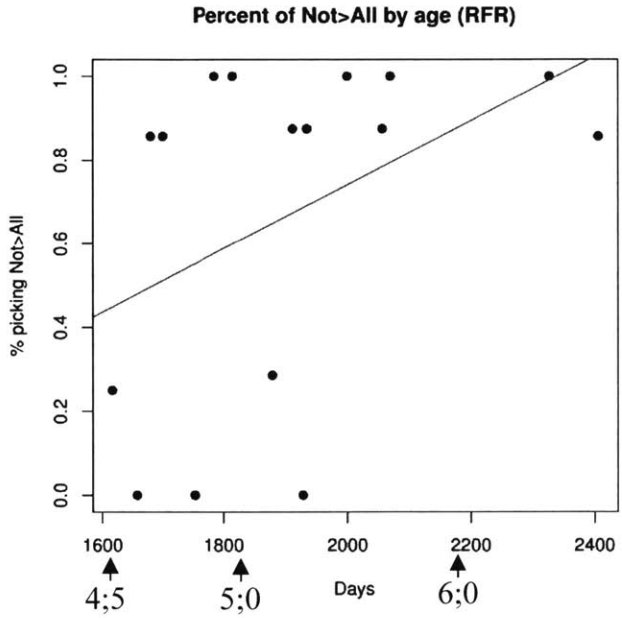


Figure 7: Correlation between % choosing "Not > All" pictures on the RFR condition

Did the isomorphic adults not go through a maturation process? It is assumed not to be the case; and as reviewed in §5.2.4.1, there are ways to compute the RFR effects and stick to the "All > Not" interpretation at the same time. It is an accommodation strategy that accounts for at least some of the isomorphic responses by adults.

In the context of developmental path and sensitivity to prosodic cues (and its consequences), let us discuss the results of pre-test sessions in McDaniel and Maxfield (1992). The aim of their experiment was to test the hypothesis that once children become sensitive to contrastive stress, they correctly understand Principle B; given that there are utterances in adult speech such as "Guess who I chose? I chose *me*," where Principle B is violated but *me* with stress can co-refer the speaker in adult grammar. They found that there was a correlation between scores on contrastive stress sensitivity and performance on the knowledge of Principle B in children. Our

interest is in the scores on contrastive stress sensitivity and age that are reported in their study. They tested 35 children ranging from three years old to six years old. Their test items include act-out tasks to investigate whether the children can whisper to Grover when instructed as "Goofy is whispering to Grover. Now YOU whisper to him.", and whether they can pet Grover when instructed as "Grover is patting Bert. Now YOU pat HIM." The children's scores on contrastive stress sensitivity ranged from 2 to 9 (out of 10), and the average score for three year olds was 4.7, for four year olds was 4.7, for five year olds was 5.8, and for six year olds was 6.7. Please note that the sensitivity scores look very much like our results on the rates of choosing "Not > All" on the RFR condition by age. We do not know if the children in their experiment showed bimodal distribution similar to our experiment, but when averaged, it seems that the children in their experiment acquire sensitivity to contrastive stress sometime during the time of five years old, and the ratio of sensitive children increases in six year olds. I do not think it is incidental that in our experiment as well most of the children who are over 5 and a half years old, show clear sensitivity to the prosodic contour regarding RFR.

5.2.4.3 Discussion – Methodology

The method we used, an elaborated version of a picture-selection task, is suitable to test the effect of RFR for three reasons. [86] The first reason is that the experimental procedure supports the way listeners compute the effect of RFR and it does not interfere with it. The computation of the effect of RFR goes as follows; a listener generates a set of sets of alternatives (a set of questions) modulo Contrastive Topic, and then compare and choose an LF so that it satisfies the condition of having at least one sub-question remaining unresolved under the same D-tree. The experimental procedure of the picture-selection task aligns well with the way the effects are computed; and a listener hears the sentence with RFR with no prior bias at this moment, and they are guided to evaluate its meaning and choose one that fits among a set of alternatives (i.e., four pictures in the second phase). The common procedure among the computation of the RFR effect and the picture-selection task is that the listener has a set of alternatives and then they choose one situation that best describes what the sentence means. In a Truth-value judgment task, the experimental procedure does not align with the computation of the RFR

[86] This discussion is meant to contrast the picture-selection task and a Truth-Value Judgement task when applied to the phenomenon of RFR. Notice that this discussion never argues that TVJT is not suitable to test scope relations in general.

effect. In a TVJT, a completed story is presented to the listener, then they hear a target sentence with RFR. The set of alternatives are not physically present in this method, though the listener who computes the effect of RFR entertains the set of alternatives, in theory; which might result in underestimating the comprehender's ability to compute the effect.

The second reason is that with our picture-selection task, the listeners are not biased until hearing the target sentence, which enables us to confidently claim that the effect we observe comes from the fact that they compute the effect of RFR. In a TVJT, by contrast, the listener is first presented a story; and they know what the interpretation of the sentence should convey based on what they hear (i.e., they are biased toward a certain interpretation). It is possible that this bias may inflate children's performance.

The third reason is that the choice of "Not > All" pictures in our picture-selection experiment is likely considered as a negative answer to the "All > Not" pictures. This is because "All > Not" pictures are visible while the participants make a choice, and thus it remains the participants' choice. The fact that they did not choose the "All > Not" pictures suggests that they considered the "Not > All" pictures were more suitable than the "All > Not" pictures. With a picture-selection task, we can ensure this without testing a "false" sentence as in a TVJT. One of the merits with TVJTs is that it can test whether children are able to correctly reject a target sentence when it is incorrect in the light of adult grammar, despite children's yes-bias. It is also important to involve "false" items in the design of an experiment. However, previous studies on the scopal interaction between a universal quantifier and negation typically tests only "yes" items (e.g., two out of three horses jump over a fence and the target sentence goes "Every horse didn't jump over the fence"). This means that children only have to accept the target sentences in order to be regarded as adult-like. However, in their defense, we should note that there is a reason that the previous studies did not test "false" or rather, "infelicitous" items. In order to guarantee that the listener accesses the "Not > All" interpretation and favors it over the isomorphic "All > Not" interpretation, an experimenter has to prepare a situation of "All > Not", utters the target sentence, and lets the listener say "No, that is not felicitous, because the sentence should mean "Not > All" and that has an implicature that there exists someone that did the action". Such an experiment is difficult to construct, especially because "Not > All" is consistent with the "All > Not" situation (e.g., none of the horses jumped over a fence).[87]

[87] The "All > Not" situation is generally judged infelicitous upon accessing the "Not > All" interpretation because of the implicature computation as we saw in §4.5.4; scalar implicatures are computed by negating logically

One possible manipulation is to give participants choices between *fine* and *infelicitous*, rather than *right* and *wrong*. In the literature to investigate children's acquisition of scalar implicatures of various kinds, researchers have tried different methodologies to detect the distinction between *infelicitous* and *wrong*: Felicity Judgment Task; where two puppets utter target sentences which differ in its logical strength and a child chooses one that says it better (Chierchia et al. 2001, Gualmini et al. 2001, Foppolo et al. 2012), instructions such as "If you think the puppet says it well, give him this token (a treat, e.g., a star), but if you think the puppet can say it better, give him this token (not a treat, e.g., a bug)" to avoid saying "wrong" (Shetreet et al. 2012), asking participants *"Did the puppet answer well?"* instead of asking if he was "right" (Papafragou and Musolino 2003), or Covered Box Task (Huang and Snedeker 2009, Bill et al. 2014).

When a participant chooses a "Not > All" picture over an "All > Not" picture in a picture-selection task, we can obtain evidence that the participant regards the "Not > All" picture as a better description than the "All > Not" picture, without having them say *no* to an "All > Not" situation in a TVJT. In Experiment 2, we ran a TVJT version of the experiment for a comparison.

5.3 Experiment 2 – Truth-value judgment task

5.3.1 Method

Since we wanted to compare the two methodology as an evaluation tool for the effect of RFR, this follow-up experiment uses the methodology of TVJT and tests sentences with RFR. The PowerPoint slides and sound files from Experiment 1 were used. In the instructions of the experiment, we asked the participants to indicate whether the girl is doing a good job or being silly; an example of the introduction of the experiment is described in (6).

(6) Here you will see some stories with interesting things... like animals and planes. Each story has two pictures. Sometimes, something happens... like, the animals might get treats, but there might be stories where nothing happens after all! So you have to watch

stronger alternatives; asserting ($\neg\forall$) generates a logically stronger alternative ($\neg\exists$), and the negation of it returns existential implicature ($\neg(\neg\exists) \rightarrow \exists$).

the stories carefully. OK? After each story, this girl will tell you what she thinks happened. Sometimes she does a good job, but sometimes she might be saying something silly. So you have to tell me if she's doing a good job or being silly. OK, let's practice...

The instructions make sure that the participants know that some stories might not have any change ("All > Not" situation is supported). In order for the participants and the experimenter to know that the scene has advanced even when no change has happened, the slides advance with an animation like a picture-card show. As we discussed in §5.2.4.3, the "Not > All" LF is not contradictory to the "All > Not" situation. For this reason we did not want to use the word "wrong/incorrect"; instead used "silly" for non-match cases. Using the word "right/correct" might prime the notion of "wrong", so we used a description "doing a good job" for matching cases instead. Figure 8 shows the experimental procedure for a "good job" item.

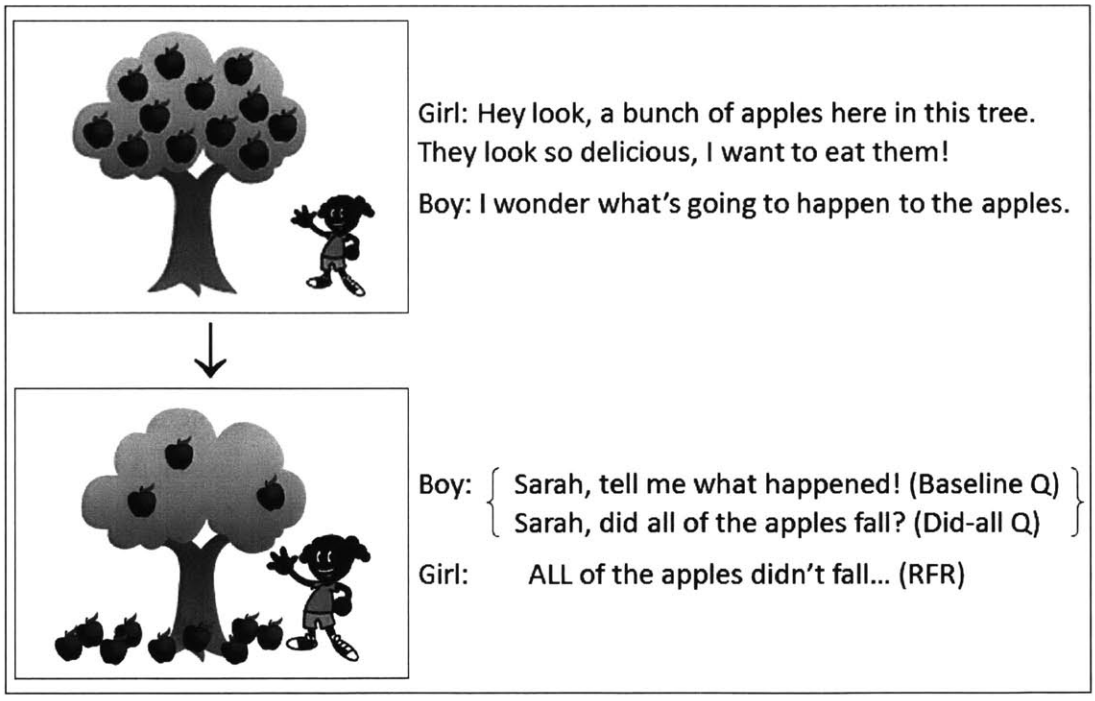


Figure 8: Experimental procedure of Experiment 2

We crossed question type (Baseline-Q vs. Did-All-Q) and the scope relation in the outcome picture (All > Not vs. Not > All). The target sentences were always read with RFR. All the

conditions were within-subjects factor. Note that with RFR, the implicature computation is imposed on the listener; and the "All > Not" interpretation, which resolves all the other sub-questions under the relevant D-tree, is excluded from the possible interpretations. That way our experimental design allows us to argue that the RFR sentences (presumably only "Not all but some" interpretation) are infelicitous relative to the "All > Not" situation; and thus we coded the response "silly" to the "All > Not" items as correct, and the response "good job" to the "Not > All" items as correct.

