ABSTRACT

Title of Document: COMMITMENT AND FLEXIBILITY IN THE DEVELOPING PARSER

Akira Omaki, Doctor of Philosophy, 2010

Directed By: Professor Colin Phillips, Department of Linguistics

This dissertation investigates adults and children's sentence processing mechanisms, with a special focus on how multiple levels of linguistic representation are incrementally computed in real time, and how this process affects the parser's ability to later revise its early commitments. Using cross-methodological and cross-linguistic investigations of long-distance dependency processing, this dissertation demonstrates how paying explicit attention to the procedures by which linguistic representations are computed is vital to understanding both adults' real time linguistic computation and children's reanalysis mechanisms.

The first part of the dissertation uses time course evidence from self-paced reading and eye tracking studies (reading and visual world) to show that long-distance dependency processing can be decomposed into a sequence of syntactic and interpretive processes. First, the reading experiments provide evidence that suggests that filler-gap dependencies are constructed before verb information is accessed. Second, visual world experiments show that, in the absence of information that would allow hearers to predict

verb content in advance, interpretive processes in filler-gap dependency computation take around 600ms. These results argue for a predictive model of sentence interpretation in which syntactic representations are computed in advance of interpretive processes.

The second part of the dissertation capitalizes on this procedural account of filler-gap dependency processing, and reports cross-linguistic studies on children's long-distance dependency processing. Interpretation data from English and Japanese demonstrate that children actively associate a fronted wh-phrase with the first VP in the sentence, and successfully retract such active syntactic commitments when the lack of felicitous interpretation is signaled by verb information, but not when it is signaled by syntactic information. A comparison of the process of anaphor reconstruction in adults and children further suggests that verb-based thematic information is an effective revision cue for children.

Finally, distributional analyses of wh-dependencies in child-directed speech are conducted to investigate how parsing constraints impact language acquisition. It is shown that the actual properties of the child parser can skew the input distribution, such that the effective distribution differs drastically from the input distribution seen from a researcher's perspective. This suggests that properties of developing perceptual mechanisms deserve more attention in language acquisition research.

COMMITMENT AND FLEXIBILITY IN THE DEVELOPING PARSER

By

Akira Omaki

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Colin Phillips, Chair
Professor Jeffrey Lidz
Professor Norbert Hornstein
Professor Amy Weinberg
Professor Robert DeKeyser, Dean's Representative

© Copyright by
Akira Omaki
2010

Acknowledgements

Now that I am finally writing the acknowledgment section of my thesis, all I can think of is this one question: How did I get here? When I try to answer it, I can only come up with names after names, who did what for me and when, not knowing why they have been so unconditionally supportive of my (sometimes bumpy) journey through the Linguistics program in Maryland. It feels almost dishonest to say "I wrote this dissertation" because - just like everything else that people think "I" achieved in the past 5 years - it would have been impossible without all the wonderful people around me.

First and foremost, I must thank Colin Phillips for all his support. I do not pretend that working with Colin was never frustrating - it is a natural response when my ideas or experiment plans get shot down regularly in weekly meetings, I get pages and pages of critical comments on my slides and papers (a lot of them I had to translate in my mind - e.g., "This is promising." → "Hey, this looks pretty good!"), and at times, he is so stubborn on his ideas that I had to give in and follow his advice against my will. But after almost all these temporarily frustrating occasions, I ended up saying "oh boy, Colin was right *again*." He knew exactly what my problem was: I was always full of ideas, but oftentimes I was not carefully thinking through the logic to see pitfalls in the experimentation or argumentation, and he knew exactly what kind of learning experience I would need in order to break out of my comfort zone to overcome my weaknesses. This is all due to many of Colin's superb abilities as a researcher and supervisor: his ability to synthesize information and ideas across different issues and fields to predict what will happen, his passion and drive for patiently listening to his students and improving his students' research, his true altruistic desire to respond to students' needs, and his amazing

ability to think straight and work efficiently while sleep-deprived, which is why he can respond to his students' emails at 3am (he can definitely beat a certain presidential candidate in the department of "being productive and making important decisions at 3am"). He was particularly encouraging and supportive in my final year at Maryland, while I was going through the stressful phases of job hunting and dissertation projects. I cannot be more grateful that I had a chance to work with him and learn from him, and in a way, I feel like I was not able to learn as much as I should have, given the insane amount of support he has provided. All I can do now is to pay this forward to my future students, as I would like them to have the kind of learning experience I was lucky to have with Colin. And I will certainly not forget the most important lesson I have learned from him: When my experiments did not work and I was distressed, he said that we do not learn anything new when things work, because when things work, that only confirms what we already knew. We can only learn new things when things do not work - that is what advances our knowledge.

I am truly indebted to all the support and encouragement that Jeff Lidz has given me over the last 5 years, and I feel extremely lucky that he came to Maryland at the same time as I did. If I had only worked with Colin, I may have been at times too discouraged or frustrated to come up with solutions on my own, but it is on these occasions that Jeff patiently listened to me, suggested innovative solutions, or reminded me why my ideas were interesting and worth pursuing. Working as a TA for his undergraduate course on language acquisition was extremely delightful, as I could not only learn from his crystal clear lectures, but I was also able to have little meetings to discuss all kinds of things with him twice a week, walking between Marie Mount Hall and Hornbake. It was through

these interactions with him that I came up with half of my dissertation projects (especially Chapters 4 and 6). In many ways, I could really sympathize with his general approach to science - encouraging new and innovative ideas, while being unafraid to be wrong, and I really wish to be a wonderful researcher like him in the future.

I am also thankful for the other three thesis committee members, Norbert Hornstein, Amy Weinberg, and Robert DeKeyser, for their patience and support for my dissertation work. Despite my late submission of the pre-defense thesis draft, they have provided healthy encouragement and skepticism that really pushed my thinking to go so much further than where it was before the defense. Norbert made me think more about the importance of representational details in my psycholinguistic research, and this also reminded me how often he brought my attention back to the fact that syntax is just plain interesting and fun. Amy has reminded me of the importance of explicitness in explaining cognitive processes, and her precision helped me clarify some issues that I could not solve before. I owe them all so much for improving the quality of my dissertation.

I am indebted to Howard Lasnik for all the syntax training he gave me in Maryland. Whenever I felt a little tired of experimental work, I sought refuge in theoretical syntax (in addition to soccer and Ling Pong). Howard was always there for me when I needed his help with my syntax work, and the fundamental training in syntax I got from him has helped my work in psycholinguistics in many ways. I'd also like to thank David Poeppel. Although I was only able to take one course from him (though I certainly wanted more), I was extremely lucky to have the chance to work on a child ERP review paper with him. That was the most fun and exiting paper writing experience I have ever

had in my life, and I will never forget that even the generally painful experience of writing can be turned into a joy.

I am also thankful that I had two wonderful mentors in Maryland - Masaya Yoshida and Takuya Goro. Masaya influenced my approach to syntax and psycholinguistics in many ways, and many discussions I had with him were essential for understanding the importance of syntactic theories in psycholinguistics. Takuya was my mentor on language acquisition, and he taught me everything I wished to know about doing child language research. His insights and critical comments on my child experiment ideas and NSF dissertation grant proposal were truly indispensable, and I am indebted to his kindness and efforts to make my child research happen in Mito, Japan.

I was blessed to have so many wonderful friends and colleagues who enriched my everyday life in and out of the lab. I was extremely lucky to be co-author on projects with wonderful colleagues and friends. Ellen Lau was simply wonderful to work with. She is always very careful and yet encouraging and fun, and she gave me lots of practical advice on "getting things done" during my dissertation writing phase. It was a real pleasure to explore the eye-tracking-during-reading method with her, and I am so happy and proud that our work became the first eye-tracking-during-reading experiment to happen in Maryland Linguistics. Chizuru Nakao was my syntax buddy - I could not have written any of my syntax papers without all the discussions we had as well as her analytical insights. She was also a lab night owl like me, and I will never forget those fun nights I spent in the lab, discussing syntax (and occasionally singing Metallica and Dream Theater songs) with her. Johannes Jurka gave me a chance to work on a Japanese syntax project, which in a way I was always afraid of tackling. Not everything went well in this

project, but all the fun time we had together talking about Japanese islands gave me courage to face problems in the Japanese syntax literature. Dave Kush also made it possible for me to look further into Swedish islands, which was a topic that I had always been fascinated by but was always too intimidated to work on by myself. I will never forget his incredible analytical insights or our GLOW trip to Nante. Stacey Trock was my pre-school study mate for a long time. Although the project we worked on together stopped in the adult pilot phase, she taught me all the important practical skills in making stimuli for child experiments, the most important of which is the use of Microsoft clipart. Matt Wagers introduced me to the world of R in analyzing eye-tracking data, as well as to the issue of memory in psycholinguistics that I had not known the importance of before interacting with him.

My graduate research was supported by many wonderful research assistants. First of all, Imogen Davidson White - I really do not know how to thank her enough. With her amazing ability to learn things quickly, she assisted me so well with all aspects of my research in my final year at Maryland and in Hiroshima, and most importantly, she was a real joy to work with. Alan Mishler was so kind and helpful in assisting me with the ASL D6 Remote eye-tracker. Our journey with this eye-tracker was not always fun, but for sure I could not have gone through all that stress if I was working on it just by myself. I also need to thank him for his excellent editorial assistance in proofreading my papers. Angela Stanley was very helpful with CHILDES coding and child eye-tracking experiments. Adina Rubinoff, Hana Quon, Mike Shvartsmann and Jennifer Merickel provided assistance with my ISCAN eye-tracking experiments. I am especially indebted to Hana Quon, who patiently joined my first journey with visual-world eye-tracking.

Kassie Gynther and Tiff Boatman were wonderful assistants who helped with all aspects of my infant lab research, which unfortunately did not make it into the dissertation. I also thank Rebecca Baier for facilitating my work in the infant lab.

I'm grateful that I had wonderful classmates, Tim Hunter, Chris Dyer, Shiti Malhotra, and Jobin Mathew, who went through the first-year course work together with me. Tim Hunter in particular enriched the theoretical linguistics side of my life so much, and I will not forget all those fun hours I spent talking about syntax (and his healthy skepticism about psycholinguistics!) with him.

Annie Gagliardi and Brian Dillon have been wonderful friends and colleagues for me. Even though we never got to work on a project together, we shared a lot of interests and had so many interesting discussions in and out of the lab.

Maki Kishida has been a wonderful friend and roommate, and I cannot thank her enough for maintaining the apartment well when I was too busy to attend to it.

Ewan Dunbar, WingYee Chow, and Shevaun Lewis were wonderful office mates in my final year, and I appreciate that they did not complain about me dominating the white board, or hanging a "Noren" (a Japanese door curtain) in front of our office door.

Other friends, colleagues and professors who have enriched my personal and intellectual life in Maryland include Pedro Alcocer, Diogo Almeida, Sachiko Aoshima, Elika Bergelson, Tonia Bleam, Greg Cogan, Alex Drummond, Veronica Figueora, Rob Fiorentino, Suzanne Freynik, Tomo Fujii, Michael Gagnon, Valentine Hacquard, Tim Hawes, Sarah Helfinstein, Shannon Hoerner, So-one Hwang, Bill Idsardi, Yakov Kronrod, Sol Lago, Brad Larson, Chris LaTerza, Terje Londahl, Rebecca Mckeown, Utako Minai, Phil Monahan, Giovanna Morini, Valerie Van Mulukom, Hajime Ono,

Kaori Ozawa, Leticia Pablos, Dan Parker, Lisa Pearl, Ariane Rhone, Bridget Samuels, Matthias Scharinger, Jon Sprouse, Megan Sutton, Eri Takahashi, Jeff Walker, Alexis Wellwood, Alexander Williams, and Ming Xiang.

Friends and researchers outside Maryland have also given me a lot of support. Nobuaki Akagi, Jon Brennan, Youngon Choi, Jenny Culbertson, Judith Degen, Matt Husband, Christina Kim, Yukiko Koizumi, Ivana Mitrovic, Miki Obata, Hugh Rabagliati, Barbara Schulz, and Sharon Unsworth, thank you all for your encouragement. People who I worked with in Hawaii remained very supportive of my work and development after I left Hawaii. Bonnie Schwartz, Kamil Deen and Amy Shafer have been particularly supportive and always showed interest in my research at Maryland. I also learned a lot from visitors at Maryland, especially Jun Abe, Myung-Kwan Park, and Alan Munn, who kindly showed interest in my work. The dissertation benefited from discussions and comments from Janet Fodor, Richard Larson, Geraldine Legendre, Paul Smolensky, Jesse Snedeker, and John Trueswell.

I am indebted to Hiromu Sakai for making it possible to conduct part of my dissertation research in Hiroshima. It was a real joy to interact with students in Hiroshima too, and I learned a lot from the warm and kind research community there.

I learned an incredible amount from teaching students at Maryland. I taught as a sole instructor for Ling240 (Language and Mind: Introduction to Linguistics) and Ling499 (Psycholinguistics seminar), and I am particularly thankful for students in Ling499 who gave me a fresh perspective on many aspects of my dissertation work.

Kathi Faulkingham, Rob Magee, and Kim Kwok have been very kind and helpful with all sorts of paperwork that I needed for my travels, grant management and payroll. I

always appreciated Kathi's quick response to my requests, as well as Kim's patience and thoughtfulness when she was dealing with my NSF grant applications and management. It is due to their support that I was able to concentrate on my academic work.

My research was supported by the Summer Research Fellowship and the Wylie Dissertation Fellowship from the Graduate School at the University of Maryland, as well as the NSF #BCS-0954651 Doctoral Dissertation Improvement Grant. I am grateful for the financial support for my research.

I am also indebted to Lake Artemesia in Berwyn Heights. I went there almost every week, and made more trips when I was feeling stressed and tired. Whenever I took a walk around the lake, it always lifted my spirits. Some people say Lake Artemesia is not 'pretty enough' to visit, but I never understood this, nor do I care what others say. What I appreciate is the fact that Lake Artemesia is so accessible, and just accepts anybody who visits there. Not to mention the cute little birds and animals that live there, but adults and children of all backgrounds go there for relaxation and recreation, enjoying jogging, fishing, sitting on benches and watching the sunset. Seeing all this together in one place just makes me happy, reminding me that diversity and generosity is one of the reasons why I like being in this country.

Finally, I am indebted to my family, Yasuko Omaki, Chuichi Omaki, Nobuaki Omaki, and Hiroshi Omaki. Thank you for raising me to be who I am today. They have supported my academic endeavors and accepted me the way I am for all these years, even if they do not understand what I am doing. I really owe them everything.

Table of Contents

Acknowledgements vi
Table of Contents xiv
List of Tables xvii
List of Figures xviii
Chapter 1: Introduction 1
1.1 Incremental sentence interpretation in adult language comprehension 3
1.1.1 Incremental sentence interpretation in the visual world 5
1.1.2 Active computation of long-distance dependencies 11
1.2 Incremental sentence interpretation and revision difficulties in children 16
1.2.1 Ambiguity resolution and kindergarten-path effects 16
1.2.2 Incremental structural misanalysis and interpretation persistence 20
1.2.3 Quality of error signals and kindergarten-path effects 23
1.3 Outline of the dissertation 26
Chapter 2: Syntactic processes of verb-independent gap creation 32
2.1 What triggers active gap creation? 33
2.2 Experiment 1 (self-paced reading) 42
2.2.1 Method 44
2.2.2 Results 47
2.2.3 Discussion 52
2.3 Experiment 2 (eye-tracking during reading) 56
2.3.1 Method 57
2.3.2 Results 64
2.3.3 Discussion 67
2.4 General discussion 70
2.5 Conclusion 76
Chapter 3: Active interpretation of wh-dependencies in the visual world 78
3.1 A critical review of Sussman and Sedivy (2003) 80
3.2 Experiment 3: Replication of Sussman and Sedivy (2003) 85
3.2.1 Method 85
3.2.2 Results 89
3.2.3 Discussion 93
3.3 Experiment 4: Two-event story design 94
3.3.1 Method 95
3.3.2 Results 99
3.3.3 Discussion 102
3.4 Conclusion 105
Chapter 4: Kindergarten-path effects in English and Japanese wh-processing 109
4.1 Verb bias and kindergarten-path effects 115
4.2 Experiment 5 (English question-after-story with verb manipulation) 124
4.2.1 Method 124
4.2.2 Results 130
4.2.3 Discussion 132

4.3 Experiment 6 (English truth value judgment study with adults) 134
4.3.1 Method 136
4.3.2 Results 137
4.3.3 Discussion 137
4.4 Experiment 7 (Japanese question-after-story, ambiguous sentence) 139
4.4.1 Method 139
4.4.2 Results 140
4.4.3 Discussion 141
4.5 Experiment 8 (Japanese question-after-story, error signal manipulation) 143
4.5.1 Method 144
4.5.2 Results 150
4.5.3 Discussion 151
4.6 General discussion 153
4.7 Conclusion 159
Chapter 5: Active processing of reconstruction and binding 160
5.1 Kindergarten-path effects in reflexive binding? 161
5.2 Experiment 9 (self-paced reading) 169
5.2.1 Method 169
5.2.2 Results 171
5.2.3 Discussion 175
5.3 Experiment 10 (eye-tracking during reading study) 181
5.3.1 Method 181
5.3.2 Results 187
5.3.3 Discussion 192
5.4 Conclusion 197
Chapter 6: Where did parents tell their children that wh-phrases should be attached? .. 198
6.1 Parsing and distributional learning 203
6.2 Analysis of wh-dependency distribution in child-directed speech 210
6.2.1 Corpus information and analysis procedure 210
6.2.2 Results 214
6.2.3 Discussion 223
6.3 Conclusion 229
Chapter 7: Conclusion 231
7.1 Summary of empirical findings and specific conclusions 232
7.2 Broader implications for psycholinguistic research 239
7.3 Broader implications for language acquisition research 241
7.3.1 Incremental processing and input distribution for long distance reflexives 241
7.3.2 Active processing may assist grammar acquisition 243
Appendix A: List of target items used in Experiment 1 245
Appendix B: List of target items used in Experiment 2 251
Appendix C: List of target items used in Experiment 3 257
Appendix D: List of target items used in Experiment 4 262
Appendix E: List of target items used in Experiment 5 263
Appendix F: List of target items used in Experiment 7 269
Appendix G: List of target items used in Experiment 8 271
Appendix H: List of target items for Experiments 9 and 10 274

Appendix I: Wh-dependencies in child directed speech..... 279
References..... 282

List of Tables

Table 1. Experiment 2: Participant Mean Reading Times in Milliseconds (Standard Error) and Percent Regressions 64
Table 2. Experiment 3: Mean fixation proportion across pariticipants in 100ms windows 1-4 within the verb region 92
Table 3. Experiment 4: Fixation proportions across subjects in 100ms windows 1-5 within the verb region 102
Table 4. Experiment 10: Participant Mean Reading Times in Milliseconds and Percent Regressions (Standard Error) 188
Table 5. Details of corpora used in the current study 210
Table 6. Distribution of argument wh-dependencies 215
Table 7. Distribution of adjunct wh-dependencies 215
Table 8. Distribution of main clause verb type in adjunct wh-dependencies 218
Table 9. Distribution of wh-phrase types and main verb types 220
Table 10. The input distribution and its interaction with the child grammar and parser 222

List of Figures

Figure 1. A sample scene used in Tanenhaus and colleagues' study (Tanenhaus et al., 1995). The left picture shows the one-referent context (i.e., only one apple in the scene), and the right picture illustrates the two-referent context (i.e., there are two apples in the scene). 7
Figure 2. Mean reading time (ms) for the transitive non-island and island conditions. Error bars indicate standard error of the mean. 49
Figure 3. Mean reading time (ms) for the intransitive non-island and island conditions. Error bars indicate standard error of the mean. 50
Figure 4. A sample display used in Sussman and Sedivy (2003); picture made by AO .. 81
Figure 5. Proportion of fixations to displayed items in the wh-condition 90
Figure 6. Proportion of fixations to displayed items in the yes/no-condition 90
Figure 7. Comparison of fixations to the target theme (spider). The dotted line indicates the onset of the verb. 91
Figure 8. A sample display used in Experiment 4 (the initial phase on the left, the final phase on the right). The pictures in this sample display are a cake (target theme), some dishes (competitor theme), a sponge (distractor), Emily (subject) and a fork (instrument). 98
Figure 9. Fixation proportion to displayed items in the wh-condition 99
Figure 10. Fixation proportion to displayed items in the yes/no-condition 100
Figure 11. Comparison of fixations to the target object ('the cake'). The vertical dotted line represents the verb onset. 100
Figure 12. A sample story sequence. 126
Figure 13. Mean proportion of main clause responses in Experiment 5. The error bars indicate standard errors of the mean. 131
Figure 14. Mean proportion of main clause responses in Japanese Question-after-Story experiments (Experiments 7 and 8). The error bars indicate standard errors of the mean. 141
Figure 15. Mean reading time (ms) for the wh-argument gender match and mismatch conditions. Error bars indicate standard error of the mean. 173
Figure 16. Mean reading time (ms) for the wh-predicate gender match and mismatch conditions. Error bars indicate standard error of the mean. 174

Chapter 1: Introduction

One of the central goals of psycholinguistics is to understand how sentence comprehension mechanisms rapidly recover syntactic structures from strings of words and assign interpretations to them. Adult language comprehension research to date has demonstrated that the sentence processing mechanism makes incremental hypotheses about possible syntactic structures and interpretations while drawing upon a variety of different sources of linguistic and non-linguistic information. This requires that hearers be able to flexibly retract such incremental commitments when the selected hypothesis is disconfirmed by additional information. Such a reanalysis processes could be extremely taxing for the comprehension system. In this sense, the ability to make incremental syntactic and interpretive commitments is a double-edged sword: incrementality allows efficient processing at the risk of incurring reanalysis costs.

This dissertation is concerned with the nature of both edges of that sword. It address the following question: During incremental sentence interpretation, how are multiple levels of linguistic representations computed, and what properties of incremental sentence interpretation processes give rise to differences in revision difficulties? Adult psycholinguistic studies have used a variety of real-time measures to document the presence of incremental interpretive processes and integration of multiple sources of information, but many of these studies focused on examining what kind of information can be used to resolve structural and referential ambiguities. The first goal of this dissertation is to provide an explicit procedural account of how multiple levels of representations are computed in the course of generating a sentence interpretation. To this

end, I use a variety of time course measures to examine computation of long distance dependencies in adult sentence processing. I will present cross-linguistic evidence that processing of long distance dependencies can be decomposed into a sequence of computations at distinct levels of representations. As such, this work illustrates the importance of cross-methodological and cross-linguistic investigations in understanding the real-time status of linguistic operations and representations.

The second goal of this dissertation is to understand the development of incremental sentence interpretation and reanalysis mechanisms. Given that children are more limited than adults in both their linguistic knowledge and their processing resources, one might imagine that children would be less likely to make risky incremental processing decisions in favor of reducing the risk of incurring reanalysis costs. However, the past decade of child sentence processing research has demonstrated that children are in fact eager to make incremental syntactic and interpretive commitments, and that when they encounter later information that disconfirms such incrementally constructed representations, children tend to perseverate and struggle to retreat from their earlier commitments. This difference in revision capacity between the child parser and adult parser raises an interesting ontogenetic question about how such properties mature in the child sentence processing mechanism. To address this question, this dissertation reports cross-linguistic experimental studies on English and Japanese wh-dependency processing in children. Using the explicit procedural account of filler-gap dependency processing that I develop based on adult sentence processing, I argue that actively constructed syntactic representations can be retracted when such representations are disconfirmed later by error signals attributed to the verb information.

1.1 Incremental sentence interpretation in adult language comprehension

In our daily lives, language comprehension occurs so fast and effortlessly that one might be led to think that it must be an extremely simple process. However, the complexity of comprehension processes becomes clear once we take into account the fact that language consists of many distinct levels of representations (e.g., acoustic, phonetic, phonemic, lexical, morphological, syntactic, semantic, prosodic, pragmatic and discourse representations), and that hearers must somehow map the language input to these different representational levels in order to achieve language comprehension. Here, we focus on the sentence interpretation mechanism, which minimally consists of *syntactic processes* and *interpretive processes*: The sentence interpretation mechanism must compute the most likely syntactic structures for the incoming string of words, and it must build thematic and semantic representations that can be integrated into the discourse representation, at which point hearers can use such information to establish reference to the actual world. There is still an on-going debate on how exactly these processes are implemented, but there is a broad consensus that these two processes must take place during sentence interpretation (Altmann & Steedman, 1988; Boland & Cutler, 1996; Crocker, 1996; Frazier, 1987; Kuperberg, 2007; Lewis & Vasishth, 2005; Tanenhaus & Trueswell, 1995; for a review, see van Gompel & Pickering, 2007). This set of procedures is summarized in (1).

(1) A (minimal) set of procedures involved in sentence interpretation

a. *Syntactic processes:* Generate and select structural hypotheses that are most likely to match the input.

b. *Interpretive processes:* Compute thematic relations and semantic representations, and integrate them into pragmatic and discourse representations.

Despite the complex set of operations and representations involved in the sentence interpretation procedures in (1), one of the most consistent findings in adult language comprehension research is that interpretation occurs rapidly as the sentence unfolds, in part because the parser incrementally computes structural representations and possible meanings of a sentence (e.g., Altmann & Kamide, 1999; Frazier & Rayner, 1982; Marslen-Wilson, 1973; Pickering, Traxler, & Crocker, 2000; Staub & Clifton, 2006; Sturt, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Incrementality is a design feature that could plausibly facilitate language comprehension in a variety of ways. For example, the ability to structurally organize the input in a word-by-word fashion provides the parser with a number of advantages. First, holding unstructured material in memory is costly for the memory system (e.g., Miller, 1956), and therefore the ability to incrementally build a unified structural representation should help to reduce the memory burden for the parser (Frazier & Fodor, 1978). Second, given the constraints on working memory capacity, it is computationally efficient to reduce the number of simultaneous operations (e.g. Just & Carpenter, 1992). If subsequent input is recognized while the previous input is still being processed, then the increased number of goal operations may create a larger burden on the comprehension mechanism. Third, incrementally hypothesized representations place tight constraints on what kind of subsequent input is possible, due to rules and constraints of the grammar. In other words, incrementality allows the comprehension mechanism to constrain the space of hypotheses

and predict what kind of input may be coming up later in the sentence. This in turn allows the hearer to efficiently integrate the bottom-up information into the previously built representation, regardless of noise in the environment.

1.1.1 Incremental sentence interpretation in the visual world

There has been a surge of interest in studies that investigate how eye gaze patterns on a visual scene reflect incremental sentence interpretation. This so called 'visual world' paradigm was originally devised by Cooper (1974), who demonstrated that as the sentence unfolds, listeners tend to look at pictures that correspond to the referring expressions they heard in the auditory presentation of sentences. This paradigm was revived and popularized by Tanenhaus and colleagues (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995, and their subsequent work). Studies that have used this method in sentence processing have mainly focused on how listeners incrementally establish reference during sentence interpretation while integrating linguistic and visual information in the scene.

For example, Altmann and Kamide (1999) investigated the incrementality of thematic assignment by testing how quickly the selectional restriction information of a verb can constrain visual attention to the referent of the potential direct object of the verb in the visual scene. This experiment presented an image of a scene with a boy, a cake, and a few toys on a computer display, while auditory stimuli like (2) were presented.

(2) a. The boy will *eat* the cake.

b. The boy will *move* the cake.

In (2a), the verb *eat* can take only one of the objects in the scene as its argument, namely the cake, unlike the verb *move* in (2b) which is compatible with any of the objects in the display. Altmann and Kamide found a higher probability of saccades towards the target object (i.e., the cake) before the onset of the object noun in the *eat* condition than in the *move* condition (.54 vs. .38), and crucially, this difference in the probability of saccades to the target object emerged even before the verb offset (.29 vs. .22), despite the fact that the verb duration was only 400ms. This can be taken to indicate that the verb's selectional information was accessed even during the presentation of the verb to predict the upcoming object NP, which resulted in *anticipatory looks* towards the most plausible object in the visual display (see also Altmann, 2004; Altmann & Kamide, 2007; Boland, 2005; Nation, Marshall, & Altmann, 2003).

The resolution of temporary PP attachment ambiguity is an example of incremental integration of multiple sources of information that has drawn much attention in both adult and child sentence processing research. For example, an eye-tracking study by Tanenhaus and colleagues (Tanenhaus et al., 1995) presented an act-out instruction in (3) to adults, together with an array of objects as shown in Figure 1.

(3) Put the apple on the towel in the box.

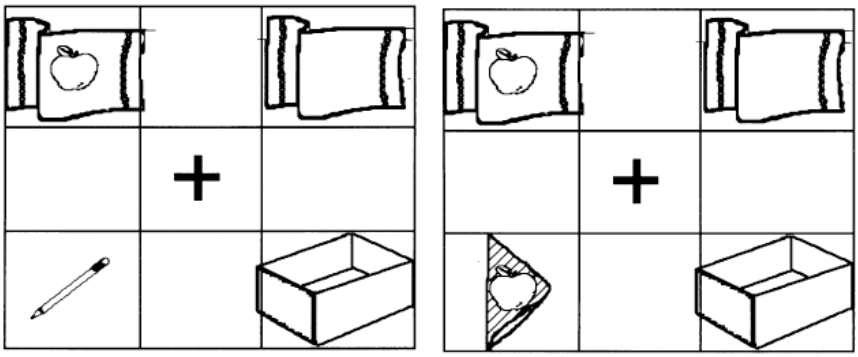


Figure 1. A sample scene used in Tanenhaus and colleagues' study (Tanenhaus et al., 1995). The left picture shows the one-referent context (i.e., only one apple in the scene), and the right picture illustrates the two-referent context (i.e., there are two apples in the scene).

In the sentence shown in (3), the first prepositional phrase (PP) *on the towel* is structurally ambiguous in that it can serve as a noun phrase (NP) modifier specifying the location of the NP *the apple*, or as a verb phrase (VP) argument specifying the destination for the *putting the apple* event. Here, the verb *put* can only accommodate one destination PP, and therefore the ultimate destination must be *in the box.*[1] However, at the point of hearing the ambiguous PP, the hearer does not have access to the disambiguation information from the second PP. In this sense, this sentence contains a temporary syntactic ambiguity at the point of processing the first PP, and one of two different

[1] Another possible interpretation is that *on the towel (that is) in the box* as a whole serves as the ultimate destination. However, given that there is no box that contains a towel in the display shown in Figure 1.1, the only felicitous interpretation is the one in which *in the box* is analyzed as the destination, and the first PP *on the towel* is analyzed as an NP modifier specifying where the apple is.

interpretations can arise, depending on which structural option is adopted.

Tanenhaus and colleagues used two dependent measures to investigate the sentence interpretation processes involved in PP attachment ambiguity resolution for sentences like (3): (a) measures of moment-by-moment eye fixations on the objects in the scene, which can be used to infer the time course of interpretive processes; and (b) act-out performance, which reflects the ultimate sentence interpretation that results from incremental syntactic ambiguity resolution. There were two main findings in this study. First, in a one-referent context where there was only one apple in the scene (the left side of Figure 1), adults initially looked toward the empty towel as soon as the first PP *on the towel* was presented, which suggests that this PP was immediately analyzed as a destination PP. However, in the two-referent context, in which one apple was on a towel and another apple was on a napkin (the right side of Figure 1), it is pragmatically more likely that the definite NP *the apple* should be followed by a modifier (Heim, 1982; Crain & Steedman, 1985). Here, the presentation of *on the towel* led adults to fixate on the apple that was on a towel instead of on the empty towel, suggesting that adults did not analyze the first PP *on the towel* as the destination PP but rather as an NP modifier. These findings corroborate the earlier observation that English-speaking adults have a preference to incrementally analyze the PP as an argument of a ditransitive verb like *put* (e.g., Britt, 1994; Carlson & Tanenhaus, 1988; Spivey-Knowlton & Sedivy, 1995), while the referential information and pragmatic inferences can be quickly integrated to override this VP attachment bias and select the NP-modifier analysis. This highlights the fact that the sentence comprehension mechanism uses multiple sources of information to maximize the chance of adopting the most likely interpretation given the utterance

situation (Altmann, 1998; Frazier, 1987; Gibson & Pearlmutter, 1998; Tanenhaus & Trueswell, 1995; for a summary, see Crocker, 1999; Kamide, 2008).

These visual world studies present strong evidence that syntactic and interpretive processes occur extremely quickly and involve the integration of multiple sources of information. This paradigm has an additional advantage, in that this is one of the few on-line measures of sentence comprehension that can be used with children (e.g., Nation, Marshall, & Altmann, 2003; Trueswell, Sekerina, Hill, & Logrip, 1999), as will be reviewed in more detail in Section 1.2. On the other hand, these studies do not reveal what kind of representations were computed or how they were computed during the course of generating an interpretation. There are two reasons for this. The first has to do with the nature of the dependent measure. Given that fixation patterns on a visual scene reflect how listeners establish reference to the world, the visual world eye-tracking measures reveal the *product* of the interpretive processes. Based on this type of measure, however, it is difficult to unpack the procedural details of the sentence interpretation processes. For example, it is not clear what kind of representation might have contributed to Altmann and Kamide's finding that the presentation of *the boy will eat* generates looks towards the *cake*. It could be that the comprehension mechanism generated a syntactic and thematic representation that allowed a prediction of the upcoming noun, or it could simply be that the lexical content of the verb *eat* increased the activation of [+edible] features and led to fixation on eating-related objects in the scene (for visual world studies on semantic associate activation in noun recognition, see Dahan & Tananhaus, 2005; Huettig & Altmann, 2005; Yee & Sedivy, 2006).

The ambiguity resolution study by Tanenhaus and colleagues (Tanenhaus et al.,

1995) demonstrates that referential information is rapidly integrated into sentence interpretation processes to select an analysis that is most congruent with the scene, but this finding does not reveal how exactly the referential information was used, or at which level of representation. For example, it could be the case that the parser generated both structural possibilities and used the referential information to select one, or it could be the case that the referential information directly influenced the structure generation process itself and only one structural hypothesis was ever considered by the parser. This suggests that time course measures that probe mental processes independent of reference to the world would be useful for shedding light on sub-processes that constitute sentence interpretation.

Another reason why the studies reviewed above shed little light on the real time computation of linguistic representations has to do with the nature of the linguistic constructions that were examined in many of these studies. In Altmann and Kamide's study, the main interest was in the extent to which information from the verb can guide subsequent referential processes. However, once the verb is accessed, many levels of linguistic representation can be computed instantaneously. Verbs contain not only semantic and conceptual features, but also rich argument structure information that directly influences the syntactic structures that can immediately follow. In Tanenhaus and colleagues' PP attachment study, at the point of processing the structurally ambiguous PP in *Put the apple on the towel*…, listeners can compute all the syntactic and interpretive processes that result in a complete proposition, which in turn can be integrated into the discourse/conceptual representation. Therefore, in order to fully identify the set of procedures and representations that constitute sentence interpretation, it is useful to

examine a linguistic environment in which different levels of representations can be computed at different points in time. As explained in the next section, investigations of long distance dependency processing are instrumental in addressing this question.

In summary, visual world experiments on sentence interpretation provide a useful testing ground for how listeners incrementally establish reference to the world; but, partly due to the nature of the measure itself, and the constructions that have been previously examined in this research tradition, the procedural details of linguistic computation have been partially masked in the existing literature. In the next section, I review reading research on long distance dependency processing that could complement the findings from visual world eye-tracking studies.

1.1.2 Active computation of long-distance dependencies

Grammars of human languages contain many syntactic rules and constraints that are used to relate two non-adjacent constituents of a sentence. These so-called long distance syntactic dependencies have received much attention in psycholinguistics research for various reasons, but one primary motivation is that long-distance dependency processing provides clear evidence for incremental syntactic processes. For example, in sentences that involve backward anaphora (4), the sentence processing mechanism must find an antecedent that can be co-referential with the pronoun in the fronted subordinate clause.