The experiment consists of 8 target items, namely 2 items per condition – admittedly it is a small number for each condition; but we did not particularly expect there to be an effect for the question type, considering that Experiment 1 for both adults and children did not have any effect. The main purpose of this experiment is to compare children's responses on "good job" items and "silly" items to the percentage of "Not > All" responses in Experiment 1. The filler sentences did not contain negation, and half of them are designed to be true and the other half to be false. Under the same conditions as Experiment 1, experiments took place at local daycares and at Boston Children's Museum. Each session typically took about ten minutes per participant.

5.3.2 Results and discussion

We recruited 19 children for this experiment. Children who skipped two or more trials (n = 1), who wrongly answered 3 or more filler items out of 6 (n = 1), and who was a yes-sayer (n = 1) were excluded. Total of 16 children (from 4;2 to 6;8, M = 5;2) are included in the analysis. Compared to the RFR condition in Experiment 1, the profile of the participants are comparable (n = 16, from 4;5 to 6;7, M = 5;2). No child participated in both experiments.

The accuracy rates plotted by the question type and truth are illustrated in Figure 9. The accuracy rates for the BaselineQ-False (infelicitous) condition was 28%, for the All-Q-False condition was 38%, for the BaselineQ-True (felicitous) condition was 66%, and for the All-Q-True condition was 63%. See Appendix 6 for the list of the stimuli and accuracy rates for each item. The error bars indicate 95% confidence intervals in both Figure 9 and 10.

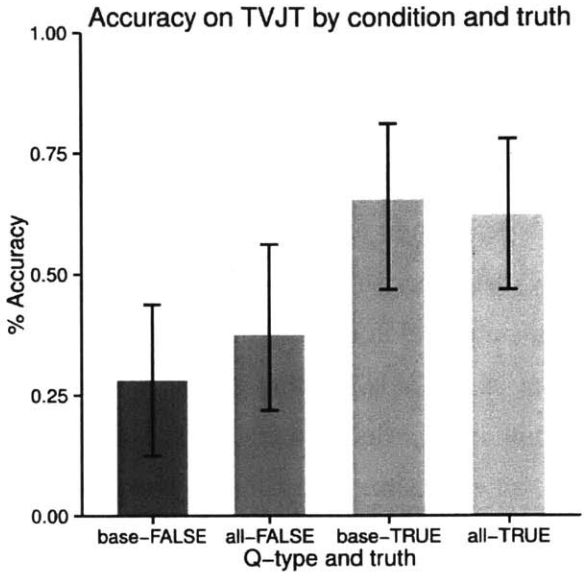


Figure 9: Accuracy rates in Experiment 2 by question type and truth

Table 3: Statistical analysis of Experiment 2

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.2580	0.4870	-0.530	0.596
qTypeBase	-0.4097	0.5713	-0.717	0.473
truth1	1.1657	0.5675	2.054	0.040 *
order2	-0.6502	0.4902	-1.326	0.185
qTypeBase: truth1	0.5122	0.7967	0.643	0.520

Statistical analysis using GLMEM reveals a main effect of truth (p = .04).[88] This was expected given children's yes-bias. As discussed in §5.2.4.3, only when a child can correctly reject an incorrect sentence can we safely claim that the child has the grammatical knowledge being evaluated in an experimental setting.

Figure 10 compares the accuracy rates averaged across the truth in Experiment 2 and the rates of choosing "Not > All" pictures on the RFR condition in Experiment 1. Experiment 2

[88] The order of the presentation was not considered to enter into interaction with the other two potential factors. Also, slopes for the random factor of subjects did not include estimates for truth.

only used RFR sentences, so the two experiments are comparable; and moreover, the stories and sentences in the experiments were exactly the same. The two experiments only differ in its presentational method (hearing the sentence, then decide on a picture vs. knowing the outcome, then decide on the felicity of the sentence). The averaged accuracy rates on the BaselineQ condition in Experiment 2 is 47%, on the AllQ condition in Experiment 2 is 50%, on the BaselineQ condition in Experiment 1 is 68%, and on the AllQ condition in Experiment 1 is 71%. Treating the two experiments as between-subjects sub-experiments of one large experiment (i.e., treating the difference in methodology as a between-subjects factor – to which some readers might object), statistical analysis using GLMEM does not observe an effect (the effect of the experiment type is $p = .118$); although the power analysis reveals that the effect size (Cohen's d) is .436, which is regarded as a medium-to-large effect. Follow-up experiments to increase the sample size, an extension to testing the corresponding Falling condition and to adult population, need to be performed for additional statistical relevance to the dataset.

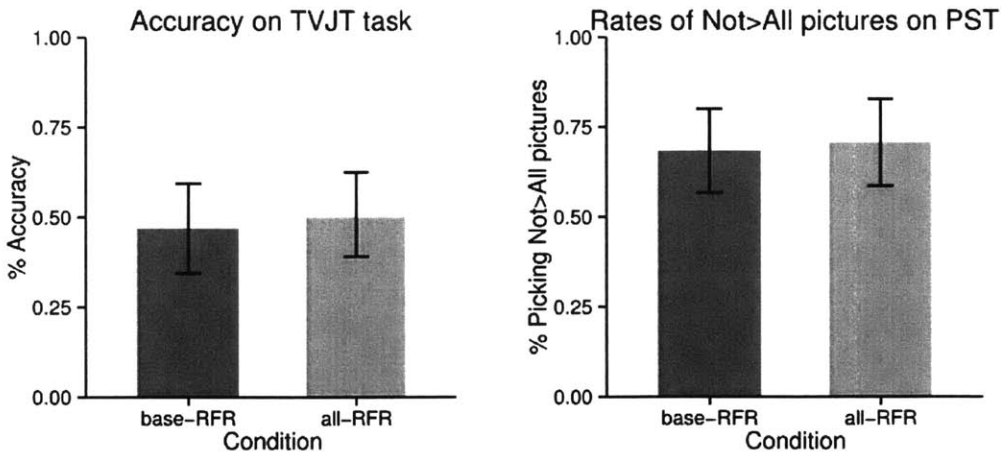


Figure 10: Comparison between TVJT (truth collapsed) and picture-selection task on the RFR condition

The accuracy rates on the True conditions were fairly high in our experiment (66% and 63%). Those results, by themselves, are meaningful. Previous literature in the relevant field (scopal relations between a universal quantifier and negation) employed TVJTs and their rates of "yes" responses are taken to be an indicator of children's ability to access the non-isomorphic interpretation. That being said, the results of the True conditions in our Experiment 2 are comparable to the long-accepted measures of children's competence with non-isomorphic

reading. As we reviewed in §4.5.2, several studies have shown that children are capable to say "yes" to a "Not > All" situation for about 60% of the time when an affirmative description precedes; and for about 80% of the time when the first half of the session involves sentences such as "*Not every horse jumped over the pig*", which had primed children with the "Not > All" interpretation before hearing the target sentence. Experiment 2 in our study, as we have noted already, did not bias the participants toward either of the interpretations prior to hearing the target sentence. The results can be taken as a strong indicator that children have the competence with accessing the "Not > All" interpretation, which they compute solely based on the target sentence with the RFR prosody.

5.4 Conclusion

In this chapter, a new set of data was presented that shows preschool children do access the non-isomorphic reading in sentences with a universally quantified subject and negation. Investigation on whether children can get the non-isomorphic reading in sentences with a universally quantified subject and negation has been debated since Musolino's (1998) research; and several studies have revealed that children have the competence in computing the non-isomorphic interpretation, when the processing load is lightened. The results of the current study add an important new datum to the literature showing that children do access the interpretation even when (i) implicature computations evoked by the RFR prosody are involved, and (ii) the participants are not biased toward either of the readings prior to the target sentence. The results obtained are outcomes of careful manipulations such as having a trained speaker to have clearer effects, employing dialogue to create a felicity circumstance (where CT congruence is met), providing no bias toward a certain interpretation prior to the target sentence, and aligning with the way the RFR effects are computed by using a picture-selection task.

The discussion on methodology suggests that picture-selection tasks are more suitable to test this kind of phenomenon. The results suggest the necessity of comparing different methodologies but have similar materials that are used across experiments. The comparison between Experiment 1 and Experiment 2 shows that there is likely to be an effect of the methodological difference; however, further research will be performed in hopes to support this initial finding.

We have tested whether children are sensitive to the difference in prosody and compute the effects of RFR. Our results show that children arrive at different interpretations solely depending on the difference in prosody, and that they interpret the sentences in the same direction as adults do; as well as, there is roughly the same quantitative extent and variability across children participants as for adults. Then why did the study by Ianucci and Dodd (1980) see a smaller percent of choosing "Not > All" pictures than in our study, even though I&D's study also used picture-selection tasks, which we argued is a suitable methodology to test this phenomenon? This may be largely due to the lack of a felicitous environment where the RFR sentence is uttered. The RFR sentences call for CT congruence; and therefore, using a dialogue (interlocutors exchange questions and answers) to provide an appropriate QUD to obey the CT congruence is important when setting up a felicitous context.

Some notes on the attempts I made with other methodologies might be helpful to support this idea. Though all the "attempts" are far from being conclusive and not to be reported formally, making some remarks on the results of two experiments to test the effect of RFR might be of some interest. First, a TVJT with a typical protocol was tried out. Stories were made to express either a "Not all but some" or "None" situation. At the end of the story, a puppet utters a target sentence with the Falling or RFR prosody, and the child has to judge if the puppet is doing a good job or being silly. Even though the experimenter prompts the puppet's utterance, it was more similar to a monologue of the puppet. Among the 6 children we tested, 3 were isomorphic, 2 were flexible (i.e., accepting both "Not > All" and "All > Not" situations), and 1 accepted "Not > All" and rejected "All > Not" for the RFR prosody. For the second attempt, a guessing game was designed. We thought that the situation for the participant to figure out should not be revealed to them in order to let them compute the implicature regarding the RFR. In a typical story, some event happens in the next room, and Dora, who knows the outcome, would report to the participant what happened in the next room, using the Falling or RFR prosody. The participants were encouraged to guess what the outcome was (choices among "none" and "not all but some"). We tested a few participants, but the task seemed too difficult.

The current study, specifically the Picture-selection task, has improved on those attempts, in that the experimental manipulation was made to support the CT congruence, by leaving the relevant question open, using a dialogue and making "Did all?"-question present/easily accommodated. If we consider the CT congruence as a part of QAC, then the results from the RFR study can be characterized similarly to the study on *only*; children's comprehension becomes adult-like when QAC is satisfied.

The results of the experiments reported in this chapter kept showing the absence of the effect of question type. I propose two possible accounts for not observing the effect of question type; (i) the prosodic cue triggering implicature computation was so strong that the question type effect was drowned out, and (ii) it was an experimental confound; in this particular experiment, what kind of sub-questions should be accommodated under the baseline question was so obvious that there was virtually no cost for accommodation of sub-questions when the baseline question was presented. In other words, the stories were so simple that the listeners might have been able to predict even at the first phase of the story what kind of event would happen in the second phase. This intuition acquires support from some remarks children made during the experiments. When the first phase ended with the boy's utterance "I wonder what's going to happen to the *NP*s.", some children went ahead and tried to guess, e.g., "The apples are going to fall!", "The vases will break!", "The trees are going to grow!", and so on.[89]

In conclusion, our results remarkably demonstrate that a child's knowledge and processing are similar in many of the defining features of an adult's knowledge and processing. Indeed, children know and use the link between prosody and semantics in a similar manner to adults.

[89] See Appendix 7 for the results from the L2 population suggest that question types affect processing of the RFR sentences. With the broad question-RFR pairs, more errors were produced than with the "Did all"-question-RFR pairs.

Chapter 6

Discussion

6.1 Overall summary

In this thesis, I have presented results from experiments on two case studies: sentences with *only* and sentences with the RFR contour. The current research on *only*, I believe, makes four main contributions. (i) It shows that children have command of Subj-*only*, in contrast to many previous studies. (ii) Crain's puzzle is more generally restated as "Why is an object-question easier to accommodate than a subject-question?", and this question is not limited to children but also extended to adults. (iii) Children are very sensitive to QAC. And (iv) the role of QAC for children is to strongly guide the constituent taken to be the associate of *only*. For adults, congruence facilitates processing, while incongruence interferes with processing.

The second case study on the prosody-semantics link with RFR makes two main contributions. (i) It shows that children arrive at different scope interpretations solely based on the difference in prosody, which triggers computation of an implicature disambiguating scope. And (ii) the direction of the effect and the variability in picture choice were shown to be similar for both the child and adult participants. The case study of the RFR contour raised several questions for follow-up experiments (as discussed in §5.3.2 regarding methodological issues, and as will also be discussed in §6.5), which will be left for future research.

The notion that connects the two case studies is QAC. The case study of *only* directly tested whether QAC affects the degree of comprehension when the interpretation of the sentence depends on the position of the focus (F) mark. The results confirmed that QAC plays an important role in determining where to assign F for children, and in facilitating processing in adults' comprehension. Like the *only* study, the experiments on RFR also tested two kinds of question type, but did not yield the question type effect observed with *only*. One might suspect

that QAC did not play a role in the RFR case. In fact, the circumstance where the effect of QAC was tested was different in the second case study from the first one; in the RFR case, the interpretation of the sentence depended on implicature computation evoked by the particular prosody associated with CT. It is possible that the phonological cue triggering implicature computation is stronger than the QAC cue. This is the converse situation from the first case study, where the QAC cue was stronger than the syntactic cue. Evidence showing that QAC actually has an effect on comprehending sentences with RFR for an L2 population, though preliminary, is reported in Appendix 7.3.

6.2 Summary – the case study of *only*

Our experiments on *only* showed that children are sensitive to QAC, and they know the meaning of *only* regardless of where it attaches. We arrived at this conclusion based on the experimental result that children are capable of interpreting Subj-*only* sentences in an adult-like way when QAC is satisfied, and are much less accurate in understanding VP-*only* sentences when the Q-A pairs are incongruent. The experiments conducted with adults showed parallel results; the asymmetry between Subj-*only* and VP-*only* was replicated in RT measures, and the processing differences between directly congruent and incongruent Q-A pairs were demonstrated in measures of accuracy and RTs. The results, all in all, suggest that the reason why children in previous studies have exhibited difficulty interpreting Subj-*only* sentences has to do with the processing load required for accommodation of a sub-question. The problem is not a deficiency in grammatical competence about Subj-*only*. Furthermore, some production data to indicate that children do use Subj-*only* sentences without problems was provided. These results are summarized as contributions (i) and (iii) stated at the beginning of this chapter.

Children's and adults' responses to incongruent Q-A pairs in the present study raises an interesting discussion about what the proper characterization of Crain's puzzle is. Children's primary strategy for determining the associate of *only* was based on QAC, whereas adults' primary strategy was based on the syntactic position of *only*. This shows that children's command of *only* is not fully adult-like for both types of *only* (i.e., importantly, even for VP-*only*). In other words, what is not adult-like about comprehension of sentences with *only* is children's knowledge of the FCC-*only*; for adults, the FCC-*only* is not a violable condition, but

for children, it does seem possible to ignore it when it conflicts with QAC. This difference in the status of the FCC-*only* is what divides children's behavior and adult-like behavior. For children, QAC constitutes a stronger cue to the associate of *only* than the FCC-*only* does. As long as children succeed in determining the associate of *only*, they can compute the meaning of the sentence with *only* wherever *only* attaches. The experiments with adults showed that QAC affects adult sentence processing as well. For incongruent trials, comparison between RTs of "correct" responses (complying with the FCC-*only*) and those of "incorrect" responses (ignoring the FCC-*only*, but following QAC) showed that participants take longer to answer if they provide a correct response. This suggests that complying with the FCC-*only* involves more processing complexity, and that the strategy of respecting QAC and ignoring the FCC-*only* could be less costly. In turn, this provides support for the idea that children's primary strategy when faced with incongruence is to pursue the less costly option. These results can be summarized as contribution (iv) stated at the beginning of this chapter.

6.3 Future directions – the case study of *only*

Contribution (ii), stated at the beginning of this chapter, is more of an open-ended statement, which requires further research, though a proposal was made, to note the link between the accommodation under the theory of D-tree and the semantics of multiple *wh*-questions. The current study contributes to our understanding of Crain's puzzle in that it poses a new research question to investigate; one way to approach the puzzle is to ask what makes the accommodation of an object-question easier than that of a subject-question. This is now a more general issue, for it applies to adult sentence processing, as well as to children's linguistic knowledge. The results of our experiments have shown that the asymmetry in difficulty between Subj-*only* and VP-*only*, in fact, follows from an asymmetry in processing load between the accommodation of a subject-question and that of an object-question — a contrast also evident in the experiments with adults. A proposal was made to account for the observed asymmetry in the cost of accommodating a subject vs. object question in a broad focus context; a broad question is by default broken down into a set of object-questions in our case, which is sorted according to the Topic of the sentence (i.e. the subject). The proposal gives rise to a testable hypothesis; if the Topic is the sorting key which determines how a broad question is

decomposed, then a situation where the object is made Topic and the subject is focused/new, a broad question should be preferably broken down into a set of *subject*-questions. Then, Subj-*only* should be more readily comprehensible, compared to VP-*only*, even when the overt question preceding the sentence is a broad question. It is difficult to construct an experimental design to test such a hypothesis, but it is possible do so. In addition to the idea of Topic as a sorting key, it was pointed out as further support to the proposal that a set of object-questions and a set of subject-questions are regarded as parallel to the semantics of superiority-obeying and superiority-violating multiple *wh*-questions.

As noted at the end of Chapter 2, the nature of the scalar presupposition of *only* plays an important role in determining where to assign F, in addition to QAC. Evidence obtained with adult participants shows that the ease or difficulty of satisfying the scalar presupposition affects the processing of sentences with *only* (Hackl et al. 2014, paper in progress). Testing whether this is the case with child participants is needed in future research.

Another contribution of the current research is to show that QAC should be considered as an important factor in constructing a TVJT, in general. This study has demonstrated that children are very sensitive to QAC, with congruence between the prompting question and the target sentence affecting their performance to a great extent. As such, one has to pay careful attention to focus structure when deciding whether a broad focus question should be used to prompt the puppet's utterance in a TVJT.

6.4 Summary – the case study of RFR and scope

Previous experiments on the relative scope of universals and negation showed that children can access inverse scope readings when an experimental manipulation is done to lighten processing load. The experiments on RFR conducted here showed that children can access inverse scope even without such experimental manipulations. It was shown that children arrived at the inverse scope reading when the interpretation is required by an implicature associated with RFR, even when the context prior to the target sentence does not bias one reading over the other. The results from the experiments with children and those with adults conform to each other pretty well, which suggests that children's knowledge and processing of sentences with RFR are similar in many respects to those of adults.

The RFR contour is not peculiar to sentences with a universally quantified subject and negation. More generally, RFR contributes a contrastive meaning, requiring that there to be an unresolved sub-question under the same D-tree. Because children successfully settle on inverse scope readings on the basis of this requirement, the prediction is that children should also access a contrastive meaning in simpler RFR sentences without scope operators. However, it is possible that this would not be as clear in the experimental results. The reason for that is that adding a contrastive meaning to a plain sentence such as "*[JOHN]$_{CT}$ came...*" does not affect the truth conditions of the sentence: uttering the sentence with a neutral contour or an RFR contour results in the same assertion. By contrast, the sentence with a universal quantifier and negation is ambiguous between "All > Not" and "Not > All" readings and, accordingly, the contrastive meaning associated with the RFR prosody has the effect of disambiguating the truth-conditions. The clear difference between the two interpretations may have had a role in making the effect of RFR clearly show up for the child participants.

On a side note, it appears that the meaning associated with the RFR contour in the theoretical literature can be associated with phonetically distinct intonational patterns: in addition to RFR itself, a rise-fall contour without a subsequent rise, and a neutral contour with a subsequent rise seem to be interpreted in a similar way. In the experiment, I have followed the theoretical literature in looking at the RFR contour itself. We have well-constructed theories of RFR (Büring's and Constant's) that integrate naturally with more general theories of the semantics-pragmatics interface (Rooth's QAC and Roberts' QUD stack, as well as Büring's D-tree). It might be true that RFR itself is not ubiquitous,[90] but speakers have strong intuitions about RFR.[91]

One puzzling result from our current experiment is that we did not observe the effect of question type. In the experiment, two kinds of question type were tested (Baseline-Q "*Tell me what happened.*" vs. All-Q "*Did all of the NPs VP?*"). It was expected that the RFR sentences preceded by a "Did all" question would increase the rates of "Not > All" compared to the RFR sentences preceded by a baseline question, at least with children. The rationale is the following;

[90] To be exact, RFR with a sentence that contains a universal quantifier and negation might not be found everywhere. However, "rising at the end" as a part of CT itself is very common in daily conversation exchange, and it actually has an implicature that there should be at least one sub-question under the same D-tree that is not answered after the assertion.

[91] I suspect this can be understood as an analogy that not many people produce weird but grammatical sentences that would appear in linguistics textbooks, but when asked to judge, people do have intuitions and the intuition is shared by most people.

given the isomorphic preference that child participants would have, our measure of interest is how much of an increase in rate of "Not > All" readings we observe with RFR compared to the baseline of the Falling contour. I had expected the implicature computation with RFR to involve some processing cost, and therefore the computation with RFR would be easier with the felicitous Q-A pair given the CT congruence (i.e., "Did all" question and RFR answer) than with a baseline question. Two possible accounts for not observing the question type effect are that (i) the prosodic cue triggering implicature computation was so strong that the question type effect was drowned out, and/or (ii) in this particular experiment, it was so obvious what kind of sub-questions should be accommodated under the baseline question that there was virtually no cost for accommodation of sub-questions when the baseline question was presented.

6.5 Future directions – the case study of RFR and scope

Our results provided experimental support for theoretical analyses of RFR and CT in English, as these predict that the RFR contour should disambiguate scope, as we observed. There are other languages that exhibit similar effects of RFR, where CT marking on a universal quantifier in a negated sentence strongly favors the "Not > All" interpretation. Phonological and morphological realizations of CT vary, but similar phenomena to RFR, characterized as phonological cues affecting scope relations, are reported in German (Büring 1997, Krifka 1998, Sauerland and Bott 2002), Japanese (Hara 2006, Nakanishi 2007, Tomioka 2010a, b, a.o.), Hungarian (Gyuris 2002, É. Kiss and Gyuris 2003, Jackson 2008), and Greek (Giannakidou 1998, 2000, Baltazani 2002). Experimental research investigating the extent to which listeners make use of CT marking in those languages is left for future work.[92]

[92] Since the realizations of CT differs across languages, it will be interesting to see what L2 learners of each language infer from hearing the sentences with the CT contour in their second language. Their L1 grammar has the notion of CT connected with a certain phonological realization and they have access to the mechanism to derive the correct interpretation, but the phonological realization in the L2 might be different from what it is in the L1. By testing whether L2 leaners can access the correct interpretation, we can know if they make use of the abstract notion of CT or if they cannot make use of it because of the actual realization being different from what it is in the L1 and they do not connect the input with the notion of CT. Preliminary results to investigate this with Japanese-speaking L2 leaners of English are reported in Appendix 7.3.

Related to that, Sugawara and Wexler (2014) studied Japanese-speaking children with sentences with CT *-wa* and Thematic Topic (TT) *-wa*, as in (1).[93] The two kinds of topic differ in their prosody in that the focused element with CT *-wa* bears a high pitch contour followed by post-focal reduction, whereas the pitch contour of the contents following the TT does not undergo post-focal reduction (for comparison of pitch tracks, see Nakanishi 2007).