(4) When <u>he</u> was fed up, the {boy | girl} visited …

Here, there are many candidate antecedents for this pronoun: the antecedent could be the subject position of the main clause *the boy/girl*, but it could also be a noun in the later part of the sentence, or it could be that the antecedent is in a previous utterance. However, in the absence of a preceding utterance, the main clause subject position is the first position that could present a potential antecedent. Reading studies that have investigated the processing of sentences like (4) have reported evidence for reading disruption when the main clause subject mis-matched the pronoun in gender (*the girl*) compared to when the subject and the pronoun matched in gender (*the boy*) (van Gompel & Liversedge, 2003; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007; for a related finding in Japanese, see also Aoshima, Yoshida, & Phillips, 2009). This so-called gender mismatch effect suggests that the sentence processing mechanism attempts to incrementally associate the pronoun with the first potential antecedent position, even though this position may not provide the correct antecedent. This type of incremental syntactic commitment that the parser makes without waiting for further bottom-up information is called *active dependency completion.* It is also worth noting that the parser could have first accessed the lexical content of this subject position to check the gender property before initiating the association process. But the fact that the reading disruption was nevertheless observed suggests that incremental association between the pronoun and the subject NP position occurs prior to accessing the lexical features of the subject NP *the boy/girl.*

A leading source of evidence for active dependency completion in long-distance dependency processing comes from studies of 'filler-gap' dependencies of the kind shown in (5). Here, the object NP *the city* (called the *filler*) is dislocated from the post-

verbal thematic position (called the *gap*), and the parser needs to associate the filler and the gap in order to assign a thematic interpretation.

(5) The city that the author visited ____ was named for an explorer.

Past research on filler-gap dependency processing has established that the parser actively creates a gap before there is sufficient bottom-up evidence to confirm that particular gap location (*Active gap filling:* Crain & Fodor, 1985; Fodor, 1978; Frazier & Flores D'Arcais, 1989). For example, Stowe (1986) observed a *filled gap effect* in (6), i.e., slower reading times at the direct object position *us* in the wh-fronting condition (6a) than in a control condition that did not involve wh-fronting (6b). This pattern of reading time data suggests that the parser had already posited the object gap before checking whether the direct object position was occupied.

(6) a. My brother wanted to know who Ruth will bring us home to ____ at Christmas.

b. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

Converging evidence comes from an eye-tracking experiment by Traxler and Pickering (1996), who manipulated the semantic fit between the filler and the potential verb host, as in (7).

(7) We like the city / book that the author *wrote* unceasingly and with great dedication about _____ while waiting for a contract.

Traxler and Pickering found a plausibility mismatch effect at the critical verb in (7), i.e., the eye gaze duration at the optionally transitive verb *wrote* increased when the filler was an implausible object of the verb (i.e., *the city*), compared to when the filler was a plausible object of the verb (i.e., *the book*). This suggests that at least as early as the verb position, the parser postulates a gap and analyzes the filler as the object of the verb.

The evidence for active gap creation may appear to make the same point as the visual world study by Altmann and Kamide (1999) discussed above, but there is one critical difference: as discussed above, it is unclear what source of information led to Altmann and Kamide's evidence for the direct object recognition, as it could reflect lexical feature activation rather than a hypothesis about the structure of the utterance. However, the plausibility mismatch effect clearly indicates the presence of syntactic integration of the filler and subsequent semantic composition, which is a pre-requisite for the detection of the semantic misfit between the filler and the verb. There is ample time course evidence for active object gap creation in many languages, using a variety of dependent measures such as reading time and gaze duration measures (Crain & Fodor, 1985; Frazier, 1987; Frazier & Clifton, 1989; Phillips, 2006; Pickering & Traxler, 2001, 2003; Wagers & Phillips, 2009), cross-modal priming (Nicol, 1993; Nicol & Swinney, 1989), as well as event-related potentials (Garnsey, Tanenhaus, & Chapman, 1989; Gouvea, Phillips, Kazanina, & Poeppel, 2009; Hestvik, Maxfield, Schwartz, & Shafer, 2007; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005).

In these ways, reading research on active dependency completion can shed light on how the syntactic representation of long distance dependencies can be incrementally constructed. On the other hand, the time course evidence for structure building processes

does not illustrate how exactly the syntactic representations interact with other levels of linguistic representation to generate the ultimate interpretation. In this regard, the visual world paradigm is useful in providing time course evidence for how sentence interpretation unfolds in time. As a first step to explicitly illustrating the entire process of linguistic computation during sentence interpretation, I will use both reading methods and the visual world paradigm to investigate how different levels of representations are computed to arrive at an interpretation of long distance dependencies. I focus on the computation of filler-gap dependencies, presenting time course evidence that suggests that in filler-gap dependency computation, syntactic processes can take place before interpretive processes. This process is summarized in (8).

(8) a. Actively construct the syntactic dependency between the filler and the gap.

b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

The English reading studies reported in Chapter 2 use insights from filler-gap dependency processing in head-final languages (e.g., Japanese, in Aoshima, Phillips, & Weinberg, 2004), presenting time course evidence that lends support to the separation of syntactic (8a) and semantic processes (8b) in filler-gap dependency completion. Chapter 3 reports visual world experiments on filler-gap dependency completion. Taking together the time course data in reading (Chapter 2) and referential processes (Chapter 3), I will argue that the syntactic processes are completed before the interpretive processes (8b) in real time computation of filler-gap dependencies. These findings provide the starting

point for subsequent studies on corresponding processes in children.

1.2 Incremental sentence interpretation and revision difficulties in children

1.2.1 Ambiguity resolution and kindergarten-path effects

While adult psycholinguistic studies in the past decades have extensively investigated incremental properties of the sentence comprehension mechanism, the developmental research on incremental sentence processing capacity in children is relatively new. As discussed in section 1.1.1, the incremental property of the parser can help to reduce the burden on short term memory and maximize computational efficiency, and for this reason it may be seen as a design feature that is especially desirable for children, since many aspects of their working memory capacity are limited compared to that of adults (Diamond, 2006; Gathercole, Pickering, Ambridge, & Wearing, 2004). However, incrementality in sentence processing is risky in that early structural analyses and the resulting interpretations could turn out to be incorrect and need to be retracted, as discussed above in the PP attachment study by Tanenhaus and colleagues (1995).

A seminal study by Trueswell, Sekerina, Hill and Logrip (1999) extended the experiment by Tanenhaus and colleagues to children to investigate whether children can incrementally build structures and interpretations by integrating multiple sources of information like adults do. An example of temporary PP attachment ambiguity that is relevant to their study is repeated here (3).

(3) Put the apple on the towel in the box.

The findings by Trueswell and colleagues (1999) were rather striking: Although children

showed the same general preference as adults to incrementally analyze and interpret the first PP as a destination, children behaved differently from adults in two ways. First, the VP-attachment preference continued to be observed even when additional referential information was provided in the two-referent context. Second, more strikingly, children were unable to revise this initial analysis even after the second PP was presented, and preserved the destination interpretation of the first PP (see also Weighall, 2008). The observation that children perseverate and fail to revise their initial interpretation has been dubbed the 'kindergarten-path effect.'

Subsequent studies investigated whether children's reanalysis ability could be enhanced by manipulating various linguistic and non-linguistic cues that had proven to be useful for adults. The manipulation was realized in a variety of ways, such as by drawing explicit attention to the NP modifier analysis via elicited production (Hurewitz, Brown-Schmidt, Thorpe, Gleitman, & Trueswell, 2000), or by withholding the visual context until after verbal instructions are provided (Weighall, 2008; cf. Meroni & Crain, 2003). However, the original kindergarten-path effect was consistently replicated in all these studies. To the extent that measures of eye movement and act-out performance in these studies reflect sentence interpretation, a generalization that can be drawn from these findings is that the first interpretation that children arrive at tends to be the only interpretation that they can entertain. (For related findings in scope and binding ambiguities, see Leddon & Lidz, 2006; Musolino & Lidz, 2006; cf. Gualmini, 2004).

This thesis attempts to better understand the nature of kindergarten-path phenomena, as they present interesting questions for developmental cognitive science and language acquisition. First, understanding the nature of kindergarten-path effects could

help us understand the nature of developing domain-general cognitive architectures, such as the executive function mechanism. Executive function refers to a set of cognitive processes that are necessary for inhibiting automatic responses to stimuli and controlling adaptive and efficient responses to novel or difficult situations. Based on observations that executive function mechanisms do not fully develop until late adolescence (Diamond, 2002; Davidson, Amso, Anderson, & Diamond, 2006), it has been proposed that the pervasiveness of kindergarten-path effects could be attributed to immature executive function mechanisms (Choi & Trueswell, 2010; Novick, Trueswell, & Thompson-Schill, 2005). If this is true, then understanding the nature of kindergarten-path effects could shed further light on the nature and development of the executive function mechanism itself (for discussion, see Mazuka, Jincho, & Onishi, 2009).

Second, the lack of flexibility and revision capacity could potentially interfere with children's acquisition of new grammatical rules. In order to correctly acquire the target grammar that generates the input in the environment, children must be able to assign a target-like syntactic representation to the input. However, if children's sentence processing mechanisms operate according to their developing, non-adult-like grammars, a correct linguistic representation may not be assigned to the input (Valian, 1990). This suggests that it is important for children to have some flexible ability to revise their initial structural analyses in order to correct their parse and acquire the target-like grammar.

Third, even if children can sometimes retract their incremental commitments, the frequency with which such mis-analyses occur could have a significant consequence for language learning processes, because acquisition of the target grammar must be sensitive to the distributional features of the input at least to some extent. For example, one may

suppose that the null subject parameter (i.e., whether a language allows a null subject in a finite clause; Chomsky, 1981; Rizzi, 1982) could be learned by a single observation of an utterance with a null subject in a finite clause. However, if language learning proceeded in such a deterministic way, then English-speaking children might have difficulties setting the null subject parameter properly (i.e., the negative setting, because English disallows null subjects). One reason is that English-speaking adults' utterances sometimes include sentences without an overt subject (e.g., *Smells like dinner is ready)*; and it is also possible that the noise in the environment may simply prevent children from hearing the overt subject. On the other hand, if we suppose that children learn the null subject parameter based on the distribution of overt subjects in finite clauses, then the large majority of English sentences will have an overt subject and children can correctly acquire the target grammar (for more discussions of difficulties associated with deterministic, 'trigger' based learning models, see Gibson & Wexler, 1994; Fodor, 1998; Yang, 2002, 2004). However, if children's parsers have strong biases in incremental syntactic analyses and fail to retract such incremental commitments, then this raises the possibility that the input distribution might be skewed and may not be 'correctly' represented in the child's mind. In the last two decades, there has been a surge of interest in experimental and computational modeling research on the role of input distribution in language acquisition (Gathercole & Hoff, 2007; Gerken, 2006; Gomez, 2002; Hudson-Kam & Newport, 2006; Miller & Schmitt, 2006; Pearl & Lidz, 2009; Rowland & Pine, 2000; Xu & Tenenbaum, 2007; Thompson & Newport, 2007; Tomasello, 2003; Yang, 2002, 2004), but many of these studies have ignored the fact that children's perceptual mechanisms are biased in many ways and that they show non-adult-like properties. For

these reasons, understanding how children's (in)ability to recover from mis-analyses affects the input distribution should shed new light on the interaction between the input distribution and language acquisition processes.

1.2.2 Incremental structural misanalysis and interpretation persistence

Child sentence processing research so far has focused on a small number of ambiguity resolution contexts, and consequently relatively little research has investigated what kind of sentential environments and their associated syntactic and interpretive procedures give rise to kindergarten-path effects. In order to further our understanding of the child reanalysis mechanism, this dissertation examines the hypothesis that children's revision difficulties may be confined to environments in which (a) children's syntactic and interpretive commitments were completed and confirmed by bottom-up evidence, and (b) error signals are not based on information from the verb.

This hypothesis is motivated by an observation that even adults tend to have revision difficulties in contexts where a temporary misanalysis causes comprehenders to make interpretive commitments that are later disconfirmed. For example, Christianson, Hollingworth, Halliwell, and Ferreira (2001) presented comprehension questions after garden-path sentences like (9), in which the verb *dressed* is used as a reflexive verb but the subject of the second clause *the baby* can be temporarily misanalyzed as its direct object.

(9) While Anna dressed the baby spit up on the bed.

The off-line comprehension measures revealed that adults often retained the thematic

interpretation that Anna was dressing the baby. This suggests that the interpretive commitment that resulted from the temporary syntactic analysis persisted in their ultimate interpretation of the sentence, even after this temporary syntactic analysis is ultimately revised (for a review of these findings, see Ferreira & Patson, 2007). An eye-tracking during reading study by Sturt (2007) presented time course evidence for an interpretive persistence effect in sentences like (10), which contains a temporary structural ambiguity in that *the South Pole* can be temporarily analyzed as the direct object of the main clause verb *found* or the subject of the embedded clause.

(10) a. The explorers found (that) the South Pole was actually right at their feet.

b. The explorers found (that) the South Pole was actually impossible to reach.

In sentences like (10), when the overt complementizer *that* is absent, readers generally prefer to analyze the structurally ambiguous NP *the South Pole* as the direct object of the main clause verb *found* (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Pickering, Traxler, & Crocker, 2000; Rayner & Frazier, 1987). The study manipulated the presence or absence of *that* to examine the effect of temporary structural ambiguity. Moreover, in order to probe a potential interpretive persistence effect resulting from the temporary misanalysis, the study also manipulated the embedded clause content to be consistent or inconsistent with the direct object analysis of *the South Pole.* In (10a), for example, finding that the South Pole was at their feet entails finding the South Pole, and for this reason the thematic interpretation that results from the temporary direct object analysis is consistent with the ultimate interpretation. In (10b), on the other hand, finding that the

South Pole could not be reached strongly indicates that the explorers did not find the South Pole, and therefore the ultimate interpretation contradicts the temporary thematic interpretation that results from the direct object analysis of *the South Pole*.

The eye-movement measures revealed that in first-pass reading processes, readers spent more time reading the disambiguating region *was actually* in the temporarily ambiguous condition without the overt complementizer *that*, demonstrating that the readers initially analyzed *the South Pole* as the direct object of *found*. Moreover, after they had already started reading subsequent regions, readers also spent more time re-reading the disambiguating region *was actually* in the ambiguous inconsistent condition (10b) than in the ambiguous consistent condition (10a). This suggests that the readers had difficulties retracting the initial interpretive commitment, and thus struggled to resolve the conflict between the initial interpretation (e.g., the explorers found the South Pole) and the ultimate interpretation (e.g., the South Pole was unreachable).

What these studies highlight is that interpretive commitments based on temporary syntactic mis-analyses of the input are difficult to retract. This is a feature that is shared by sentences like (3) which demonstrated the original kindergarten-path effects (Trueswell et al., 1999).

(3) Put the apple on the towel in the box.

Here, as soon as the first PP *on the towel* is processed, the argument structure of the ditransitive verb *put* can be completed. In other words, before encountering the second PP *in the box*, children can construct complete syntactic, semantic and discourse

representations that are supported by the bottom-up input. Thus, the fact that the interpretive commitments constructed from *Put the apple on the towel* cannot be retracted could reflect the faithfulness to the interpretative commitments that are confirmed by bottom-up information.

Given these observations, kindergarten-path effects may be generalized as a constraint to preserve interpretive commitments that are confirmed by bottom-up information. This approach to kindergarten-path effects raises a new question about children's reanalysis mechanism: What about cases in which syntactic commitments are made without relying on critical bottom-up information? Recall from the discussion in Section 1.1.2 that in processing of long distance dependencies, the parser actively hypothesizes syntactic structures in advance of bottom-up lexical information that becomes available at the tail of the dependency. However, it is also important to note that such active commitments may be disconfirmed by late arriving bottom-up information. If children's revision failures reflect general problems with retracting early commitments, then children are expected to demonstrate kindergarten-path effects in long distance dependency processing as well. On the other hand, if kindergarten-path effects can be re-interpreted as a reflection of faithfulness to bottom-up information, then it is reasonable to expect that active commitments that were made without bottom-up evidence can be easily retracted in favor of analyses that could be constructed by late arriving bottom-up information.

1.2.3 Quality of error signals and kindergarten-path effects

Another important factor that contributes to the severity of reanalysis difficulties is the quality of error signals. For example, Fodor and Inoue (1994) have observed that in

temporary garden-path sentences like (11), the intuitive reanalysis difficulty in (11a) appears to be larger than in (11b).

(11) a. Susan put the book that she'd been reading all afternoon in the library.

b. Susan put the book that she'd been reading all afternoon into her briefcase.

Fodor and Inoue explains this contrast as follows: In (11a), readers generally consider *in the library* as the location in which the reading event happened, but this analysis would render the sentence syntactically ill-formed because no destination PP has been assigned to the verb *put*. In (11b), on the other hand, the PP *into her briefcase* cannot be used as a modifier of the reading event, and therefore it is easier to associate this PP with the verb *put*. This observation highlights that an early indication of the local semantic incongruity in (11b) can facilitate the subsequent revision process.

So far, the studies that have documented children's resistance to reanalysis despite the presence of overt error signals have mostly focused on PP attachment ambiguity resolution in sentences like *Put the frog on the napkin in the box*. Here, the critical error signal is a syntactic one: If the first PP *on the napkin* is incrementally analyzed as the destination for the putting event and remains so despite the second PP *in the box,* then the second PP cannot be integrated into the structure. Trueswell and colleagues' study demonstrates that this syntactic error signal does not serve as an effective revision cue for children, but no study has investigated whether children's revision success would vary depending on the source of error signals. This dissertation presents the first attempt to address this issue by comparing the effectiveness of syntactic error signals and semantic

error signals that are based on verb information. It has been shown in ambiguity resolution contexts that children are extremely sensitive to verb information (e.g., Snedeker & Trueswell, 2004), and therefore it is feasible that children are sensitive to verb information in making syntactic reanalysis decisions as well.

In summary, this dissertation examines children's reanalysis abilities in the context of two distinct kinds of incremental commitments: incremental commitments that are based on bottom-up information that has become available in the input (as in previous PP attachment ambiguity resolution studies), and *active* incremental commitments that are not yet supported by bottom-up information. More specifically, I investigate whether (a) children process filler-gap dependencies actively like adults do, and (b) if so, whether children can recover from active syntactic and interpretive commitments. Chapter 4 examines a cross-linguistic comparison of wh-dependency interpretation in children. The studies in this chapter capitalize on the difference in verb order between English and Japanese to demonstrate that children actively complete wh-dependencies. Using this property, the experiments examine the quality of various error signals when the incremental syntactic analysis of wh-dependencies is disconfirmed by additional information. In Chapter 5, we extend this investigation to global ambiguities in binding of reflexive pronouns, for which children show a selective interpretation preference (Leddon & Lidz, 2006). This chapter reports experiments with adults and attempts to present time course evidence that sheds light on the active antecedent search and recovery processes. The results from these chapters demonstrate that children can retract active syntactic commitments based on later arriving bottom-up information so long as the verb-based interpretation of earlier constituents can be preserved.

1.3 Outline of the dissertation

The rest of the dissertation is organized as follows. Chapters 2 and 3 present a series of experiments that establish an explicit procedural model of incremental syntactic and interpretive processes in filler-gap dependency computation. Chapter 2 focuses on English-speaking adults' wh-dependency processing in reading sentences with wh-argument fronting, and argues that syntactic processes of wh-dependency completion proceed actively without waiting for lexical information from the tail of the dependency. There is existing evidence that suggests that an English subject gap can be created actively without relying on bottom-up lexical information (Lee, 2004), but it is still unclear whether the parser consults other types of lexical information, such as verb transitivity information, when it actively creates an object gap. Previous adult psycholinguistic studies have demonstrated that speakers of verb-final languages are able to create an object gap and complete wh-dependencies without having access to verb transitivity information. However, this may only reflect a language-specific adaptation to the demands of verb-finality, as some existing work suggests that the English parser creates an object gap only after checking verb properties to confirm that the verb can host an object position (Staub, 2007). In two adult reading experiments (self-paced and eye-tracking) that manipulate verb transitivity, I report evidence for reading disruption when the verb is intransitive, while no such reading difficulty arises when the critical verb is embedded inside a syntactic island which blocks dependency completion (Ross, 1967). These results show that in both English and verb-final languages, the parser can complete the syntactic process of object gap creation without relying on lexical information from the tail of the dependency. Given that interpretive processes cannot be completed until

the readers and listeners have access to lexical information, this evidence establishes that syntactic processes can be completed before interpretive processes in the real time computation of filler-gap dependency processing.

Chapter 3 focuses on the interpretive processes associated with wh-dependency processing, and I present time course evidence that the interpretive processes in filler-gap dependency computation take place after syntactic processes are completed. This chapter presents a visual world eye-tracking experiment design that can be used to investigate real time measures of wh-dependency interpretation in adults. The visual world paradigm allows the researcher to measure moment-by-moment fixations on visual stimuli during presentation of spoken sentences, and can be used to provide a more precise time course estimate of how quickly listeners integrate the semantic and discourse representation to establish reference to the world (Altmann & Kamide, 1999; Cooper, 1974; Tanenhaus, et al., 1995). Sussman and Sedivy (2003) conducted a visual world experiment using *wh*-questions like *What did Jody squash the spider with?* and reported anticipatory looks towards the direct object referent (i.e., spider) during the presentation of the verb, which they interpreted as reflecting active dependency completion. However, a careful analysis of the time course of fixations in this study, and consideration of the linking hypothesis between dependency formation and eye movement, casts doubt on this conclusion. The adult eye-tracking experiments reported in this chapter provide a more precise time course estimate for filler-gap dependency processing and subsequent interpretive processes. Taking together the findings in Chapter 2 and the visual world results in Chapter 3, I argue that the syntactic process of object gap creation occurs before the lexical information from the verb is fully accessed, and that subsequent interpretive

processes associating the wh-phrase and the verb can be completed within around 600ms after the verb is presented, at least in the types of constrained environments used in visual world studies.

Chapters 4 and 5 attempt to uncover the details of children's reanalysis mechanisms by exploring the hypothesis that children can retract active commitments that were not supported by bottom-up evidence. Chapter 4 focuses on active dependency completion and reanalysis difficulties in the processing of bi-clausal wh-questions like (12) in English and Japanese.

(12) Where did Emily tell someone that she hurt herself?

The experiments reported in this chapter use a series of story-based off-line comprehension studies that examine how children and adults ultimately interpret wh-questions like (12) in English and Japanese. The results show that children (a) preferentially associate the fronted wh-phrase with the first verb phrase in the sentence, and (b) fail to revise this incremental syntactic commitment when the revision cue is a syntactic error signal; but (c) children successfully entertain an alternative interpretation when the first verb in the sentence is semantically or pragmatically incompatible with the wh-phrase. This constitutes the first demonstration that semantic and pragmatic information from the verb has a privileged status in leading children to retract their active syntactic commitments.

Chapter 5 extends our investigation of reanalysis difficulties to the incremental processing of binding relations. Using ambiguous sentences like (13a), Leddon and Lidz

(2006) demonstrated that 4-year old children are able to accept the reading in which *herself* is bound by *Ms. Cruella* (the 'surface' interpretation), but fail to accept the reading in which *herself* is bound by *Janie* (the 'reconstruction' interpretation). In (13b), on the other hand, children only accepted the reconstruction interpretation, which is the only grammatically permissible interpretive option.

(13) a. Miss Cruella$_1$ knew which painting of herself$_{1/2}$ Janie$_2$ put up ____ .

b. Mr. Monkey$_1$ figured out how proud of himself$_{*1/2}$ Andy$_2$ was ____ .

Leddon and Lidz attributed their finding to reanalysis difficulties: for (13a), children incrementally adopt the surface interpretation that becomes available first in the sentence, and fail to consider the other grammatically permissible reconstruction interpretation that becomes available later in the sentence. In this sense, this study potentially presents further evidence for interpretative persistence. However, the Truth Value Judgment Study by Leddon and Lidz did not present time course data, and therefore it is unclear whether the surface interpretation preference can indeed be characterized as an incremental interpretive commitment. This chapter reports time course evidence from self-paced reading and eye-tracking experiments with adults that lends indirect support to the kindergarten-path account for (13a) proposed in Leddon and Lidz (2006). For sentences like (13b), however, I suggest that the parser actively establishes a binding relation between the reflexive and the main clause subject, but that children successfully recover from this active commitment based on the bottom-up error signal in the embedded clause.

Chapter 6 examines potential consequences of children's revision difficulties for language acquisition processes. Grammatical features that vary cross-linguistically must be acquired based on the input, and for that reason the input must contain sufficient evidence for language-specific grammatical rules and constructions. However, the actual distribution of certain constructions may not be properly represented in a child's mind if she consistently assigns a structure based on the parser's biases and fail to reanalyze it when necessary. For example, the studies reported in Chapter 4 show that English-speaking children are strongly biased towards the main clause interpretation in processing a globally ambiguous wh-question like (14).

(14) Where did Emily tell someone that she hurt herself?

However, the Russian counterpart of (14) is unambiguous (15), in that only the main clause interpretation is grammatically permissible (Stepanov, 2000; Stepanov & Stateva, 2006).

(15) Gdje Emili skazala komu-to chto ona ushiblas'?
Where Emily said some-person that she hurt+REFL

Here, English-learning children must detect the availability of the embedded clause interpretation in (15) in order to distinguish their target grammar from that of Russian, and it is reasonable to expect that the actual English input contains sufficient evidence for the availability of the embedded clause interpretations. However, given that children are

generally biased towards the main clause analysis in (14), the distribution of the actual input may be skewed in such a way that the data demonstrating the embedded clause interpretation in the effective input (so called 'intake': Corder, 1967; Fodor, 1998) could seem rarer than they actually are.

In order to examine the nature of this 'skewed-input' problem for language acquisition, Chapter 6 presents an investigation of the distribution of English wh-questions like (14) in child-directed speech from CHILDES (MacWhinney, 2000). The corpus analysis shows that the active dependency completion biases in children could potentially skew the input distribution to the extent that the input may not clearly indicate the availability of the embedded clause interpretation, but that children's ability to use verb information to retract such active dependency completion commitments would help to create clear evidence for the English-type long distance dependency formation rule. These results illustrate that the input distributional pattern could vary depending on properties of the children's perceptual mechanisms, and that it is important to consider the role of children's perceptual mechanisms in understanding the role of input distribution in language acquisition.

Chapter 7 reviews the overall findings in this dissertation regarding procedural models of linguistic computation as well as reanalysis processes in adult and child sentence processing mechanisms. I discuss the broader implications of the procedural account of long distance dependency computation and reanalysis mechanism for psycholinguistic theories, as well as the implications of children's active processing and reanalysis mechanisms for first language acquisition research.

Chapter 2: Syntactic processes of verb-independent gap creation

As discussed in Chapter 1, there has been abundant psycholinguistic evidence that as a sentence unfolds, complex processes of sentence interpretation take place rapidly as the listener/reader integrates multiple sources of information. On the other hand, less progress has been made toward understanding the explicit procedures for computing the multitude of linguistic representations that are involved in sentence interpretation. Chapters 2 and 3 take filler-gap dependency computation as a case study to unpack the procedural details of linguistic computation during sentence interpretation. Specifically, taking together the time course evidence for syntactic processes in reading studies (Chapter 2) and visual world eye-tracking studies (Chapter 3), I will argue that interpretation of filler-gap dependencies consists of the following sequence of distinct operations (1).

(1) a. Actively construct the syntactic dependency between the filler and the gap.
b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

This chapter focuses on the processes described in (1a), namely whether the syntactic representation of filler-gap dependency can be constructed independent of the lexical information that becomes available at the tail of the dependency. As the review in the next section illustrates, while the separation of the two operations is not clearly supported by previous work on verb-medial languages like English, there is evidence for the

separation in processing of filler-gap dependencies in head-final languages. The reading experiments reported in this chapter examine whether English object gap creation is conditioned by verb transitivity information, and demonstrate that gap creation processes take place regardless of whether the verb can syntactically host an object NP or not.

2.1 What triggers active gap creation?

As reviewed in Chapter 1, past research on filler-gap dependency processing has established that the parser postulates a gap before there is sufficient bottom-up evidence to validate that analysis (*Active gap filling:* Crain & Fodor, 1985; Fodor, 1978; Frazier & Flores D'Arcais, 1989). For example, Stowe (1986) observed a *Filled gap effect* in (2), i.e., slower reading times at the direct object position *us* in the wh-fronting condition (2a) than in a control condition that did not involve wh-fronting (2b). This pattern of reading time data suggests that the parser had already posited a gap before checking whether the direct object position is occupied.

(2) a. My brother wanted to know who Ruth will bring us home to ____ at Christmas.

b. My brother wanted to know if Ruth will bring us home to Mom at Christmas.

Converging evidence comes from an eye-tracking experiment by Traxler and Pickering (1996), who manipulated the semantic fit between the filler and the potential verb host, as in (3).

(3) We like the city / book that the author *wrote* unceasingly and with great dedication about _____ while waiting for a contract.

Traxler and Pickering found a plausibility mismatch effect at the critical verb in (3), i.e., the first fixation duration and first pass time measures at the optionally transitive verb *wrote* increased when the filler was an implausible object of the verb (i.e., *the city*), compared to when the filler was a plausible object of the verb (i.e., *the book*). This suggests that at least as early as the verb position, the parser postulates a gap and analyzes the filler as the object of the verb. In fact, there is ample time course evidence for active object gap creation in many languages, using a variety of dependent measures such as reading time and gaze duration measures (Crain & Fodor, 1985; Frazier, 1987; Frazier & Clifton, 1989; Phillips, 2006; Pickering & Traxler, 2001, 2003; Wagers & Phillips, 2009), cross-modal priming (Nicol, 1993; Nicol & Swinney, 1989), visual world eye-tracking (Sussman & Sedivy, 2003) as well as event-related potentials (Garnsey, Tanenhaus, & Chapman, 1989; Gouvea, Phillips, Kazanina, & Poeppel, 2009; Kaan, Harris, Gibson, & Holcomb, 2000; Phillips, Kazanina, & Abada, 2005).

The work mentioned above may suggest that filler-gap dependency completion in an object position is triggered only after the parser gains access to the verb and confirms the subcategorizaiton and thematic properties of the verb. This is compatible with a lexical/verb-driven view of active gap creation, which attributes active gap completion to the parser's motivation to assign thematic interpretations to structurally unintegrated constituents as soon as possible. This approach is seen, for example, in a head-driven parsing mechanism proposed by Pritchett (1992), in which the thematic requirement of the verb plays a critical role in driving syntactic structure building processes. If in fact filler-gap dependency completion is driven by verb information, then it also follows that lexical access to the content of the verb must precede the completion of the syntactic

dependency. On this view, as soon as the syntactic dependency is completed, thematic assignment is also expected to take place instantaneously, given that the completion of thematic assignment process is what triggers syntactic dependency completion.

On the other hand, existing evidence discussed above is also compatible with the view that the parser is prioritizing completion of the syntactic dependencies rather than satisfying thematic requirements of the verb, as originally proposed in the Active Filler Hypothesis of Frazier and Clifton (1989). The motivation for early structural integration could come from various sources. For example, maintaining the filler in memory while it is structurally unintegrated has been shown to impose a burden on the working memory system (Chen, Gibson, & Wolf, 2005; Gibson, 1998, 2000; Gordon, Hendrick, & Levine, 2002). Moreover, it is possible that the active gap creation processes are indeed driven by the motivation to assign a thematic interpretation to the filler as soon as possible, rather than satisfying the thematic requirement of the verb (Aoshima, Phillips, & Weinberg, 2004; Pickering & Barry, 1991). On this view, the parser should prioritize integrating the filler into the first grammatically permissible structural position that can potentially receive a thematic role. Here, the motivation for active gap creation is thematic assignment, but crucially the gap creation process is guided by syntactic constraints of the grammar.

Research on subject gap creation in English as well as studies that investigated object gap creation in verb-final languages have presented evidence that active dependency completion does not depend on the lexical information of the tail of a dependency. For example, Lee (2004) used sentences like (4) to attest a filled gap effect in the subject NP position.

(4) a. That is the laboratory which, on two different occasions, Irene used a courier to deliver the samples to ___.

b. That is the laboratory to which, on two different occasions, Irene used a courier to deliver the samples ___.

Here, the content of the wh-filler is manipulated in such a way that the wh-filler can plausibly be a subject (4a) or not (4b). The results showed a longer reading time at the subject NP *Irene* in (4a) than in (4b), suggesting that the parser had postulated the subject gap before encountering the actual subject NP. However, it is not surprising that the parser actively creates a subject gap without having access to verb information, given that a subject is present in any sentence, regardless of verb properties. In this sense, if verb information was to play a role in the parser's attempt to create a gap, the critical empirical evidence should come from object gap creation, where the presence or absence of an object gap position critically relies on properties of the verb.

Evidence for pre-verbal object gap creation has been reported for verb-final languages like Japanese in which the object gap position linearly precedes the verb. For example, Aoshima and colleagues examined processing of long-distance scrambling sentences in which an embedded dative object NP was dislocated to sentence initial position, and found a filled gap effect at a pre-verbal dative object position (Aoshima et al., 2004). Using similar sentences, Nakano and colleagues reported evidence for an antecedent priming effect for the scrambled NP at a pre-verbal gap position (Nakano et al, 2002). These data indicate that the parser can in principle complete filler-gap dependencies before accessing verb information.

In verb-medial languages, on the other hand, no such evidence for pre-verbal object gap creation has been reported to date. The absence of evidence for preverbal gap creation could be explained by two factors. First, this may reflect a real architectural difference between languages in processing strategy. As mentioned above, maintaining the filler in memory while it is structurally unintegrated or when it has not received a thematic interpretation has been argued to impose a burden on working memory (Gibson, 1998; Gordon, Hendrick, & Levine, 2002, Haarman & Cameron, 2005; King & Just, 1991). Pre-verbal object gap creation in verb-final languages may thus reflect the parser's adaptation to the demands of processing these languages. Alternatively, the parser may be architecturally constrained to assign a thematic interpretation to the filler as soon as possible (Aoshima, Phillips, & Weinberg, 2004; Pickering & Barry, 1991). On this view, the parser should prioritize integrating the filler into the first grammatically permissible structural position that can receive a thematic role. Given that filler-gap dependencies are potentially unbounded, waiting for the verb before constructing the ultimate object gap position could impose a large processing burden on speakers of verb-final languages.

In verb-medial languages like English, verbs become available relatively earlier in the sentence, such that the average working memory cost of waiting for the verb would be less than in verb-final languages. The advantage of waiting for the verb information is that the parser can reduce the likelihood of making risky commitments, because the verb may turn out to be intransitive and disallow an object NP analysis for the filler. In English, therefore, the parser may create an object gap position only after the verb is confirmed to be transitive. Postverbal gap creation would still constitute active gap filling, in the sense that the ultimate gap position may turn out to be somewhere later than

the object position in the sentence (as in the sentences that are used to illustrate the filled gap effect in (2)). Let us call this a *conservative active gap filling* mechanism, since the bottom-up information from the verb still plays a critical role in the parser's decision on whether to postulate an object gap or not. This view of active gap filling is rather standard for explaining filler-gap dependency completion in verb medial languages like English. For example, McElree and colleagues have argued that the dependency completion process is triggered when the parser accesses information from the verb and initiates the retrieval process for the filler that is stored in working memory (McElree & Griffith, 1998; McElree, Foraker, & Dyer, 2003; see also Lewis & Vasishth, 2005; Pickering & Barry, 1991).

On the other hand, pre-verbal object gap creation in verb-final languages may reflect a language-general property of the processing architecture, although evidence for such mechanisms may be simply more difficult to obtain in verb-medial languages. In the English filler-gap case, for example, in any parser that adopts some form of left-corner strategy, the presence of the subject NP allows the parser to predict the presence of a VP (Abney & Johnson, 1991; Crocker, 1996; Gibson, 1991; Kimball, 1975; Resnik, 1992; Shieber & Johnson, 1993, Stabler, 1994). Given that a VP can contain an object NP position, the parser could project a VP with an object NP slot and assign the filler to this object position before confirming whether the upcoming verb is a transitive verb or not.[2] Let us call this a *hyper-active gap filling* mechanism, because this involves a more risky

[2] Alternatively, the hyper-active gap filling expectation could be instantiated as an expectation for verbs that can assign an accusative case or a thematic role to an internal argument. This alternative mechanism still involves the expectation of transitivity that is generated prior to encountering the verb. Thus, I will treat this lexical implementation of hyper-active gap filling mechanism as equivalent to the model in which an object gap position is hyper-actively postulated.

predictive structure building process than is standardly assumed for active object gap creation in English. Filler retrieval and structural integration is still integral to the hyper-active gap filling mechanism, but the crucial difference is in what information triggers the retrieval and integration, and consequently, at what point in the sentence this process is executed.

It is important to note that either of these two active gap filling mechanisms is compatible with the existing data on active object gap creation reviewed above. A filled gap effect only indicates that the gap had been created before the actual object NP is processed, but given that both hyper-active gap filling and conservative active gap filling mechanisms assume that object NP gap creation happens before or on the verb, this result is predicted by both accounts. A plausibility mismatch effect indicates that when the verb is potentially transitive, then the semantic fit between the filler and the verb is immediately assessed. This is also predicted by both accounts. The assessment of the semantic relation between the filler and the verb requires the parser to access the content of the verb, by which point the object gap position should have been created on either account. Thus, neither paradigm allows us to tease apart the two hypotheses on what kind of information is sufficient for triggering object gap creation.

In the current study we test the predictions of two hypothesized mechanisms for active object gap creation processes. If English speakers construct the gap site before encountering the verb, just like speakers of verb-final languages, then English speakers risk the possibility that the verb transitivity information might not ultimately license this structure. Therefore, disruption should be observed in filler-gap configurations when the verb turns out to be intransitive, relative to transitive verbs (e.g., *The party that the*

student arrived/planned…). According to the conservative active gap filling mechanism outlined above, the parser waits for a transitive verb before postulating the corresponding gap structure. If this is the mechanism used by English speakers, one should not expect to see disruption at an intransitive verb, since no gap that would require a transitive verb would have been posited in advance of the verb.

Two previous studies are relevant to the two hypotheses about active object gap creation in English. Previous work by Pickering and Traxler (2003) examined the effect of subcategorization frequency in optionally transitive verbs (e.g., *Those are the lines/props that the author spoke [about]*…). It was found that readers did not take differences in subcategorization frequency into account in deciding where to posit a gap, as there was a strong preference to posit a gap in the verb object position (NP complement) even with verbs that more frequently take a PP complement. The absence of a subcategorization preference in active object gap creation could be taken to indicate that verb information is not relevant for object gap creation processes. However, it should be noted that all of the verbs in Pickering and Traxler's study could grammatically accommodate an NP complement, and that the parser may therefore have relied on the transitivity information of the verb to create an object gap. Therefore, this finding is compatible with the predictions of the two proposed mechanisms for active object gap creation.

To our knowledge, the only previous test of these two active object gap creation hypotheses is in Experiment 3 of Staub (2007). The test sentences in this experiment (5a-d) manipulated the transitivity of the verb (*called* vs. *arrived*) and sentence structure (relative clause with a gap vs. simple declarative with no gap). The filler was manipulated

to be an implausible object of the transitive verb (*gadget-called*). Under the hyper-active gap filling hypothesis, the parser in effect predicts the presence of a transitive verb; therefore, the reading processes in the gap conditions should be disrupted when the verb turns out to be intransitive, and processing should also be disrupted when the verb is transitive, because of the plausibility mismatch effect. On the other hand, the conservative active gap filling mechanism postulates a gap only after checking whether the verb is capable of hosting an object NP, and therefore reading disruption is predicted only in the transitive gap condition due to the plausibility mismatch effect.

(5) a. The gadget that the manager *called* occasionally about ...

b. The manager *called* occasionally about the gadget …

c. The party that the student *arrived* promptly for …

d. The student *arrived* promptly for the party …

Staub (2007) found longer first-fixation durations in the transitive gap condition (5a) than in the transitive no-gap condition (5b), but no such difference was observed between the intransitive gap and no-gap conditions (5c) and (5d). This pattern of data supports the prediction of the conservative active gap filling hypothesis, suggesting that the parser does not create an object gap until it checks the transitivity information of the verb. However, the reading times in the no-gap, baseline conditions in this experiment may have been complicated by other factors. For example, the gap conditions (5a) and (5c) contain an extra NP (i.e., the head of the relative clause) prior to the critical verb region in comparison to the no-gap conditions (5b) and (5d), and the difference in the

number of NPs leads to a difference in the amount of contextual information. It has been argued that increased contextual information can facilitate processing for subsequent lexical items (Kutas & Federmeier, 2000; Hale, 2003; Levy, 2008), and for this reason, lexical access for the intransitive verb in the gap condition may have become faster and masked the potential reading time slowdown associated with the structural manipulation. Also, reading times near the beginning of a sentence are often more variable and tend to be slower, so the fact that the critical word was so early in the sentence in the no-gap conditions (word 3 compared to word 6 in the gap conditions) may have masked slowdowns in the gap conditions. The current study attempted to control for these methodological concerns to provide a better test of the predictions of the hyper-active and conservative active gap filling accounts.

2.2 Experiment 1 (self-paced reading)

Experiment 1 was a self-paced reading study that was designed to test the predictions of the hyper-active and conservative active gap filling hypotheses, while addressing methodological concerns about previous work. We employed the transitivity mismatch paradigm used in Staub (2007) in order to test whether a verb transitivity manipulation affects reading time at the verb. Critically, in the baseline conditions, the critical verb was embedded inside a relative clause structure, a syntactic 'island' domain that prohibits long-distance dependency formation (Ross, 1967; for a review, see Szabolcsi & den Dikken, 2003). A sample set of stimuli is shown in (6).

(6) a. Transitive, Non-island

The city that the author wrote regularly about was named for an explorer.

b. Transitive, Island

The city that the author who wrote regularly saw was named for an explorer.

c. Intransitive, Non-island

The city that the author chatted regularly about was named for an explorer.

d. Intransitive, Island

The city that the author who chatted regularly saw was named for an explorer.

A number of previous studies have shown that the parser respects island constraints in real-time syntactic processing, such that it avoids actively constructing filler-gap dependencies that span syntactic island boundaries (Kluender & Kutas, 1993; McElree & Griffith, 1998; McKinnon & Osterhout, 1996; Stowe, 1986; Traxler & Pickering, 1996; Wagers & Phillips, 2009; Yoshida, 2006). The relative clause island condition thus provides a baseline measure of reading times for the critical transitive and intransitive verbs, independent of processes of filler-gap dependency completion. The use of island configurations allows us to address the methodological concerns with previous work. First, this design allows us to retain the filler and the gap surrounding the island domain, such that the same amount of contextual information from the lexical items is present in advance of the critical verb region across the four conditions. Second, the critical region is closely matched across conditions (word 6 in the non-island conditions, word 7 in the

island conditions), and it is also placed away from the early portion of the sentence.

Furthermore, following Staub's design, we selected transitive verbs that are implausible hosts for the filler. Under this design, the hyper-active gap filling hypothesis predicts a reading time slowdown in both the non-island transitive (6a) and the non-island intransitive (6c) conditions relative to their baseline conditions (6b) and (6d), but for a different reason in the two cases. In the transitive condition, the slowdown would reflect a plausibility mismatch effect triggered by the semantic misfit between the filler and the verb. In the intransitive condition, the slowdown would result from a transitivity mismatch effect due to the mismatch between the expected subcategorization property of the verb (i.e., transitive) and the actual subcategorization property of the verb. On the other hand, the conservative active gap filling hypothesis predicts an interaction: A reading time contrast should be observed between the non-island transitive condition (6a) and the island transitive condition (6b) due to the plausibility mismatch effect, but no such reading time contrast should be observed between the two intransitive conditions (6c) and (6d), given that the parser should not actively create an object gap in either condition. Note that the lexical difference in the critical verb region across conditions is not problematic, since the critical contrast is between non-island and island conditions within each verb type.

2.2.1 Method

Participants

We recruited 32 native speakers of American English from the University of Maryland community. They all received a course credit or were paid $10 for their participation and were naïve to the purpose of the experiment.

Materials

We used 28 sets of four sentences like (6a) to (6d), which are all listed in Appendix A. The transitive non-island and island conditions were taken from the implausible semantic fit conditions in Omaki and Schulz (in press), who used a modified version of the plausibility manipulation materials from Traxler and Pickering (1996). Omaki and Schulz replicated Traxler and Pickering's plausibility mismatch effect with native and non-native speakers alike, confirming that the semantic fit between the filler and the verb affects the reading time for the verb when the verb is in a gap filling (i.e., non-island) environment, but not when the verb is inside a relative clause island. Critically, it was also found that the implausible verb-filler combination in a non-island environment (e.g., *city-wrote*) leads to a significant slowdown at the verb compared to its island counterpart with the same implausible verb-filler combination. Thus, even though the current experiment did not include a plausible counterpart of the implausible transitive verb condition, we can be confident that a reading time contrast between the transitive non-island and island conditions results from the semantic misfit between the filler and the verb. In other words, the finding in Omaki and Schulz's study supports the notion that island conditions in general can be used as baseline conditions for any reading disruption associated with active object gap creation. The intransitive conditions were modeled after the transitive conditions by replacing the optionally transitive verb with unergative or unaccusative intransitive verbs (Levin & Rappaport Hovav, 1995).

The non-island and island conditions differed in the number of relative clauses: The non-island condition had only one relative clause (*the city that the author wrote/chatted regularly about*) such that the object position of the verb *wrote/chatted* was

the first potential gap position after the embedded subject was encountered, whereas in the island conditions the critical verb was embedded inside another relative clause *the author who wrote/chatted regularly*, such that linearly this was still the first verb but grammatically the filler should not be accessible to the verb due to the relative clause island constraint. Thus, the first verb served as the critical region for testing the plausibility and transitivity mismatch effects. All the transitive verbs were optionally transitive, such that the sentences in the island conditions were all ultimately grammatical. The subcategorization frequency of the optionally transitive verbs was not controlled, since Pickering and Traxler (2003) have demonstrated that plausibility mismatch effects are attested for optionally transitive verbs regardless of subcategorization frequency. In all four conditions the same adverb immediately followed the verb, making it possible to observe potential spill-over effects. The 28 sentence sets were counter-balanced across four lists so that each participant saw only one version of the target items and consequently read 7 tokens of each condition. In addition, 72 fillers of similar length and complexity were constructed and added to each list.

Procedure

The self-paced reading task was implemented on the Linger software developed by Doug Rohde (http://tedlab.mit.edu/~dr/Linger/). We used a word-by-word, non-cumulative moving window presentation (Just, Carpenter, & Woolley, 1982). In this design, each sentence initially appears as a series of dashes, and these dashes are replaced by a word from left to right every time the participant presses the space bar. In order to ensure that the participants were paying attention while reading the sentences, all

sentences were followed by yes-no comprehension questions, and feedback was provided if the questions were answered incorrectly. At the beginning of the experiment, participants were instructed to read at a natural pace and to answer the questions as accurately as possible. Seven practice items preceded the self-paced reading experiment, and the order of presentation was randomized for each participant. The experiment took approximately 30 minutes.

Analysis

The data from two items were excluded from analyses due to coding errors, and only trials in which the comprehension question was answered accurately were included in the analysis. Self-paced reading times for the target sentences were examined for each successive region, although the words after the auxiliary *was* were combined into a single region because these lay beyond the critical regions and were unlikely to show effects relevant for the critical manipulation. Reading time data that exceeded three standard deviations from the group mean at each region and in each condition were excluded, affecting 1.7% of the data. The participant mean (F1) and item mean (F2) of the remaining reading time data for each region were submitted to a repeated measures 2 × 2 ANOVA with the factors structure type (non-island vs. island) and verb type (transitive vs. intransitive). In the critical regions, planned comparisons were conducted to test for systematic differences between non-island and island conditions within each verb type.

2.2.2 Results

Comprehension accuracy. The mean comprehension question accuracy for experimental items across participants and items was 93.0%. For the non-island conditions, the transitive items were answered with an accuracy of 93.7% (*SE* = 1.9), and

the intransitive items with an accuracy of 94.6% (*SE* = 1.4). For the island conditions, the transitive items were answered with an accuracy of 91.5% (*SE* = 1.7), and the intransitive items with an accuracy of 92.0% (*SE* = 2.2). The mean accuracy did not differ reliably across conditions (*F*s < 1), although the fact that the mean accuracy for island conditions was numerically lower may reflect the complexity difference between non-island and island conditions.

Reading time data. The critical regions where a potential plausibility or transitivity mismatch effect was expected consist of Region 7 (i.e., the verb *wrote/chatted* in the example sentence in (6)) and the following Region 8 (i.e., the adverb *regularly* in the example sentence in (6)), in which spill-over effects could be observed. Regions 1 through 6 were predicted to show no difference across conditions, since they were lexically matched. Regions 9 through 11 could reveal reading time differences after the filler-gap dependency is completed (Region 9 hosts the true gap site), and with a possible additional difference in the island conditions, due to the complexity associated with the extra relative clause in these conditions.

The region-by-region mean reading time for the transitive conditions is presented in Figure 2, and the mean region-by-region reading time for the intransitive conditions is presented in Figure 3.

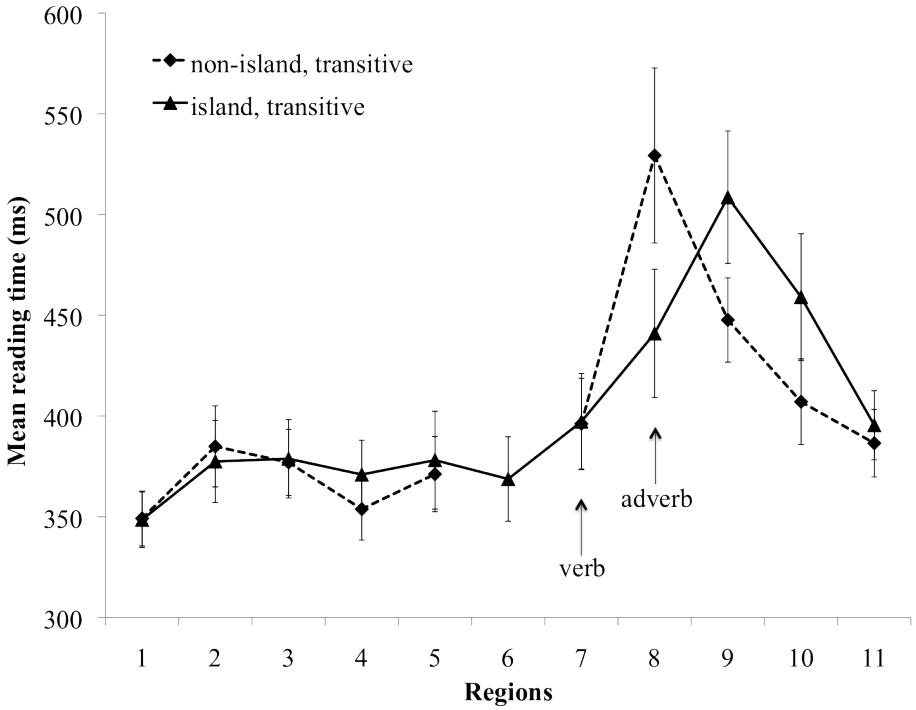


Figure 2. Mean reading time (ms) for the transitive non-island and island conditions. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The$_1$ city$_2$ that$_3$ the$_4$ author$_5$ (who)$_6$ wrote$_7$ regularly$_8$ about/(saw)$_9$ was$_{10}$ [named for an explorer]$_{11}$.

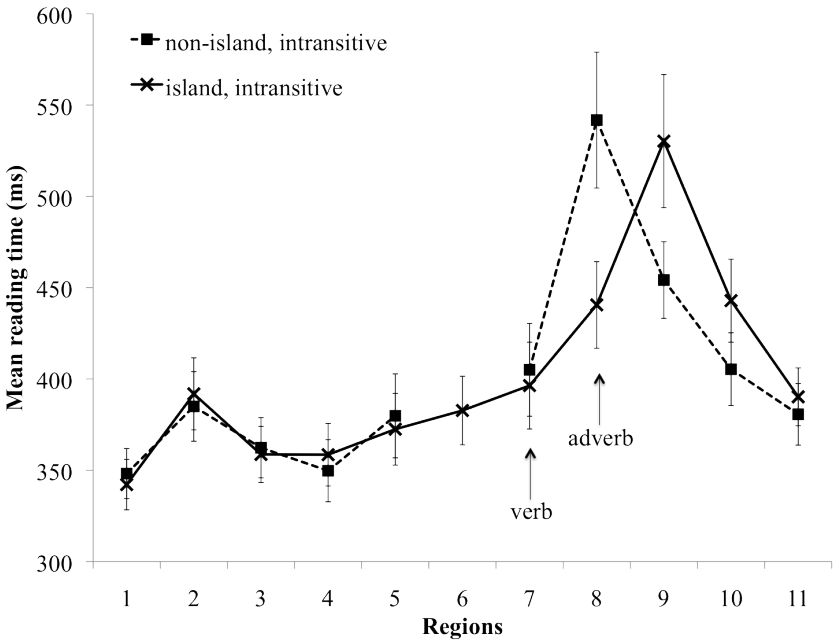


Figure 3. Mean reading time (ms) for the intransitive non-island and island conditions. Error bars indicate standard error of the mean.

Sample sentence (words in parentheses appear only in island conditions; words in brackets represent one region):

The$_1$ city$_2$ that$_3$ the$_4$ author$_5$ (who)$_6$ chatted$_7$ regularly$_8$ about/(saw)$_9$ was$_{10}$ [named for an explorer]$_{11}$.

The statistical analysis of reading time data revealed some spurious effects in non-critical regions (Regions 3 and 4), but crucially the predicted effects in the critical regions were larger than those effects.

In the non-critical Regions 1 to 6, there were no significant differences in Regions 1, 2, 5 and 6 (Fs < 1). In Region 3 there was a main effect of verb type, $F1(1, 31) = 10.30$, MSE = 9671, $p < .005$, $F2(1, 25) = 6.63$, MSE = 13687, $p < .05$, due to slower reading times in the transitive conditions than in the intransitive conditions (381 ms vs.

358 ms). Region 4 showed a main effect of the island manipulation in the participant analysis only, $F1(1, 31) = 5.20$, MSE = 5391, $p < .05$, $F2(1, 25) = 1.79$, MSE = 2166, $p > .1$, due to slower reading times in the island conditions (352 ms vs. 365 ms). Since these regions were lexically matched across conditions, we conclude that these must be spurious effects. But given that the effects were small and occurred well ahead of the critical regions, these unexpected effects were unlikely to have impacted the observations in the critical regions.

At the critical verb in Region 7 there were no significant differences (Fs < 1). The following spill-over region (Region 8) revealed no main effect of verb type (Fs < 1), but there was a main effect of structure type, $F1(1, 31) = 16.90$, MSE = 287354, $p < .0005$, $F2(1, 25) = 14.79$, MSE = 230119, $p < .005$, reflecting the fact that the non-island conditions produced significantly slower reading times than the island conditions (529 ms vs. 435 ms). There was no significant interaction of verb type and structure type (Fs < 1). A pair-wise comparison revealed that the reading times in the non-island condition were significantly slower than the reading times in the island condition for transitive sentences (529 ms vs. 441 ms), $t1(31) = 3.15$, $p < .005$, $t2(25) = 2.54$, $p < .05$, as well as for intransitive sentences (542 vs. 441 ms), $t1(31) = 4.05$, $p < .0005$, $t2(25) = 3.64$, $p < .005$.

Region 9 consisted of a second verb in the island conditions and a preposition in the non-island conditions. We observed a main effect of structure type in Region 9, $F1(1, 31) = 13.53$, MSE = 150149, $p < .005$, $F2(1, 25) = 13.61$, MSE = 108850, $p < .005$, as well as in Region 10, $F1(1, 31) = 12.75$, MSE = 64094, $p < .005$, $F2(1, 25) = 5.78$, MSE = 48875, $p < .05$, in these cases due to slower reading times in the island conditions (Region 9: 519 ms vs. 451 ms, Region 10: 451 ms vs. 406 ms). Region 11

revealed no significant differences (Fs < 1).

2.2.3 Discussion

In Experiment 1, we tested the predictions of two hypotheses about active object gap creation. The hyper-active gap filling hypothesis predicted the presence of reading disruption at intransitive verbs, because encountering an intransitive verb in a filler-gap context would be incompatible with the object gap structure constructed earlier. On the other hand, the conservative active gap filling hypothesis predicted no such reading disruption, because the parser should first consult the transitivity information of the verb to decide whether to posit an object gap or not. As a baseline for estimating the degree of disruption at the verb due to plausibility mismatch and transitivity mismatch, we used relative clause island constructions, which block the association of the filler with the critical verb. In the region following the verb, we observed slower reading times for both plausibility-mismatched transitive verbs and intransitive verbs in non-island conditions than in corresponding island conditions. Previous work has shown that these items with plausibility mismatched transitive verbs in a non-island environment demonstrate longer reading times than their plausible non-island or plausible/implausible island counterparts (Omaki and Schulz, in press), and here we replicated the finding of a slowdown after the optionally transitive verb in the implausible non-island condition relative to the implausible island condition. This slowdown can be interpreted as the result of active association of the filler with the transitive verb, which in these stimuli resulted in a verb-object plausibility mismatch. In the island condition, the verb was inaccessible as a potential gap position and thus this implausible verb-object combination could not be formed. The slowdown observed in the intransitive non-island condition relative to the

intransitive island condition can be interpreted as a *transitivity* mismatch. This suggests that the parser does not wait for bottom-up evidence from the verb that the verb can syntactically license a gap, but rather attempts to construct the dependency before this information is available. This slowdown cannot reflect the cost of maintaining the filler in working memory, because a filler is also being maintained at this position in the baseline island condition.

In Regions 9 and 10, the island conditions were read more slowly for both levels of verb type. Region 9 corresponds to the word that licensed the true gap site across all conditions, and hence this slowdown could reflect a difference in the so-called integration cost (Gibson, 1998; 2000) between non-island and island conditions. Previous work on filler-gap dependency processing has demonstrated that increased complexity and length differences result in increased processing difficulties at the gap site, as measured by reading time (Gibson & Warren, 2004) and reduced accuracy in speeded acceptability judgment tasks (McElree et al., 2003).

Note that it is unlikely that the reading time contrast between non-island and island conditions in Region 8 is related to the overall complexity of the constructions used in our stimuli, given that on all accounts that we are aware of, island domains have been argued to be syntactically more complex and more taxing for working memory resources (Deane, 1991; Kluender, 1999, 2004; Kluender & Kutas, 1993; Hofmeister & Sag, 2010). The fact that the putatively less complex non-island conditions were read more slowly allows us to attribute the slowdown to processes that uniquely occur in the non-island conditions, namely filler-verb association.

In summary, the presence of both a plausibility mismatch effect and a transitivity

mismatch effect lends support to the hyper-active gap filling hypothesis, and argues against a conservative active gap filling hypothesis under which transitivity information is consulted before attempting to create an object gap. This finding directly contrasts with that of Staub (2007), who did not find evidence for a transitivity mismatch effect. One possible reason for this discrepancy is the difference in the baseline conditions used in the two studies. As noted above, the lack of the filler NP in Staub's baseline condition may have unexpectedly created a difference in the amount of helpful contextual information in the two conditions and hence may have complicated the reading time prediction for the critical verb region. Moreover, there was a large difference in the word position of the critical verb across conditions, and this may have additionally complicated the reading time patterns. In the current study, both of these complicating factors were controlled. Both the target and baseline conditions contained the same filler (and hence the same contextual information) in advance of the critical verb region, and the word position of the critical verb was closely matched across conditions. Given that our study used a verb transitivity manipulation design similar to Staub (2007), the removal of these complicating factors seems to be the most likely explanation for the presence of verb transitivity mismatch effect.

However, there are other methodological differences between our Experiment 1 and Staub (2007) that could account for the difference in the findings. First, our intransitive materials consisted of two types of intransitive verbs: We mainly used unergative verbs which only take a semantic agent as an argument, but we also used unaccusative intransitive verbs that only take a theme/experiencer as an argument (Perlmutter, 1978; Levin & Rappaport-Hovav, 1995). On the other hand, Staub's

intransitive condition used only unaccusative intransitive verbs. These two types of intransitive verbs are generally treated as being on a par with one another, in that they are generally incompatible with an overt direct object NP; but in some restricted contexts unergative verbs are capable of hosting a cognate NP in its complement position (e.g., "laugh a big laugh"; see Keyser & Roeper, 1984). It is possible that this special property of unergative verbs may have led the parser to treat it in the same way as transitive verbs in our experiment, whereas unaccusative intransitive verbs admit no such exceptions.

Second, Staub (2007) used an eye-tracking during reading method and reported evidence for active gap creation on the transitive verb region, whereas in our self-paced reading experiment, evidence for reading disruption for transitive and intransitive verbs (i.e., the slowdown in non-island conditions compared to island conditions) was not observed until the spill-over adverb region. Spill-over effects are extremely common in self-paced reading experiments and it is thus common to attribute spill-over effects to processes triggered in a preceding region, but in our experiment there is an alternative explanation for the effect in the adverb region that would not require hyper-active gap filling. For the intransitive condition, the slowdown in the adverb region could indicate that the parser had expected the presence of a preposition, which would allow structural integration of the filler. Under this alternative account, the slowdown is not due to a transitivity mismatch on the verb, but rather to a word category expectation mismatch in the adverb region that was triggered by the verb itself. This account is entirely consistent with the conservative active gap filling hypothesis, since the parser's expectation regarding filler-gap dependency completion is based on the information from the verb, and this may be the reason why a reading disruption was found at the verb region for

transitive verbs but not for intransitive verbs. (For another eye-tracking demonstration of the plausibility mismatch effect directly on the transitive verb, see Traxler & Pickering, 1996). On the other hand, an eye-tracking during reading method generally provides better temporal precision than the self-paced reading method used in our study (Rayner, 1998; Rayner & Pollatsek, 2006). Thus, if we were to use an eye-tracking during reading method while maintaining our structural manipulations, we might find evidence for a transitivity mismatch effect on the verb region as well, which would lend support to our interpretation of the self-paced reading results as evidence for the hyper-active gap filling hypothesis.

In sum, results from Experiment 1 are consistent with the predictions of the hyper-active gap filling hypothesis. This is contrary to the results reported in Staub (2007), which may attributable to to the more closely matched baseline conditions in our experiment. However, there were two methodological differences between our study and Staub's study that call for caution in this conclusion: namely, the broader range of intransitive verbs used in our study, as well as the fact that we used a dependent measure that does not provide as good temporal precision as the eye-tracking during reading method. Experiment 2 was designed to address both of these concerns.

2.3 Experiment 2 (eye-tracking during reading)

The goal of Experiment 2 was to further test the predictions of the hyper-active gap filling and conservative active gap filling hypotheses while addressing the methodological concerns raised in Experiment 1. In order to make our experimental methodology as close as possible to the method used by Staub (2007), we made two methodological changes from Experiment 1. First, we constructed new sets of stimuli that

used only the unaccusative intransitive verbs that were used in Staub (2007). Given that unaccusative intransitive verbs are syntactically incapable of hosting an overt direct object NP, this class of intransitive verbs provides a stronger test of the transitivity mismatch effect. Second, we used an eye-tracking during reading method instead of the self-paced reading task used in Experiment 1. Eye-tracking provides better sensitivity to the temporal dynamics of reading processes, and in the domain of filler-gap processing, time course evidence for gap filling such as the plausibility mismatch effect is attested in the first fixation duration on the verb region itself (Traxler & Pickering, 1996). Given that the hyper-active gap filling hypothesis concerns structure building processes that are predicted to lead to syntactic incongruities when readers encounter a specific verb type, it is critical that we use a dependent measure that can elicit a mismatch response on the verb region itself.

2.3.1 Method

Participants

We recruited 44 native speakers of American English from the University of Maryland community. All had normal or corrected-to-normal vision, and were naïve to the purpose of the experiment. They received course credit or were paid $10 for their participation, which lasted around 40 minutes.

Materials

We used 24 sets of four sentences like (7), which are all listed in Appendix B.

(7) a. Intransitive, non-island

The airport that the ambassador departed rapidly from during the unrest was closed to most traffic.

b. Intransitive, island

The airport that the ambassador who departed rapidly had visited during the unrest was closed to most traffic.

c. Transitive, non-island

The airport that the ambassador left rapidly for during the unrest was closed to most traffic.

d. Transitive, island

The airport that the ambassador who left rapidly had visited during the unrest was closed to most traffic.

This experiment used the same transitivity mismatch logic as Experiment 1 and manipulated the verb transitivity type (intransitive vs. transitive). However, in this experiment the semantic fit between the filler and the transitive verb was controlled to be plausible, such that no reading disruption was expected at the transitive verb in the non-island condition. As in Experiment 1 we manipulated structure type (non-island vs. island), using conditions with relative clause island structures as baseline conditions. Relative clause islands provide an effective baseline, since they include the same filler NP and other lexical material as the non-island condition, while preventing dependency completion at the critical verb. As in Experiment 1, the transitive verbs were optionally transitive and the true gap position occurred outside the island domain, allowing the

sentence to continue grammatically. The 24 sentence sets were counter-balanced across four lists, such that each participant saw only one version of each of the target sentences. The target sentences were combined with 108 fillers of similar length and complexity.

The two hypotheses on active gap creation mechanisms predict different reading time patterns for the transitive and intransitive conditions. The hyper-active gap filling hypothesis states that the parser can use pre-verbal information to create an object gap before encountering the verb, and hence reading disruption is predicted in the intransitive non-island condition in comparison to the corresponding island condition. No such reading disruption is expected in the transitive non-island condition because the transitive verb expectation is satisfied, and hence no reading time contrasts should be observed between the two transitive conditions. On the other hand, the conservative active gap filling hypothesis states that the object gap creation process occurs only after checking whether the verb can host an object NP. In other words, in the transitive non-island condition gap creation should occur, but in the intransitive non-island condition no gap should be created and hence there should be no disruption reflecting the need for reanalysis. Thus, this hypothesis predicts no reading time contrast between the two transitive conditions or between the two intransitive conditions.

Procedure

An SR Research (Mississauga, Ontario, Canada) Eyelink 1000 eye-tracker was used to record eye movements. The participant's head was stabilized by a chin rest and a forehead rest. The position of the right eye only was monitored at a sampling rate of 1000 Hz. The eye-tracker display allowed a maximum of 100 characters per line. Some filler sentences were displayed on two lines, but all target sentences were displayed on one

line. Stimuli were displayed on a 17-inch monitor, and participants were seated 60 cm from the computer screen.

Before the experiment started, participants were seated in front of the eye-tracker and received instructions for the experiment. A calibration routine was performed at the beginning of the experiment, and the experimenter monitored the calibration accuracy throughout the experiment, recalibrating when necessary. The experiment started with written instructions on the display and four practice trials. At the beginning of each trial, a black square was displayed on the left side of the monitor, which corresponded to the location of the beginning of the sentence. The text was displayed after the participant successfully fixated on the square. After reading each sentence, the participant pressed a button to remove the sentence display. Each sentence was followed by a yes-no comprehension question, and the participant answered the comprehension question by pressing a left or right button. The entire experiment lasted approximately 40 minutes.

Data analysis

Data from six participants were removed due to calibration errors. Trials in which participants answered the comprehension question incorrectly were removed from the eye movement analyses. For the remaining data an automatic procedure was used to pool short contiguous fixations. The procedure incorporated fixations of less than 80 ms into larger fixations when they occurred within one character of each other and deleted any remaining fixations of less than 80 ms, because little information can be extracted during such short fixations (Rayner & Pollatsek, 1989). Unusually long fixations greater than 800 ms were also removed, because they usually reflect tracker losses or other anomalous events. This procedure resulted in the exclusion of 2.63% of all fixations.

For the purpose of analysis of the eye movement data, the sentences were divided into the following regions.

(8) a. Non-island conditions

The airport that/ the ambassador/ {departed | left}/ rapidly/ {from | for} during the unrest was closed to most traffic.

b. Island conditions

The airport that/ the ambassador who/ {departed | left}/ rapidly/ had visited during the unrest was closed to most traffic.

We report eye movement data in the following three regions: a) the pre-verb region (*the ambassador* in (8a), *the ambassador who* in (8b)), in order to ensure that there were no unexpected reading behavior differences that might compromise the interpretation of the data from the critical region; b) the verb region, which is the critical region in which potential transitivity mismatch effects might be observed; and c) the post-verb region, which corresponds to the post-verbal adverb and could be used to probe for potential spill-over effects. The data in the remaining regions are not reported, because reading times at these regions are not critical for distinguishing the competing hypotheses, and after the post-verb region, the lexical items were not held constant across conditions and therefore any observed differences would be difficult to interpret. Each region started with the space before the first word in the region, and ended on the last character of the last word in the region.

Note that the island conditions contained one extra word, i.e., the relative pronoun

(e.g., *who*), which could have affected reading times in the pre-verb region as well as in regression measures for subsequent regions. Furthermore, the length of the critical verbs in the transitive and intransitive conditions was not controlled, as the critical comparison was between non-island and island conditions within each level of the verb type factor. In fact, the mean length of the verbs was longer in the intransitive condition (7.58 characters) than in the transitive condition (6.33 characters), $t(23) = 2.63$, $p < .05$. Thus, the critical observation that is relevant for testing the current hypotheses comes from the structure type manipulation within each verb type, as well as from the interaction of structure type and verb type, rather than from a main effect of verb type, which may simply reflect the different properties of the lexical items across conditions.

Following the data analysis procedures used in Staub (2007), four reading time measures were computed for the three regions of interests: *first fixation duration, first pass time, regression path time,* and *percent regressions* (Rayner, 1998; Rayner & Pollatsek, 2006; Staub & Rayner, 2007). First fixation duration is the duration of the very first fixation in a region, regardless of whether there is a single word or multiple words in that region. This measure is often used as an index of lexical difficulty (e.g., Reichle, Rayner, & Pollatsek, 2003) but is also informative about the earliest syntactic processes that immediately follow lexical access (e.g., Frazier & Rayner, 1982; Sturt, 2003). In fact, previous eye-tracking studies on filler-gap dependency completion have shown that the plausibility mismatch effect can be observed in the first fixation duration on the verb (Staub, 2007; Traxler & Pickering, 1996), and it is thus reasonable to expect a transitivity mismatch effect in this measure in the current study.

The *first-pass reading time* is calculated by summing the fixations in a region

between the time when the eye-gaze first enters the region from the left and the time when the eye-gaze exits the region either to the left or the right. First-pass reading times also index early lexical and syntactic processes associated with a region, but given that they consist of multiple fixations on the same region, they may also reflect slightly later processes than the first fixation measure. We follow the convention in the eye-tracking literature and use the term *gaze duration* interchangeably with first-pass reading time when the region consists of a single word.

Regression path times are the sum of fixations from the time when the eye-gaze first enters a region from the left to the time when the eye-gaze exits the region to the right. Regression path time is identical to first-pass reading time if the eye-gaze first exits the region to the right, but if the eye-gaze exits the region to the left, then regression path times are longer than the first-pass time as they include all fixations in previous regions as well as re-fixations on the region before exiting the region to the right. Thus, regression path times are likely to reflect slightly later processes, such as integration of the critical region with the preceding context. The *percent regressions* indicate the probability that a reader made a regressive eye movement after fixating a given region. This measure includes only regressions made during the reader's first pass through the region, and does not include regressions made after re-fixating the region.

These four eye movement measures from the pre-verb, verb, and post-verb regions were submitted to a 2 × 2 repeated measures ANOVA with verb type and structure type as within-participants factors. When there was a main effect of structure type or a significant interaction of verb type and structure type in the critical verb region,

a planned comparison was conducted to test for systematic differences between the island and non-island conditions within each verb type.

2.3.2 Results

Mean comprehension accuracy for the experimental items was 91.9% across the four conditions, and did not differ across the four conditions (*F*s < 1). Table 1 presents the participant means on each measure for each region as well as the standard errors of the participant means.