(1) a. Zoosan minna-wa e-o kak-anak-atta-yo. (TT)
elephant everyone-TOP picture-ACC draw-NEG-PAST-PRT
'Every elephant didn't draw pictures.' (^{ok}All > Not, *Not > All)

b. Penginsan [minna-wa]$_F$ e-o kak-anak-atta-yo. (CT)
penguin everyone-TOP picture-ACC draw-NEG-PAST-PRT
'Every penguin didn't draw pictures.' (*All > Not, ^{ok}Not > All)

S&W found that Japanese-speaking children (n = 23, ranging 4;3 to 6;2, *M* = 5;3) interpreted the sentences with CT contour to mean what the sentences with TT contour would mean; in other words, they interpreted the sentences to mean "All > Not" regardless of the prosodic contour. If the mechanism of deriving the "Not > All" interpretation is shared by Japanese CT and English RFR, as Nakanishi (2007) and Constant (2014) suggest, the results from Japanese experiments might be seen as a conflict with the results of our English RFR experiment. One way to account for why the results look different is that the verbs used in the Japanese experiment were transitive, instead of intransitive as in the English RFR experiment, and this might have caused a complication. We were testing another condition with scrambling, and in order to make the sentences in the CT and TT conditions and in the scrambling condition as similar as possible, we had to use transitive sentences. Since this manipulation introduced another argument in the target sentence, the LF representation became more complex than with an intransitive sentence. Another potentially relevant factor is that the experiment did not support the felicitous environment where a CT sentence would be uttered. As we discussed at length in Chapter 4, CT sentences are uttered naturally in a dialogue, specifically as an answer to a question asking about the constituent that bears CT-marking. In S&W, we were not aware

[93] For this experiment, I would like to thank Erika Trent for her tremendous help and excellent skills in designing materials and carrying out the experiments.

of this point and we did not use an appropriate dialogue.[94] Given these potential confounds, it is worth asking again whether Japanese-speaking children are sensitive to the difference in prosody between the CT and TT conditions in a more carefully designed experimental setting.

6.5 Concluding remarks

I started this thesis by asking how children are different from adults in their linguistic ability and why: is it grammar or some extra-grammatical factor that makes children appear to be different from adults? Furthermore, what are the aspects of grammar that are not different between children and adults? – This thesis investigated the role of QAC in child language acquisition and adult sentence processing by specifically looking into two case studies: sentences with *only* and sentences with RFR. The results from our experimental investigations have turned up some answers to our starting questions. It was revealed that both children and adults are sensitive to QAC, with congruence facilitating processing, and incongruence interfering with processing. However, the extent to which QAC has influence upon comprehension is different for children and adults: for children, the QAC cues outrank the syntactic condition on how to assign F in a sentence with a focus sensitive operator, which is essentially an inviolable condition for adults. The results also revealed that QAC cues do not as strongly affect comprehension as prosodic cues that call for implicature computation; it was shown that both children and adults are sensitive to the RFR contour and able to compute the associated implicatures to disambiguate the "Not > All" interpretation in sentences with a universally quantified subject and negation. Interestingly, the different levels of QAC (felicitous Q-A and infelicitous Q-A) did not affect the extent of computing the implicatures. This suggests that with regard to the ability to compute the effect of RFR, children are not different from adults. Though, it is worth noting that our analysis of the correlation between the rate of choosing "Not > All" on the RFR condition and children's age is likely to indicate that children older than five and a half years old almost always chose the "Not > All" pictures, while

[94] Yet another possible account is that the subject "NP *minna-wa*" might have caused another complication. A study by Hattori et al. (2006) used a simple "*minna-wa*" and the children in their study yielded higher rates of "Not > All" responses. However, this study was inconclusive mainly because the experiment only used stories that supported the "Not > All" reading in their TVJT method. This means that only "true" stories were presented, and this could have caused a bias toward the "Not > All" reading and increased the correct responses.

children younger than five and a half years old have mixed populations regarding choices of pictures. The results indicate that younger children might be different from adults regarding the ability to compute the implicatures necessary for RFR, though this is yet to be confirmed with a larger number of participants.

The results of the current studies, overall, have added new data to the literature. As was pointed out, there are multiple directions to pursue to investigate the questions that this set of results has opened up. Further research should be performed to develop a better understanding of what characterizes children's linguistic capacity.

Appendices

Appendix 1: Experimental materials in Experiment 1 of Chapter 2

Below is the list of the sentences and accuracy rates per item for the baseline experiment with children in the case study of *only*. All the sentences were preceded by a broad question – "Kermit, (can you) tell me what happened?" The T's and F's indicate the intended truth value designated in the experiment, and the percentage (%) on the right indicates its accuracy rates for each item, for the younger and older groups. The younger group consists of children younger than 5;2, while the older group consists of children who are over 5;2. The criterion was chosen just because it bisects the set of children tested (5;2 was the mean age of the participants). The same trends are found if the age cutoff is set to other criteria, e.g., three groups as 4 year-olds, 5 year-olds and 6 year-olds, and so on.

Table 1: List of items and accuracy rates by the item

Target sentences	**Correct truth value**	**Accuracy: Younger (n=19, 4;0-5;1, *M*=4;6)**	**Accuracy: Older (n=19, 5;4-6;9, *M*=5;9)**
Only the cat got ice cream.	F	5%	32%
Only the mouse got bread.	F	5%	47%
Only the frog got juice.	T	21%	57%
Only the horse got candy.	T	21%	47%
The cat only got ice cream.	T	89%	74%
The dog only got cake.	T	100%	79%
The bird only got cheese.	F	89%	68%
The horse only got candy.	F	84%	68%
Filler sentences			
Pooh got juice.	F	100%	100%
The elephant got candy.	F	89%	95%
The frog got juice.	T	95%	100%
The bird got jello.	T	63%	79%

As you might notice, the accuracy rates of the filler item "The bird got jello" yielded as low as 63% with younger children. In this item, Winnie the Pooh got jello and cake, a mouse got cake, and a bird got jello and candy. The target sentence "The bird got jello" is true, but the children who rejected the sentence gave a justification that suggests that they applied exhaustification over the items the bird got (e.g., "Because, the bird got candy too"). It is not totally unpredicted that any sentence (in general) is under the influence of exhaustification, and furthermore I suspect that the tendency was more prominent in the experiment since the participants heard the word "only"; which forces them to apply exhaustification, even though that filler sentence in question does not contain the word "*only*" in the experiment. Having heard the word *only* might have "primed" children to apply exhaustification to the sentence. This line of explanation acquires support when we look at the accuracy rates more closely – we prepared two lists of pseudo-randomized experimental items, and in one list the filler item appeared in the middle of the session (after hearing sentences with *only* several times), and in the other list this item appeared at the beginning of the session (after going through practice items, which were meant to familiarize participants with interacting with the puppet, but before hearing any sentences with *only*). The accuracy rates across all ages were 65% and 77% from the first and second group, respectively.

Appendix 2: Experimental materials in Experiments 2 of Chapter 2

Below are the lists of the sentences and adult-like response rates per item. Table 2 is for Experiment 2A, where the attachment site of *only* was a within-subjects factor and the question type was a between-subjects factor. Depending on which question precedes the sentence, the congruence varies. We see that even though a target sentence itself is exactly the same, the accuracy rates on the congruent condition and those on the incongruent condition are dramatically different.

Table 2: List of items and accuracy rates by the item, Experiment 2A – Question type is constant

Target sentences	Correct truth value	Subj-question (n=12, 4;0-6;8, *M*=5;1)		Obj-question (n=12, 4;0-6;11, *M*=5;0)	
		Congruence	Accuracy	Congruence	Accuracy
Only the cat got ice cream.	F	Congruent	50%	Incongruent	0%
Only Cookie Monster got bread.	F	Congruent	58%	Incongruent	0%
Only the monkey got juice.	T	Congruent	92%	Incongruent	17%
Only Elmo got candy.	T	Congruent	92%	Incongruent	8%
Tweety only got ice cream.	T	Incongruent	50%	Congruent	100%
The turtle only got cake.	T	Incongruent	50%	Congruent	100%
The cat in the hat only got cheese.	F	Incongruent	8%	Congruent	92%
The horse only got candy.	F	Incongruent	17%	Congruent	92%

Table 3: List of items and accuracy rates by the item, Experiment 2B -- Attachment site is constant

		Target sentences	Correct truth value	Congruence	Accuracy
Subj-*only* (n=12, 4;7-6;1, *M*=5;3)	Subj-Q	Only the cat got ice cream.	F	Congruent	67%
		Only Cookie Monster got bread.	F	Congruent	58%
		Only the monkey got juice.	T	Congruent	83%
		Only Elmo got candy.	T	Congruent	75%
	Obj-Q	Only Tweety got ice cream.	T	Incongruent	18%
		Only the turtle got cake.	T	Incongruent	33%
		Only the cat in the hat got cheese.	F	Incongruent	33%
		Only the horse got candy.	F	Incongruent	17%
VP-*only* (n=12, 4;5-6;5, *M*=5;3)	Subj-Q	The cat only got ice cream.	T	Incongruent	75%
		Cookie Monster only got bread.	T	Incongruent	83%
		The monkey only got juice.	F	Incongruent	50%
		Elmo only got candy.	F	Incongruent	67%
	Obj-Q	Tweety only got ice cream.	F	Congruent	92%
		The turtle only got cake.	F	Congruent	100%
		The cat in the hat only got cheese.	T	Congruent	83%
		The horse only got candy.	T	Congruent	83%

Appendix 3: Experimental materials in Chapter 3

The list of the target and filler sentences for Experiment 1, 2A and 2B is shown below. Since the list of the target items were created using a Latin Square design, each sentence was tested to be Subj-*only* and VP-*only*, and true and false, for the equal number of the time. Half of the filler sentences were designed to be false and the other half to be true. The items of Experiment 1-prime are carried over from those items, only differing the position of *only*; Experiment 1-prime placed *only* between *gotten* and the food item for the Obj-*only* condition. The items in the Subj-*only* condition in Experiment 1-prime were the same as the Subj-*only* items in the other experiments. I will not list the items for Experiment 1-prime.

(1) Target sentences

a. (Only) the cat has (only) gotten ice cream.
b. (Only) Pooh has (only) gotten bread.
c. (Only) the dog has (only) gotten cake.
d. (Only) the lion has (only) gotten candy.
e. (Only) the goose has (only) gotten juice.
f. (Only) Mickey has (only) gotten jello.
g. (Only) the elephant has (only) gotten cheese.
h. (Only) the mouse has (only) gotten ice cream.
i. (Only) the mouse has (only) gotten bread.
j. (Only) the cat has (only) gotten cake.
k. (Only) the lion has (only) gotten candy.
l. (Only) the horse has (only) gotten juice.
m. (Only) the frog has (only) gotten jello.
n. (Only) the penguin has (only) gotten cheese.
o. (Only) the horse has (only) gotten ice cream.
p. (Only) the bird has (only) gotten bread.
q. (Only) the bird has (only) gotten cake.
r. (Only) Mickey has (only) gotten candy.
s. (Only) the lion has (only) gotten juice.
t. (Only) the elephant has (only) gotten jello.

u. (Only) the frog has (only) gotten cheese.
v. (Only) the bird has (only) gotten ice cream.
w. (Only) the horse has (only) gotten bread.
x. (Only) the penguin has (only) gotten cake.

(2) Filler sentences

a. No animal but the bird has gotten cake.
b. No animal but the cat has gotten candy.
c. No animal but the lion has gotten juice.
d. No animal but the cat has gotten ice cream.
e. Pooh has gotten nothing but bread.
f. The dog has gotten nothing but cake.
g. The elephant has gotten nothing but jello.
h. The frog has gotten nothing but cheese.
i. Even the dog has gotten ice cream.
j. Even the lion has gotten bread.
k. Even the cat has gotten cheese.
l. Even the goose has gotten ice cream.
m. Pooh has gotten even candy.
n. The dog has gotten even bread.
o. The lion has gotten even ice cream.
p. The goose has gotten even cake.
q. Mickey has gotten jello, and the cat has, too.
r. The elephant has gotten cheese, and the lion has, too.
s. The mouse has gotten juice, and candy too.
t. The goose has gotten bread, and jello too.
u. Mickey has gotten cake, and Pooh hasn't.
v. The dog has gotten candy, and the frog hasn't.
w. The bird has gotten juice, but not cheese.
x. The frog has gotten jello, but not ice cream.
y. Every animal but the penguin has gotten cheese.
z. Every animal but the horse has gotten juice.
aa. The bird has gotten everything but bread.
bb. The frog has gotten everything but cake.

cc. The crocodile has gotten juice.

dd. The unicorn has gotten candy.

ee. The frog has gotten cheese.

ff. The mouse has gotten cake.

gg. The horse has gotten jello.

hh. The pig has gotten ice cream.