Table 1. Experiment 2: Participant Mean Reading Times in Milliseconds (Standard Error) and Percent Regressions

Measure	Pre-verb region	Verb region	Post-verb region
First fixation duration			
Non-island, intransitive	226 (7)	299 (10)	271 (9)
Island, intransitive	222 (6)	270 (8)	259 (8)
Non-island, transitive	229 (8)	277 (8)	268 (11)
Island, transitive	236 (8)	266 (8)	258 (9)
First-pass time			
Non-island, intransitive	349 (19)	379 (13)	340 (15)
Island, intransitive	461 (21)	345 (20)	330 (21)
Non-island, transitive	367 (22)	319 (11)	308 (14)
Island, transitive	468 (29)	316 (14)	321 (16)
Regression path time			
Non-island, intransitive	538 (43)	528 (38)	545 (40)
Island, intransitive	762 (48)	527 (54)	497 (43)
Non-island, transitive	529 (29)	386 (20)	553 (79)
Island, transitive	706 (47)	520 (44)	529 (45)
Percent regressions			
Non-island, intransitive	26.7 (3.9)	14.4 (2.3)	24.0 (3.5)
Island, intransitive	32.1 (3.6)	11.7 (2.7)	21.2 (3.1)
Non-island, transitive	31.0 (3.7)	24.0 (3.3)	26.4 (3.8)
Island, transitive	26.3 (4.1)	28.4 (3.6)	25.9 (2.9)

In the pre-verb region, the first fixation duration measure showed no significant differences (*F*s < 2.5, p > .13). The first pass reading time measure showed a main effect

of the structure type manipulation, $F1(1, 37) = 31.39$, MSE = 426763, $p < .0005$, $F2(1, 23) = 43.88$, MSE = 245985, $p < .0005$, due to slower reading times in the island conditions (358 ms vs. 464 ms), and there was no main effect of verb type or interaction of the two factors (Fs < 1). Similarly, in the regression path duration measure, we found a main effect of structure type, $F1(1, 37) = 36.67$, MSE = 1518807, $p < .0005$, $F2(1, 23) = 29.08$, MSE = 812343, $p < .0005$, due to slower reading times in the island conditions (534 ms vs. 734 ms), and there was no main effect of verb type or interaction of the two factors (Fs < 2.5, $p > .12$). This structure type effect was expected, since the island conditions contained an extra relative pronoun. In the percent regressions measure, there was a marginally significant interaction in the participants analysis, $F1(1, 37) = 2.92$, MSE = .098, $p = .096$, but this effect was not present in the items analysis, $F2(1, 23) = 2.07$, MSE = .065, $p = .16$.

Turning to the critical verb region, in the first fixation duration there was a main effect of structure type in both the participants and the items analyses, $F1(1, 37) = 5.17$, MSE = 15614, $p < .05$, $F2(1, 23) = 5.09$, MSE = 8395, $p < .05$, as well as a marginal effect of verb type in the participants analysis, $F1(1, 37) = 3.77$, MSE = 6641, $p = .06$, $F2(1, 23) = 2.20$, MSE = 4215, $p > .1$, with no significant interaction between the factors (Fs < 1.7, $p > .2$). A pair-wise comparison was conducted to test whether the structure type manipulation affected the first fixation duration for both verb types. This comparison revealed that the reading time in the non-island intransitive condition was significantly longer than in the island intransitive condition (299 ms vs. 270 ms), $t1(37) = 2.31$, $p < .05$, $t2(23) = 2.49$, $p < .05$, but that the difference in first fixation duration between the transitive non-island and island conditions did not reach significance (277

ms vs. 266 ms; *t*s < 1.1).

On the gaze duration measure in the verb region, we observed a main effect of verb type, *F1*(1, 37) = 13.91, MSE = 73856, $p < .005$, *F2*(1, 23) = 10.73, MSE = 45071, $p < .005$, due to slower reading times in the intransitive conditions than in the transitive conditions (362 ms vs. 318 ms), and there was a significant interaction of the two factors in the items analysis, *F1*(1, 37) = 1.81, MSE = 8736, $p > .1$, *F2*(1, 23) = 4.37, MSE = 11062, $p < .05$. A pair-wise comparison revealed that the reading time in the non-island intransitive condition was significantly longer than in the island intransitive condition in the items analysis, though the effect was marginal in the participants analysis (379 ms vs. 345 ms), *t*1(37) = 1.7, $p = .097$, *t*2(23) = 2.26, $p < .05$. However, the difference in gaze duration between the transitive non-island and island conditions did not reach significance (319 ms vs. 315 ms; *t*s < 1).

On the regression path time in the verb region, there was a marginally significant effect of structure type, *F1*(1, 37) = 3.30, MSE = 169918, $p = .078$; *F2*(1, 23) = 3.63, MSE = 80697, $p = .07$, and a main effect of verb type that was significant in the participants analysis but only marginal in the items analysis, *F1*(1, 37) = 4.26, MSE = 209916, $p < .05$; *F2*(1, 23) = 3.70, MSE = 83712, $p = .067$. There was also a significant interaction of structure type and verb type in the items analysis though the effect was marginal in the participants analysis, *F1*(1, 37) = 2.87, MSE = 170952, $p = .099$,

F2(1, 23) = 5.41, MSE = 143218, $p < .05$. A pair-wise comparison revealed that the structure type effect was present for transitive verb sentences but not for intransitive verb sentences: The regression path time in the transitive non-island condition was significantly faster than in the transitive island condition (386 ms vs. 520 ms), *t*1(37) =

2.77, $p < .01$, $t2(23) = 2.92$, $p < .01$, whereas there was no difference between the intransitive non-island and island conditions (528 ms vs. 527 ms; ts < 1).

On the percent regressions in the verb region, there was a significant effect of structure type, $F1(1, 37) = 11.44$, MSE = .656, $p < .005$, $F2(1, 23) = 21.25$, MSE = .417, $p < .001$, but there were no other significant effects (Fs < 1.6, $p > .2$). Planned comparisons revealed that the structure type effect was present for both verb types: fewer regressive eye movements were made in the intransitive non-island condition than in the intransitive island condition (14.4% vs. 24.0%), $t1(37) = 2.13$, $p < .05$, $t2(23) = 3.66$, $p < .005$, and the same pattern was observed in the corresponding transitive conditions, (11.7% vs. 28.3%), $t1(37) = 3.26$, $p < .005$, $t2(23) = 3.33$, $p < .005$.

In the post-verb region there was no significant effect of verb type or structure type, and no interaction of the two factors in any of the dependent measures (all Fs < 2.6, $p > .11$).

2.3.3 Discussion

The results of this experiment can be summarized as follows. First, the first fixation duration for intransitive verbs in a structure that would allow a gap (non-island condition) was significantly longer than when the same verb appeared within an island configuration. This effect was not observed when the critical verb was transitive. The fact that there was a reading disruption for intransitive verbs but not for transitive verbs is consistent with the prediction of the hyper-active gap filling hypothesis: if the parser creates an object gap and integrates the filler into the object position before having access to verb transitivity information, reading disruption in the non-island intransitive condition should result from the mismatch between the predicted transitivity and actual transitivity

of the verb. A similar pattern was observed in gaze duration times, although differences here were less robust.

Under the proposed hyper-active gap mechanism, first fixation duration is the most appropriate dependent measure for attesting the transitivity mismatch effect, because this mechanism effectively predicts the presence of transitive verb. Therefore, the evaluation of this prediction should be possible as soon as the parser gains access to the lexical content of the verb, and this effect can be reasonably expected in the first fixation duration measure, which has generally been shown to reflect the earliest phase of lexical access (Rayner, 1998; Staub & Rayner, 2007). Thus, the intransitive vs. transitive contrast in first fixation duration provides support for the hyper-active gap filling hypothesis, contrary to the conclusion reached in Staub (2007).

Regression path times at the verb region showed a slightly more puzzling pattern: regression path times were much shorter for the transitive non-island condition than the other three conditions. Although speculative, one possible interpretation of this pattern rests on the fact that the regression path time is likely to be longer for the island conditions than for non-island conditions simply because island conditions contain an extra word (i.e., the relative pronoun *who*). Additionally, the island condition sentences mayalso possibly incur a higher processing cost due to their greater complexity (Deane, 1991; Kluender & Kutas, 1993), which seems to be confirmed by the general increase of probability of regressions in the island conditions. Therefore, the fact that the non-island transitive was faster than its corresponding island condition whereas the non-island intransitive was not faster than its corresponding island condition may in fact reflect a relative slowdown in the non-island intransitive condition due to a transitivity mismatch

effect. However, more studies with additional conditions would be needed to confirm this interpretation of the regression path results.

One may argue that there is a possible alternative account for our findings that is compatible with the conservative active gap filling hypothesis. We have so far treated unaccusative intransitive verbs as strictly intransitive. However, syntactically speaking, unaccusative intransitive verbs also contain an object NP position, in which the subject receives a thematic role (Burzio, 1986). Based on this syntactic property, one could argue that the parser first accesses the lexical information of the verb and identifies it as an unaccusative intransitive and then assigns the subject NP to the object position. At this point, the wh-filler that is retrieved would clash with the subject NP that already occupies the object NP position. On this view, the reading disruption that we found in our study may be another instance of a filled-gap effect. However, this account seems incompatible with the nature of the conservative active gap filling mechanism. This account presupposes that the association of the subject NP with the object position must take place as soon as the parser accesses the verb and identifies its unaccusativity. Note that at this point, the verb's thematic role is already assigned to its internal argument. This contrasts with the standard cases in which a filled-gap effect is observed, where the thematic role of the verb is not assigned to any overt NP in the input that has been received. In this context, it is unclear why the conservative active gap filling mechanism should attempt to retrieve the wh-filler even after the object position is already filled, given that the access to the unaccusativity of the verb should already indicate that the wh-filler cannot be integrated at the verb position. The reason why the conservative active gap filling mechanism is attractive is that it reduces the likelihood of mis-creation of a

gap by confirming whether the verb can host an object NP or not. In this sense, the assumptions that are required for this account are incompatible with a leading motivation for the conservative active gap filling hypothesis. On the other hand, the filled-gap account for reading disruption at unaccusative intransitive verbs is perfectly compatible with the hyper-active gap filling mechanism. Recall that on this view, the gap position is already filled upon accessing the lexical content of the verb, including its unaccusativity. Thus, if a subject NP is retrieved upon accessing unaccusativity, a filled-gap effect results, because the wh-filler will have already been assigned to the object NP position. Therefore, regardless of how we conceive of the role of unaccusativity in our target sentences, there is stronger support for the hyper-active gap filling mechanism than for the conservative gap filling mechanism.

2.4 General discussion

Both Experiment 1 and Experiment 2 demonstrated evidence for reading disruption at an intransitive verb when the verb was in a potential gap-filling environment. The reading disruption that can be attributed to a transitivity mismatch effect was observed at the same region as the region that revealed a plausibility mismatch effect (Experiment 1), and this reading disruption for an intransitive verb was observed as early as the first fixation on the intransitive verb (Experiment 2). These results lend support to the hyper-active gap filling hypothesis, which claims that in English filler-gap dependency processing, object gap creation can be initiated based on pre-verbal information and can thereby lead the parser to expect a transitive verb. This is indeed what has been proposed for the filler-gap dependency processing mechanism in head-final languages (Aoshima et al., 2004; Nakano et al., 2002), but the current work suggests

that the same mechanism extends to the processing of filler-gap dependency in verb-medial languages like English as well.

The view that object gap creation is triggered by pre-verbal information contrasts with a standard view of object gap creation in English filler-gap dependency processing that object gap creation is driven by properties of the verb (e.g., McElree et al., 2003; Pickering & Barry, 1991). In fact, the hyper-active gap filling mechanism suggests an alternative interpretation of existing data on active object gap creation. For example, a plausibility mismatch effect found in Traxler and Pickering (1996) has been taken to suggest that filler-retrieval occurs after accessing the transitivity information on the verb, and that subsequent structural integration of the filler leads to the implausible verb-object composition, which in turn results in reading time slowdown. However, under the hyper-active gap filling account, prior to the verb the reader analyzes the filler as a direct object of the upcoming verb, and given the combination of the subject NP and the hypothesized object NP, the reader may already expect a certain class of transitive verbs that would be semantically compatible with the filler noun phrase. In other words, plausibility mismatch effects could be reconsidered as a reflection of a violation of lexical expectations, which result from predictive structural analysis. In fact, it is worth noting that the ERP experiment reported by Garnsey and colleagues (Garnsey et al.,1989) found an N400 amplitude difference at the verb between sentences in which the filler was a plausible object of the verb and sentences in which the filler was an implausible object. Recent views of N400 effects suggest that N400 amplitude differences may index differences in pre-activation of the lexical or conceptual representation of the verb (for a summary, see Federmeier, 2007; Lau, Phillips, & Poeppel, 2008; cf. Hagoort, 2003).

Taken together, this research suggests that the re-interpretation of the plausibility mismatch effect as a violation of a pre-verbal expectation is feasible.

It is also conceivable that hyper-active gap creation only involves a lexical prediction without predictive structure building process. For example, hyper-active gap filling could be instantiated as an instruction for the lexicon, which leads to a lexical expectation for verbs that can assign an accusative case or a thematic role to an internal argument. Strictly speaking, pre-verbal activation of verbs with such transitive features can be achieved without constructing a syntactic representation for the NP object gap. Given that case or thematic information of the verb is directly relevant to constraining the upcoming structural representation, this view can be seen as practically equivalent to the view that an NP object gap position is predictively constructed. It is important to note, however, that with this lexical expectation view, the predictive lexical activation process must somehow be restricted to verbs or predicates that can immediately follow the subject NP, and that other lexical categories like prepositions could not be pre-activated at the point of processing the subject NP. If prepositions can be pre-activated as well, then even when the verb turned out to be intransitive, there should be no reading disruption because intransitive verbs can be followed by a predictively activated preposition which may introduce a PP adjunct (e.g., *arrived for…*).

The present study has focused on filler-gap dependency processing, but the current conclusion is consistent with a broader class of models of sentence processing that propose that the parser utilizes a variety of sources of linguistic and contextual information to predictively build structural representations (Altmann & Kamide, 1999; Gibson, 1998; Hale, 2003; Kimball, 1975; Levy, 2008). On the other hand, the present

study does not reveal what kind of pre-verbal information is critical for triggering object gap creation in advance of the verb. One possible source that was already discussed in the Introduction is the grammatical knowledge of phrase structure rules, which suggest that the upcoming VP representation can contain an object NP slot. However, it is equally feasible that the parser could use non-grammatical information in predictively positing the object gap, such as differences in relative conditional probabilities derived from the lexical and contextual information contributed by the combination of the filler noun phrase and the subject. For example, even when a clause appears to resemble a gap structure like a relative clause, with a certain combination an adjunct gap may seem much more plausible than an object gap analysis (e.g., *the day that*… can continue as involving an adjunct gap as in *the day that I was born,* or an object gap as in *the day that I have been looking forward to*). Further studies are needed to investigate what kind of information contributes to such predictive object gap creation processes.

One may argue that there is an alternative interpretation of the results that still assumes that verb information plays a critical role in filler-gap dependency formation in English. For example, it is possible that filler retrieval processes are automatically activated as soon as the parser accesses the categorial information of the verb without accessing the transitivity information of the verb. Under this alternative account, the transitivity mismatch effect arises because the filler that was 'blindly' retrieved based on the verb categorial information mismatches the subcategorization property of the verb that is accessed later (see van Gompel & Liversedge, 2003, for a similar proposal for a gender mismatch effect in pronominal processing). Although this category-driven account of filler retrieval is compatible with our findings, we argue that there is little

evidence for this view. First, this account assumes that contents of lexical information are ordered, such that categorial information is temporarily accessed before the subcategorization property of the verb. However, there is little evidence to support such ordered access to category vs. other contents of a verb (Farmer, Christiansen, & Monaghan, 2006 is one rare case, but see Staub, Grant, Clifton, & Rayner, 2009 for a counterargument), whereas there is an abundance of psycholinguistic and neurolinguistic research demonstrating extremely fast access to various contents of lexical items (e.g., Federmeier, Segal, Lombrozo, & Kutas, 2000; Dambacher, Kliegl, Hofmann, & Jacobs, 2006; Hauk, Davis, Ford, Pulvermuller, & Marslen-Wilson 2006; Staub & Rayner, 2007; Tanenhaus, 2007; Almeida & Poeppel, submitted). On the other hand, there has been a recent surge of empirical work demonstrating that structure building processes can proceed predictively based on various types of top-down linguistic and contextual information, as discussed above (e.g., Konieczny, 2000; Kamide et al., 2003; Delong, Urbach, & Kutas, 2005; van Berkum, Brown, Zwisterlood, Kooijman, & Hagoort, 2005; Lau, Stroud, Plesch, & Phillips, 2006; Yoshida, 2006; Staub & Clifton, 2006; Yoshida, Dickey, & Sturt, in press), and the current work demonstrating extremely early object gap creation processes can be seen as another instance of such predictive structure building processes. However, further work is needed to more firmly establish that the hyper-active gap filling hypothesis is a better account for the pattern of results observed across a variety of paradigms than this alternative category-driven approach.

The current finding may also seem to contradict findings by Boland, Tanenhaus, Garnsey, and Carlson (1995) and Pickering and Traxler (2001). These authors tested the processing of filler-gap dependencies in sentences that contain verbs like *persuade* or

remind that can have both an NP slot and a clausal complement slot in their argument structure, and found no evidence for reading disruption when the filler was semantically incompatible with the direct object NP slot but compatible with the complement slot. According to the hyper-active gap filling account, encountering a *persuade*-type verb should not result in a transitivity mismatch effect since *persuade* does make available an object position, but one may wonder whether it should result in a plausibility mismatch effect when the filler is a semantically incompatible object, since an object-gap structure is hypothesized to be predictively constructed before the verb.

We can see two ways of reconciling these findings with the results presented here. First, the plausibility mismatch slowdown observed for simple transitive verbs may largely reflect the cost of reanalyzing the predicted structure to one that is compatible with the new input, which may vary depending on the argument structure of the verb. Revision may be costly in the cases where the verb is intransitive or mono-transitive and does not provide sufficient information for the parser to anticipate an alternative structural position for the filler, whereas in the *persuade/remind* cases, the revision may be less costly because the argument structure of the verb clearly indicates the presence of an upcoming clause in which the filler can be integrated. (For related discussions on variance in reanalysis costs as a function of alternative structural possibilities, see Fodor & Inoue, 1994; Gorrell, 1995; Sturt & Crocker, 1996, 1997; Weinberg, 1993). Second, the predicted filler-gap structure may be more abstract than we have indicated so far. Rather than specifically predicting an object gap when the filler and relative clause subject are encountered, the parser may simply predict an argument gap position somewhere inside the complement domain of an upcoming VP representation, such that a

gap in either the NP slot or in the clausal complement slot of *persuade*-class ditransitive verbs would be consistent with the prediction. The current results are compatible with either account.

Finally, the conclusion that the same filler-gap dependency completion procedure is used across head-initial and head-final languages suggests that the parser's structure building procedures, at least for filler-gap dependency completion, may not be qualitatively different across languages. However, this still leaves us with many open questions. For example, the evidence for hyper-active gap filling so far has come from Japanese (Aoshima et al., 2004; Nakano et al., 2002) and English (the present paper), but this line of work obviously needs to be extended to other languages to test the robustness of the hypothesis. Moreover, predictive dependency formation processes are observed in domains other than filler-gap dependency processing (e.g., resolution of backward anaphora; Aoshima, Yoshida, & Phillips, 2009; van Gompel & Liversedge, 2003; Kazanina, Lau, Lieberman, Yoshida, & Phillips, 2007), but it is not known whether there is cross-linguistic variation in other predictive structure building processes, and if such cross-linguistic variation exists, what its cause might be. These questions about the correspondence between linguistic properties and psycholinguistic processes must be addressed by further cross-linguistic investigations of parsing.

2.5 Conclusion

The two experiments reported in this chapter demonstrated that in filler-gap dependency processing, the parser completes the syntactic representation for filler-gap dependencies before accessing the lexical content of the tails of the dependencies. In the presence of filler-gap dependency, intransitive verbs consistently led to reading

disruption, a pattern that was replicated in self-paced reading measures as well as in eye movement measures. These findings suggest that the parser postulates the object gap at least as soon as it encounters the subject NP and is licensed to predict the presence of a VP, which could in principle accommodate an object NP. This allows us to conclude that the parser uses top-down information to compute the structural representation for filler-gap dependencies regardless of the word order differences across languages.

Let us now revisit the sequence of steps involved in filler-gap dependency computation that we proposed at the outset of this chapter (1).

(1) a. Actively construct the syntactic dependency between the filler and the gap.

b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

The findings in the present chapter lend support to the separation of the syntactic processes (1a) from the lexical and semantic processes in (1b). In the next chapter, we will examine how the syntactic representations that are completed early on feed the subsequent interpretive processes.

Chapter 3: Active interpretation of wh-dependencies in the visual world

In Chapter 2, it was argued that the syntactic process of gap creation is triggered independently of information from the verb, regardless of the word order properties of the language. As reviewed in Chapter 1, however, constructing syntactic dependencies is not sufficient for generating a sentence interpretation, which also includes interpretive processes that compute semantic, pragmatic and discourse-level representations. This chapter focuses on the interpretive processes involved in filler-gap dependency computation. The experiments reported in this chapter focus on how the syntactic representations constructed in (1a) are mapped onto linguistic representations computed in the interpretive processes in (1b).

(1) a. Actively construct the syntactic dependency between the filler and the gap.

b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

The experimental hypothesis explored in this chapter is this: If the syntactic dependency is completed early such that a wh-filler is integrated into the object gap position before the verb information is accessed, the semantic and subsequent interpretive processes over the verb+object combination should be initiated as soon as the verb information is processed. On the other hand, if a sentence did not involve an early association of the wh-filler and the verb, then at the point of encountering the verb, the

object of the verb is not available for the interpretation mechanism, hence the interpretive computation for the verb+object combination must wait until the object NP is encountered. In other words, using the sentences without filler-gap dependencies as a baseline, we attempt to establish a time course estimate for how long it takes to complete the interpretive component of filler-gap dependency computation in (1). The time course data on interpretive processes in this chapter thus complement the time course evidence for the syntactic processes reported in Chapter 2, and together provide a plausible timeline for the overall filler-gap computation procedure in real time.

In order to test this hypothesis, this chapter uses a visual world eye-tracking method that examines how listeners incrementally compute the semantic and discourse representations for filler-gap dependencies. As reviewed in Chapter 1, visual world eye-tracking measures have been extremely informative in understanding real-time language comprehension in relation to objects in the scene (Cooper, 1974; Tanenhaus et al., 1995; for reviews, see Henderson & Ferreira, 2004; Tanenhaus & Trueswell, 2005). In this paradigm, participants are presented with objects in the scene or images on a computer display, together with spoken language stimuli. Moment-by-moment fixations on the visual stimuli are monitored during the presentation of the language stimuli, such that the eye movement measures can be used to infer how the language stimuli are comprehended. Given that fixations on the scene reflect reference processes that are guided by semantic and discourse representations, this dependent measure is suitable for understanding how incremental interpretive processes unfold over time.

Sussman and Sedivy (2003) conducted a visual world experiment to examine the time course of interpretive processes in filler-gap dependency computation. The study

presented wh-questions like *What did Jody squash the spider with ___ ?* and reported anticipatory looks towards the direct object referent during the presentation of the verb, which the authors interpreted as reflecting active dependency completion. However, this chapter carefully re-examines the time course of fixations and the linking assumptions that connect active dependency interpretation and eye movements, and uses an improved experiment design to provide a more accurate timeline for the interpretive procedures.

This chapter is organized as follows. In section 3.1, we will review Sussman and Sedivy's work on filler-gap dependency processing. A critique of this study forms the starting point for the eye-tracking experiments reported in this chapter. Section 3.2 presents Experiment 3, which attempts to replicate Sussman and Sedivy's findings in our lab setting. Section 3.3 presents Experiment 4, which presents an improved fixation-based measure of active dependency interpretation. Section 3.4 presents a general discussion of the implications of the present findings for the mechanism of active dependency completion and for the advantage of cross-methodological investigations of psycholinguistic processes.

3.1 A critical review of Sussman and Sedivy (2003)

Sussman and Sedivy (2003) created a visual world experiment design for the investigation of active dependency completion, so it is useful to review their study in detail. Sussman and Sedivy presented on the computer screen a 3 × 3 grid display with four pictures in the corners (see Figure 4), e.g., milk (distractor), spider (competitor), Jody (subject), and shoe (target), while participants listened to a recording of a story, such as the following:

Jody was eating breakfast one morning when she saw a big hairy spider creeping across the table towards her. Jody, whose terrible arachnophobia had caused her to seek therapy a few years ago, drew on the techniques of relaxation and anxiety management that her psychologist had taught her. Instead of screaming or freaking out, she calmly took off her shoe and slammed it down on top of the spider. She ate the rest of her Fruit Loops in peace.

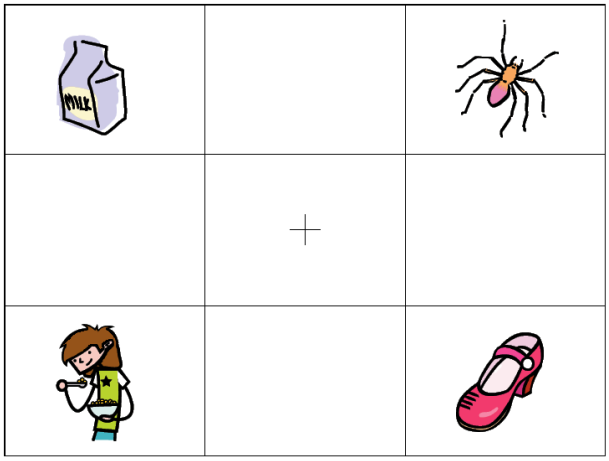


Figure 4. A sample display used in Sussman and Sedivy (2003); picture made by AO

After the story, participants heard a question about the story like (2) and answered the question based on what happened in the story.

(2) a. Wh-question: What did Jody squash the spider with?

b. YesNo-question: Did Jody squash the spider with her shoe?

The fixation proportion analysis of eye-movements in the verb region (i.e., while participants listened to the verb *squash*) revealed that a) there were more fixations on the theme object *spider* in the *wh*-question condition (2a) than in the baseline yes-no question

condition (2b), starting around 50ms after the verb onset, and b) there were more looks to the theme *spider* than to the instrument *shoe*, even though participants had at that point only heard the question only up to the verb and there was no direct bottom-up evidence that the object position contained a gap. Sussman and Sedivy interpreted these earlier looks towards the theme as evidence for active dependency completion triggered by the *wh*-dependency.

These data could be taken to suggest that interpretive processes in filler-gap dependency computation can be completed within 50ms after the verb onset, but there are reasons to think that this is unusually fast. Let us consider in (3) what cognitive processes need to take place between syntactic dependency completion and the point at which the interpretation of filler-gap dependencies could trigger eye movement.

(3) Processes that link active gap creation and eye movements

Syntactic processes:

a. Complete the syntactic representation for a filler-gap dependency.

Interpretive processes:

b. Access the lexical information of the verb, and construct semantic and discourse representations based on the verb+object combination.

Referential processes:

c. Relate the semantic representation to memory representation of the events in the story.

d. Program and launch an eye movement towards the correct object.

Let us now examine how much time is necessary to execute each step illustrated in (3). As discussed in Chapter 2, the syntactic process of dependency completion could be completed at least as soon as the categorial information of the verb is identified before other contents of the verb are accessed. It is not clear how much time it takes to complete the entire process of (3b) and (3c), but it has been argued that lexical access alone takes ~200ms (Almeida & Poeppel, submitted; Sereno & Rayner, 2003; cf. Dahan, Magnuson & Tanenhaus, 2001).[3] With respect to (3d), the psychophysics literature on eye movement suggests that it takes at least 150-200ms to program eye-movement (Matin, Shao & Boff, 1993; Saslow, 1967) and perhaps even as long as 400-500ms in a visual

[3] It is conceivable that the lexical access portion may take less than the ~200ms estimate here, given that in this experiment the potential referents are extremely limited, and each lexical item corresponding to the pictures as well as verbs that could describe relations between them may be highly activated. See Allopena, Magnuson, and Tanenhaus (1998) for discussion.

world language processing experiment if an object in a display is mentioned without any preceding cues (Altmann & Kamide, 2004). Taken together, it seems likely that the early eye movements towards the theme object in Sussman and Sedivy's wh-condition must have been programmed before the verb onset.

One possible explanation for the extremely early anticipatory looks to the theme object is that the set-up of the story and visual display may have allowed participants to compute some of the steps in (3) even before the verb was encountered. In the example story described above, for instance, the story involves only one primary event (i.e., squashing the spider), and given that the subjects were looking at the pictures of the spider, Jody, the shoe and a distractor picture, and knew that they were going to hear a question about one of those four pictures, participants may have been able to narrow down the set of possible verbs and predict that a verb like *squash* was likely to follow. This prediction about an upcoming verb may be plausibly generated as soon as the wh-phrase (what) and the subject (Jody) are processed, because at this point it should be clear to participants that the question is about something that Jody did in the story.[4] This could explain why the increase of looks towards the theme object occurred extremely early in Sussman and Sedivy's study, but it also indicates that the early identification of the theme might be an artifact of the experiment design and may not reflect processes involved in normal wh-dependency interpretation, where the upcoming verb cannot be predicted in advance.

[4] Sussman and Sedivy included stories in which multiple events occur, but given that the relevant set of pictures was already previewed by the participants while listening to the story, it may still not have been difficult to infer the content of the question.

The two experiments with adults reported in the next sections attempt to address the methodological concerns raised above. Experiment 3 (Section 3.2) attempts to replicate Sussman and Sedivy's study in our lab in order to examine the reliability of their timing estimates. Experiment 4 (Section 3.3) addresses the concern about the use of a task-taking strategy by using an improved experiment design.

3.2 Experiment 3: Replication of Sussman and Sedivy (2003)

3.2.1 Method

Participants

17 members of the University of Maryland community were either paid or given course credit to participate in this experiment. All were native speakers of American English, and either had uncorrected normal vision, or wore soft contact lenses or eye glasses.

Materials

The present study adopted exactly the same design as Sussman and Sedivy's study, in which participants were presented with a narrative while they looked at a 3 × 3 grid with four pictures in the corners, and heard either a *wh*-question or a yes-no question about the story, which they were instructed to answer aloud. We used the ten sets of narratives and questions provided in the appendix of Sussman and Sedivy (2003). Each *wh*-question had a structure in which the *wh*-phrase *what* is extracted from a PP (five instrument PPs with the preposition *with*, and five locative PPs, with three occurrences of *in*, one of *under* and one of *on*) as in *What did Jody squash the spider with?* Yes-No questions had an overt PP object, as in *Did Jody squash the spider with her shoe?*, and

the answers to the target yes-no questions were always "yes." The target narratives and sentences are listed in Appendix C.

The experiment included new 10 filler narratives of a similar structure. Following Sussman and Sedivy, we constructed five filler *wh*-questions that contained a direct object question (e.g., *What did Jody squash?*),[5] as well as five filler yes-no questions with the same structure as the target yes-no questions but with a correct answer of "no." The target questions were counter-balanced across participants such that each participant heard only one version of a target question.

The experiment used 3 × 3 grid displays for each narrative, using freely available clip-art images, as shown in Figure 4 above. Each display contained a competitor object that corresponded to the theme of the verb (e.g., a spider), as well as a target object that corresponded to the object of the preposition (e.g., a shoe). Items were counter-balanced using a Latin Square design to ensure that the target answer to a wh-question was equally likely to appear in any of the four corners of the display. The narratives and questions were read with normal prosody by a female native speaker of American English, and recorded with a sampling rate of 44.1 kHz. The sound files were incorporated into a QuickTime movie together with the 3 × 3 grid displays, such that the onset of the sound files and the onset of the visual stimulus presentation would be synchronized. The sounds were presented to participants via two speakers positioned on either side of the viewing

[5] We did not have access to the actual filler sentences that Sussman and Sedivy used, and based on their description of the filler items we could not tell whether their filler *wh*-questions contained a prepositional phrase as in *What did Jody squash with her shoe?*, as opposed to simply terminating the sentence after the object gap as in *What did Jody squash?* We chose the latter option in our filler stimuli, though this may not have been an ideal choice in that this makes the targets the only *wh*-questions with PPs and skews the distribution of sentences.

monitor. The onsets of each word region in the target sentences were determined by inspection of each sound file in Praat? for the purpose of data analysis.

Procedure

Participants were seated in front of a 20-inch display with their eyes between 20 and 25 inches away from the monitor. They were instructed that in each trial they would first see a grid display with four pictures while listening to stories, then a 1000ms display of a fixation cross, and finally the four pictures again together with auditory presentation of a question. Participants were instructed to answer the questions aloud[6] and to keep their eyes on the computer screen during the experiment so that they could remember the story well and answer the questions correctly. After the instructions, participants were shown one practice item, and then they put on an ISCAN ETL-410 lightweight video-based head-mounted eye tracking system. The eye camera captured an infrared image of the eye at 60 Hz and determined monocular eye position by monitoring the locations of the center of the pupil and the corneal reflection. A scene camera was mounted on the side of the visor, providing an image of the participant's field of view. Each participant underwent a brief five-point calibration procedure prior to the experiment.

The twenty narratives and questions were divided into four blocks, and the presentation of the stimuli was controlled by Psyscope X. Calibration was carefully monitored throughout each trial. Minor adjustments were occasionally made between trials. Moreover, at the end of each block the experimenter instructed the subjects to look

[6] The fixation cross before the question phase was inserted to ensure that the participants were paying attention to the display before the question phase began. This was not included in the original Sussman and Sedivy study, but we decided to include this in order to more closely match the procedure for Experiment 4, which we conducted prior to Experiment 3.

again at the five calibration points in order to re-check calibration accuracy. The entire experimental procedure, including giving informed consent, instructions, calibration, the experiment, and debriefing took approximately 30 minutes.

Data analysis

The data were collected from videotape records using Supercoder (Hollich, 2005). For each subject and trial, coders used the crosshairs generated by the eye tracker to establish which of the five objects in the display (i.e., four pictures and the cross in the center) was fixated at each time frame (30 per second), beginning at the onset of the target trials. Fixations were coded on each trial from the onset of the target picture display following the fixation cross until the pictures disappeared at the end of the trial. The subject's gaze had to remain on the same quadrant for more than one frame in order to be counted as a fixation. If blinking occurred, fixation data was lost, typically for one to three frames. Fixations during this interval were attributed to the previous object being fixated. For the fixation proportion analysis, four word regions (Subject, Verb, NP and Preposition) were defined. Because the stimulus sentences varied in length, we followed Sussman and Sedivy and re-synchronized the manually coded data to the onset of each respective word region for each trial, such that for all trials the relevant region onset was controlled to be the same. Moreover, since it takes approximately 200ms to program an eye movement, the word regions were shifted backwards by 200ms, such that in each trial the word region started 200ms after the actual word onset and ended 200ms after the actual word offset. Next, each region was divided into a series of 100ms analysis

windows based on the average duration of each region.[7] This approach allows us to minimize the loss of time course information while retaining a sufficient number of observations to produce robust effects (Altmann & Kamide, 2004; Boland, 2005; but cf. Mirman, Dixon, & Magnuson, 2008). For each 100ms analysis window, we calculated fixation proportions across the five objects in the display (i.e., the four pictures and the fixation cross).

3.2.2 Results

Out of 170 target trials, 9 trials were lost due to technical failures or experimenter errors. We first calculated the frame-by-frame proportion of fixations to the objects of interest and plotted the average proportion across participants to the following figures: Figure 5 shows the fixation proportions in the wh-condition, Figure 6 shows the fixation proportions in the yes-no condition, and Figure 7 shows the proportion of looks to the target vs. the theme object (i.e., the spider in the example display) in the two conditions. In all figures the zero time point is aligned to the onset of the verb region since the verb region is the earliest point at which the filler-gap dependency can be interpreted.

[7] This is a rather non-conservative definition of 100ms windows, as it means that for some of the trials, the later 100ms windows already include fixations after the onset of the next word (cf. Boland, 2005). Also, for some of the word regions the average duration was not a multiple of 100ms units. In such cases, the last 100ms window was defined inclusively, i.e., the last 100ms window included frames from the next word region as well.