Appendix 4: Experimental materials in the experiment for adult control in Chapter 5

Below is the list of the sentences and the proportion of choosing "Not > All" for each item in the Falling and RFR conditions in the adult control experiment.

Table 4: List of items and "Not > All" rates in the Falling and RFR conditions - Adults

Target sentences		**Not > All rates in Falling condition (n=12)**	**Not > All rates in RFR condition (n=12)**
Baseline Question	All of the flowers didn't grow.	33% (4/12)	75% (9/12)
	All of the apples didn't fall.	33% (4/12)	67% (8/12)
	All of the planes didn't fly.	8% (1/12)	83% (10/12)
	All of the doors didn't open.	8% (1/12)	67% (8/12)
Did-All Question	All of the vases didn't break.	50% (6/12)	82% (9/11)
	All of the boxes didn't open.	17% (2/12)	67% (8/12)
	All of the trees didn't grow.	17% (2/12)	73% (8/11)
	All of the socks didn't dry.	17% (2/12)	67% (8/12)

Below are the entire stories for each item. We deliberately avoided to have the word "all" in Phase 1. There are four different girls appearing (one at a time): Sarah, Hannah, Ashley and Martha. As shown below, there appears to be more trials with Sarah in the target items, but it was arbitrary.[95]

[95] We designed the experiment so that it can be extended to eye-tracking experiments, for which we prepared 24 kinds of stories, as well as 60 fillers including 36 stories where girls do not appear. We assigned each girl an equal

Table 5: Experimental stories for each item

First phase		Second phase	
Introduction (girl)	**(boy)**	**Prompt (boy)**	**Target sentences**
Today is a beautiful day to pick some flowers in this field. But look, they're so small. Maybe I should come back next week.	I wonder what's going to happen to the flowers.	Sarah, tell me what happened.	All of the flowers didn't grow.
Hey look, a bunch of apples here in this tree. They look so delicious, I want to eat them!	I wonder what's going to happen to the apples.	Sarah, tell me what happened.	All of the apples didn't fall.
Hey look, a bunch of cool planes here in the airfield I'm visiting.	I wonder what's going to happen to the planes.	Sarah, tell me what happened.	All of the planes didn't fly.
Hey, look at those mysterious doors here in this hallway.	I wonder what's going to happen to the doors.	Sarah, tell me what happened.	All of the doors didn't open.
Hey, look at these nice vases I have in my house. Oh no, it looks like my cat got onto the shelf!	I wonder what's going to happen to the vases.	Ashley, did all of the vases break?	All of the vases didn't break.
We've just had a party, and I have these interesting boxes left in my room.	I wonder what's going to happen to the boxes.	Hannah, did all of the boxes open?	All of the boxes didn't open.
Hey, look at the trees on this Christmas tree farm. But they're so small. Maybe I should come back next week.	I wonder what's going to happen to the trees.	Sarah, did all of the trees grow?	All of the trees didn't grow.
I have a lot of laundry today. I just finished washing the socks but they're still wet.	I wonder what's going to happen to the socks.	Hannah, did all of the socks dry?	All of the socks didn't dry.

number of stories. And a subset of the materials (which use verbs that are frequent in child corpus) is extracted to construct these experiments reported in Chapter 5. The extraction happened to contain more stories with Sarah than with others. To the author's knowledge, the unbalanced distribution of the character did not interfere with the performance of the participants. All the girls were narrated by the same female voice, and all the utterances by the boy were narrated by the same male voice.

Appendix 5: Experimental materials in the experiment for children in Chapter 5

The list of the sentences and the proportion of choosing "Not > All" for each item in the Falling and RFR conditions is found below. The rates are standardized, in that the responses to choose "All Did" pictures or "Irrelevant" pictures are excluded from the analysis. The fraction indicates the actual numbers of the items that were responded with "Not > All" choice over the number of the items that were answered with either "Not > All" or "All > Not". Looking at Falling condition, it seems that the third sentence in Baseline Question condition "*All of the planed didn't fly*" marks an especially low "Not > All" rate among others. One may wonder that the story might have biased the "All > Not" interpretation for one reason or other, but when we look at the corresponding rate in the RFR condition, it is proved that that was not the case (rather, the rate of choosing "Not > All" in the RFR condition is slightly higher than the average). The entire stories for each item are found in Appendix 4.

Table 6: List of items and "Not > All" rates in the Falling and RFR conditions – Children

Target sentences		**Not > All rates in Falling condition (n=16, 4;4-6;10, *M*=5;3)**	**Not > All rates in RFR condition (n=16, 4;5-6;7, *M*=5;2)**
Baseline Question	All of the flowers didn't grow.	33% (4/12)	64% (9/14)
	All of the apples didn't fall.	31% (4/13)	69% (11/16)
	All of the planes didn't fly.	14% (2/14)	73% (11/15)
	All of the doors didn't open.	38% (6/16)	67% (10/15)
Did-All Question	All of the vases didn't break.	31% (4/13)	64% (9/14)
	All of the boxes didn't open.	36% (5/14)	69% (11/16)
	All of the trees didn't grow.	25% (3/12)	67% (10/15)
	All of the socks didn't dry.	29% (4/14)	85% (11/13)

Appendix 6: Experimental materials in Experiment 2 in Chapter 5

The list of the sentences and the accuracy rates per item is found below. We used the same pictures and sound files as in Experiment 1. The entire stories for each item are found in Appendix 4.

Table 7: List of items and accuracy rates in Experiment 2

Target sentences		**Picture (Truth)**	**Accuracy (n=16)**
Baseline Question	All of the flowers didn't grow.	All > Not (F)	31% (5/16)
	All of the planes didn't fly.	All > Not (F)	25% (4/16)
	All of the apples didn't fall.	Not > All (T)	63% (10/16)
	All of the doors didn't open.	Not > All (T)	69% (11/16)
Did-All Question	All of the boxes didn't open.	All > Not (F)	38% (6/16)
	All of the socks didn't dry.	All > Not (F)	38% (6/16)
	All of the vases didn't break.	Not > All (T)	56% (9/16)
	All of the trees didn't grow.	Not > All (T)	69% (11/16)

Appendix 7: Inference of RFR with Japanese speakers

We have seen that English-speaking children and adults compute the effect of English RFR. This appendix will report preliminary results from experiments carried out with Japanese-speaking population, who study English as a foreign language. The experiment reported in Appendix 7.3 examines whether Japanese-speaking leaners of English are sensitive to the prosodic difference and aware of its consequences, by running the experiment that is basically the same as the experiments that we ran with English-speaking populations. The results suggest that Japanese-speaking leaners of English, although they have never explicitly taught the effect of RFR in classroom, show some sensitivity to the prosody and its effect on scope in the same direction as English-speaking populations. In order to understand the results better, let us look at results from two kinds of survey first in Appendix §7.1 and §7.2. The first survey is concerned with the property of scope rigidity and flexibility in Japanese. The second survey asked the Japanese students to translate English sentences into Japanese. We see from the results of these

surveys that Japanese has tendency of scope rigidity as has been discussed for long in the literature, and the tendency was replicated with our experimental materials. The results also show that the environment where the inverse scope is possible is when a universal quantifier precedes a Topic marker, *-wa*.

Appendix 7.1 Survey 1: Scope rigidity and flexibility in Japanese

Methods

Since early research on generative approaches to Japanese, it has been noted that Japanese exhibits a property of scope rigidity (Kuroda 1965, Kuno 1973, and Hoji 1985, and subsequent research). The property is widely accepted, but recent studies suggest that Japanese might not be so rigid, or that Japanese-speaking children exhibit more flexibility than would adults (Goro 2007). The survey I ran with Japanese-speaking college students examines interpretations they get from sentences with various kinds of word orders.

Methods are the following. The participants were handed a sheet of questionnaire with sets of Japanese sentences and pictures showing different outcomes. They were asked to check cell(s) that the sentence would indicate, and they were explicitly told that the meaning of the sentence might be ambiguous and if they found multiple pictures suitable, they can check multiple cells.

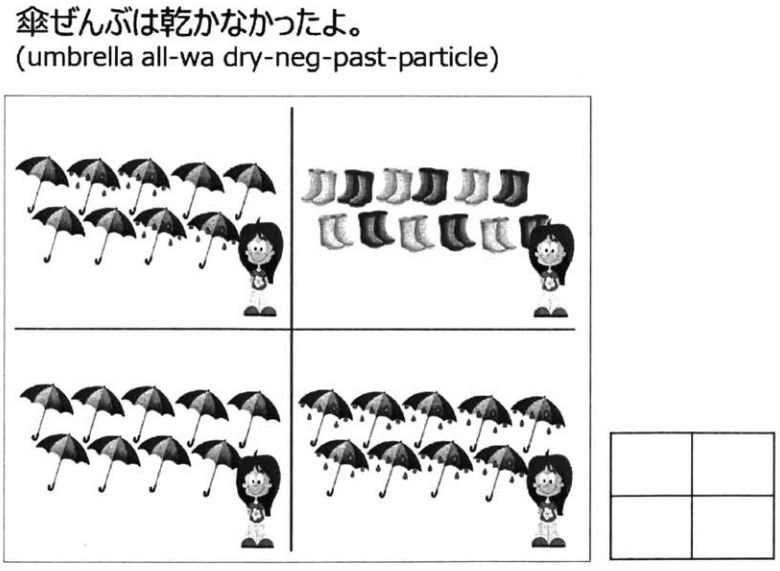


Figure 1: An example of how the sentence, pictures and checking cells are in formation

The pictures used were taken from the experiments with English-speaking populations (Chapter 5). There were six conditions tested in all. Among them, four conditions used floated quantifiers. The position of case marker and the universal quantifier (case marker *all* vs. *all* case marker) formed a factor, and the kind of case marker (*-ga* vs. *-wa*) formed another factor; the four conditions were created by crossing these factors. The rest of the two conditions had the universal quantifier preceding the noun phrase, and differs in the kind of case marker (*-ga* vs. *-wa*). The six conditions are summarized below.

(3) Examples corresponding to "*All of the umbrellas didn't dry.*"

Floated quantifier (FQ)

a.	Kasa	-ga	zembu	kawaka-nak-atta	-yo.
	umbrella	-NOM	all	dry-NEG-PAST	PRT
b.	Kasa	zembu	-ga	kawaka-nak-atta	-yo.
	umbrella	all	NOM	dry-NEG-PAST	PRT
c.	Kasa	-wa	zembu	kawaka-nak-atta	-yo.
	umbrella	-TOP	all	dry-NEG-PAST	PRT
d.	Kasa	zembu	-wa	kawaka-nak-atta	-yo.
	umbrella	all	TOP	dry-NEG-PAST	PRT

Non-floated

a.	Subete-no	kasa	-ga	kawaka-nak-atta	-yo.
	All-GEN	umbrella	-NOM	dry-NEG-PAST	PRT
b.	Subete-no	kasa	-wa	kawaka-nak-atta	-yo.
	All-GEN	umbrella	-TOP	dry-NEG-PAST	PRT

The volunteer participants were undergraduate students at Mie University, who attended either Introduction to Language Acquisition course or Introduction to Sentence Processing course. They participated voluntarily at the end of one of the classes. Students had no background in linguistics prior to taking the courses, and at the time of the survey, some minutes on scope phenomenon in English had been introduced. They were told to consult their own intuitions in Japanese. The four FQ conditions were between-subjects factors, and total of 87 students participated (for (3a), n = 22, (3b), n = 22, (3c), n = 22, and (3d), n = 21). Each questionnaire contained 5 target sentences as well as 3 filler sentences which did not contain negation. The

two non-FQ conditions were within-subjects factors, and total of 131 students participated. Each questionnaire contained just two target sentences (one with *-ga* and one with *-wa*).[96]

Results

Below is the summary of the results. With the left three word orders (NP-*ga* all, NP all-*ga*, and N-*wa* all), we see that the prominent interpretation is unambiguously "All > Not". Non-floated FQ accompanied with *-ga* also strongly suggests that the sentence is unambiguously "All > Not. The two sentence types (the 4th and the 6th from the left), both of which have the universal quantifier preceding *-wa*, have more responses that indicate the "Not > All" interpretation.