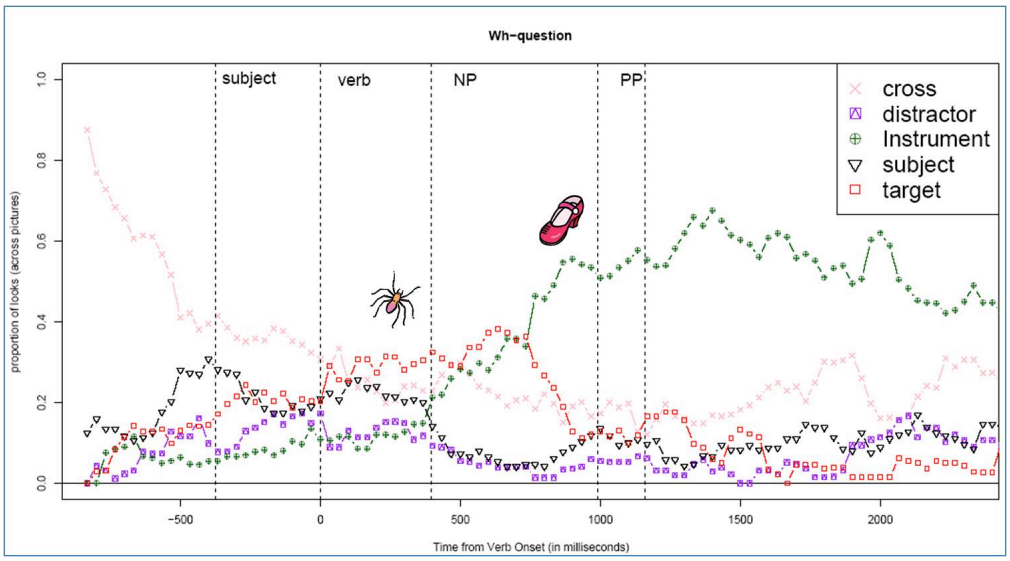


Figure 5. Proportion of fixations to displayed items in the wh-condition

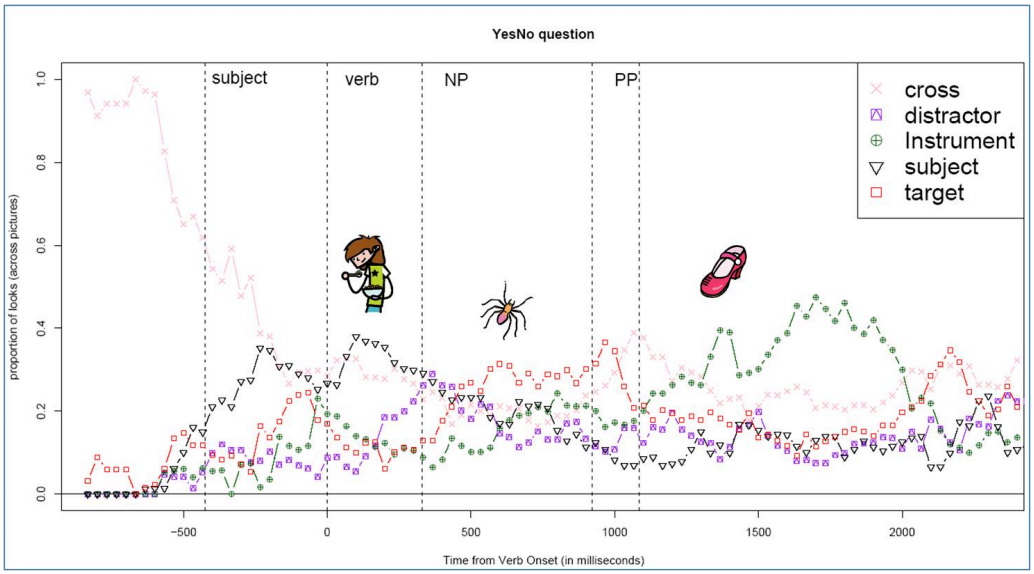


Figure 6. Proportion of fixations to displayed items in the yes/no-condition

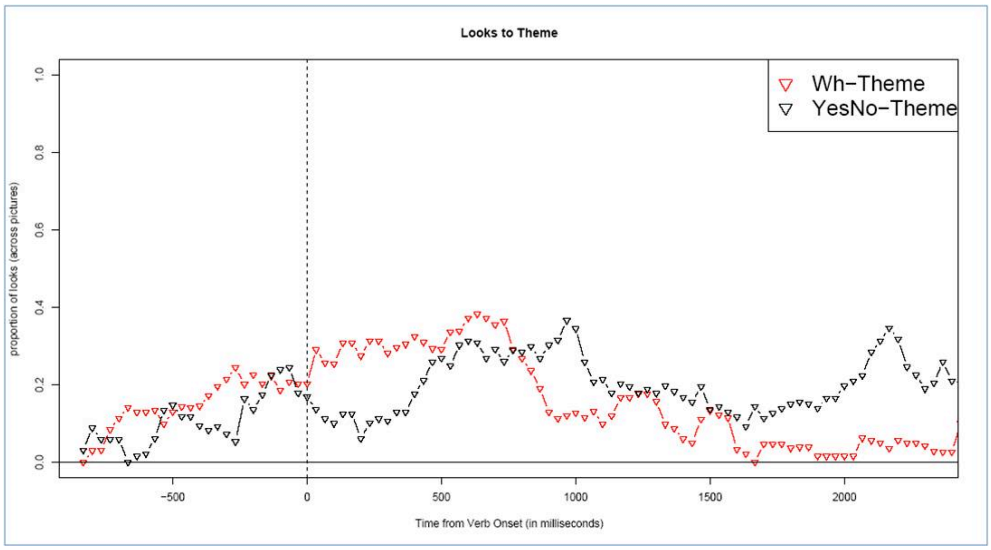


Figure 7. Comparison of fixations to the target theme (spider). The dotted line indicates the onset of the verb.

Qualitatively, these graphs indicate fixation patterns that are very similar to those found by Sussman and Sedivy. In the wh-condition (Figure 5), there was an increase of looks towards the target theme during the verb region, starting roughly around 2 frames (i.e., 66ms) after the verb onset. In the subsequent NP region ("the spider"), the proportion of looks to the target remained high but was gradually superseded by looks towards the PP object, which corresponds to the answer to the target wh-questions. In the yes/no question condition (Figure 6), however, the picture of the subject received the highest proportion of looks among the four objects during the subject NP and verb regions. The theme picture received the largest proportion of fixations after the onset of the direct object region, and the looks towards the instrument showed a noticeable increase only after the onset of the PP object. A comparison of the proportion of looks towards the target picture in the two conditions (Figure 7) clearly indicates the difference in fixation proportion during the verb region. However, the graph also highlights the fact that the looks towards

the target were already fairly high in the wh-condition even prior to the onset of the verb. This suggests the possibility that participants might have maintained fixation on the target object from the subject region onward.

We next calculated the average proportion of fixations on the theme image for the four 100ms time windows (T1-T4) starting at the verb onset, which roughly covered the entire verb region. This is summarized in Table 2.

Table 2. Experiment 3: Mean fixation proportion across pariticipants in 100ms windows 1-4 within the verb region

	T1. 0-100ms theme	T2. 100-200ms theme	T3. 200-300ms theme	T4. 300-400ms theme
Wh	0.31	0.3	0.29	0.33
Y/N	0.10	0.12	0.17	0.25

The average fixation proportions were submitted to a repeated measures ANOVA with question type (wh vs. yes-no) as an independent variable.[8] There was a main effect of question type in the first 100ms time window [T1: $F(1, 16) = 6.66$, $p < .05$], and a marginal effect of question type in the second 100ms time window [T2: $F(1, 16) = 2.86$, $p = .11$], but there was no main effect in the third and fourth 100ms time window ($Fs < 2$, $p > .2$). These results suggest that only the first 100ms window in the verb region resembled the pattern of results reported in Sussman and Sedivy, and that in the later

[8] Sussman and Sedivy included picture type (theme vs. instrument) and region (pre-verbal vs. verbal vs. post-verbal) as additional independent variables. However, we excluded these factors from our analyses since our main concern is in the pattern of fixation on the theme object in relation to the onset of the verb region.

parts of the verb region there was no significant difference between the wh- and yes-no conditions.

3.2.3 Discussion

This experiment attempted to replicate Sussman and Sedivy's visual world measure of active gap creation. Our fixation proportion analysis replicated Sussman and Sedivy's main finding, namely that during the verb region (more specifically, in the first 100ms window) the theme object received more fixations in the wh-condition than in the yes-no condition, and also that the theme object received more fixations than the PP object did. We argued in Section 3.1 that not all the cognitive processes involved in filler-gap dependency interpretation and eye movement to the theme object can be completed as early as Sussman and Sedivy's data indicated (i.e., within 50ms). In our data analysis we shifted the word regions by 200ms, and therefore the 0-100ms analysis window actually corresponds to 200-300ms after the actual verb onset. This allows more time to complete relevant cognitive processes than in the 50ms window that Sussman and Sedivy's data suggested, but it remains the case that in order to launch a significantly larger amount of looks to the theme in the wh-condition, the lexical access to the verb and subsequent interpretive processes must be completed within 100ms (see Section 3.1 for a review), which still seems rather implausible.

In Section 3.1 we suggested that such extremely early looks might be possible because the constraining visual display shown during the story could have allowed participants to predict the upcoming verb as soon as they heard the wh-filler and the subject NP. Our present data seem to be compatible with this interpretation, as we see a significant difference between the wh- and yes-no conditions only in an early part of the

verb region. If this alternative interpretation of Sussman and Sedivy's findings is on the right track, then we predict that manipulating or eliminating potential sources of such question expectations in the experimental design should yield increased looks to the theme object only in later time windows than in Sussman and Sedivy's data and our data in Experiment 3. The next experiment tests this prediction.

3.3 Experiment 4: Two-event story design

Experiment 4 addresses the possibility that the anticipatory looks observed in Sussman and Sedivy's experiment may have been due to a verb expectation that resulted from the combination of the story content and the display that constrained a possible set of events that could be questioned about. In order to address this possibility, this experiment adopted essentially the same design as Sussman and Sedivy, but each story included two events that required two distinct verbs (e.g., *eat* vs. *wash*) as well as two distinct theme pictures (e.g., *cake* vs. *dishes*) from both events in the display. Given that the first potential gap position followed the verb and that either event could be queried with the same sentence up to the subject, including two events in each story should force participants to recognize the verb first before computing the relevant interpretive processes and launching an eye-movement towards the correct theme object. If anticipatory fixations on the theme during the verb region are observed more in the *wh*-condition than in the yes-no condition, despite there being two events that could have been queried, then we can be more confident that this truly reflects the time course of interpretive processes in filler-gap dependency processing. Moreover, this experiment was also designed with the intention of testing children's wh-dependency processing (a

study currently in progress), and for this reason various aspects of the design were deliberately made child-friendly.

3.3.1 Method

Participants

Twenty-four members of the University of Maryland community were either paid or given course credit for their participation in this experiment. All were native speakers of American English, and either had uncorrected normal vision or wore soft contact lenses or eye glasses.

Materials

Questions. We constructed a new set of ten *wh*-questions and yes-no questions (listed in Appendix D) as well as ten fillers, but the structure of the questions as well as the type of fillers was essentially the same as in Experiment 3, with two minor differences. First, in order to make the task more interactive and child-friendly, the question now took the form of an indirect question like *Can you tell me…*asked by a third-person character, Dora the Explorer (see below for more details). Second, the verb form was converted into progressive *was V-ing* to make the verb duration longer so that there would be a sufficient number of time frames for us to observe eye movements triggered by the verb before the onset of the theme object NP. As in Experiment 3, the target items were counter-balanced across the two lists such that each participant heard only one version of the target questions.

Story and display design. The story and display design for this experiment differed from the previous experiment. First, in order to make this experiment more engaging and child-friendly, we made the character in the display narrate the story, and

included animation of objects. For example, in a scene where the character is described to be eating a cake with a fork, the image of the fork moved to the position of the cake. The characters never moved since it was already clear that they were involved in all events. Moreover, each animation left a "trace" of what happened: for example, after a character ate a piece of cake, one piece of the cake disappeared and the fork that was used to eat the cake became slightly dirty. This was intended to help participants to encode a visual record of the story and thereby to facilitate relevant anticipatory looks.

Second, as noted above, we instructed the participants that the goal of the task was to answer the questions asked by Dora the Explorer, who observed the story with the participants. Dora appeared after the story phase and commented briefly on the story before yielding the screen to a fixation cross, which was presented for 1000ms. Then the last display from the story phase reappeared, together with the auditory presentation of Dora's question about the events in the story. The list of two events used in each target story is summarized with the questions in Appendix D. A sample narrative and target questions are presented below (the animation is described in square brackets):

Hi, my name is Emily. Today I'd like to eat some cake, but I also need to wash some dishes. Mmm, what should I do first? I think I'm gonna eat the cake. For that I need a fork. [the fork moves to the cake and a piece of the cake disappears] Mmm, that cake was yummy! Now it's time to wash the dishes. I'm gonna need to use a sponge. [sponge bubbles up and moves to the dishes. The dishes become shiny] Oh, those dishes are so clean, I did a great job today!

[Dora appears] That was a good story. Now let me ask you a question:

[Dora disappears; fixation cross appears for 1000 ms]

Wh-question: Can you tell me what Emily was eating the cake with?

Yes-no question: Can you tell me if Emily was eating the cake with the fork?

The third difference, which was crucial to this experiment, was that all the narratives contained two events, which were described with two different verbs and two instruments (in the example above, eating a cake with a fork vs. washing the dishes with a sponge), but had the same agent. Accordingly, the display contained pictures of each set of objects (theme and instrument) from the two events as well as a picture of the character, resulting in a display with five pictures. The beginning and the end phase of a sample display is shown in Figure 8.

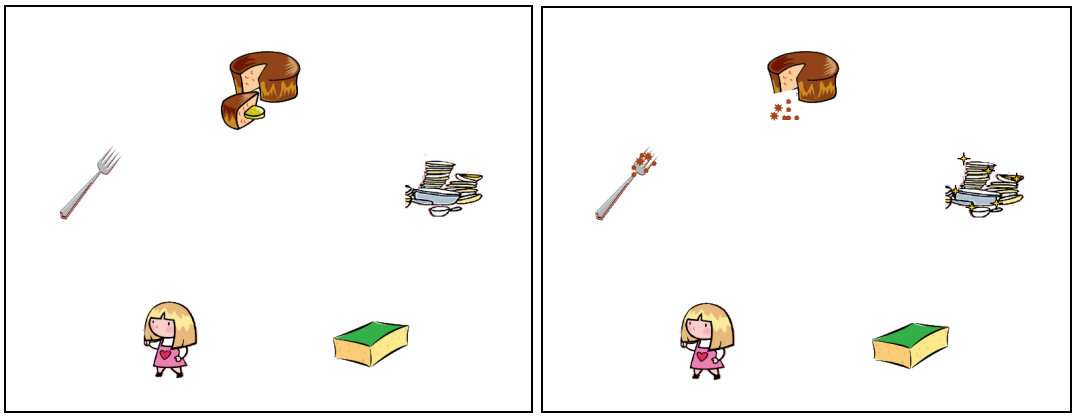


Figure 8. A sample display used in Experiment 4 (the initial phase on the left, the final phase on the right). The pictures in this sample display are a cake (target theme), some dishes (competitor theme), a sponge (distractor), Emily (subject) and a fork (instrument).

The displays were again constructed with freely available clipart images. The position in which each object type appeared was counter-balanced across items such that they could appear in any position on the display. In this experiment the grid lines were removed so as to make it seem more natural for the objects to interact with each other.

Each narrative was recorded by a female native speaker of American English with a sampling rate of 44.1 kHz, and Dora's lead-in and questions were recorded by another female native speaker of English. The narratives and questions were read with normal but careful prosody. The sound files were incorporated into a QuickTime format movie file together with the pictures in the same manner as in Experiment 3.

Procedure and data analysis

The experimental procedure and the data analysis method were identical to those in Experiment 3.

3.3.2 Results

Out of 240 target trials, 14 trials were lost due to technical failures or experimenter errors. As in Experiment 3, we calculated the frame-by-frame proportion of fixations to the objects of interest and plotted the average proportion across participants in the following figures: Figure 9 shows the fixation proportion in the wh-condition, Figure 10 shows the fixation proportion in the yes-no condition, and Figure 11 compares the proportion of looks to the theme object (i.e., cake in the example display). In all figures the zero time point on the x-axis is aligned to the onset of the verb region, and the word regions were shifted backwards by 200ms such that in each trial the word region started 200ms after the actual word onset and ended 200ms after the actual word offset.

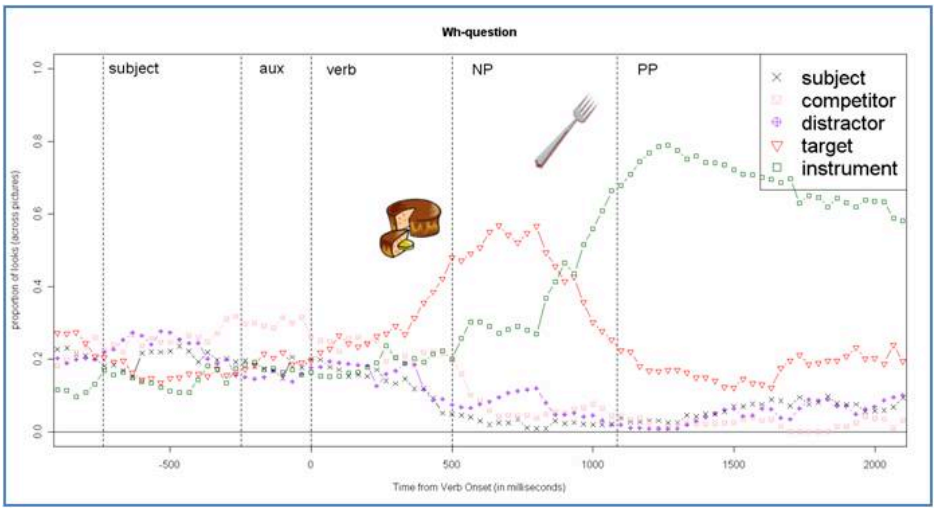


Figure 9. Fixation proportion to displayed items in the wh-condition

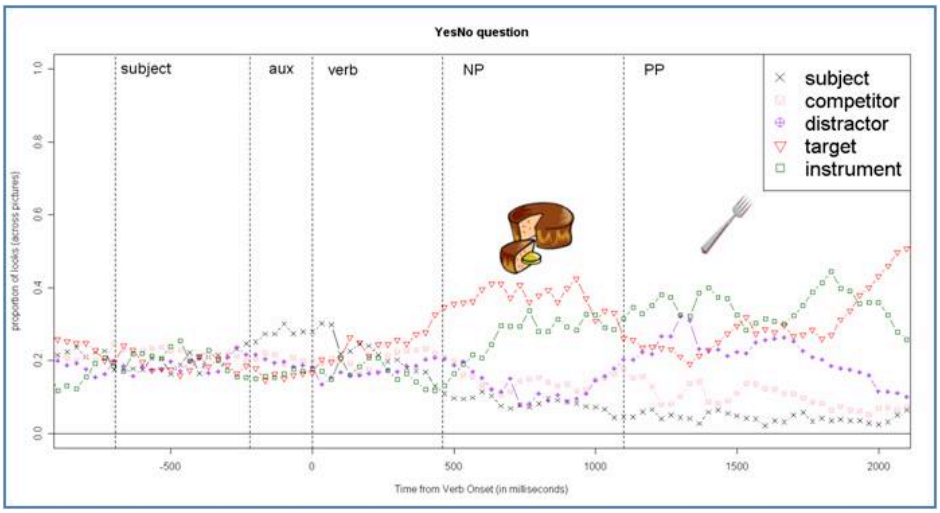


Figure 10. Fixation proportion to displayed items in the yes/no-condition

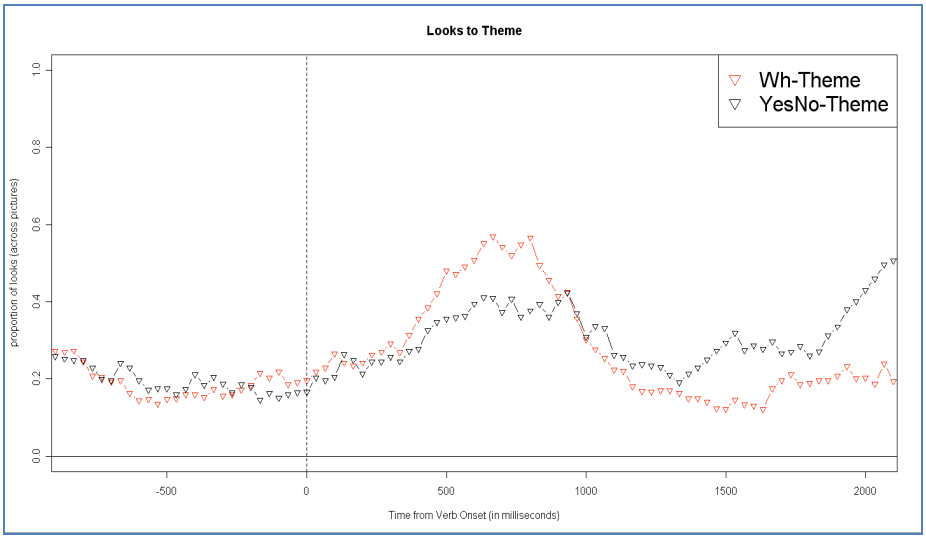


Figure 11. Comparison of fixations to the target object (‘the cake’). The vertical dotted line represents the verb onset.

The pattern of fixations in these figures presents a similar but slightly different picture from that of Experiment 3. As Figure 9 shows, in the wh-condition the fixations on the target object started to increase only around the end of the verb region, unlike in

Experiment 3, where the looks to the theme started to increase much earlier. The fixations on the theme continued to increase up to the middle of the NP region. Starting around the end of the NP region, the instrument picture received the highest proportion of fixations, and this fixation pattern continued after the onset of the preposition. The pattern of fixations in the yes-no condition (Figure 10) is very similar to that of the wh-condition in terms of the timing of the increase in looks to the theme as well as to the instrument picture, but the amount of looks to either of these objects was much reduced relative to the wh-condition. Figure 11 shows the looks towards the target theme object in both conditions, and here the difference in fixation proportion between the two conditions is clearly highlighted from around 400ms to 1000ms after the verb region onset.

In this experiment, there were two possible theme targets that participants could in principle fixate upon in the verb region: either the correct theme (i.e., the cake) or the incorrect theme (i.e., the dishes). Fixations on the incorrect theme are plausible if participants used a question expectation strategy discussed in Section 3.2 and randomly chose to fixate on a theme object from either of the two events. For this reason, the average proportion of fixations on the correct theme object and the incorrect, competitor theme object was calculated for five 100ms time windows (T1-T5) that roughly corresponded to the entire verb region. This is summarized in Table 3.

Table 3. Experiment 4: Fixation proportions across subjects in 100ms windows 1-5 within the verb region

	T1. 0-100ms		T2. 100-200ms		T3. 200-300ms		T4. 300-400ms		T5. 400-500ms	
	correct theme	wrong theme	correct theme	wrong theme	correct theme	wrong theme	correct theme	wrong theme	correct theme	wrong theme
Wh	0.24	0.22	0.31	0.19	0.37	0.19	0.44	0.18	0.52	0.08
Y/N	0.24	0.19	0.27	0.23	0.32	0.20	0.36	0.16	0.36	0.12

For each analysis window, the average fixation proportion was submitted to a 2 × 2 repeated measures ANOVA with question type (wh vs. yes-no) and picture type (correct theme vs. incorrect theme) as independent variables. There was a main effect of picture type in the third, fourth and fifth 100ms window [T3: $F(1, 23) = 13.10$, $p < .005$; W4: $F(1, 23) = 24.81$, $p < .005$, T5: $F(1, 23) = 70.21$, $p < .005$] but only the fifth window (400-500ms) showed a significant interaction of question type and picture type [$F(1, 23) = 5.91$, $p < .05$]. A pair-wise comparison in this analysis window shows a significant difference in fixations to the correct-theme between the wh- and yes-no question conditions [$F1(1, 23) = 6.26$, $p < .05$]. This pattern of results suggests that overall the correct theme received more fixations than the incorrect, competitor theme starting around 200-300ms after the verb region onset, but crucially the difference between wh- and yes-no conditions did not emerge until 400-500ms after the verb region onset.

3.3.3 Discussion

The present visual world experiment examined active filler-gap dependency interpretation in a context in which verb recognition was necessary in order to compute

interpretive processes and launch an eye movement towards the correct theme object. In our fixation proportion analysis of the verb region, we found more looks to the correct theme than to the incorrect theme, suggesting that the recognition of the verb led to fixations on the correct theme object during the presentation of the verb. More importantly, in the verb region we also observed a larger proportion of anticipatory looks to the theme in the wh-condition than in the yes-no condition, but crucially this difference did not emerge until the 400-500ms time window after the verb region onset, which corresponded to 600-700ms after the actual verb onset.

The present findings clearly contrast in the timing of anticipatory looks in Experiment 3, where the difference in the amount of anticipatory looks emerged within 100ms after the verb onset. Given that in the current experiment design the identification of the correct theme object was impossible until the verb was recognized, we can reliably infer that 600-700ms was the minimally necessary amount of time in order to complete the interpretive processes for the filler-gap dependency. In fact, the linking assumptions between filler-gap dependency computation and eye movements stated in (3a-d), repeated here for convenience, can be reasonably executed in 600-700ms, considering that the syntactic processes in (3a) could be completed before the verb is recognized (Chapter 2), and that the lexical access at the verb (3b) and programming of eye-movements (3d) plausibly require roughly 400ms together.

(3) Processes that link active gap creation and eye movements

Syntactic processes:

a. Complete the syntactic representation for a filler-gap dependency.

Interpretive processes:

b. Access the lexical information of the verb, and construct semantic and discourse representations based on the verb+object combination.

Referential processes:

c. Relate the semantic representation to memory representation of the events in the story.

d. Program and launch an eye movement towards the correct object.

Interestingly, the time course estimate of filler-gap dependency computation that we have established based on eye movement measures appears to correlate with the time course of filler-gap dependency processing in ERP studies. There are two ERP findings related to filler-gap dependency processing. First, Garnsey, Tanenhaus and Chapman (1989) reported that when there was a semantic misfit between the wh-filler and the verb, an N400 was observed at the verb position. Second, in the absence of any syntactic or semantic misfit, a P600 is observed at the verb where a filler-gap dependency can be completed (Kaan et al., 2000; Phillips et al., 2005; Gouvea et al., 2009). We could interpret the N400 effect as an index of lexical access processes illustrated in (3b), and the P600 effect may plausibly correspond to the discourse integration and referential processes described in (3b) and (3c). However, to what extent this superficial correspondence is on the right track ultimately depends on the nature of the N400 and

P600. There is a debate about whether N400 reflects pre-activation of lexical features (Federmeier & Kutas, 1999; Lau, Phillips, & Poeppel, 2008) or semantic and discourse integration of lexical items (Hagoort, 2003; Otten & van Berkum, 2007). Similarly, the P600 has been traditionally seen as an index of syntactic integration processes (e.g., Friederici, 2002; Hagoort, 2003), but recent evidence suggests that the P600 is modulated by semantic and discourse-level expectations and therefore reflects discourse integration processes (Kuperberg, 2007; Nieuwland & Kuperberg, 2008; Osterhout, Holcomb, & Swinney, 1994). However, the observation of correlation here at least suggests the possibility that cross-methodological investigations of psycholinguistic processes may shed light on the much debated nature of N400 and P600 effects.

In summary, this timing difference in anticipatory looks between Experiments 3 and 4 not only lends support to the view that the anticipatory fixations on the theme in Sussman and Sedivy's study were driven by predictive processes prior to the recognition of the verb, but also allows us to confidently argue that the anticipatory looks observed in Experiment 4 indeed index the active interpretative processes involved in filler-gap dependency computation.

3.4 Conclusion

This chapter reported a series of visual world eye-tracking experiments that examined the time course of interpretation in filler-gap dependency computation. Experiment 3 used the stimuli from Sussman and Sedivy (2003) and replicated their finding that anticipatory fixations on the referent of the object NP increased within 200-300ms from the onset of the verb. However, we argued that this is implausibly too early given the set of cognitive processes that are necessary to arrive at a sentence

interpretation in filler-gap dependency processing and launch an eye movement to establish reference. We hypothesized that this early identification of the object NP was caused by the combination of the display, story content, the wh-phrase and subject NP which together allowed listeners to predict the upcoming verb. This would essentially allow listeners to complete early interpretive processes as soon as the subject NP is processed. Experiment 4 addressed this possibility by presenting stories with two events that can be queried with the same string of words up to the verb region. Given that listeners cannot predict which event the question is going to concern, this design effectively forces listeners to identify the verb first. The fixation results demonstrated evidence for active interpretation of wh-dependencies within 600ms from the verb onset, suggesting that the entire interpretive processes in filler-gap dependency computation take 600ms to execute. We discussed the close correspondence between the time course of filler-gap computation established by these eye movement measures and ERP evidence for filler-gap dependency processing.

The present findings and our approach to the studies reported in Chapters 2 and 3 have theoretical and methodological implications. One theoretical implication is that the sentence interpretation mechanism integrates linguistic and non-linguistic information to *predictively* process the input. In Chapter 2, we discussed whether the object gap creation process in English that proceeds independently of the lexical information from the verb is truly predictive or not. The predictive account attributes the verb-independent object gap creation to the use of top-down, grammatical knowledge (e.g., phrase structure rules) that allows the parser to predict that the verb phrase is a potential gap licensing position, which may effectively lead the parser to expect a transitive verb. On the non-predictive

account, the parser first accesses the categorial information of the verb in advance of the transitivity information, and the access to the categorial information triggers the filler retrieval process. We discussed various reasons why the predictive account was more attractive, but did not present a decisive argument. However, the studies reported in Chapter 3 lend further support to the predictive account. Experiment 4 with the two-event story design suggested that the interpretive processes initiated by recognition of the verb take place within 600ms from the verb onset, but Experiment 3, which used stimuli from Sussman and Sedivy (2003), showed that the interpretive processes can be completed within 200-300ms from the verb onset. Given that the set of operations involved in filler-gap dependency interpretive processes must be the same in Experiments 3 and 4, this time course discrepancy suggests that some of the interpretive processes were predictively executed in Experiment 3. We argued that the extremely constraining story content along with the display plausibly allowed listeners to predict the upcoming verb as soon as the wh-phrase *what* and the subject NP were processed, such that a subset of the interpretive processes could be initiated earlier. Thus, the present observation presents strong empirical support for the view that sentence interpretation processes do in fact involve predictive processes.

One methodological implication of this finding is that it is extremely important to carefully consider the linking assumptions between cognitive processes of interest and the dependent measures in order to establish precise time course evidence for real-time language comprehension mechanisms. To take the visual world paradigm as an example, there are an infinite number of reasons for fixations on the images or objects in the scene. In order to infer the mechanism of language comprehension processes based on such

fixations, the experiment design must be constructed in such a way that fixations on the scene are clearly triggered by the cognitive processes of interest (for discussions, see Boland, 2004; Tanenhaus & Trueswell, 2005). In Chapters 2 and 3, we attempted to be explicit about the link between what kind of psycholinguistic processes must occur in filler-gap dependency computation, how much time it should take to execute each of those processes, how much time it should take for those processes to be reflected in the dependent measures of choice, and why certain dependent measures should be most suitable for attesting the cognitive process of interest. As a consequence, we were able to establish time course evidence for the real-time mechanism of filler-gap computation as a whole, and this approach is promising in that it can be extended to other linguistic computation to gain more precise understanding of how mapping between different levels of linguistic representations might take place in real time. Moreover, being explicit about the mechanism of language processing and the dependent measures allowed us to begin to see a link between diverse time course measures of language processing. For example, we saw above that the time course of filler-gap dependency processing measured in eye movement seemed to correlate with the time course of filler-gap dependency processing in ERP studies, and this suggests the possibilities that we can use cross-methodological evidence to further our understanding of real time computation of linguistic representations (for related discussions, see Sereno & Rayner, 2003).

Chapter 4: Kindergarten-path effects in English and Japanese wh-processing

As discussed in Chapter 1, incremental syntactic and interpretive processes are useful for maximizing the efficiency of sentence interpretation; but on the other hand, incremental commitments may be disconfirmed by later information in the sentence and need to be retracted, which could potentially cause a large processing burden.[9] Chapters 2 and 3 have focused on understanding the real time mechanism of filler-gap dependency computation, and have provided time course evidence from various on-line measures that the parser actively constructs syntactic representations before accessing the lexical information at the tail of the dependency. Furthermore, it has been shown that subsequent interpretive processes construct semantic and discourse representations in a few hundred milliseconds after the syntactic representation for the dependency is constructed.

Chapters 4 and 5 examine the consequences of active syntactic processing for children's reanalysis mechanisms. There has been a surge of interest in the development of reanalysis mechanisms in children since the seminal study by Trueswell and colleagues (Trueswell et al., 1999) that demonstrated that children only entertain the destination interpretation of the first PP *on the napkin* in (1), and later fail to retract this incremental commitment (the so-called 'kindergarten-path effect').

(1) Put the frog on the napkin in the box.

[9] The work reported in this chapter was done in collaboration with Imogen Davidson White and Takuya Goro.

Subsequent studies on children's reanalysis abilities have focused on uncovering the underlying cognitive mechanism that is responsible for kindergarten-path effects. For example, it has been proposed that children's revision difficulties can be attributed to the immature executive function mechanism (Choi & Trueswell, 2010; Novick et al., 2005). Executive function refers to a set of cognitive processes that are necessary for inhibiting automatic responses to stimuli and controlling adaptive and efficient responses to novel or difficult situations. While the exact nature of this mechanism is still debated, various cognitive control studies have reported that executive function abilities do not fully develop until late adolescence (Diamond, 2002; Davidson, Amso, Anderson, & Diamond, 2006). Thus, it is possible that kindergarten-path effects arise from children's inability to inhibit early, erroneous syntactic commitments in order to allow new syntactic and interpretative processes to unfold (for discussion, see Mazuka, Jincho, & Onishi, 2009). If this is the case, then as the executive function system matures in children, the kindergarten-path effects are expected to disappear. The studies that explore this hypothesis are beginning to examine to what extent children's performances in general inhibition control tasks correlate with their ability to revise their syntactic commitments in sentences like (1) (e.g., Choi, 2010).

These investigations of developing cognitive mechanisms that support language comprehension are clearly valuable, but on the other hand, it is important to note that kindergarten-path effects have been observed in a very restricted set of linguistic environments, such as the garden-path sentence with a PP attachment ambiguity exemplified in (1). There are several reasons why it would be useful to further explore what kind of linguistic environment causes kindergarten-path effects. First, it is possible

that children's revision difficulties may be confined to PP attachment ambiguities and may not generalize to all cases in which revisions are required. The studies that attribute kindergarten-path effects to the immature executive function mechanism generally assume that there is continuity between the child and adult sentence reanalysis mechanisms, but this assumption merits empirical investigation. Second, although the immature executive function may indeed play a role in kindergarten-path effects, there may be additional interacting factors. For example, children's inability to retract their incremental commitments may be attributable to their general processing biases to prioritize bottom-up information in their syntactic analyses. As reviewed in Chapter 1, it has been shown that children are sensitive to bottom-up information such as verb biases in making their initial syntactic ambiguity resolution decisions, whereas they seem to be particularly insensitive to top-down information such as pragmatic information in resolving syntactic ambiguities (Snedeker & Trueswell, 2004). Thus, kindergarten-path effects may only reflect children's faithfulness to the analysis and interpretation that was fully constructed using bottom-up information up to the first PP. On the other hand, the observation that bottom-up information is prioritized in children's syntactic ambiguity resolution is mostly based on resolution of PP attachment ambiguities, and it is unclear to what extent this observation generalizes to other environments.

For these reasons, it is clearly useful to expand the empirical domain of investigation and further examine the nature of the linguistic environments that give rise to kindergarten-path effects. Specifically, this chapter examines the interaction of bottom-up information and children's reanalysis mechanism by testing whether actively constructed syntactic analyses can be retracted. Recall from Chapters 2 and 3 that we

have used time course evidence to establish a sequence of processes involved in active filler-gap dependency computation in (2).

(2) a. Actively construct the syntactic dependency between the filler and the gap.
b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

As (2) illustrates, the syntactic representations for filler-gap dependencies are constructed without relying on bottom-up, lexical information at the tail of the dependencies. Chapter 2 argued for this using time course evidence for reading disruption in English when the verb was intransitive. Moreover, as discussed in Chapter 2, evidence for pre-verbal active dependency completion in verb-final languages like Japanese is even more powerful because the gap position clearly precedes the verb (Aoshima et al., 2004; Nakano et al., 2002). If children actively construct filler-gap dependencies like adults in advance of critical bottom-up information about the gap position, this allows us to investigate a new question about the interaction of bottom-up information and children's reanalysis capacity: When actively constructed filler-gap dependencies are disconfirmed by later arriving bottom-up information about the gap, are children able to successfully retract their initial syntactic commitments? If children are successful at retracting active syntactic commitments when the error signal consists of bottom-up information that is critical for identifying the gap position, then it would suggest that kindergarten-path effects do not reflect general difficulties in retracting early commitments, but rather reflect what type of information children prioritize in constructing and re-constructing

syntactic representations.