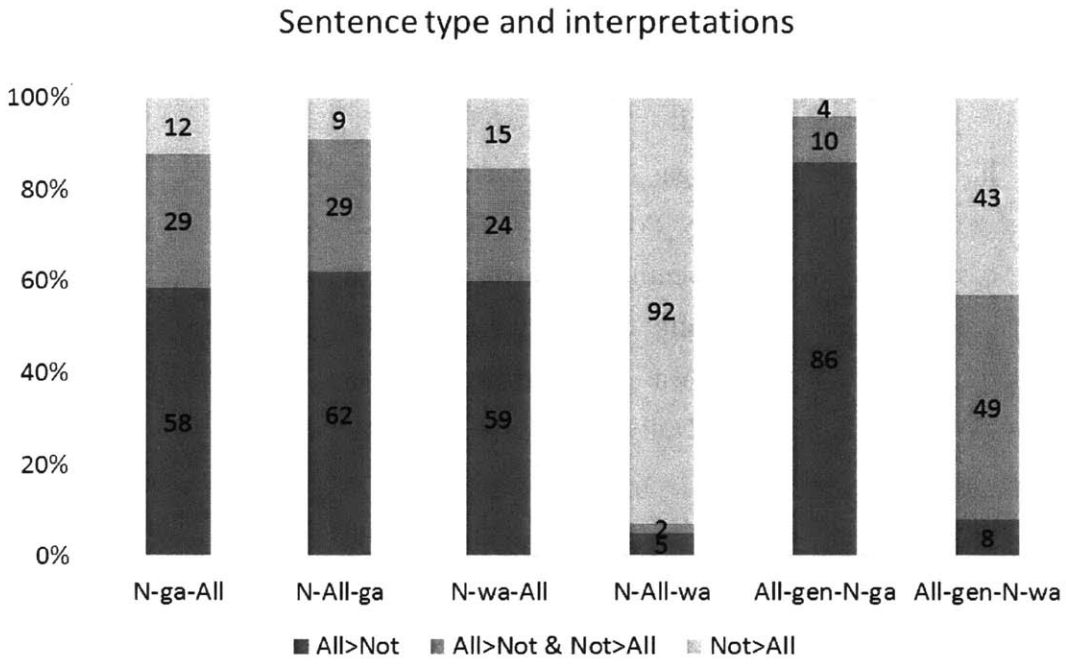


Figure 2: Results from survey 1

Looking at the left three FQ conditions, there are 25-30% of the answers which indicated both the "All > Not" pictures and "Not > All" pictures. I had explicitly raised a possibility that the

[96] I did not include sentences only with a universal quantifier in its subject position such as the pair below, because I thought the results would be obvious. However, it would make a better comparison if those sentences are tested as well – something to try for a follow-up experiment.

(i)	Subete-ga	kawaka-nak-atta-yo.	
	All-NOM	dry-NEG-PAST-PRT	(All > Not)
(ii)	Subete-wa	kawaka-nak-atta-yo.	
	All-TOP	dry-NEG-PAST-PRT	(Not > All)

sentence might be ambiguous, so I would take 25-30% as indicating the maximal percentages of the population who can get ambiguity in such sentences. We see that the forth FQ condition, "NP all-*wa* … Neg" strongly prefers the "Not > All" interpretation. Interesting thing to note is that the result of the non-FQ with -*wa* condition. Half of the responses indicate that the sentence would be ambiguous. The participants typically answered with the "ambiguous" responses with a note on prosody (e.g., "with -*wa* stressed, 'Not > All', but otherwise 'All > Not'"). About 40% of the responses indicated that the sentence would be unambiguously "Not > All".

In sum, the survey tells us that with the nominative marker -*ga*, the dominant interpretation that people access is the "All > Not" reading, regardless of the word order. With the Topic marker -*wa*, when the universal quantifier is floated behind the Topic marker, the sentence does not behave differently from the sentences with -*ga*. On the other hand, when the universal quantifier precedes the Topic marker -*wa*, preferred reading is "Not > All", and specifically, with the non-FQ word order "All-GEN N-*wa* … Neg", the sentence becomes highly ambiguous. This supplemental survey does not aim to construct a theory to explain all these behaviors, but at least the results give us ideas on what kind of sentence types give rise to which interpretation.

Appendix 7.2 Survey 2: Translation from English to Japanese

Methods

The next question is how Japanese speakers interpret English sentences with "All of the NPs … Neg", when it is in the written form. The participants were handed a sheet of paper, on which several pairs of a question and an answer are presented in English. The participants were told to translate them into natural conversation in Japanese. The volunteer participants were undergraduate students at Mie University, who attended Introduction to Sentence Processing course. Students had no background in linguistics prior to taking the courses, and at the time of the survey, no materials in semantics and pragmatics were introduced. The total of 53 students participated. Each questionnaire contained 5 target sentences as well as 3 filler sentences which did not contain negation. No pictures accompanied the sentences. The target sentences were taken from the experiments with English-speaking population. Their responses were coded and categorized for analysis.

Results

Below is the summary of results. The 6 categories at the top is "All > Not"-oriented responses. The 3 categories in the middle is "Not > All"-oriented responses, and the one at the bottom is the responses that had the structure of "All-GEN NP-wa … Neg", which is highly ambiguous as we saw in Appendix 7.1.

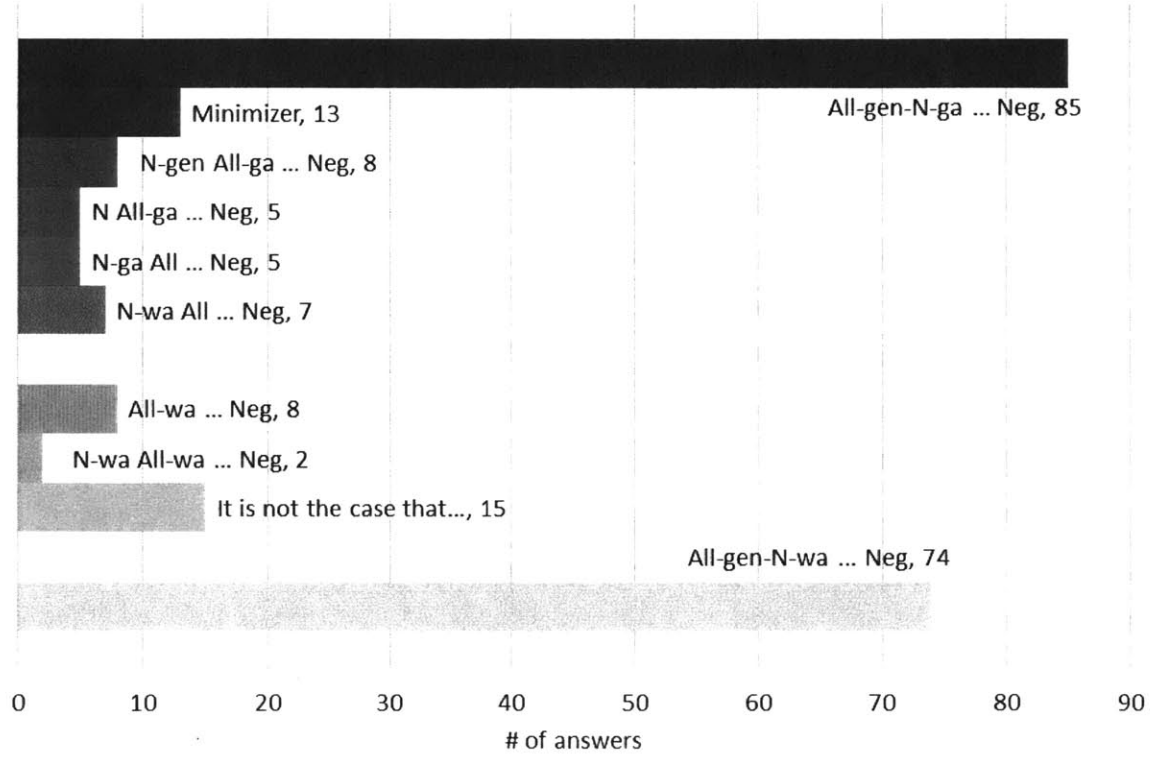


Figure 3: Results of survey 2

The results show that when the sentence is presented in written form, i.e., without any phonological cues, the participants' preferred interpretation seems to be "All > Not". The results seem to be exhibiting Japanese-speaking population's general preference for the surface scope. In other words, the preference for the "All > Not" interpretation could be characterized as a transfer from their first language.

Appendix 7.3 Japanese learners of English are sensitive to QAC and the RFR effect

Japanese speakers in typical school-settings generally never get explicitly taught that English sentences could be scopally ambiguous, let alone the connection between prosody and scope.

Also, reading and writing are the most common and focused methods to teach English at Japanese schools. All this leads to these questions; are Japanese-speaking learners of English sensitive to the difference in prosodic contour? Are they sensitive to the different levels of QAC? Do their rates of "Not > All" increase with RFR?

Methods

We used the same materials and procedure to the experiments with English-speaking populations. Unlike the children experiment, the participants were handed an answer sheet with four cells for each item, and asked to indicate a picture that would match what the girl was talking about. They were told to mark one picture, not multiple pictures. The PowerPoint slides with sound were played at the front of a classroom, and the students observed the story and marked their responses. Items were not repeated twice, and the sound files were played at natural speed, which together might have made the task challenging to the Japanese-speaking participants. There were 4 conditions as in the experiments with English-speaking populations, but one thing has changed; I placed the two kinds of contours in blocked sections to investigate a potential within-subjects effect of contour. That is, the half of the participants heard the sentences with the Falling contour first then sentences with RFR later, and the half of the participants heard the sentences with RFR first then sentences with Falling later.

The volunteer participants were undergraduate students at Mie University, who attended a course that is not related to linguistics. Students had no or little background in linguistics. The total of 99 students participated (for the experiment "Falling → RFR", n = 50, and for the experiment "RFR → Falling", n = 49). A session contained 8 target sentences (4 with Falling and 4 with RFR) as well as 6 filler sentences which did not contain negation.

Results and discussion

Some participants were excluded from the analyses, and the exclusion criteria are the following. Those who answered incorrectly to filler sentences more than half of the time (n = 9) were excluded. Similar to the experiments with English-speaking populations, sentences without negation were used as fillers. Half of the fillers used an emphatic past (e.g., "*All of the helicopters did fly.*"), and some participants might have got confused by that. Also, those who picked "All did" or "irrelevant" pictures on the target trials, which always contained negation, more than half of the time (n = 26) were excluded, because I suspect those participants cannot be regarded that they understood the sentences. As I noted earlier, the relatively harsh

environment might have deflated the performance to some extent; for example, they were only able to hear the trials once, and the session were played on a screen at the front of a classroom (they were not tested individually, but in groups of 15-25 participants at once). The total of 64 participants (30 on "Falling → RFR", 34 on "RFR → Falling") are included in the analyses.

Firstly, the responses were coded as "negated" vs. "non-negated" – given that there were many participants who missed negation, it would be interesting to see what the rates of picking either "Not > All" or "All > Not". The rates can be paraphrased as the rates of correctly hearing the negation in the sentence. Figure 4 below shows the results by condition. The error bars indicate 95% confidence intervals.

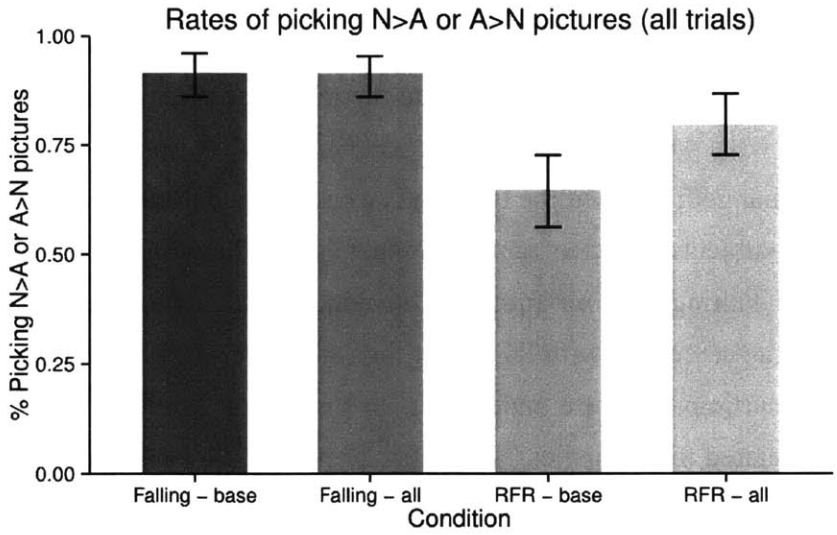


Figure 4: Rates of picking either "Not > All" or "All > Not"

The rates for the BaselineQ-Falling condition was 91%, for the AllQ-Falling condition was 91%, for the BaselineQ-RFR condition was 65%, and for the AllQ-RFR condition was 80%. Using GLMEM, a main effect of contour (p = .046), a marginally significant effect of question type (p = .055), and a significant interaction (p = .023) were observed.

The main effect of contour means that more errors on the RFR conditions were observed. This would suggest that for Japanese-speaking learners of English, RFR was perceived as a marked prosody and thus comprehending RFR required more processing load than the Falling counterparts did. This, in turn, suggests that Japanese-speaking learners of English are indeed sensitive to the difference in contour. The significant interaction means that more errors were

made on the combination of indirectly congruent Q-A pairs. This would suggest that for Japanese-speaking learners of English, the pair of indirectly congruent Q-A pairs was perceived differently from the other Q-A pairs; specifically, they got more confused by this combination. This result suggests that Japanese-speaking learners of English are sensitive to the felicity condition of RFR, which we discussed in §4.4. The effect was something that we did not see in the experiments with English-speaking populations in Chapter 5.

Secondly, let us look at the kinds of responses, to see how much "Not > All" responses were obtained. In the choice analysis, only the responses that picked either "All > Not" or "Not > All" are included to run a binomial analysis. The following is from the first halves of each blocked experiments, for the responses were not affected or primed by the other contour.

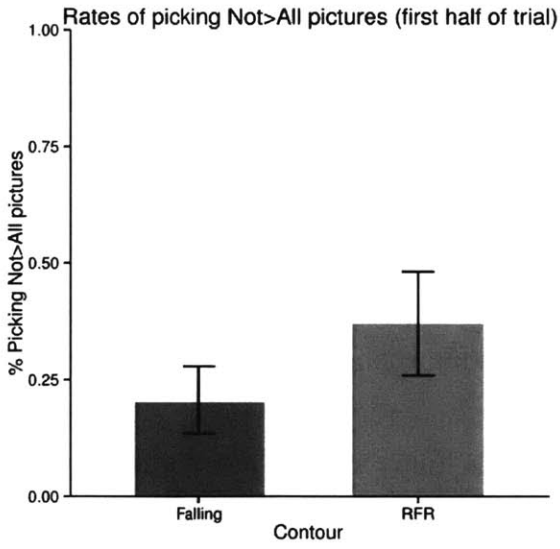


Figure 5: Rates of picking "Not > All" pictures in the first half of the trial

The rates of picking "Not > All" pictures on the Falling condition (question types combined) was 20%, and on the RFR condition (question types combined) was 37%. Using GLMEM, a main effect of contour ($p = .044$) was observed.

Note that the main effect is found in the same direction as English-speaking children and adults. The results suggest that Japanese-speaking learners of English seem to compute the effect of RFR in the same direction as English speakers do.[97] However, it is also important to

[97] The effect of RFR is not taught at school. This, combined with the results, suggest that the participants are forming Interlanguage (Selinker 1972), and the results would support that there is some (partial or indirect) access

note that the rates of "Not > All" was (only) 37% on the RFR condition. That means that majority of the responses indicated "All > Not" interpretations, which can be accounted for by the L1 transfer of preference of surface scope, which has been witnessed in Appendices §7.1 and §7.2.

We observed a question type effect, which was not observed in the experiments with English-speaking populations. Why did it emerge in the L2 study and not in the L1 studies? I suspect that the overall poor competence in L2 is in fact revealing the pragmatic effect. The rationale is the following. From the choice analysis, we see that the participants potentially have the knowledge of computing the effect of RFR. However, since their L2 performance is poor enough (compared to English-speaking children and adults), they cannot use their knowledge regarding prosody-semantics link; they have so much to process, and so they cannot afford to exercise the knowledge of prosody-semantics link. Only under such circumstances did the pragmatic manipulation, i.e. the question type, emerge. The discussion in §3.3 presented a phenomenon where a subtle difference in processing cost can be observable only by the population with small working memory/cognitive resources but not by the population with large working memory/cognitive resources. It is possible that the current results from L2 fall into the same category; in other words, the question type manipulation was too subtle for proficient L1 speakers (adults and children) to exhibit different behaviors. By contrast, the question type manipulation was realized as a big enough obstacle for developing L2 speakers to exhibit the different behaviors.

As I noted earlier, the results from the set of surveys and experiment with L2 population are preliminary and follow-up experiments will be needed to confirm the arguments.

---

to UG in second language acquisition. I am not committed to make claims regarding the inferences that the current results would make to the theory of second language acquisition, for now.

Bibliography

Akmajian, Adrian, and Ray Jackendoff. 1970. Coreferentiality and stress. *Linguistic Inquiry* 1(1): 124-126.

Alxatib, Sam. 2013. *'Only' and association with negative antonyms*. PhD dissertation, Massachusetts Institute of Technology.

Arnold, Jennifer E. 2008. THE BACON not *the bacon*: How children and adults understand accented and unaccented noun phrases. *Cognition* 108: 69-99.

Baauw, Sergio, Esther Ruigendijk, and Fernando Cuetos. 2004. The interpretation of contrastive stress in Spanish-speaking children. In *Proceedings of Generative Approaches to Language Acquisition 2003 (GALA2003)*, 103-114.

Baltazani, Mary. 2002. *Quantifier scope and the role of intonation in Greek*. PhD dissertation, University of California, Los Angeles.

Barner, David, Neon Brooks, and Alan Bale. 2011. Accessing the unsaid: the role of scalar alternatives in children's pragmatic inference. *Cognition* 118: 84-93.

Bayer, Josef. 1996. *Directionality and logical form*. Dordrecht: Springer Science+Business Media.

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

Beck, Sigrid. 2006. Intervention effects follow from focus interpretation. *Natural Language Semantics* 14:1-56.

Bill, Cory, Jacopo Romoli, Florian Schwarz, and Stephen Crain. 2014. Indirect scalar implicatures are neither scalar implicatures nor presuppositions (or both). Poster presented at the 27th Annual CUNY Conference on Human Sentence Processing.

Bolinger, Dwight. 1972. Accent is predictable (if you're a mind-reader). *Language* 48(3): 633-644.

Bonomi, Andrea, and Paolo Casalegno. 1993. Only: Association with focus in event semantics. *Natural Language Semantics* 2(1): 1-45.

Büring, Daniel. 1997. The great scope inversion conspiracy. *Linguistics and Philosophy* 20: 175–194.

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

Büring, Daniel, and Katharina Hartmann. 2001. The syntax and semantics of focus-sensitive particles in German. *Natural Language & Linguistic Theory* 19: 229-281.

Carlson, Lauri Henrik. 1982. *Dialogue games: An approach to discourse analysis*. PhD dissertation, Massachusetts Institute of Technology.

Chierchia, Gennaro, Stephen Crain, Maria Teresa Guasti, Andrea Gualmini, and Luisa Meroni. 2001. The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In *BUCLD 25: Proceedings of the 25th annual Boston University Conference on Language Development*, 157-168. Somerville, MA: Cascadilla Press.

Conroy, Anastasia Marie. 2008. *The role of verification strategies in semantic ambiguity resolution in children and adults*. PhD dissertation, University of Maryland.

Constant, Noah. 2012. English rise-fall-rise: a study in the semantics and pragmatics of intonation. *Linguistics and Philosophy* 35: 407-442.

Constant, Noah. 2014. *Contrastive topics: Meanings and realizations*. PhD dissertation, University of Massachusetts Amherst.

Crain, Stephen, Weijia Ni, and Laura Conway. 1994. Learning, parsing, and modularity. In *Perspectives on sentence processing*, eds. Charles Clifton, Jr., Lyn Frazier, and Keith Rayner, 443-467. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crain, Stephen and Rosalind Thornton. 1998. *Investigations in universal grammar: A guide to research on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Dahan, Delphine, Michael K. Tanenhaus, and Craig G. Chambers. 2002. Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language* 47: 292-314.

Davidson, Donald. 1984. *Inquiries into truth and interpretation*. Oxford: Clarendon Press.

Demuth, Katherine, Jennifer Culbertson, and Jennifer Alter. 2006. Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech* 49(2): 137-174.

É. Kiss, Katalin, and Beáta Gyuris. 2003. Apparent scope inversion under the rise fall contour. *Acta Linguistica Hungarian* 50(3-4): 371-404.

Endo, Mika. 2004. Developmental issues on the interpretation of focus particles by Japanese children. In *BUCLD 28: Proceedings of the 28th annual Boston University Conference on Language Development*, 141-152. Somerville, MA: Cascadilla Press.

Erteschik-Shir, Nomi. 1997. *The dynamics of focus structure*. Cambridge: Cambridge University Press.

Erteschik-Shir, Nomi. 2007. *Information structure: The syntax-discourse interface*. Oxford: Oxford University Press.

von Fintel, Kai. 1994. *Restrictions on quantifier domains*. PhD dissertation, University of Massachusetts Amherst.

Foppolo, Francesca, Maria Teresa Guasti, and Gennaro Chierchia. 2012. Scalar implicatures in child language: give children a chance. *Language Learning and Development* 8: 365-394.

Fox, Danny. 2012. Handouts in lectures on the semantics of questions, MIT, September, 2012. Available at: http://web.mit.edu/linguistics/people/faculty/fox/talks.html

Giannakidou, Anastasia. 1998. *Polarity sensitivity as (non)veridical dependency*. Amsterdam: John Benjamins.

Giannakidou, Anastasia. 2000. Negative... concord? *Natural Language & Linguistic Theory* 18(3): 457-523.

Goro, Takuya. 2007. *Language-specific constraints on scope interpretation in first language acquisition*. PhD dissertation, University of Maryland.

Grice, Paul. 1975. Logic and conversation. In *Syntax and semantics*, eds. Peter Cole and James Morgan, 41-58. New York, NY: Academic Press.

Groenendijk, Jeroen, and Gerardus Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers*. PhD dissertation, University of Amsterdam.

Gualmini, Andrea, Stephen Crain, Luisa Meroni, Gennaro Chierchia, and Maria Teresa Guasti. 2001. At the semantics/pragmatics interface in child language. In *Proceedings of the 11th Semantics and Linguistic Theory Conference (SALT XI)*, 231-247. Ithaca, NY: Cornell University.

Gualmini, Andrea, Sarah Hulsey, Valentine Hacquard and Danny Fox. 2008. The Question-Answer Requirement for scope assignment. *Natural Language Semantics* 16: 205-237.

Gualmini, Andrea, Simona Maciukaite, and Stephen Crain. 2003. Children's insensitivity to contrastive stress in sentences with *only*. In *Proceedings of the 25th Annual Penn Linguistics Colloquium Volume 8*, 87-100.

Guasti, Maria Teresa, Gennaro Chierchia, Stephen Crain, Francesca Foppolo, Andrea Gualmini and Luisa Meroni. 2005. Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes* 20(5): 667-696.

Guerzoni, Elena. 2003. *Why even ask? On the pragmatics of questions and the semantics of answers*. PhD dissertation, Massachusetts Institute of Technology.

Gussenhoven, Carlos. 1983. *A semantic analysis of the nuclear tones of English*. Bloomington, IN: Indiana University Linguistics Club (IULC).

Gyuris, Beáta. 2002. *The semantics of contrastive topics in Hungarian*. PhD dissertation, Eötvös Loránd Tudományegyetem.

Hackl, Martin, Erin Olson, and Ayaka Sugawara. 2014. Processing *only*: Scalar presupposition and the structure of ALT(S). Presentation at the 45th annual meeting of the North East Linguistic Society (NELS 45).

Hackl, Martin, Ayaka Sugawara, and Ken Wexler. 2015. Question-Answer (In)Congruence in the acquisition of *only*. In *BUCLD 39: Proceedings of the 39th annual Boston University Conference on Language Development*, 204-217. Somerville, MA: Cascadilla Press.