Alternatively, it may be the case that the critical factor in explaining reanalysis success and failures does not depend on the distinction between actively hypothesized structures and incrementally confirmed structures, but is rather determined based on the nature of error signals. For example, the error signal used in Trueswell and colleagues' test sentence (*Put the frog on the napkin in the box*) was a syntactic error signal. Due to the design of the scene, the second PP *in the box* could not be taken as a modifier of the preceding NP *the napkin*. For this reason, *in the box* must be analyzed as the destination for the *putting* action, and consequently the initial analysis of the first PP *on the napkin* must be retracted. This, however, leaves open the possibility that error signals of different sources, such as lexical or semantic information, could serve as a more effective revision cue. This is a reasonable possibility given that adult reanalysis mechanisms also show differential sensitivity to various error signals (e.g., Fodor & Inoue, 1994; Frazier & Clifton, 1998).

To address these questions, this chapter presents a cross-linguistic comparison of wh-dependency interpretation in children. We capitalize on the difference in verb order between English and Japanese to demonstrate that children actively complete wh-dependencies, and using this property, we will further examine the effectiveness of different types of bottom-up error signals about the gap position that disconfirm actively constructed wh-dependencies. If children generally prioritize critical bottom-up information about the gap position, then we predict that bottom-up error signals in general should serve as an effective revision cue regardless of the type of error signals. On the other hand, it is possible that children prioritize different sources of information in

making reanalysis decisions. We attempt to shed light on this question by presenting syntactic error signals ('filled-gap') and verb-based error signals (incompatibility between the wh-phrase and the verb). Both are bottom-up error signals about the gap location, but as the review in Section 4.1 will show, children have been observed to be particularly sensitive to verb information in resolving syntactic ambiguities. It is thus plausible that children may be more sensitive to verb information in retracting active syntactic commitments as well.

The remainder of this chapter is organized as follows. Section 4.1 reviews previous work that suggests that verb information surprisingly fails to trigger a successful reanalysis. We argue that this is due to the difficulties of retracting interpretations assigned to an earlier constituent in the sentence, and motivate the use of bi-clausal ambiguous wh-adjunct questions in English and Japanese, where a syntactic revision can be triggered by verb information or syntactic information without altering the interpretation assigned to the fronted wh-phrase. Section 4.2 reports an English Question-after-Story experiment with adults and children (Experiment 5), and demonstrates that adults and children generally prefer to associate the fronted wh-adjunct *where* with the main clause VP, although this preference is modulated by properties of the main clause verb. Section 4.3 reports a Truth Value Judgment Experiment with English-speaking adults (Experiment 6) to test whether the interpretive preference observed in Experiment 6 reflects a bias to select one of the two possible interpretations, or a bias to preserve the initially assigned interpretation. It is shown that adults only accept the preferred interpretation as true despite the fact that the other interpretation is perfectly consistent with the story, suggesting that the interpretive preference reflects persistence of the initial

interpretation. Section 4.4 extends the Question-after-Story experiment in Section 4.2 to Japanese in which the verb order is the opposite of English (Experiment 7), and demonstrates that Japanese children and adults show an opposite preference to associate the wh-phrase with the embedded clause VP. Taken together with the English results, we argue that the ultimate interpretations of these constructions uniformly reflect association of the wh-adjunct with the first VP in the sentence, which suggests that the ultimate interpretation results from active gap creation at the first possible thematic position for the fronted *where*. Experiment 8 reported in Section 4.5 manipulates the Japanese sentences used in Experiment 7 to examine the effectiveness of verb-based error signals and syntax-based error signals. The results demonstrate that verb-based cues are effective while syntactic error signals fail to trigger a successful revision. Broader implications of these findings are discussed in Section 4.6, and Section 4.7 concludes the chapter.

4.1 Verb bias and kindergarten-path effects

One source of information that seems to have a privileged status in child sentence processing mechanisms is verb information. For example, Snedeker and Trueswell (2004) used an eye-tracking during act-out task similar to that of Trueswell and colleagues (1999), and presented instructions like (3) with PP attachment ambiguity, while manipulating the verb types in addition to the referential context.

(3) {Tickle | choose | feel} the cow with the stick.

The three verb types were determined based on a norming study on how likely each verb was to co-occur with an instrument PP: verbs that frequently take an instrument PP (e.g.,

tickle), verbs that rarely do so (*choose*), and equi-biased verbs (*feel*). Snedeker and Trueswell found that the verb biases directly correlated with children's PP attachment preferences, but replicated the earlier finding that children's PP attachment decisions were not influenced by manipulation of referential cues. (For a similar verb bias effect on ambiguity resolution in young children, see Kidd & Bavin, 2007).

These studies demonstrate that children are sensitive to verb information in making PP attachment decisions, and suggest the possibility that this may be a general feature of the child sentence processing mechanism that goes beyond ambiguity resolution. In the domain of children's revision capacity, however, there is evidence that suggests that verb information may not serve as an effective revision cue in garden path sentences. Choi and Trueswell (2010) conducted an experiment similar to Snedeker and Trueswell's study in Korean, a verb-final language in which the temporal order of verb and PP is the reverse of that of English (English: V precedes PP, Korean: PP precedes V), as shown in (4).

(4) naypkhin-ey kaykwuli-lul {nohu-sey-yo | cipu-sey-yo}
napkin- Loc/Gen frog-Acc put-Hon-SE pick up-Hon-SE
"Put / pick up the frog on the napkin"

In this sentence, the *-ey* particle attached to the first NP is morphologically ambiguous between a locative marker, which effectively produces a destination interpretation of *on the napkin*, and a genitive case marker, which leads to an NP modification interpretation, specifying the location of the following NP (i.e., *frog on the napkin*). In other words,

unlike in English PP attachment ambiguity sentences like (3), the structural ambiguity in Korean arises as soon as the first and second NPs are processed, i.e., before the verb information becomes available. Choi and Trueswell manipulated the semantic fit between the verb and the first NP to disambiguate the sentence in (4) to the destination analysis (i.e., *nohu-sey-yo* 'put') or NP modifier analysis (i.e., *cipu-sey-yo* 'pick up').[10] Choi and Trueswell found that both children and adults fixate on the empty napkin in the scene as soon as the ambiguous PP *naypkhin-ey* is presented, which suggests that a destination interpretation is initially assigned to the NP marked with the *-ey* particle. However, when the verb turned out to be incompatible with the destination analysis of the ambiguous phrase (e.g., *pick up*), adults were able to incorporate the verb information to revise this initial analysis and perform the correct actions (e.g., pick up the frog that is sitting on the napkin and hold it in the air), but children were unable to revise the initial destination interpretation of the first NP and coerced the destination interpretation and performed an incorrect action, such as picking up the frog and then putting it on the napkin.

This presents an interesting puzzle: on the one hand, verb information seems to exclusively drive children's PP attachment ambiguity resolution in English (Snedeker & Trueswell, 2004). On the other hand, verb information in Korean does not lead the parser to revise the commitment made early in the sentence (Choi & Trueswell, 2010). However, there may be another competing constraint that made it seem as if verbs could not trigger reanalysis. In Choi and Trueswell's Korean stimuli, the syntactic reanalysis

[10] Note that due to the possibility of argument drop in Korean, the *put* condition can be compatible with the NP modifier analysis if the destination argument is unpronounced. However, given that there were multiple possible destinations in the scene, it is unlikely that the destination argument is dropped in the context of Choi and Trueswell's experiment.

process would not only require changing the syntactic analysis of where the first NP should be attached, but would also require hearers to change the destination interpretation of the first NP *naypkhin-ey* (‘on the napkin’) to the locative interpretation that specifies the location of the following NP *kaykwuli-lul* (‘frog’). Here, it is possible that what Korean children were unable to do is the revision of the interpretive commitment that was driven by the preference to analyze the *-ey* particle as a destination marker. As reviewed in Chapter 1, retracting interpretive commitments that result from temporarily entertained syntactic analyses is often observed to be difficult even for adults (Ferreira & Patson, 2007; Sturt, 2007). For example, Christianson, Hollingworth, Halliwell, and Ferreira (2001) presented comprehension questions after garden-path sentences like *While Anna dressed the baby spit up on the bed,* in which the verb *dressed* is used as a reflexive verb but the subject of the second clause *the baby* can be temporarily misanalyzed as the direct object of the verb *dressed*. The results of the comprehension measures revealed that adults retained the interpretation that Anna was dressing the baby, suggesting that the interpretive commitment that resulted from the temporary syntactic analysis was not fully retracted in their global interpretation of the sentence, even after the initial structural analysis itself was revised. Given this demonstration that even adults struggle to retract interpretive commitments, the ideal test of children’s sensitivity to verb information in reanalysis process should use a sentential context in which the syntactic analyses could be changed by the error signals without altering the interpretation of the early constituents in the sentence.

The present study explores this possibility by examining children’s processing of ambiguous bi-clausal wh-questions in English and Japanese. Examples from English and

Japanese are given in (5):

(5) a. Where did Emily tell someone that she hurt herself?

b. Doko-de Emily-chan-wa *pro* ashi-o kegashita to itteta-no?
where-at Emily-Dim-Top she foot-Acc hurt Comp was telling-Q

These sentences are globally ambiguous, in that the fronted wh-adjunct *where* can be associated with one of two possible thematic positions, namely, either the main clause VP (*tell someone*) or the embedded clause VP (*hurt herself*). Wh-fronting is generally an obligatory rule of English wh-questions, but Japanese is a *wh*-in-situ language and hence wh-phrases can generally stay in their thematic positions. In (5b), the wh-phrase is fronted to the beginning of the sentence via scrambling, by which many constituents other than subjects and verbs can be fronted within and across clauses (Harada, 1977; Saito, 1985). This allows us to construct a closely matched sentence pair as in (5), but notably, the order of verbs in Japanese (*hurt-told*) is the opposite of that of English (*told-hurt*) due to verb-finality, allowing us to examine the importance of the temporal order of verbs across languages.

There are three features of this construction that make these sentences an appropriate testing ground for the role of verb information in initial syntactic analysis and reanalysis. First, as discussed in the previous chapters, it is widely accepted that the adult parser completes filler-gap dependencies in an active fashion, postulating a gap before there is sufficient bottom-up evidence that confirms that analysis (*Active gap filling:* Crain & Fodor, 1985; Frazier & Clifton, 1989; Garnsey, Tanenhaus, & Chapman, 1989).

Evidence from cross-modal picture priming studies suggests that 5-year-old children also actively process filler-gap dependencies. For example, Love (2007) presented a sentence like *The zebra that the hippo had kissed ___ on the nose ran away*, and found that a) the alive vs. non-alive decision was made much more quickly due to associate priming when they presented a picture that corresponded to the filler (e.g., zebra) at the onset of the verb rather than an unrelated picture (e.g. camel), and b) when these pictures were presented at the onset of the subject inside the relative clause region (i.e., hippo), there was no difference in reaction time. This pattern of results suggests that the head of the relative clause was activated at the point of the verb, despite the risk that this may not be the correct gap position (see also Roberts, Marinis, Felser, & Clahsen, 2007). If children actively complete filler-gap dependencies, then in the context of sentences like (5), it is plausible that children favor wh-association to the first VP in the sentence, namely, the main clause VP *tell someone* in English. Moreover, the order of verbs in a verb-final language like Japanese is the opposite of English (Japanese: embedded-main, English: main-embedded), and for this reason, if children actively create a gap and assign interpretation to the Japanese counterpart in (5b), then we predict that Japanese children should prefer wh-association with the embedded clause VP (*hurt herself*). In fact, adult reading time studies on long-distance scrambling in verb-final languages like Japanese have shown that the parser attempts to complete filler-gap dependencies in the embedded clause (Aoshima, Phillips, & Weinberg, 2004; Nakano, Felser, & Clahsen, 2002). Thus, measuring interpretive preferences across languages allows us to further examine whether the child parser actively processes filler-gap dependencies (Experiments 5 to 7).

A second attractive feature of (5) is that the VP region can be manipulated in a

variety of ways to investigate the effectiveness of various revision cues (Experiment 8). In particular, the Japanese version (5b) provides an ideal testing ground because the embedded clause VP can be manipulated much more flexibly than the main clause VP in English. Moreover, the fact that the fronted wh-phrase is an adjunct allows a flexible manipulation of the error signal. In the case of wh-argument fronting studied in Chapter 3, because the verb syntactically and semantically selects an obligatory argument, the gap position is clearly indicated by a missing argument in the string. On the other hand, wh-adjuncts are not selected by verbs and can rather freely attach to any VPs that meet the semantic and pragmatic requirement of the wh-adjuncts. Using this flexible feature of wh-adjuncts, we will present two kinds of bottom-up error signals: the wh-association with a VP can be blocked by presenting an overt locative PP (a filled-gap in the sense of Stowe, 1986), or it can be blocked by manipulating the lexical content of the verb, such that association of the wh-adjunct with this VP would yield no felicitous interpretation (verb-based reanalysis cues). This allows us to test the effectiveness of these different revision cues while keeping the syntactic environment constant.

The third desirable feature of the construction in (5) is that even when the initial active wh-attachment to the first verb in the sentence is later disconfirmed by additional information, there is another verb later in the sentence that can accommodate the wh-phrase. Importantly, even if the wh-attachment site changes from the first VP to the second VP, the interpretation of the wh-phrase itself remains constant as a wh-locative. This feature thus allows us to examine whether the lack of verb-driven reanalysis in Choi and Trueswell (2010) was related to the fact that the reanalysis process required a revision of the interpretive commitment that was made when resolving the morphological

ambiguity on the first NP.

The experiments reported in this chapter use two types of story-based interpretation tasks that are suitable for probing the ultimate interpretation that hearers adopt. Experiments 5, 7, and 8 use a Question-after-Story task (de Villiers, Roeper, & Vainikka, 1990; de Villiers & Roeper, 1996) and examine whether English-speaking adults and children (Experiment 5) as well as Japanese-speaking adults and children (Experiments 7 and 8) demonstrate evidence for active gap creation, i.e., a bias to attach the wh-phrase to the first VP in ambiguous wh-questions like (5). Experiment 6 uses a Truth Value Judgment Task with adult English speakers in order to examine whether the results from question-answering responses actually reflect parsing biases rather than question-answering strategies, and also to explore the possibility that the use of sentences that are designed to force one of the two possible interpretations may help reveal the hearers' reanalysis abilities.

It is important to note that children's comprehension of sentences like (5a) has been tested in an important series of studies by de Villiers and her colleagues (de Villiers et al., 1990; de Villiers, Roeper, Bland-Stewart, & Pearson, 2008; Roeper & de Villiers, 1992). Regarding children's main clause vs. embedded clause interpretation preference, rather mixed results were found in earlier small-scale studies that tested a combination of various wh-question structures with mixed main clause verb types and contextual manipulation (de Villiers et al., 1990; Roeper & de Villiers, 1992). However, a large-scale study with 703 typically developing children (age range 4 to 9-years-old) revealed that children showed an 82% preference for the embedded clause interpretation to the sentence *How did the boy say he hurt himself?* (de Villiers et al., 2008). This is a strong

demonstration of an embedded clause bias, although de Villiers and colleagues did not provide an explanation for why the embedded clause interpretation is preferred. This evidence may call into question whether children actively process filler-gap dependencies, or whether this task is even appropriate for probing incremental wh-attachment. However, the fact that de Villiers and colleagues (de Villiers et al., 2008) only had one token of this type of wh-question raises the possibility that this may have been driven by properties of this one specific item. In particular, it is important to note that the previous work by de Villiers and her colleagues has mostly used the verb *say* as the main clause verb. Based on Snedeker and Trueswell's (2004) demonstration that children's PP attachment decisions are affected by verb biases, Experiment 5 explored the possibility that the previously reported embedded clause preference in de Villiers and colleagues (de Villiers et al., 2008) may also be tied to the use of this particular verb *say*. In this experiment, we constructed a series of stories that were designed to make both main clause and embedded clause interpretations available and further manipulated the verb type. We used the bare *say*, *tell someone* and *say to someone* (6) that were equally compatible with the stories, and examined whether adults and children's wh-attachment is sensitive to properties of the main clause VP.

(6) Where did Emily { say | tell someone | say to someone } that she hurt herself?

If the embedded clause interpretation preference observed in de Villiers and colleagues' study was driven by properties of the main clause verb *say*, then we predict that different interpretive preferences may arise with *tell someone* or *say to someone*.

4.2 Experiment 5 (English question-after-story with verb manipulation)

4.2.1 Method

Participants

We recruited 36 adult native speakers of American English from the University of Maryland community. Forty five children who were between the ages of 4;7 and 6;5 and acquiring English as their native language also participated in the study. The children were recruited at a preschool at the University of Maryland or from College Park, MD and its surrounding communities. Data from 9 children were excluded due to fussiness (n=2) or providing more than two incorrect or irrelevant answers (n=7), and the remaining 36 children were included in the final sample (mean age; 5;5). The adult participants were given course credit or $10 for their participation in a one-hour experiment session that included the present study and other experiments. The age range for children was determined based on two factors: a) we wanted to ensure that children would be able to sit patiently through the study and listen to the relatively complex stories used in our experiment, and b) we wanted an age range that was roughly equivalent to the ones used in previous studies on PP attachment ambiguity (e.g., Trueswell et al., 1999) and filler-gap processing (e.g., Love, 2007) so that our results would be comparable to the previous findings.

Materials

Story design. The same 8 stories were used in all four experiments reported in this study, so we will elaborate on the details of the story design here. Our story stimuli were cartoon movies made from a sequence of clipart image animations, and these movies incorporated features used in an earlier visual world eye-tracking study (Sussman &

Sedivy, 2003; see Chapter 3 for discussion), so that the same stimuli could also be used as an eye-tracking experiment in the future (Omaki, in progress). Each story display contained images of four locations of a type familiar to children (e.g., a playground), and they all appeared in roughly the same size. A sample story and a target question is given in (7) (see below for details of the question design), and a sample sequence of events on the visual display is shown in Figure 12.

(7) **[Intro phase]** Emily likes to play outdoors.

[1st location] One day she was swinging on the swings, and she jumped off the swings from really high up! She balanced herself really well on the landing, so she didn't fall.

[2nd location] When Emily got bored with the swings, she decided to go climb a really tall tree. She got up pretty high, but suddenly one of the branches broke, so she fell off the tree and hurt herself! But Emily was a brave girl, so she got back up right away and didn't cry at all. Emily wanted to tell her friends how pleased she felt about not crying after hurting herself, so she went to find some of her friends at the library.

[3rd location] She found a friend of hers there, but the librarian came out and told her that in the library they must be very quiet! So, Emily couldn't talk to her friend and felt disappointed. But she had a good idea: She could go to the swimming pool to see more friends, because at the pool she can talk as much as she wants!

[4th location] When she got to the pool and found her friends, she said to them, "I hurt myself falling out of the tree, but didn't cry at all!" Emily was happy she could finally tell someone about her day, and her friends were impressed by how brave she was.

[Question] Where did Emily tell someone that she hurt herself?

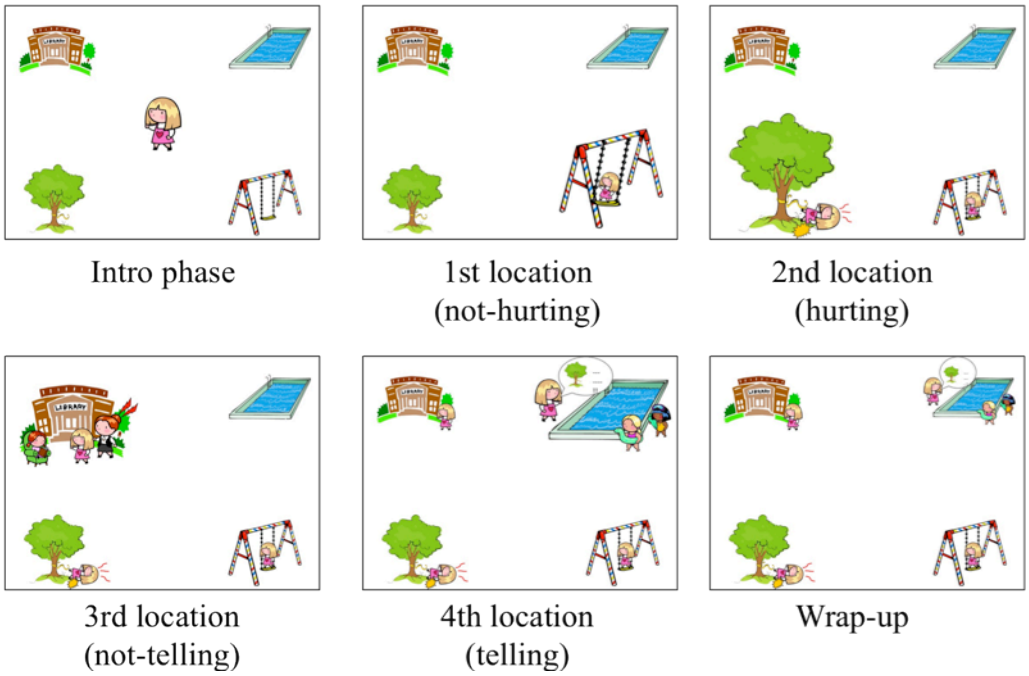


Figure 12. A sample story sequence.

In each story, a character visited the four locations on the computer display, and during the character's visit, the relevant location was magnified so that the event that happened in that location was clearly visible to participants. The stories consisted of six phases. In the introduction phase, a character was introduced in the center of the display and the theme of the story was narrated. In the next four phases, the character visited each location and underwent failure and success in achieving the intended activities, and after

the story ended, the fourth location shrank to its original size (wrap-up phase) and the question was presented.

The first two locations and the last two locations are relevant for either main clause event or the embedded clause event, and in each pair of locations, the relevant event almost occurs in the first location but eventually fails to occur (e.g., in the example story, Emily almost talked to her friend at the library), and it is only in the second location that the relevant event occurs (e.g., in the example story, Emily finally talked to her friends at the pool). This feature was introduced to make the location-event parsing for each target event memorable, and it was also expected that having two potential locations for a certain event would make the wh-questions more felicitous. In order to further facilitate remembering what happened in each location, we also left visual 'traces' of what happened in each location. In the 'failed attempt' locations, the image of the character standing in front of the location is left, but in the locations in which the intended activity successfully occurred, the visual trace of that event remained on the display (e.g., the hurting scene or the speech bubble with an image of the tree in Figure 12).

In order to control for the possibility that participants' answers might reflect a recency bias (i.e., they might only provide an answer that relates to the last event in the story), in half of the stories the first two locations were relevant for the embedded clause event, and in the other half of the stories the first two locations involved the main clause event. In either sequence, the first set of events provided an important motivation for the next set of events. In the story in (7) above, for example, the 'telling' events were motivated by how tough Emily felt when she hurt herself and did not cry. In stories that

started with 'telling' events, the story was constructed in such a way that 'telling' was a pre-requisite for the next set of events (e.g., an astronaut needed to consult a scientist to get advice on his space mission to find aliens).

The quadrants in which relevant events occur were randomized across stories to prevent the participants from predicting which quadrant would correspond to which type of event. The English story scripts and questions are made available in Appendix E, and a sample movie file can be found at (http://www.ling.umd.edu/labs/acquisition/stimuli/).

Question design. We constructed 8 bi-clausal wh-questions with *where* as shown in (6), which is repeated below for convenience.

(6) Where did Emily { say | tell someone | say to someone } that she hurt herself?

This construction is globally ambiguous and allows two possible interpretations, one in which the wh-adjunct *where* is associated with the main clause VP, and one in which the wh-phrase is associated with the embedded clause VP. All of the target sentences contained an overt complementizer *that*, as our pilot work with adult native speakers showed that the presence of an overt compelmentizer is critical for making both main clause and embedded clause interpretations equally accessible. We manipulated the verb type (*say, tell someone, say to someone*) as a between-participants factor instead of a within-participants factor, so as to avoid any potential priming effects from one verb type to another. There were 12 children in each verb condition, and the resulting mean age range for each verb condition was 5;4 for *say*, 5;7 for *tell someone*, and 5;3 for *say to someone*, respectively. There was no reliable difference in mean age across the three

conditions ($F < 1$).

In addition to these 8 target wh-questions, we constructed 8 unambiguous filler wh-questions using *why*. Fillers were designed such that they could be used for the same set of stories as the target sentences. We chose *why* questions as fillers because these questions were slightly more challenging to answer than *where* questions, since in the latter case, there were visual cues on the display that participants could use to find where certain events occurred, whereas there were no such direct visual cues that would help participants remember reasons why certain events occurred in the story. In order to balance the type of locations that were asked about, all the *why* questions targeted locations that were not questioned about in the *where* questions.

The target and filler questions were distributed across two lists, such that half the participants saw items 1 to 4 in the target question form and items 5 to 8 as filler questions, and the other half of the participants saw items 5 to 8 in target question form and items 1 to 4 in filler question form. Each list had two versions with a different order of stories, but in each version the target question trials were interspersed with filler question trials.

Procedure

Children were told that they were going to play a quiz game with a puppet, and that they were going to watch cartoon movies and hear a question after each story. They first saw two short practice trials to make sure they understood the task, and then saw 4 target and 4 filler trial movies. The sound was presented through a speaker attached to the computer. An experimenter presented the movies using digital media player software, and the movie presentation was paused at the end of each story, so that the experimenter

asked a question using the puppet and noted down the answer. The experimenter was trained to produce the questions with a neutral prosody that would not bias the hearers to one interpretation or the other. During the story phase, the experimenter provided brief comments on events happening in each location to make the task interactive and help keep children's attention. The comments were made on events in each location so as to balance the degree of saliency across the four locations. After a child answered a question, a brief positive feedback movie was played on the computer to encourage children to keep paying attention. The experiment took approximately 20 minutes.

For adult participants, an experimenter explained the task using one practice trial, and the rest of the experiment was automated in a single movie file. For adult participants, the experimenter stayed outside the testing room during the experiment. In this version, the computer presented a question automatically after each story, and participants had roughly 6 seconds to write down the answer in a short phrase (e.g., 'at the pool') on an answer sheet. With this procedure, the experiment took approximately 15 minutes.

4.2.2 Results

The dependent measure was the proportion of main clause responses (i.e., answering 'at the pool' to the example scenario in (7)). Each verb condition had two lists of items, but the data from the two lists were collapsed in the analysis, since there was no significant difference across two lists of stimuli ($F < 1$). The proportion of main clause responses for all conditions is summarized in Figure 13.

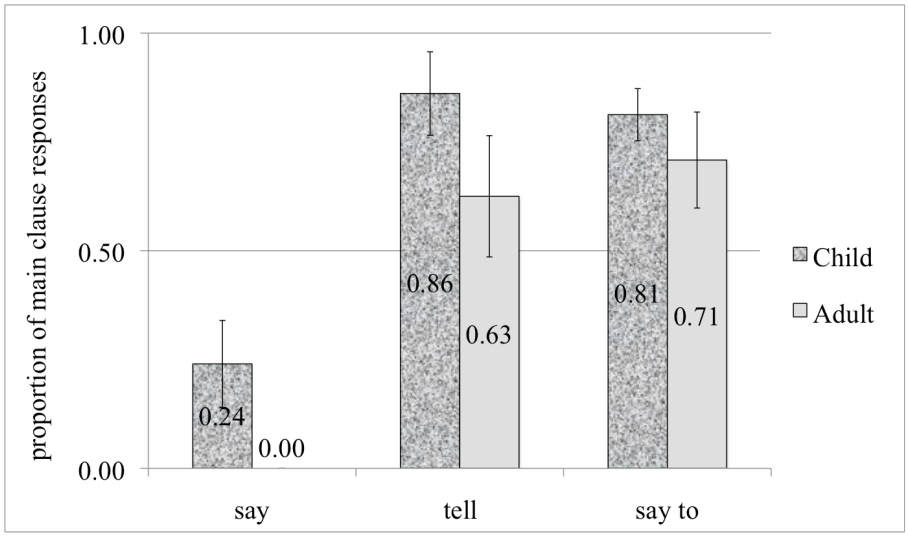


Figure 13. Mean proportion of main clause responses in Experiment 5. The error bars indicate standard errors of the mean.

Overall, children generally provided more main clause responses than adults did, but children and adults showed the same pattern of responses in each verb-type condition. In the *say* condition, both children and adults produced few main clause answers (child: 24%; adult: 0%), but they preferred the main clause interpretation in the *tell someone* and *say to someone* condition (children: 86% for *tell someone*, 81% for *say to someone*; adults: 63% for *tell someone*, 71% for *say to someone*). To assess the reliability of this pattern, the mean proportion of main clause responses were submitted to an ANOVA with age group (child vs. adult) and verb type (*say* vs. *tell someone* vs. *say to someone*) as between-participants factors. We found that there was a main effect of age, $F(1, 66) = 6.12$, $p < .05$, as well as a main effect of verb type $F(2, 66) = 29.58$, $p < .001$, but there was no significant interaction of the two factors, $F < 1$.

4.2.3 Discussion

The results from this experiment demonstrate that both children and adults show very similar preferences in interpreting ambiguous wh-questions. Previous work by de Villiers and colleagues did not include adult participants, and this is therefore the first demonstration that children and adults show an extremely similar interpretive preference in answering ambiguous bi-clausal wh-questions. Our results replicated the previous finding from de Villiers and colleagues (de Villiers et al., 2008) in that both adults and children preferred the embedded clause interpretation when the main clause verb is bare *say*; however, we also found that when the main clause VP was changed to *tell someone* or *say to someone,* both adults and children preferred the main clause interpretation. These results have two implications for adults and children's wh-attachment ambiguity resolution mechanism. First, wh-attachment decisions are strongly influenced by properties of the first VP in the sentence. Our finding thus extends the previous demonstration of verb primacy in PP attachment ambiguity resolution to the domain of wh-attachment ambiguity resolution, suggesting that this is a fairly robust property of the sentence processing mechanism in adults and children.

Second, the main clause interpretation seems to be more widely preferred, since the embedded clause interpretation preference was restricted to the bare *say* condition. This suggests that wh-attachment to the first VP in the sentence is preferred over attachment to the second VP. This may reflect active gap creation processes for the fronted constituent, although the present data do not provide decisive evidence that the main clause interpretation was assigned incrementally (see below for further discussion). The exceptional embedded clause interpretation in the bare *say* condition raises a

question as to what makes that condition different from the other two verb conditions. We do not have a definitive answer to this question, but we tentatively suggest that the embedded clause interpretation preference in the bare *say* condition may result from the fact that the English verb *say* can be used as an evidential marker, especially when it appears in isolation (Aikhenvald, 2004; Simons, 2007). In other words, when children and adults hear *Where did Emily say that she hurt herself?*, children could be analyzing the main clause subject and verb as a description of the source of information, such that the resulting structural representation does not involve a sentential complementation, as in *According to Emily, where did she hurt herself?* If this is the representation constructed by the sentence processing mechanism, then there is only one attachment site (i.e., *hurt herself*) for the wh-phrase and consequently the 'embedded clause' interpretation becomes the only interpretation available in this sentence. Under this account, it is still unclear why this structure is preferred over the actual bi-clausal sentence representation that more closely matches the string, but we leave this question open for future research.

One may instead argue that the main clause interpretation preference in the *tell someone* and *say to someone* conditions only reflects pragmatic biases that are derived from the choice of the sentence form. If the speaker intended to ask for the location of the embedded clause event, this speaker could have just asked a mono-clausal question (*where did Emily hurt herself?*), but the fact that the bi-clausal sentence was used may already indicate that the speaker wanted to gather information about the event described by the main clause predicate. However, the contrast between the bare *say* condition and the *tell someone* or *say to someone* conditions casts doubt on this explanation, because

this account predicts that bi-clausal questions should always yield a main clause interpretation preference.

It is important to note that the current evidence for main clause interpretation preferences does not necessarily indicate that children and adults' interpretive preferences follow from active association of the wh-phrase with the main clause VP. For example, it is possible that adults and children initially give equal consideration to both interpretations of the ambiguous questions, and that they then use verb biases to select one of the two possible answers. If this were the case, then the Question-after-Story task might not provide an appropriate measure of the actual sentence comprehension processes, because the dependent measure would only reflect biases in deciding which answer is more appropriate in the particular Question-after-Story experiment setting. We address this question in Experiment 6.

4.3 Experiment 6 (English truth value judgment study with adults)

In order to examine whether the main clause wh-attachment preference observed in Experiment 5 only reflects biases in selecting a possible answer to a question, rather than active gap creation processes that yield an interpretative commitment to the main clause interpretation, Experiment 6 used a Truth Value Judgment Task (TVJT) in which participants were forced to judge the truth value of one of the two possible interpretations (Crain & Thornton, 1998). An example target sentence is shown in (8).

(8) The place where Emily said to someone that she hurt herself was the {pool | tree}.

This sentence is derived from the target questions used in Experiment 6, and it involves a relative clause *the place where…* which essentially contains the same wh-attachment ambiguity that can be resolved by association with the first VP (*said to someone*) or the second VP (*hurt herself*). At the end of the sentence, one of the two locations that correspond to the possible interpretations is provided (e.g., *pool* for the saying event, *tree* for the hurting event), such that participants were forced to evaluate whether the first or second VP interpretation was true in the story. If the participants in Experiment 5 entertained both possible interpretations and then selected one of them as a more plausible answer using the verb information, then we predict that in sentences like (8) participants should accept both statements as true. On the other hand, recall from the discussion in the introduction to this chapter that the first interpretation that adults incrementally generate tends to persist in the ultimate interpretation (Ferreira & Patson, 2007; Sturt, 2007). This suggests that based on the rejection of the possible interpretation that becomes available later in the sentence, we should be able to infer that the sentence interpretation mechanism has incrementally committed to the interpretation that became available first in the sentence (for a similar use of the TVJT method to shed light on the parsing mechanism, see Crain, Ni, & Conway, 1994; Musolino & Lidz, 2006). Thus, if the main clause preference seen in question-answering responses in Experiment 5 truly reflected active commitments to the main clause interpretation, then we would predict that the sentence that forces participants to evaluate the truth value of the embedded clause interpretation would be rejected. This experiment tested adult native speakers, who are presumably much more competent at retracting incremental commitments, in order to

examine how robustly the early interpretative commitment remains in their ultimate comprehension of the target sentence.

4.3.1 Method

Participants

We recruited 32 adult native speakers of American English from the University of Maryland community. The participants were given a course credit or $10 for their participation in a one-hour experiment session that consisted of the present study and other experiments.

Materials

This experiment used the same 8 stories as Experiment 5, but instead of presenting questions, declarative sentences like (8) were presented after each story. In the target sentences, we only used one verb type, *say to someone,* for which we saw the clearest main clause preference in Experiment 5. For fillers, we constructed unambiguous relative clause sentences about locations other than the ones that the target sentences described (e.g., The place where Emily couldn't talk to her friends is the library). Half of the fillers were manipulated to be true descriptions of the stories, and the other half were manipulated to be false descriptions of the stories.

We treated the main clause vs. embedded clause description as a between-participants factor in order to avoid any priming effects between the two sentence types. The target sentences and fillers were distributed across two lists in the same way as in Experiment 5.

Procedure

The same overall procedure as in the adult version of Experiment 5 was used, the only difference being that participants circled TRUE or FALSE written on the answer sheet instead of writing down answers to questions. The experiment took approximately 15 minutes.

4.3.2 Results

No participant provided more than one incorrect answer for the 4 filler trials, and hence all 36 participants were included in the data analysis. The mean acceptance rate for the main clause description condition was 100%, whereas the mean acceptance rate for the embedded clause description condition was 19% (SE = 9), and there was a significant difference in acceptance rate between the two conditions, $F(1, 30) = 87.41$, $p < .001$.

4.3.3 Discussion

We found that the adults' acceptance rate was perfect for the main clause description condition, whereas in the embedded clause description condition, adults accepted the grammatically permissible embedded clause interpretation on only 19% of trials. These results indicate that the TVJT replicated the first VP preference found in the Question-after-Story experiment (Experiment 5). In fact, the rate of embedded clause interpretation rejection in Experiment 6 (81%) is similar to the rate of main clause interpretation response in Experiment 5 (71%). This parallel suggests that the same mechanism underlies the question-answer responses in Experiment 5 and truth value judgment responses in Experiment 6; namely, both measures reflect active gap creation at the main clause VP which in turn yields a commitment to the main clause interpretation.

These results also present an important methodological implication for Truth Value Judgment Tasks. It is sometimes argued that participants in TVJT experiments show a strong bias for accepting any grammatically permissible interpretations that are true and pragmatically felicitous in the given context (Principle of Charity: Gualmini, 2004; Gualmini, Hulsey, Hacquard, & Fox, 2008; cf. Crain & Thornton, 1998). Since both main clause and embedded clause interpretations are not only grammatically possible but also made pragmatically felicitous in our story design, the Principle of Charity predicts that both main clause and embedded clause interpretations should be accepted. Contrary to this prediction, however, our adult participants mostly rejected the embedded clause interpretation. This suggests that although a bias for accepting true interpretations may exist, such acceptance biases clearly interact with other psycholinguistic constraints that are relevant for assigning interpretations to the test sentence (for further discussions on interpretation preferences and truth value judgment responses, see Chapter 37 of Crain & Thornton, 1998; Musolino & Lidz, 2006).

To provide a further test on whether the ultimate interpretation of the target sentences with wh-dependencies reflects active gap creation at the first potential thematic position, Experiment 7 used the Japanese counterpart of the stimuli in Experiment 5 and examined Japanese children and adults' interpretive biases. As reviewed above, the order of verbs in a verb-final language like Japanese is the opposite of English (Japanese: embedded-main, English: main-embedded). If children actively create a gap and incrementally assign an interpretation to the Japanese counterpart of the English sentences tested in Experiment 5, then we predict that Japanese children should prefer wh-attachment to the embedded clause VP (*hurt herself*).

4.4 Experiment 7 (Japanese question-after-story, ambiguous sentence)

4.4.1 Method

Participants

We recruited 16 adult native speakers of Japanese from the Hiroshima University community in Japan. Fourteen children who were between the ages of 4;9 and 6;4 and acquiring Japanese as their native language also participated in the study. The children were recruited at Ibaraki University Kindergarten and Mito Kindergarten in Mito, Japan. Two children were excluded due to providing more than two incorrect or irrelevant answers, and the remaining 12 children were included in the final sample (mean age: 5;9). The adult participants were paid 1000 yen for their participation in a one-hour experiment session that consisted of the present study and other experiments.

Materials and procedure

The stories, target questions and filler questions in Experiment 5 were translated into Japanese sentences like (9), while ensuring that the resulting stories and sentences contained words and expressions that are familiar to Japanese children.

(9) Doko-de Emily-chan-wa [*pro* ashi-o kegashita to] itteta-no?
where-at Emily-Dim-Top she foot-Acc hurt Comp was telling-Q
"Where was Emily telling someone that she hurt herself?"

This example is translated from the English example (6) in Experiment 5, and in this sentence the wh-phrase *doko-de* ('where at') is scrambled to the beginning of the sentence. Some sentences were slightly modified to adjust for grammatical differences

between English and Japanese. For example, *hurt herself* was changed to *hurt her foot* in Japanese because the Japanese verb for *hurt* does not select a reflexive pronoun. Note also that the Japanese main clause verb *itteta* (was telling) cannot be used as an evidential marker in Japanese, because there is a distinct evidential morpheme in Japanese (*-sou*), and typologically, there is a complementary distribution between languages that use a lexical strategy of evidential marking (e.g., English) and languages that use a distinct verbal morpheme for evidential marking as in Japanese (Aikhenvald, 2004). Note that the main clause predicate was presented in the past progressive form *itteta* (was telling) rather than the regular past tense form *itta* (told) in order to avoid redundancy with a homophonous embedded clause verb *itta* (went), which is used in Experiment 8. It is important to note, however, that at least according to Japanese adult native speakers' intuitions, this does not affect the interpretive preferences in (9).

These items were distributed across two lists in the way described in Experiment 5. All the target sentences used in this experiment are listed in Appendix F. The procedure was identical to that of Experiment 5, and the experiment took approximately 20 minutes.

4.4.2 Results

The mean proportion of main clause answers for children and adults is shown in Figure 14. Children showed a clear embedded clause preference (mean = 6%, *SE* = 4), as did adults (mean = 8%, *SE* = 4). An ANOVA with age (child vs. adult) as a between-participants factor revealed no statistically significant difference between the two groups, $F < 1$.

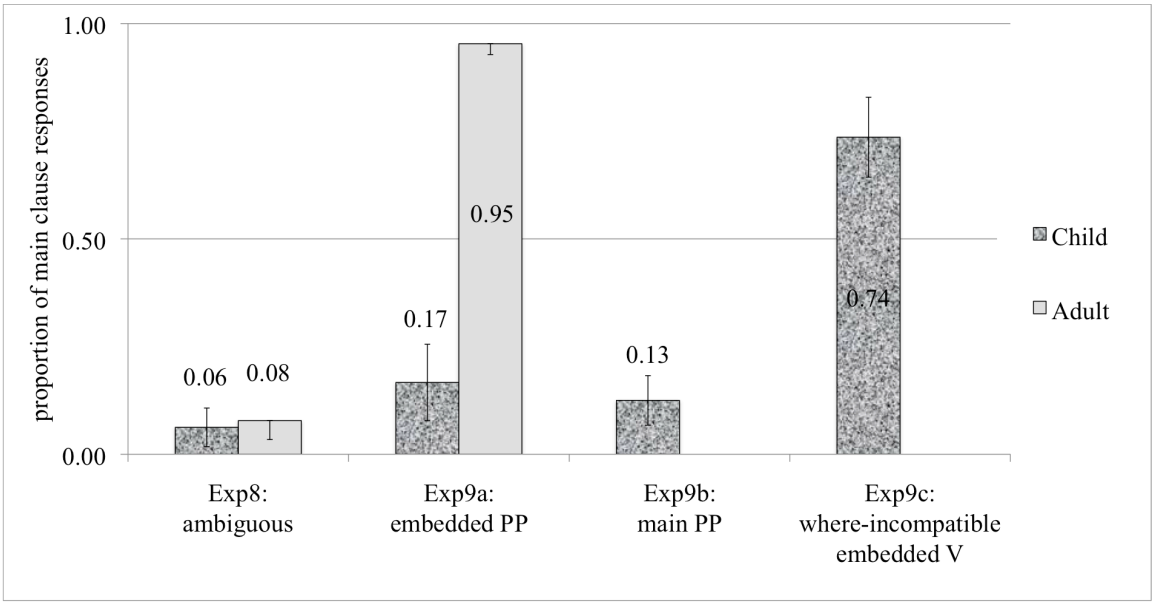


Figure 14. Mean proportion of main clause responses in Japanese Question-after-Story experiments (Experiments 7 and 8). The error bars indicate standard errors of the mean.

4.4.3 Discussion

The interpretation results demonstrated that Japanese adults and children prefer the embedded clause interpretation, unlike English-speaking adults and children in Experiments 5 and 6. Given that this experiment used the same stimuli as Experiments 5 and 6, the results indicate that the interpretive biases observed in the Japanese or English Question-after-Story experiments are not due to properties of the story design. Moreover, the embedded clause interpretation preference in Japanese cannot be due to the evidential use of the main clause predicate, because the Japanese main clause predicate *itteta* ('was telling') cannot be used as an evidential marker in the way the English bare verb *say* can. These results are compatible with findings from the online sentence processing studies with Japanese adults (Aoshima et al., 2004; Nakano et al., 2002), which showed time

course evidence for active gap creation in an embedded clause predicate region. The fact that the preferred interpretation corresponds to the one that results from the initial structural analysis suggests that our off-line interpretation measure reflects what happens in the real-time parsing of the sentences. Taken together, the present findings lead us to conclude that the Japanese child and adult sentence processing mechanism actively associates the fronted wh-adjunct with the first VP in the sentence and incrementally constructs the embedded clause interpretation.

However, there is one alternative explanation for the robust embedded clause preference. Recall that our target sentence (9), which is repeated here for convenience, included a null subject *pro* in the embedded clause subject position.

(9) Doko-de Emily-chan-wa [*pro* ashi-o kegashita to] itteta-no?
where-at Emily-Dim-Top she foot-Acc hurt Comp was telling-Q
"Where was Emily telling someone that she hurt herself?"

This null subject in Japanese is the closest counterpart to the overt pronoun that served as the subject of the embedded clause in the English target sentences. It was thus necessary to use a null subject pronoun in order to make the Japanese stimuli natural while keeping the content of the target items constant across English and Japanese. However, if the embedded subject is silent, there is no overt signal indicating the presence of the embedded clause until the overt complementizer *-to* is encountered. This raises the possibility that the listeners may misanalyze the target sentence as mono-clausal ('Where did Emily hurt herself?') and effectively ignore the main clause VP. This is rather

unlikely for adults: Our pilot work with Japanese adults also presented bi-clausal sentences with two distinct overt (non-pronominal) subjects, and in this case participants still demonstrated an embedded clause interpretation preference. This is consistent with the real-time evidence from Aoshima and colleagues (Aoshima et al., 2004) that Japanese adults actively associate the fronted wh-phrases with the embedded clause VP. However, it is possible that children may have ignored the main clause predicate for exactly this reason. We will return to this point below in Experiment 8.

4.5 Experiment 8 (Japanese question-after-story, error signal manipulation)

The results from Experiments 5 to 7 showed that both adults and children resolve the wh-attachment ambiguity by actively attaching the wh-phrase to the first VP in the sentence, regardless of the canonical word order of the language. This experiment uses this behavior to examine whether the active syntactic and interpretative commitments would lead to kindergarten-path effects despite the presence of error signals.

As reviewed in Section 4.1, previous work on child sentence processing showed that bottom-up information like verb information affects initial syntactic ambiguity resolution while top-down information like pragmatic or discourse information does not (Snedeker & Trueswell, 2004; Trueswell et al., 1999). With respect to reanalysis processes, however, verb biases did not serve as effective revision cues in a Korean sentence like (4) (Choi & Trueswell, 2010).

(4) naypkhin-ey kaykwuli-lul {nohu-sey-yo | cipu-sey-yo}
napkin- Loc/Gen frog-Acc put-Hon-SE pick up-Hon-SE
"Put / pick up the frog on the napkin"

However, this may only reflect difficulties in retracting the interpretive commitments that children made upon analyzing the ambiguous *-ey* particle as a destination marker, because changing the VP argument analysis of the ambiguous PP *on the napkin* to the NP modifier analysis results in a change of interpretation of the morphologically ambiguous *-ey* particle itself from the destination interpretation to the NP location interpretation. If this is the case, then we predict that in contexts in which the initial syntactic analysis can be revised without changing the interpretation assigned to an early constituent, children should be able to successfully revise their initial syntactic analyses.

The Japanese bi-clausal sentences from Experiment 7 provide the ideal basis for testing this question, because a) we found evidence in that experiment for active wh-attachment to the first VP (i.e., embedded clause VP) in the sentence; b) the embedded clause VP can be manipulated in various ways to block the incremental wh-attachment such that the initial syntactic analysis needs to be retracted; and c) the main clause VP can host the wh-phrase without changing the interpretation of the wh-locative adjunct itself. As a reanalysis cue, we used a syntactic error signal and a verb-based error signal in order to examine whether verb information has a special status in triggering reanalysis processes as well.

4.5.1 Method

Participants

Forty eight children who were between the ages of 4;6 and 6;5 and acquiring Japanese as their native language participated in the study. The children were recruited at Ibaraki University Kindergarten and Mito Kindergarten in Mito, Japan. Six children did not complete all the trials and were not included in the data analysis, and another 6

children were excluded due to providing more than two incorrect or irrelevant answers. The remaining 36 children were included in the final sample (mean age: 5;9).

We also recruited 16 adult native speakers of Japanese from the Hiroshima University community in Japan. The adult participants were paid 1000 yen for their participation in a one-hour experiment session that consisted of the present study and other experiments.

Materials and procedure

The experiment used three conditions that were created by modifying the target sentence used in Experiment 7. The first condition contained an overt PP headed by *-de* ('at') specifying the location of the embedded clause event to block the embedded clause interpretation. The second condition contained an overt locative PP specifying the location of the main clause event. This condition was included as a control condition to the first condition, in order to ensure that children would not just provide an answer that corresponded to the overtly mentioned location in the sentence. The third condition contained an embedded clause verb that effectively eliminated felicitous embedded clause interpretations in the given context due to lexical properties of the verb. An example sentence for each condition is shown in (10), and all the target sentences from this experiment are listed in Appendix G.

(10) a. embedded clause PP condition

Doko-de Emily-chan-wa [*pro* ashi-o ki-no-shita-de kegashita to] itteta-no?

where-at Emily-Dim-Top she foot-Acc tree-Gen-below-at hurt Comp was telling-Q

"Where was Emily telling someone that she hurt herself by the tree?"

b. main clause PP condition

Doko-de Emily-chan-wa [*pro* ashi-o kegashita to] puuru-de itteta-no?

where-at Emily-Dim-Top she foot-Acc hurt Comp pool-at was telling-Q

"Where was Emily telling someone at the pool that she hurt herself?"

c. *where*-incompatible embedded verb condition

Doko-de Emily-chan-wa [*pro* ki-kara ochita to] itteta-no?

where-at Emily-Dim-Top she tree-from fell Comp was telling-Q

"Where was Emily telling someone that she fell off the tree?"

The embedded clause PP condition was designed to test the effectiveness of syntactic revision cues. This condition was created by taking the ambiguous wh-question used in Experiment 7, and adding an overt PP headed by a postposition *-de* ('at') that specifies the location of the embedded clause VP (e.g., by the tree). In other words, the overt PP

corresponds to the embedded clause response in the ambiguous wh-question used in Experiment 7. The fronted wh-adjunct *doko-de* ('where-at') is also headed by the locative postposition *-de*, but because the locative PP position for the embedded clause VP is occupied ('filled-gap': Stowe, 1986), the attachment of the wh-phrase to the embedded clause predicate is syntactically blocked. Thus, the only possible interpretation for adults is the main clause interpretation. If children are able to use the syntactic error signal to retract their bias for an embedded clause interpretation, then they are expected to behave like adults and only allow the main clause interpretation. On the other hand, if children are unable to use the syntactic error signal to inhibit the embedded clause interpretation bias in the embedded filled-gap condition (10a), it is predicted that they should provide the embedded clause answer (e.g., by the tree) despite the fact that it was overtly mentioned in the sentence.

It is important to note here that in principle it is not impossible to attach *where* to the embedded clause VP in the filled-gap condition (10a) if there is appropriate contextual information that meets the following three conditions. First, the larger environment that contains the target location (e.g., tree) needs to be specified in the story, such that the part-whole relationship for the target location is made clear (e.g., tree in a park). Second, there needs to be another large environment that contains a counterpart of the target location (e.g., tree in the backyard), such that the two large environments can be contrasted (hurting happened by the tree in a park, but not by the tree in the backyard). Third, the protagonist must explicitly mention the target location as well as its larger environment ("I hurt myself by the tree in the park!"). None of these three conditions obtained in our story design: each location was described as an independent location

without specifying the larger environment; the target location (tree) appears only once in the story; and the protagonist never mentions a larger environment that contains the target location. Moreover, even if these felicity conditions are met, it is more natural to use *where* with a genitive marker *doko-no* that directly selects the target locative PP (e.g., *doko-no ki-no-shita-de,* which could be roughly translated as "at which tree"). It is thus pragmatically infelicitous and grammatically dispreferred to attach *where* to the embedded clause VP, and this is why the overt locative PP in the embedded clause should serve effectively as a filled-gap error signal for active wh-attachment.

The main clause PP condition (10b) was included to examine whether the mere presence of *-de* marked PPs causes children to provide a non-adult-like response. As discussed above, in the embedded filled-gap condition in (10a), children could plausibly exhibit a kindergarten-path effect and provide an embedded clause response, which corresponds to the locative PP overtly mentioned in the embedded clause. When this form of non-adult-like response is observed, however, it may reflect children's reanalysis difficulty, or it may plausibly reflect a task taking strategy by which children treat overtly mentioned locative PPs as the relevant answer to *where* questions. In order to address this possibility, we included the main clause PP condition (10b), in which the location of the main clause event is overtly expressed with a locative PP (e.g., at the pool). If children use this task-taking strategy and treat the overtly mentioned locative PP as the answer to *where* questions, we expect that children would answer the main clause event location 'at the pool' which is already mentioned in the sentence. However, if the kindergarten-path response (e.g. answering the embedded clause location 'tree') in (10a) results from children's inability to retract the active association of the wh-phrase with the embedded

clause predicate, then we expect to see an embedded clause response for (10b) as well.

The *where*-incompatible verb condition (10c) was constructed by changing the embedded clause verb of the original ambiguous sentence in Experiment 7 in such a way that no felicitous interpretation is available for the wh-association with the embedded clause verb. For example, the verb *ochita* ('fell') in (10c) is a change-of-location verb, and the combination of *doko-de* ('at where') and *ochita* ('fell') essentially asks where the falling event happened. However, for a change of location verb, our story design only provides information about the beginning or the end point of the change-of-location event (e.g., falling off the tree). Critically, there is no contextual information about where this entire falling event occurred, because the story does not specify where this tree was. The eight predicates that we used in this condition were chosen based on this criterion, and these predicates included change-of-location verbs (*itta* "went" × 5, *ochita* "fell" × 1), a change-of-state verb (*naru* "become" × 1), and an individual-level predicate (*oishikatta* "was yummy" × 1). We included more than one type of predicate in order to keep the same stories while ensuring verb incompatibility with the wh-phrase, but crucially none of these verb classes yield a felicitous embedded clause interpretation for adults. The change-of-state verb and individual-level predicate were presented together with a theme argument NP, and the change-of-location verbs were presented with a PP argument describing the origin or the direction of the change-of-location event (e.g., falling off a tree, going to a park, etc.). In this condition, if children are able to recognize that the wh-phrase is incompatible with the embedded VP and use this information to retract the embedded clause VP attachment bias, then we predict a preference for the main clause interpretation. On the other hand, if children are unable to use this verb-based error signal

to overcome the embedded clause VP attachment bias, then based on observations by Choi and Trueswell (2010), we predict that children should coerce non-adult-like interpretations and answer the location that was somewhat related to the embedded clause VP (e.g., the tree in (10c)).

Eight sentence sets with three conditions shown in (10) were constructed. We treated each condition as a between-participants factor, and consequently 12 children were randomly assigned to each of the three conditions. The resulting mean age range for the three conditions was 5;7 for the embedded clause filled-gap condition, 5;8 for the main clause filled-gap condition, and 5;11 for the *where*-incompatible verb condition, respectively, and there was no reliable difference in age across groups, $F(2, 33) = 1.70$, $p > .1$. The 12 adults were assigned to the embedded clause filled-gap condition, because as the results show, this is the condition in which children behaved against adults' intuitions. For each condition, the target items were distributed across two lists in the way described in Experiment 5, which resulted in 6 lists. The procedure was identical to that of Experiments 5 and 7, and the experiment took approximately 20 minutes.

4.5.2 Results

The mean proportion of main clause responses for each condition is shown in Figure 14. In the embedded clause filled-gap condition (10a), we found a clear contrast between children and adults, as children demonstrated preference for the non-adult-like embedded clause response (17%, $SE = 9$), whereas adults preferred the main clause response (95%, $SE = 3$) as expected. This difference was significant, $F(1, 27) = 92.36$, $p < .001$.

In the main clause filled-gap condition (10b), children showed a clear preference

for the embedded clause interpretation (13%, *SE* = 6). In the *where*-incompatible embedded verb condition (10c), children showed a preference for the main clause interpretation (74%, *SE* = 9). Since we have no Japanese adult data for the main clause filled-gap and *where*-incompatible embedded verb conditions, we report below comparisons between the three verb-type conditions among Japanese children. There was no significant difference in the proportion of main clause responses between (10a) and (10b), $F < 1$. The proportion of main clause responses in (10c) was above chance level, $t = 2.55$, $p < .05$ (2-tailed), and it was also significantly higher than in the embedded clause filled-gap condition (10a), $F(1, 23) = 19.65$, $p < .001$.

4.5.3 Discussion

This experiment manipulated the nature of error signals and examined whether verb-based error signals and syntactic error signals differ in the extent to which they prompt the retraction of an active commitment. In the embedded clause PP condition, we found that children showed a non-adult-like preference for the embedded clause interpretation and answered the location in which the event described by the embedded clause VP happened (e.g., tree), despite the fact that this location is overtly mentioned in the embedded clause such that this interpretation should not be possible. In the main clause PP condition, children showed an adult-like embedded clause preference. This suggests that the embedded clause preference observed in the embedded clause PP condition did not reflect a task taking strategy to repeat the locative PP that was provided in the sentence as an answer to *where* questions. Thus, taken together with the earlier observation in Experiment 7 that children actively associate the wh-phrase with the

embedded clause VP, the embedded clause preference in the embedded clause PP condition demonstrates a failure to reanalyze the active syntactic commitment.

On the other hand, in the *where*-incompatible embedded verb condition, Japanese children demonstrated a clear preference for the adult-like main clause interpretation, unlike in all the other conditions, in which they showed a robust preference for the embedded clause interpretation. This suggests that children were able to use the verb information to recognize that there is no felicitous embedded clause interpretation available given the context, and retract the active wh-attachment to the embedded clause VP. An interesting generalization that emerges here is that there appears to be a difference in quality of error signal: syntactic error signals did not lead children to retract an active wh-attachment to the embedded clause VP, whereas verb information that indicates the lack of felicitous interpretation did precisely that.

Finally, in Experiment 7 we discussed the possibility that children's preference for the embedded clause interpretation could be explained if children have a strong bias to misanalyze the target sentences as mono-clausal and simply ignore the main clause predicate. This is not implausible, given that bi-clausal sentences would typically include two overt subjects, but the embedded clause subject in our stimuli is a null subject, which may have made it difficult to recognize the presence of two clauses. However, the fact that children entertained the main clause interpretation in the *where*-incompatible embedded verb condition suggests that children are able to recognize these sentences as involving two clauses with two VPs. The present finding casts doubt on the alternative hypothesis that children's embedded clause interpretation in Experiment 7 can be reduced to an incorrect mono-clausal analysis of the input, although we cannot entirely rule out

the possibility that for Japanese children to recognize the presence of the main clause predicate, the wh-phrase must be incompatible with the embedded clause verb. Future research is needed to further investigate this possibility.

4.6 General discussion

The present study used a series of story-based comprehension experiments to investigate adults and children's interpretive biases in the processing of Japanese and English ambiguous wh-questions. Experiment 5 revealed that English-speaking adults and children prefer the main clause interpretation with the exception of when the main clause verb is bare *say,* while Experiment 6 used a Truth Value Judgment Task and replicated the main clause interpretation bias with English-speaking adults. Experiment 7 used the Japanese translation of the ambiguous wh-question to test a case in which the order of the verbs is the opposite of English (main-embedded in English, embedded-main in Japanese). We found that unlike English-speaking adults and children, Japanese adults and children both prefer the embedded clause interpretation, suggesting that in both English and Japanese, adults and children actively associated the wh-phrase with the first VP in the sentence. Finally, building on the finding in Experiment 7 that Japanese speakers actively attach the fronted wh-phrase to the embedded clause VP, Experiment 8 tested the effectiveness of syntactic and verb-based reanalysis cues. The results showed that children still prefer the embedded clause interpretation when the embedded clause locative PP position was occupied by an overt PP to syntactically block the wh-association with the embedded clause VP. On the other hand, when the error signal was based on semantic or pragmatic information from the verb, children were able to overcome the strong embedded clause interpretation bias and entertain the main clause

interpretation. Let us discuss broader implications of these findings for the child sentence processing mechanism.

Active gap creation in children

The observations above strongly suggest that children actively attach the fronted wh-phrase to the first VP in the sentence, and that this active commitment leads to kindergarten-path effects when there is no effective error signal. Our finding is thus consistent with the previous arguments from cross-modal picture priming studies that children attempt to actively complete the dependency between wh-phrases and their thematic positions (Love, 2007; Roberts et al., 2007). In fact, given that picture-priming results only indicate lexical activation of the filler and do not necessarily show that structural and interpretive commitments were made, the current finding provides important new evidence for the presence of active interpretation of filler-gap dependencies.

While the interpretation data reported in this chapter cannot conclusively identify the real-time mechanism that causes adults and children to demonstrate the bias for the first VP as a wh-attachment site, it is important to note that our results demonstrate a remarkable uniformity across languages with different word orders and populations with differences in cognitive capacity. The uniformity in active gap creation behaviors across verb-medial or verb-final languages has been attested by previous cross-linguistic work on wh-dependency processing (Aoshima et al., 2004; Nakano et al., 2002), but to our knowledge, the present study is the first study in which the *same stimuli* were used to elicit cross-linguistic data that establish the link between the temporal order of verbs and wh-attachment preferences in children.

If the first VP association preference reflects active gap creation processes, then the current finding also presents evidence that active gap creation is not restricted to wh-arguments. Most of the existing evidence for active gap creation either in verb-medial or verb-final languages is based on filler-gap dependencies that involve argument fronting, and partly for this reason, it has been proposed that active gap creation processes are triggered in order to saturate the verb argument structure as soon as possible (Pickering & Barry, 1991; Pritchett, 1992). However, since the wh-locative adjuncts used in our study are not selected by verbs, the present finding is more compatible with the view that active gap creation processes are triggered in order to assign a thematic interpretation to the fronted constituent itself. Moreover, Yoshida and Dickey (2008) used a self-paced reading task and found evidence for filled-gap effects using manner adverbs (e.g., *The principal asked the students how/if the teacher carefully told him that…*), further suggesting that active gap creation generally applies to fronted wh-phrases regardless of whether it is an argument or an adjunct.

Children's revision capacity

One striking finding in our study was that within the same structural environment of bi-clausal wh-questions, children's reanalysis success varied depending on what type of error signal was provided. Specifically, children failed to retract the active wh-attachment to the embedded clause VP when they encountered a syntactic error signal in the form of a 'filled-gap' locative PP, whereas they successfully retracted their active wh-attachment to the embedded clause VP when the error signal was based on the absence of felicitous interpretation due to properties of the verb. The contrast between the two sources of error signal is compatible with the previous studies that demonstrated evidence

for children's sensitivity to verb information in making PP attachment decisions (Snedeker & Trueswell, 2004; Kidd & Bavin, 2007; Trueswell et al., 1999), but there are two ways in which our demonstration of verb sensitivity goes beyond the previous work. First, this is the first demonstration that verb information can be successfully used in making syntactic (re)analysis decisions for fronted constituents that precede the verb, as the previous demonstration for verb sensitivity was based on English PP attachment ambiguity resolution where the ambiguous constituent follows the verb. In other words, the temporal order of verbs and ambiguous constituents does not change the privileged status of verb information. Second, the previous demonstration of verb effect was restricted to contexts in which the verb information was relevant for deciding whether an ambiguous PP should attach to a preceding verb or an NP, but in the present study, verb information was critical in deciding whether the fronted constituent should attach to the first VP or the second VP in the sentence. The fact that verb sensitivity was observed in diverse structural environments lends further support to the view that verb information generally plays a critical role in child sentence processing.

On the other hand, it is less clear why the filled-gap, syntactic error signal failed to cause children to retract their active syntactic commitments. One possible interpretation of this finding is that syntactic error signals in general are not effective revision cues for children. This explanation is consistent with the original demonstration of kindergarten-path effects. Trueswell and colleagues (Trueswell et al., 1999) presented sentences like *Put the frog on the napkin in the box*, and the presence of the second PP was supposed to serve as a syntactic revision cue. However, children failed to retract their earlier syntactic analyses in which the first PP was treated as a destination for the verb

put. Thus, the present finding could be seen as a novel argument that syntactic error signals in general do not effectively lead children to retract their incremental syntactic commitments.

Alternatively, it is possible that the filled-gap error signal was not effective because the fronted wh-phrase was an adjunct. Because adjuncts are not selected by verbs, there is no syntactic limit to how many adjuncts attach to VPs. This flexibility in VP-adjunction may be the reason why the filled-gap manipulation for wh-adjuncts was not effective. On the other hand, filled-gap error signals should be effective when the fronted wh-phrase is an argument of the verb. Arguments are selected by verbs or prepositions, and there is only a restricted number of argument positions available depending on properties of the verb. Thus, if the available argument positions are occupied by overt constituents, this may serve as a very clear error signal. Moreover, because wh-arguments must be selected by verbs or prepositions, there will be a syntactic head after the filled-gap that is left without an overt complement (e.g., a preposition in a sentence like *What did Emily eat the cake with ___ ?*), and the lack of an overt argument for the preposition may also facilitate the retraction of the active creation of an object gap. If the syntactic distinction between arguments and adjuncts affects the effectiveness of a filled-gap as a revision cue, we predict that if the fronted wh-phrase is a potential argument for the embedded clause predicate such as a wh-dative NP, then an overt 'filled-gap' dative NP in the embedded clause should lead children to successfully retract the embedded clause interpretation and adopt the adult-like main clause interpretation (for real-time evidence for active gap creation with fronted wh-dative NP, see Aoshima, et al., 2004). This remains to be studied in future research.

Our demonstration of kindergarten-path effects may appear to be in conflict with the results from Choi and Trueswell (2010), in which Korean children were unable to use later arriving verb information to revise PP attachment decisions that had been made earlier in the sentence. We argue, however, that there is no conflict, and in fact the present finding along with Choi and Trueswell's results together suggest that early *interpretive* commitments cannot be easily retracted. As we discussed above, a number of psycholinguistic studies with adults show that even adults generally have difficulties in retracting early interpretive commitments when the initial syntactic analysis that led to such early commitments has been disconfirmed by later arriving error signals (Ferreira & Patson, 2007; Sturt, 2006). In Choi and Trueswell's stimuli, the critical ambiguity arises due to the morphological ambiguity of the *-ey* particle, which can be analyzed as a destination marker (forcing VP attachment) or a genitive case marker (forcing NP attachment). Here, resolving the morphological ambiguity leads children to make an interpretive commitment, and we argue that it is this early interpretive commitment that cannot be undone. In our stimuli, on the other hand, the interpretation of the fronted wh-locative adjunct itself does not vary as a function of the attachment site, and this is why children were successful at using the verb information in retracting syntactic commitment to associate the wh-phrase with the embedded clause VP and eventually making adult-like wh-attachment decisions. Further studies are needed to investigate this suggestion in a more closely controlled structural environment, but this line of detailed comparative research on adult child reanalysis mechanisms promises to further our understanding of the nature of kindergarten-path effects.

4.7 Conclusion

The results of our study suggest that a) adults and children have a first VP bias in resolving wh-attachment ambiguity regardless of the canonical word order of the language; and b) children can retract their wh-attachment to the first VP when there is a verb-based reanalysis cue, but not when there is a syntactic reanalysis cue. This suggests that children can revise their initial syntactic analysis so long as it does not change the interpretation assigned to earlier constituents the sentence, and that children treat verb information as a more reliable cue for making syntactic reanalysis decisions than syntactic error signals.

Chapter 5: Active processing of reconstruction and binding

Chapter 4 focused on active filler-gap dependency processing and the reanalysis mechanism in children, and demonstrated a kindergarten-path effect that occurred when the reanalysis cue (i.e., the lack of felicitous interpretation) was indicated by syntactic information but not when it was indicated by information from the verb.[11] Based on this finding and a review of the previous demonstration of kindergarten-path effects in PP attachment ambiguity resolution, I proposed that incremental syntactic and interpretive commitments that are confirmed by bottom-up evidence cannot be easily retracted (as in Choi & Trueswell, 2010; Trueswell et al., 1999), and (b) active syntactic commitments that are not confirmed by critical bottom-up evidence can be retracted based on later arriving verb-based information (Chapter 4). This chapter extends the domain of investigation to another long distance dependency - namely, binding relations - and lends support to this generalization.

A Truth Value Judgment study by Leddon and Lidz (2006) investigated children's interpretation of anaphor reconstruction, and found an interesting asymmetry: When the sentence makes two binding interpretations grammatically available, children only adopt the first one that becomes available in the sentence, but when the sentence is manipulated to grammatically license only the second interpretation, children are able to block the first interpretation and only adopt the second, grammatical interpretation. Leddon and Lidz argued that children incrementally establish binding relations according to grammatical

[11] Part of the work reported in this chapter was conducted in collaboration with Chris Dyer, Shiti Malhotra and Jon Sprouse.

constraints, and once they adopt a grammatically permissible binding relation, they do not consider alternative grammatically permissible options.

Their findings may lend support to our generalization that incremental syntactic commitments that are made based on bottom-up information cannot be retracted. However, Leddon and Lidz did not report time course evidence to support the claim that the interpretive commitment was made incrementally. As explained below, the sentences used in their study have features that make it difficult to design a feasible real-time study with children, and for this reason this chapter presents data from adults to provide indirect support for the presence of kindergarten-path effects. I will present time course evidence from adults that partially supports Leddon and Lidz's proposal: In sentences that make two interpretations available, the one that children prefer is indeed the one that adults also consider first in the sentence. Contrary to Leddon and Lidz's conclusion, however, the parser in fact transiently entertains the ungrammatical binding relation in sentences that grammatically license only the second interpretation. It is argued that this is consistent with the fact that the parser makes active syntactic commitments without bottom-up evidence, and that children who make this active commitment can retract this on the basis of later-arriving bottom-up information from the verb.

5.1 Kindergarten-path effects in reflexive binding?

Leddon and Lidz (2006) used a Truth Value Judgment Task (Crain & Thornton, 1998) to investigate whether 4-year-old children can entertain all the grammatically permissible interpretive possibilities in sentences like (1).

(1) a. Wh-argument fronting condition

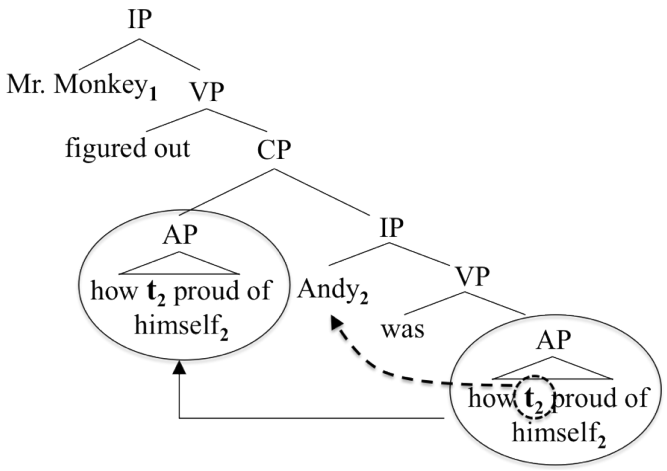
Miss Cruella$_1$ knew which painting of herself$_{1/2}$ Janie$_2$ put up ____ .

b. Wh-predicate fronting condition

Mr. Monkey$_1$ figured out how proud of himself$_{*1/2}$ Andy$_2$ was ____ .

In both (1a) and (1b), the fronted complex wh-phrases contain a reflexive, which can be bound by the embedded clause subject *Janie* (1a)*/Andy* (1b) as if the reflexive were still in the original gap position (Barss, 1986; Fox & Nissenbaum, 2004; Huang, 1993; Heycock, 1995; Takano, 1995). I will refer to this interpretive possibility as a *reconstruction interpretation*. Interestingly, however, another interpretive possibility arises depending on what type of wh-phrase is being fronted. When the fronted wh-phrase is the internal argument of the embedded clause predicate as in (1a), the reflexive *herself* can be bound by the subject of the main clause *Miss Cruella*. I will refer to this interpretive possibility as a *surface interpretation*. On the other hand, when the fronted wh-phrase consists of the predicate of the embedded clause (1b), then the reflexive inside the wh-phrase cannot be bound by the main clause subject *Mr. Monkey.* The unavailability of the surface interpretation in (1b) can be explained if we assume that the fronted wh-predicate contains an unpronounced trace left by the movement of the embedded clause subject to [Spec, TP], which in turn serves as the local c-commanding antecedent for the reflexive inside the wh-predicate (Huang, 1993). This is illustrated in (2).

(2)



In summary, both the surface and reconstruction interpretations are available when a wh-argument is fronted (1a), but when a wh-predicate is fronted (1b), only the reconstruction interpretation is available and the surface interpretation is grammatically blocked. (For a review on reconstruction and binding, see Sportiche, 2006).[12]

[12] The availability of the surface interpretation may be specific to languages like English in which a reflexive inside a *picture NP* behaves as an exempt anaphor, to which the structural constraints on binding do not apply (Pollard & Sag, 1992). In German, for example, reflexives never behave as exempt anaphors, and the German counterpart of (2a) only allows the reconstruction interpretation (Büring, 2005).

Note also that Heycock (1995) uses Condition C effects in (i) to argue against Huang's predicate internal subject hypothesis.