Hagstrom, Paul Alan. 1998. *Decomposing questions*. PhD dissertation, Massachusetts Institute of Technology.

Halliday, M. A. K. 1967. Notes on transitivity and theme in English: part 2. *Journal of Linguistics* 3(2): 199-244.

Hamblin, Charles Leonard. 1973. Questions in Montague English. *Foundations of language* 10: 41-53. Reprinted in *Montague grammar*, ed. Barbara H. Partee, 247-259. New York: Academic Press.

Hara, Yurie. 2006. *Grammar of knowledge representation: Japanese discourse items at interfaces*. PhD dissertation, University of Delaware.

Hattori, Noriko, Seiki Ayano, Dylan Herrick, David Stringer, and Koji Sugisaki. 2006. Topics in Child Japanese. In *Proceedings of the 7th Tokyo Conference on Psycholinguistics*, 103-120.

Hirschberg, Julia and Cinzia Avesani. 2000. Prosodic disambiguation in English and Italian. In *Intonation: Theory, models, and applications*, ed. Antonis Botinis, 87-95. Dordrecht: Kluwer.

Hoji, Hajime. 1985. *Logical form constraints and configurational structures in Japanese*. PhD dissertation, University of Washington.

Huang, Yi Ting, and Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive Psychology* 58: 376-415.

Ianucci, David, and David Dodd. 1980. The development of some aspects of quantifier negation in children. In *Papers and Re-ports on Child Language Development (PRCLD)* #19, 88-94. Stanford, CA: Stanford University Department of Linguistics.

Ito, Kiwako, Nobuyuki Jincho, Utako Minai, Naoto Yamane, and Reiko Mazuka. 2012. Intonation facilitates contrast resolution: Evidence from Japanese adults and 6-year olds. *Journal of Memory and Language* 66: 265-284.

Ito, Kiwako, and Shari R. Speer. 2008. Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language* 58: 541-573.

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

Jackson, Scott. 2007. *Information, truth, structure, and sound*. PhD dissertation, University of Arizona.

Jackson, Scott. 2008. The prosody-scope relation in Hungarian. In *Papers from the Veszprém conference approaches to Hungarian* 10, eds. Christopher Piñón and Szilárd Szentgyörgyi, 85-102. Budapest: Akadémiai Kiadó.

Jespersen, Otto. 1933/1964. *Essentials of English grammar*. (Eighth printing, 1994). Tuscaloosa, AL: The University of Alabama Press.

King, Jonathan, and Marcel Adam Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of memory and language* 30: 580-602.

Kjelgaard, Margaret M. and Shari R. Speer. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language* 40: 153-194.

Klinedinst, Nathan. 2005. *Scales and Only*. Master's thesis, UCLA.

Koizumi, Yukiko. 2009. *Processing the not-because ambiguity in English: The role of pragmatics and prosody*. PhD dissertation, City University of New York.

Kotek, Hadas. 2014. *Composing questions*. PhD dissertation. Massachusetts Institute of Technology.

Krifka, Manfred. 1998. Scope inversion under the rise-fall contour in German. *Linguistic Inquiry* 29: 75-112.

Krifka, Manfred. 2001. For a structured meaning account of question and answers. *Natural Language Semantics* 9: 1-40.

Krifka, Manfred. 2006. Question-answer congruence and the proper representation of focus. Slides from the 10th workshop on the semantics and pragmatics of dialogue. University of Potsdam.

Kuno, Susumu. 1972. Functional sentences perspective: A case study from Japanese and English. *Linguistic Inquiry* 3(3): 269-320.

Kuno, Susumu. 1973. *The structure of the Japanese language. Cambridge*, MA: MIT Press.

Kuno, Susumu. 1982. The focus of the question and the focus of the answer. In *Papers from the parasession on nondeclarative sentences (CLS 18(2))*, 134-157. Chicago, IL: Chicago Linguistic Society.

van Kuppevelt, Jan. 1995. Discourse structure, topicality and questioning. *Journal of Linguistics* 31(1): 109-147.

van Kuppevelt, Jan. 1996. Inferring from topics: Scalar implicatures as topic-dependent inferences. *Linguistics and Philosophy* 9: 393-443.

Kuroda, Sige-Yuki. 1965. *Generative grammatical studies in the Japanese language*. PhD dissertation, Massachusetts Institute of Technology.

Ladd, D. Robert. 1980. *The structure of intonational meaning*. Bloomington, IN: Indiana University Press.

Lambrecht, Knud. 1996. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Lehiste, Ilse. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa: An international journal of linguistics* 7: 107-122.

Liberman, Mark and Ivan Sag. 1974. Prosodic form and discourse function. *Proceedings of the tenth regional meeting of Chicago linguistic society*, 416-427. Chicago, IL: Chicago Linguistic Society.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

McDaniel, Dana, and Thomas L. Maxfield. 1992. Principle B and contrastive stress. *Language Acquisition* 2(4): 337-358.

McMahon Erin, Jeffrey Lidz and Janet Pierrehumbert. 2004. Suprasegmental cues to meaning in child-directed speech. In *Handbook of the 17th CUNY Sentence Processing Conference*. University of Maryland

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)*, 163-171. Somerville, MA: Cascadilla Proceedings Project.

Musolino, Julien. 1998. *Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in English*. PhD dissertation, University of Maryland.

Musolino, Julien. 2011. Studying language acquisition through the prism of isomorphism. In *Handbook of generative approaches to language acquisition: Studies in theoretical psycholinguistics* 41, eds. Jill de Villiers and Thomas Roeper, 319-349. Dordrecht: Springer.

Musolino, Julien, Stephen Crain and Rosalind Thornton. 2000. Navigating negative quantificational space. *Linguistics* 38-1: 1-32.

Musolino, Julien and Jeffrey Lidz. 2006. Why children aren't universally successful with quantification. *Linguistics* 44-4: 817-852.

Nakanishi, Kimiko. 2007. Prosody and scope interpretations of the topic marker *wa* in Japanese. In *Topic and focus: Cross-linguistic perspectives on meaning and intonation*, eds. Chungmin Lee, Matthew Gordon and Daniel Büring, 177-193.

Nicolae, Andreea Cristina. 2013. *Any questions? Polarity as a window into the structure of questions*. PhD dissertation, Harvard University.

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.

Noveck, Ira A. 2001. When children are more logical than adults: experimental investigations of scalar implicature. *Cognition* 78: 165-188.

Papafragou, Anna, and Julien Musolino. 2003. Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition* 86: 253-282.

Paterson, Kevin B., Simon P. Liversedge, Caroline Rowland, and Ruth Filik. 2003. Children's comprehension of sentences with focus particles. *Cognition* 89: 263-294.

Paterson, Kevin B., Simon P. Liversedge, Diane White, Ruth Filik, and Kristina Jaz. 2006. Children's interpretation of ambiguous focus in sentences with "only." *Language Acquisition* 13(3): 253-284.

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 of the 24th annual Boston University Conference on Language Development*, 583-596. Somerville, MA: Cascadilla Press.

Pintér, Lilla. 2015. Preschoolers' interpretation of the focus particle *cask* 'only' in Hungarian. In *Proceedings of ConSOLE XXIII*, 1-20.

Potts, Christopher. 2006. How far can pragmatic mechanisms take us? *Theoretical Linguistics* 32(3): 307-320.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27(1): 53-94.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2001. Relativized Minimality effects. In *The handbook of contemporary syntactic theory*, eds. Mark Baltin and Chris Collins, 89-110. Oxford: Blackwell.

Roberts, Craige. 1996/2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 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.

van Rooy, Robert. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26: 727-763.

Sano, Tetsuya. 2011. Abstract CP/IP configuration in child Japanese. In *BUCLD 36: Proceedings of the 36th annual Boston University Conference on Language Development*, 525-535. Somerville, MA: Cascadilla Press.

Sauerland, Uli, and Oliver Bott. 2002. Prosody and scope in German inverse linking constructions. In *Proceedings of the 1st international conference on speech prosody*, eds. Bernard Bel and Isabelle Marlien, 623-628. Aix-en-Provence: Laboratoire Parole et Langage.

Schwarzschild, Roger. 1999. Givenness, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics* 7: 141-177.

Selinker, Larry. 1972. Interlanguage. *IRAL; International review of applied linguistics in language teaching* 10(3): 209-231.

Selkirk, Elisabeth O. 1995. Sentence prosody: intonation, stress and phrasing. In *The handbook of phonological theory*, ed. John A. Goldsmith, 550-569. Oxford: Blackwell.

Shetreet, Einat, Julia Reading, Nadine Gaab, and Gennaro Chierchia. 2012. SOME and NOT EVERY: comparing direct and indirect scalar implicatures. Presentation at the 37th annual Boston University Conference on Language Development.

Snedeker, Jesse, and John Trueswell. 2003. Using prosody to avoid ambiguity: effects of speaker awareness and referential context. *Journal of Memory and Language* 48: 103-130.

Spivey-Knowlton, Michael J. 1996. *Integration of visual and linguistic information: Human data and model simulations*. PhD dissertation, University of Rochester.

Stalnaker, Robert. 1978. Assertion. In *Syntax and semantics vol. 9: Pragmatics*, ed. Peter Cole, 315-332. New York, NY: Academic Press.

von Stechow, Arnim. 1990. Focusing and backgrounding operators. In *Discourse particles*, ed. Werner Abraham, 37-84. Amsterdam: John Benjamins.

Sugawara, Ayaka, Martin Hackl, Irina Onoprienko, and Ken Wexler. To appear. Children know the prosody/semantic link: Experimental evidence from Rise-Fall-Rise and scope. In *Linguistic and Cognitive Aspects of Quantification*. Springer.

Sugawara, Ayaka, and Ken Wexler. 2014. Children do not accept unambiguous inverse-scope readings: experimental evidence from prosody and scrambling in Japanese. In *Proceedings of formal approaches to Japanese linguistics* 7 (FAJL7), 215-226.

Syrett, Kristen, Georgia Simon and Kristen Nisula. 2014a. Prosodic disambiguation of scopally ambiguous sentences. In *Proceedings of 43rd annual meeting of the north east linguistic society (NELS)*, 141-152.

Syrett, Kristen, Georgia Simon and Kristen Nisula. 2014b. Prosodic disambiguation of scopally ambiguous quantificational sentences in a discourse context. *Journal of Linguistics* 50: 453-493.

Tancredi, Christopher. 2016. The grammar of TOPIC, FOCUS and *Givenness*. Handout for a presentation at New York University, available at: https://sites.google.com/site/christophertancredi/talk-handouts

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217): 1632-1634.

Tomioka, Satoshi. 2010a. Contrastive topics operate on speech acts. In *Information structure: Theoretical, typological, and experimental perspectives*, eds. Malte Zimmermann and Caroline Féry, 115-138. Oxford: Oxford University Press.

Tomioka, Satoshi. 2010b. A scope theory of contrastive topics. *Iberia* 2(1): 113-130.

Vallduví, Enric and Elisabet Engdahl. 1996. The linguistic realization of information packaging. *Language* 34: 459-519.

Vermeulen, Reiko. 2009. Topics, contrast and contrastive topics in Japanese. In *Proceedings of the 5th Workshop on Altaic Formal Linguistics (WAFL5)*, 361-372.

Viau, Joshua, Jeffrey Lidz, and Julien Musolino. 2010. Priming of Abstract Logical Representations in 4-Year-Olds. *Language Acquisition* 17(1-2): 26-50.

Wagner, Michael, and Duane G. Watson. 2010. Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes* 25(7-9): 905-945.

Ward, Gregory, and Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61: 747–776.

Weist, Richard, Aleksandra Pawlak, and Karen Hoffman. 2009. Finiteness systems and lexical aspect in child Polish and English. *Linguistics* 47(6): 1321-1350.

Weist, Richard, and Andrea Zevenbergen. 2008. Autobiographical memory and past time reference. *Language Learning and Development* 4(4): 291-308.

Wells, Bill, Sue Peppé, and Nata Goulandris. 2004. Intonation development from five to thirteen. *Journal of Child Language* 31(4): 749-778.

Zhou, Peng, and 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)*, 336-346. Somerville, MA: Cascadilla Proceedings Project.

Zhou, Peng, and Stephen Crain. 2010. Focus identification in child Mandarin. *Journal of Child Language* 37: 965-1005.