(i) *How afraid of Margaret$_1$ do you think she$_1$ expects John to be ___ ?

In (i), *Margaret* does not c-command the pronoun *she* on the surface, so it is expected that the pronoun *she* could be co-indexed with *Margaret*, contrary to fact. Heycock uses this data to argue that wh-predicates obligatorily reconstruct at LF, and this can account for the fact that reflexives in this environment can only be bound by the embedded clause subject as well. However, this does not show that Huang's analysis is wrong, only that it is not sufficient for explaining all the reconstruction data. Thus, I will assume Huang's analysis in the remainder of the chapter.

For both sentence types in (1), Leddon and Lidz constructed stories that made only the surface interpretation true (i.e. main clause subject = the antecedent) or only the reconstruction interpretation true (i.e. embedded clause subject = the antecedent). After each story, a puppet described the story by stating (1a) or (1b), and participants judged whether the puppet's description was true. As expected, the adult control group accepted both the surface and reconstruction interpretations for (1a), but they only accepted the reconstruction interpretation for (1b). The 4-year-old children also only accepted the reconstruction interpretation for (1b), but for (1a), unlike adults, they generally accepted the surface interpretation while rejecting the reconstruction interpretation.

Leddon and Lidz argued that this is another instance of a kindergarten-path effect: In processing sentences like (1), children establish a binding relation at the first grammatically permissible position; i.e., the surface binding in (1a), and the reconstruction binding in (1b). In (1a), the subsequent embedded clause region could give rise to the reconstruction interpretation; Leddon and Lidz argue, however, that once children assign an antecedent to the reflexive, they do not revise their initial interpretation even when the task is designed to force them to examine the alternative interpretive possibility. In fact, it is worth noting that for (1a), the adult control group showed a higher acceptance rate for the surface interpretation (~90%) than for the reconstruction interpretation (~70%), suggesting that adults also preferred to retain the surface interpretation that became available first in the sentence.[13]

[13] Note that the fact that adults mostly accepted the dispreferred and yet grammatical reconstruction interpretation contrasts with the Truth Value Judgment Experiment (Experiment 7) reported in Chapter 4, where adults mostly rejected the grammatical embedded clause interpretation due to their strong preference for the main clause interpretation. One possible reason is the difference in the design of the Truth

If Leddon and Lidz's argument is correct, then their results lend further support to the view that incremental interpretive commitments that are based on bottom-up evidence cannot be retracted. The surface interpretation in (1a) can be constructed as soon as the reflexive is encountered, based on the bottom-up information that has become available. However, given that the Truth Value Judgment responses were made after the entire sentence was presented, there is no evidence that the surface interpretation was actually assigned incrementally during real-time sentence processing. An alternative interpretation of the data is that children do not make an incremental commitment to the surface interpretation, but rather wait until the end of the sentence, evaluate both interpretive possibilities, and then use some interpretation selection strategies and grammatical constraints to preferentially adopt the surface interpretation in (1a) and the reconstruction interpretation in (1b). If Leddon and Lidz's argument is viable, it should be possible to gather time course evidence that shows that the surface interpretation in (1a) indeed becomes available as soon as the reflexive is encountered, and that the surface interpretation is not considered when the reflexive is encountered in (1b).

There are various on-line measures of real-time sentence processing that can be used with 4-year-old children (see Sekerina et al., 2008). One primary example is visual-world eye-tracking, a kind of looking during listening paradigm (Fernald, Zangl, Portillo, & Marchman, 2008; Trueswell et al., 1999). However, there are non-trivial

---

Value Judgment Task. In our experiment in Chapter 4, the context was designed in such a way that both meanings were available, and the target sentence forced the adults to consider one of the two grammatically permissible interpretations. In Leddon and Lidz (2006), the target sentence was ambiguous, but the context was designed to be compatible with one of the two grammatically permissible interpretations. If this is the case, then modifying the story designs of Experiment 7 to force one of the two possible interpretations should cause adults to accept both grammatically permissible readings.

methodological challenges to implementing a child eye-tracking study using sentences like (1). First, it is not obvious how to construct a scene that is equally natural for both (1a) and (1b), given that the wh-phrase in (1a) involves a readily depictable object like *which picture,* whereas the wh-phrase in (1b) involves a description of a psychological state such as *proud* or *impressed*. Another challenge to implementing (1) in a visual world context is that the target reflexive region is followed immediately by the subject of the embedded clause. Thus, even if children immediately entertain the surface interpretation, by the time they program an eye movement towards the antecedent that corresponds to the surface interpretation, they will hear the beginning of the embedded clause subject. This may cause children to look instead at the image or object that corresponds to the embedded clause subject, which may mask transient consideration of the reconstruction interpretation.

In the two experiments reported in this chapter (Experiments 9 and 10), we attempt to provide time course evidence for Leddon and Lidz's argument using real-time reading data from adults. Reading paradigms provide time course data for sentence processing without the need to design a scene with images and objects that correspond to the description of the sentence. Moreover, as we pointed out, in processing the wh-fronting condition (1a), adults also showed a higher acceptance rate for the surface interpretation than for the reconstruction interpretation. The fact that adults showed a preference for the surface interpretation like children suggests that there is a common underlying cause for this preference: namely, the sentence comprehension mechanism attempts to preserve the first interpretation that was assigned based on the bottom-up information that has become available.

Among the many reading studies that have examined the time course of anaphora processing, an eye-tracking study by Sturt (2003) is particularly relevant for our purpose. This study used a sentence like (3) in which the gender feature of the reflexive was manipulated (*himself* or *herself*), while the only grammatical antecedent in the local domain was a stereotypically male or female noun (e.g., *surgeon* is stereotypically male). Moreover, the target sentence contained a male or female pronoun that referred to a name introduced in a previous sentence (e.g., Jonathan or Jennifer), but this pronoun appeared in the main clause subject position while the reflexive appeared in the embedded clause. Since reflexives must be bound by a c-commanding antecedent in the local domain, which roughly corresponds to the same clause (Principle A: Chomsky, 1981), the pronoun in the main clause subject position is grammatically inaccessible for the antecedent search.

(3) {Johathan|Jennifer} was pretty worried at the City Hospital. {**He|She**) remembered that the **surgeon** had pricked {**himself|herself**} with a used syringe needle. There should be an investigation soon.

Eye movement measures that reflect early phases of syntactic processing, such as first fixation duration or first pass reading times, revealed that readers spent more time reading the reflexive region when its gender feature mismatched the stereotypical gender of the grammatically accessible antecedent (e.g., *herself* and *surgeon* in (3)), while the reading time on the reflexive was not affected by the gender of the pronoun in the grammatically inaccessible position. However, eye movement measures that reflect later phases of

syntactic processing, like second pass reading time, indicated that readers spent more time re-reading the reflexive region when the inaccessible antecedent (*he* or *she*) mismatched the gender of the reflexive. This suggests that readers may consider the ungrammatical binding possibilities in later phases of sentence interpretation, but that at least the initial phase of antecedent search process is restricted to potential antecedents in grammatically accessible positions.

Experiment 9 uses an experiment design inspired by Sturt (2003) to test whether the time course data on reflexive processing support the assumption that Leddon and Lidz (2006) made to account for children's interpretation of (1). The present study used a self-paced reading paradigm to examine the time course of processing sentences like (4).

(4) a. Wh-argument fronting, gender match

Alice recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.

b. Wh-argument fronting, gender mismatch

Andrew recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.

c. Wh-predicate fronting, gender match

Alice recalled how pleased with herself the attractive young nanny had been during the summer vacation.

d. Wh-predicate fronting, gender mismatch

Andrew recalled how pleased with herself the attractive young nanny had been during the summer vacation.

In order to probe whether the surface interpretation is incrementally established upon encountering the reflexive inside fronted wh-phrases, this experiment manipulated the wh-type (argument vs. predicate) as well as the gender compatibility (gender match vs. mismatch) between the main clause subject and the reflexive. Previous studies on reflexive and pronominal binding have found that in a grammatically permissible binding configuration, evidence for reading disruption is observed when the gender feature of the antecedent and the anaphora mismatches (gender mismatch effect: Kazanina et al., 2007; Sturt, 2003; van Gompel & Liversedge, 2003; Xiang, Dillon, & Phillips, 2009). Recall that Leddon and Lidz (2006) argued that for children the reflexive inside the wh-argument is bound immediately, but the reflexive inside the wh-predicate is not bound because of grammatical constraints. If adults behave this way when reading sentences like (4), then the reflexive inside the fronted wh-argument mismatch condition (4b) should cause a reading time slow-down compared to the wh-argument match condition (4a), while we would expect to see no gender mismatch effect in the wh-predicate conditions (4c) and (4d).

5.2 Experiment 9 (self-paced reading)

5.2.1 Method

Participants

We recruited 25 native speakers of American English from the University of Maryland community. They all received a course credit or were paid $5 for their participation, and they were naïve to the purpose of the experiment. One participant was excluded from data analysis because their global average reading times were beyond three standard deviations from the mean. The data from the remaining 24 participants are

reported below.

Materials

We used 24 sets of four sentences like (4a) to (4d), which are all listed in Appendix H. In these sentences, the main clause subject was always either a strongly male-biased or female-biased name, taken from the Social Security Administration's report of popular names of the 1980's. The wh-phrase in the argument-fronting conditions generally consisted of "which [noun] [preposition] [reflexive]" (e.g., *which picture of himself, which rumor about herself*), whereas the wh-phrase in the predicate-fronting conditions consisted of "how [adjective] [preposition] [reflexive]" (e.g., *how angry with himself, how proud of herself*). In all four conditions, the critical reflexive region was followed by a determiner *the* along with two adjectives which did not require the use of comma (e.g., *grumpy old* [noun]). These regions were included so that a potential spill-over effect from the reflexive region could be observed. The 24 sentence sets were counter-balanced across four lists so that each participant saw only one version of the target items and consequently read 6 tokens from each condition. In addition, 96 fillers of similar length and complexity were constructed and added to each list.

Procedure

The self-paced reading task was implemented on the Linger software developed by Doug Rohde (http://tedlab.mit.edu/~dr/Linger/). We used a word-by-word, non-cumulative moving window presentation (Just, Carpenter, & Woolley, 1982). In this design, each sentence initially appears as a series of dashes, and these dashes are replaced by a word from left to right every time the participant presses the space bar. In order to ensure that the participants were paying attention while reading the sentences, all

sentences were followed by yes-no comprehension questions, and feedback was provided if the questions were answered incorrectly. At the beginning of the experiment, participants were instructed to read at a natural pace and answer the questions as accurately as possible. Seven practice items preceded the self-paced reading experiment, and the order of presentation was randomized for each participant. The experiment took approximately 30 minutes.

Analysis

Only trials in which the comprehension question was answered accurately were included in the analysis. Self-paced reading times for the target sentences were examined for each successive region, although the region after the embedded clause verb (Region 13) contained a different number of words across conditions and was also far past the critical regions, so this region was not submitted to statistical analysis. Reading time data that exceeded three standard deviations from the group mean at each region and in each condition were excluded, affecting 2.1% of the data. The participant mean (F1) and item mean (F2) of the remaining reading time data for each region were submitted to a repeated measures 2 × 2 ANOVA with the factors wh-type (argument vs. predicate) and gender compatibility (match vs. mismatch). In the critical regions, planned comparisons were conducted to test for systematic differences across wh-argument and wh-predicate conditions within each gender type.

5.2.2 Results

Comprehension accuracy. The mean comprehension question accuracy for experimental items across participants and items was 92.2%. For the wh-argument fronting conditions, the gender-match sentences were answered with an accuracy of 91%

(*SE* = 2.5), and the gender-mismatch sentences with an accuracy of 91% (*SE* = 2.2). For the wh-predicate fronting conditions, the gender-match sentences were answered with an accuracy of 95.1% (*SE* = 1.6), and the gender-mismatch sentences with an accuracy of 91.7% (*SE* = 2.5). The mean accuracy did not differ reliably across conditions ($p > .1$).

Reading time data. The critical regions where a potential gender mismatch effect was expected consist of Region 6 (i.e., the reflexive *herself/himself* in the example sentence in (4)) and the following Regions 7, 8 and 9 (i.e., the determiner *the*, and two adjectives *attractive young* in the example sentence in (4)), in which spill-over effects could be observed. Regions 1 through 5 were not completely lexically matched: Region 1 differed in the name between gender match and mismatch conditions, and Regions 3 to 5 differed between wh-argument and wh-predicate conditions due to the difference in the fronted wh-phrase. Thus, in these regions we might observe effects related to these lexical differences. Regions 10 through 12 were lexically matched and far from the critical region, and therefore should not exhibit any statistical differences.

The region-by-region mean reading time for the wh-argument conditions is presented in Figure 15, and the mean region-by-region reading time for the wh-predicate conditions is presented in Figure 16.

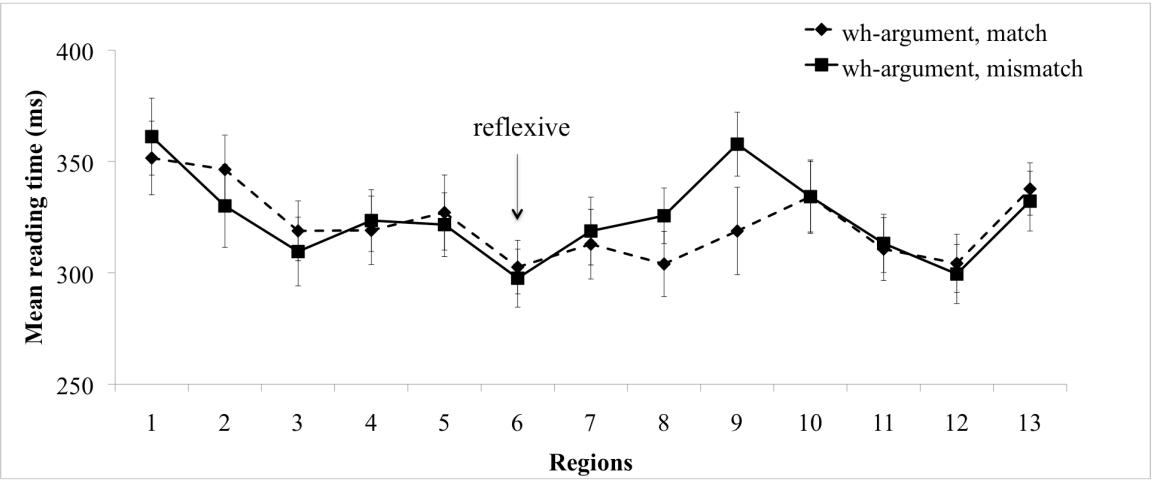


Figure 15. Mean reading time (ms) for the wh-argument gender match and mismatch conditions. Error bars indicate standard error of the mean.

Sample sentence (critical regions are highlighted in bold):

Alice/Andrew$_1$ recalled$_2$ which$_3$ drawing$_4$ of$_5$ **herself$_6$ the$_7$ attractive$_8$ young$_9$** nanny$_{10}$ had $_{11}$ damaged$_{12}$ [during the summer vacation.] $_{13}$

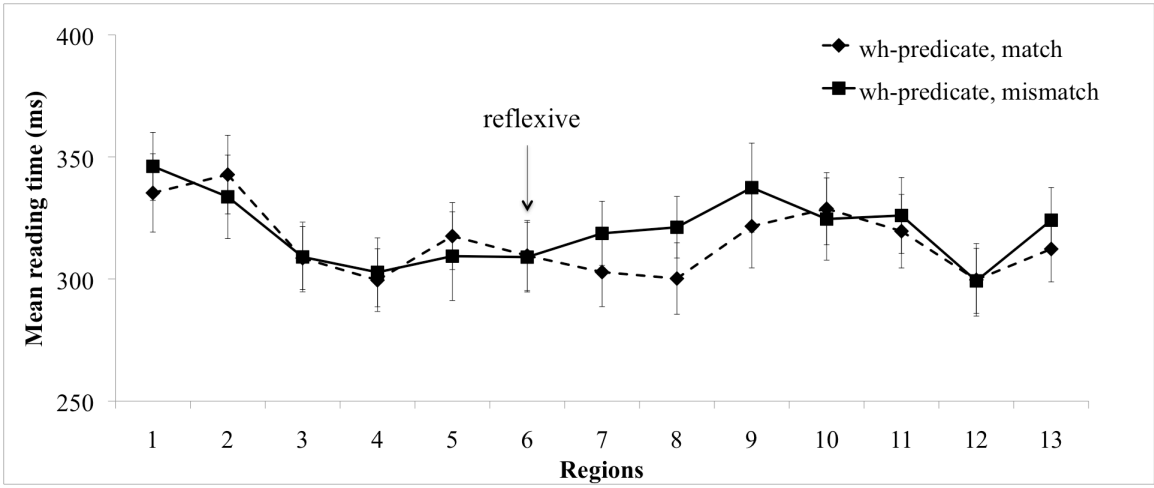


Figure 16. Mean reading time (ms) for the wh-predicate gender match and mismatch conditions. Error bars indicate standard error of the mean.

Sample sentence (critical regions are highlighted in bold):

Alice/Andrew$_1$ recalled$_2$ how$_3$ pleased$_4$ with$_5$ **herself$_6$ the$_7$ attractive$_8$ young$_9$** nanny$_{10}$ had $_{11}$ been$_{12}$ [during the summer vacation.] $_{13}$

In non-critical regions from Regions 1 through 5, we found some spurious effects as well as effects that were expected based on the lexical differences across conditions. In Region 1, there was a main effect of wh-type in by-items analysis, $F1(1, 23) = 2.28$, MSE = 5888, $p > .1$, $F2(1, 23) = 4.31$, MSE = 3665, $p < .05$. Regions 2 and 3 revealed no significant difference across conditions (Fs < 2, p > .1), and Region 4 revealed a main effect of wh-type, $F1(1, 23) = 13.5$, MSE = 9750, $p < .005$, $F2(1, 23) = 5.09$, MSE = 6870, $p < .05$. Region 5 revealed no significant difference across conditions (Fs < 2, p > .1). Given that these effects were small and well ahead of the critical regions, the noise in these regions is unlikely to affect the observation in the critical regions.

The critical reflexive region (Region 6) and the immediately following determiner region (Region 7) revealed no significant differences (Fs < 2, p > .1). The following adjective regions (Regions 8 and 9) revealed a main effect of gender (Region 8: $F1(1, 23) = 8.53$, MSE = 10918, $p < .01$, $F2(1, 23) = 18.71$, MSE = 11105, $p < .0005$; Region 9: $F1(1, 23) = 8.09$, MSE = 18031, $p < .01$, $F2(1, 23) = 6.57$, MSE = 13686, $p < .05$), but these regions revealed neither main effects of wh-type nor significant interactions (Fs < 2, $p > .1$). The planned comparison in Region 8 revealed that in the wh-argument fronting conditions the mismatch condition was reliably slower than the match conditions (argument match = 304ms, argument mismatch = 326ms), $t1(23) = 2.35$, $p < .05$, $t2(23) = 2.1$, $p < .05$, and the same pattern was observed in wh-predicate fronting conditions (predicate match 300ms; predicate mismatch 321ms), with a reliable difference in by-items analysis and a marginally significant difference in by-participants analysis, $t1(23) = 1.79$, $p = .086$, $t2(23) = 2.51$, $p < .05$. The planned comparison in Region 9 revealed that the gender mismatch condition was slower than the gender match condition in the wh-argument fronting conditions (argument match = 319ms, argument mismatch = 358ms), $t1(23) = 2.64$, $p < .05$, $t2(23) = 2.13$, $p < .05$, but the difference between the two wh-predicate conditions (predicate match = 322ms, predicate mismatch = 337ms) did not reach significance, $t1(23) = 1.44$, $p > .1$, $t2(23) < 1$. The non-critical Regions 10 through 12 did not show significant differences across conditions (Fs < 2, p > .1).

5.2.3 Discussion

For the wh-argument conditions, we observed a slow-down in the regions that followed the critical reflexive region (Regions 8 and 9, which corresponded to adjectives)

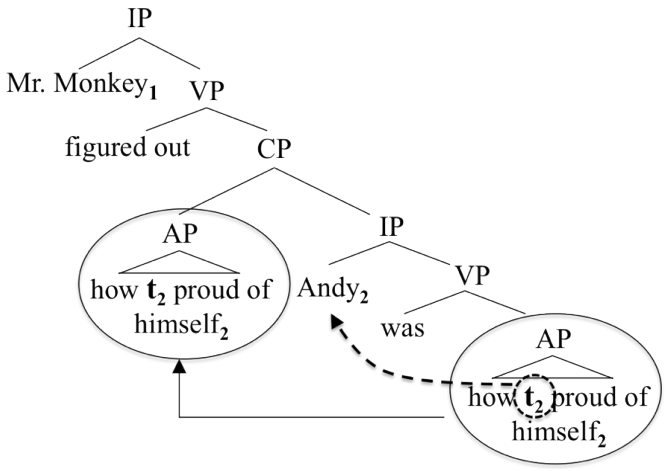
when the gender of the reflexive mismatched that of the main clause subject as compared to when the gender of the reflexive matched that of the main clause subject. The observation of a gender mismatch effect in these conditions suggests that the reflexive inside the fronted wh-argument was bound immediately by the main clause subject, and that the slow-down occurred due to the lack of agreement in gender features. For the wh-predicate conditions, however, Region 8 revealed weak evidence for a gender mismatch effect. Although the effects were weaker and not as reliable as in the wh-argument fronting conditions, this was nevertheless an unexpected result given that the reflexive inside the wh-predicate could not be grammatically bound by the main clause subject.

One possible account for the unexpected gender mismatch effect in the wh-predicate condition is that the antecedent search is unconstrained by the structural environment or the locality domain of the reflexive, and that the search process attempts to associate the reflexive with any nouns with a [+human] feature that were encountered in the input. However, this account is unlikely, since previous studies have shown that the antecedent search process for reflexives is strictly constrained by Binding Principles (Sturt, 2003; Xiang et al., 2009; cf. Badecker & Straub, 2002).[14] Furthermore, this account predicts that the incremental reflexive binding by the main clause subject should be equally available in the wh-argument and wh-predicate conditions, and hence cannot explain why the gender mismatch effect was weaker in the wh-predicate conditions. Therefore, the correct explanation for the present findings is likely to come from a view

[14] Runner, Sussman and Tanenhaus (2006) presented visual world eye-tracking evidence that suggests that reflexives inside picture NPs may not be as strictly constrained by Binding Principles as reflexives in other syntactic environments. Note, however, that what is surprising is the evidence for a binding relation in the predicate fronting condition, for which there has been no evidence so far that reflexives inside predicates do not respect Binding Principles.

that takes seriously the structural representation of sentences with wh-predicate fronting, which was presented in (2) and is repeated below for convenience.

(2)



Here, let us revisit our original predictions based on the structure shown in (2). The prediction that there should be no gender mismatch effect in the wh-predicate conditions was based on the assumption that the reflexive can only be bound grammatically by the trace inside the fronted wh-predicate, which in turn must be co-indexed with the embedded clause subject. However, it is important to take into account that at the point of processing the reflexive in real time, the parser has not encountered the embedded clause subject, and is thus unable to identify what the trace or the null subject of the fronted predicate should refer to. With this uncertainty about the reference of the null subject in mind, let us examine two possible accounts for our finding.

First, the comprehension mechanism may initially only search within the local domain (i.e., inside the fronted AP), but if the local domain contains no overt antecedent

that indicates an identifiable referent, then the parser may go beyond the locality domain and search for the next c-commanding antecedent, i.e., the main clause subject. Let us call this a *domain extension hypothesis*. On this view, the gender mismatch effect we observed in the wh-predicate fronting conditions may be analogous to what Sturt (2003) found in the second pass reading time measure, where there was evidence that readers showed sensitivity to the gender (dis-)agreement between the reflexive and non-local c-commanding subject. In our self-paced reading data, a gender mismatch effect in argument and predicate conditions was observed in the same region, but this may be due to the lower temporal precision of self-paced reading data compared to the eye-tracking during reading method used in Sturt (2003). Under this account, the weaker effect in the wh-predicate fronting conditions may reflect that the parser is less committed to the gender mismatch because the main clause subject only comes under consideration during the secondary search that goes outside the original local domain.

Another possible account is that the search domain is restricted to the fronted wh-predicate phrase (e.g., AP in (2)), but that a gender mismatch effect arises because the parser has temporarily analyzed the trace or the null subject inside the AP as co-referential with the main clause subject. Let us call this account an *active co-indexation hypothesis*. Under this hypothesis, the reflexive is bound by the null subject of the fronted predicate, but the identity of the referent for the index assigned to the trace and the reflexive could be co-referential with the main clause subject. This is informally illustrated in (5).

(5) Mr. Monkey$_1$ figured out how t_X proud of himself$_X$ … [embedded subject] $_X$

Active co-index hypothesis: Index X = 1

It is important to note here that this active co-index hypothesis can be correct if the embedded clause subject turns out to be a pronoun that is co-referential with the main clause subject (e.g., John$_1$ remembered how $t_{X=1}$ proud of himself $_{X=1}$ he $_1$ used to feel…). Under this hypothesis, the actual index assigned to the null subject is still underspecified, and hence if the actual embedded subject has a different value (e.g., 2), then the index X will automatically be replaced by this index. This active co-indexation process is a risky strategy in that the embedded clause subject may turn out not be a pronoun, but it allows the comprehension mechanism to immediately assign an external theta role of the fronted predicate. However, as discussed in previous chapters, the parser actively builds tentative structural representations for filler-gap dependencies as well, so in this sense, the active co-indexation hypothesis is not entirely unreasonable.

On this account, a gender mismatch effect arises in both wh-argument and wh-predicate fronting conditions because the antecedent in both conditions is in the grammatically accessible local domain, i.e., the main clause subject for the reflexive in the wh-argument condition, and the null subject inside the AP that is tentatively co-indexed with the main clause subject. The weaker effect in the wh-predicate conditions could be attributed to the fact that the error signal for the actively assigned co-index is presented immediately after the reflexive, i.e., the determiner of the embedded clause subject. This serves as an effective, bottom-up error signal because the trace must have an identical referent as the embedded clause subject, and if the embedded clause subject is

not a pronoun that could be co-referential with the main clause subject, the actively assigned co-index for the trace and main clause subject must be abandoned.

Importantly, the domain extension hypothesis and the active co-indexation hypothesis make different time course predictions that could potentially be teased apart in an eye-tracking during reading study. The domain extension hypothesis predicts that the gender mismatch effect in wh-predicate conditions should arise later than in the wh-argument conditions, because the search domain is extended only after the initial search in the local domain has failed to locate an antecedent. Thus, this account predicts that evidence for gender mismatch effects for the wh-argument conditions should be found in early measures of eye movement, while evidence for gender mismatch effects for the wh-predicate conditions should be found in later measures of eye movement. On the other hand, on the active co-indexation hypothesis, the grammatically accessible antecedent is in the local domain for wh-argument and wh-predicate conditions alike, and therefore this hypothesis predicts that a gender mismatch effect should be observed in an equally early measure regardless of the wh-type. Moreover, if the hypothesized co-indexation is retracted as soon as the embedded clause subject is encountered, we predict that the gender mismatch effect in the wh-predicate condition should disappear as soon as the embedded clause subject region begins. Additionally, the active co-indexation mechanism may yield a slower reading time in the embedded clause subject region for the wh-predicate gender match condition, because the fact that the embedded clause subject is not co-referential with the man clause subject would require the actively assigned index to be retracted.

5.3 Experiment 10 (eye-tracking during reading study)

5.3.1 Method

Participants

We recruited 44 native speakers of American English from the University of Maryland community. All had normal or corrected-to-normal vision, and were naïve to the purpose of the experiment. They received course credit or were paid $10 for their participation.

Materials

The target sentences used in this experiment were the same as the 24 target sentence sets used in Experiment 10 in (4), which is repeated below.

(4) a. Wh-argument fronting, gender match

Alice recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.

b. Wh-argument fronting, gender mismatch

Andrew recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.

c. Wh-predicate fronting, gender match

Alice recalled how pleased with herself the attractive young nanny had been during the summer vacation.

d. Wh-predicate fronting, gender mismatch

Andrew recalled how pleased with herself the attractive young nanny had been during the summer vacation.

These 24 sentence sets were counter-balanced across four lists so that each participant saw only one version of the target items and consequently read 6 tokens from each condition. In addition, 100 fillers of similar length and complexity were constructed and added to each list.

Procedure

An SR Research Eyelink 1000 tracker (Mississauga, Ontario, Canada) was used to record eye movements. The participant's head was stabilized by a chin rest and a forehead rest. The position of the right eye only was monitored at a sampling rate of 1000 Hz. The eye-tracker display allowed a maximum of 100 characters per line. Stimuli were displayed on a 17-inch monitor, and participants were seated 60 cm from the computer screen.

Before the experiment started, participants were seated in front of the eye-tracker and received instructions for the experiment. A calibration routine was performed at the beginning of the experiment, and the experimenter monitored the calibration accuracy throughout the experiment, recalibrating when necessary. The experiment started with written instructions on the display and four practice trials. At the beginning of each trial, a black square was displayed on the left side of the monitor, which corresponded to the beginning of the sentence. The text was displayed when the participant successfully fixated on the square. After reading each sentence, the participant pressed a button to remove the sentence display. Each sentence was followed by a yes-no comprehension question, and the participant answered the comprehension questions by pressing a left or right button. The entire experiment lasted approximately 40 minutes.

Data analysis

Data from six subjects were removed due to calibration errors. The trials in which participants provided an incorrect answer to the comprehension question were removed from eye movement analyses. For the remaining data, an automatic procedure pooled short contiguous fixations. The procedure incorporated fixations of less than 80 ms into larger fixations that occurred no more than one character's distance away and deleted any remaining fixations of less than 80 ms, because little information can be extracted during such short fixations (Rayner & Pollatsek, 1989). Unusually long fixations, those greater than 800 ms, were also removed, because they usually reflect tracker losses or other anomalous events. This procedure resulted in the exclusion of 2.41% of all fixations.

For the purpose of analysis of the eye movement data, the sentences were divided into the following regions.

(6) a. Wh-argument fronting, gender match/mismatch

{Alice|Andrew} recalled which drawing/ of herself/ the attractive/ young nanny/ had damaged during the summer vacation.

b. Wh-predicate fronting, gender match/mismatch

{Alice|Andrew} recalled how pleased/ with herself/ the attractive/ young nanny/ had been during the summer vacation.

We report eye movement data in the following four regions. Region 1 is the pre-reflexive region that includes all the words before the preposition inside the wh-phrase. This region

is analyzed to ensure that there was no unexpected reading time difference that might compromise the interpretation of the data from the critical region, but we must also be cautious in interpreting the reading time in this region because there are non-trivial differences across conditions. Specifically, the subject differs across the gender-match and gender-mismatch conditions, and the wh-phrase region differs across the wh-argument and wh-predicate conditions.

Region 2 corresponds to the critical region, which consisted of a preposition and a reflexive. The preposition was included as part of the critical region, because when readers fixate on short words like prepositions, they are likely to parafoveally perceive the reflexive. (For a discussion on the effect of parafoveal previews in eye-tracking experiments, see Clifton, Traxler, Mohamed, Williams, Morris, & Rayner, 2003). Also, our inclusion of the preposition as part of the critical region is similar to the leftward-shifting procedure used in Sturt (2003), whereby fixations within four characters before the reflexive were counted as fixations on the reflexive. However, this treatment caused the critical region to differ in character length between wh-argument and wh-predicate conditions, because the preposition type was not always matched across the two wh-types. The mean length of prepositions was 2.9 characters for the wh-argument conditions and 3.4 characters for the wh-predicate conditions, and there was a significant difference between the two conditions ($p < .05$). For this reason, the critical observation that is relevant for testing the current hypotheses comes from the presence of gender match manipulation within each wh-type as well as a significant interaction of wh-type and gender match, rather than a wh-type effect which may only reflect differences in properties of the preposition.

Region 3 is the post-reflexive region, which corresponds to the determiner and the first adjective of the embedded clause subject. This region was used to probe for potential spill-over effects. Region 4 is the embedded clause subject region, which corresponds to the second adjective and the head noun of the embedded clause subject NP. The second adjective was included as the subject region because of potential parafoveal preview of the immediately following noun. These two regions were lexically matched across conditions. The data from regions after Region 4 are not reported because the number of words and the content of lexical items in these regions are not held constant across conditions.

Five reading time measures were computed for the three regions of interests: *first fixation duration, first pass time, regression path time, percent regressions, and second pass reading time* (Rayner, 1998; Rayner & Pollatsek, 2006; Staub & Rayner, 2007). First fixation duration is the duration of the very first fixation in a region, regardless of whether there is a single word or multiple words in that region. This measure is often used as an index of lexical difficulty (e.g., Reichle, Rayner, & Pollatsek, 2003) but is also informative about the earliest syntactic processes that immediately follow lexical access (e.g., Frazier & Rayner, 1982; Sturt, 2003).

The *first-pass reading time* is calculated by summing the fixations in a region between the time the eye-gaze first enters the region and the time when the eye-gaze exits the region either to the left or the right. First-pass reading times also index early lexical and syntactic processes associated with a region, but given that they consist of multiple fixations on the same region, they may also reflect slightly later processes than the first fixation duration. Following the tradition in the eye-tracking literature, the term *gaze*

duration will be used interchangeably when the region consists of a single word.

Regression path times are the sum of fixations from the time when the eye-gaze enters a region to the time when the eye-gaze exits the region to the right. Regression path time is identical to first-pass reading time if the eye-gaze exits the region first to the right, but if the eye-gaze exits the region to the left, then regression path times are longer than the first-pass time as they include all fixations in the previous regions as well as re-fixations on the region before exiting the region to the right. Thus, regression path times are seen to reflect slightly later processes such as integration of the critical region with the preceding context. The *percent regressions* indicate the probability that a reader makes a regressive eye movement after fixating the region. This measure includes regressions made only during the reader's first pass through the region, and does not include regression made after re-fixating the region. Note that for the pre-reflexive region, regression path time will be identical to the first-pass reading time and percent regressions will be zero, because there is no region to the left that eye-gaze can regress back to.

Following Sturt (2003), we also calculated the *second pass reading time,* which is the sum of fixations made on a region after that region has been exited either to the left or ot the right for the first time. This measure is informative about processes that occur later than processes that occur in the first encounter of the region. For this measure, trials in which a region was not re-fixated contributed a value of 0ms to the cell mean.

These five eye movement measures from the pre-reflexive, reflexive, post-reflexive and embedded clause subject regions were submitted to 2 × 2 repeated measures ANOVA with wh-type and gender agreement as within-participants factors. When there

was a main effect of gender agreement or significant interaction of wh-type and gender agreement in both by-participants and bi-items analyses, a planned comparison was conducted to test for systematic differences between gender match and mismatch conditions within each wh-type.

5.3.2 Results

Comprehension accuracy for the experimental items was 91.9% across four conditions, and did not differ across the four conditions (Fs < 1). Table 4 presents the participant means on each measure for each region as well as the standard errors of the participant means.

Table 4. Experiment 10: Participant Mean Reading Times in Milliseconds and Percent Regressions (Standard Error)

Measure	Pre-reflexive region	Reflexive region	Post-reflexive region	Embedded subject region
First fixation duration				
wh-argument, match	204 (7)	249 (9)	248 (6)	246 (8)
wh-argument, mismatch	207 (6)	261 (9)	261 (10)	237 (8)
wh-predicate, match	202 (6)	250 (9)	243 (8)	248 (6)
wh-predicate, mismatch	212 (6)	273 (10)	250 (9)	247 (8)
First-pass time				
wh-argument, match	967 (46)	349 (14)	352 (13)	442 (19)
wh-argument, mismatch	993 (46)	398 (20)	408 (23)	452 (23)
wh-predicate, match	867 (45)	376 (19)	376 (19)	468 (20)
wh-predicate, mismatch	913 (51)	421 (16)	376 (18)	439 (21)
Regression path time				
wh-argument, match	967 (46)	424 (33)	435 (25)	559 (44)
wh-argument, mismatch	993 (46)	510 (38)	645 (54)	540 (27)
wh-predicate, match	867 (45)	435 (24)	491 (34)	632 (47)
wh-predicate, mismatch	913 (51)	565 (56)	540 (34)	574 (61)
Percent regressions				
wh-argument, match	0	9.0 (2.0)	12.8 (2.6)	10.4 (2.5)
wh-argument, mismatch	0	17.1 (2.8)	22.5 (3.3)	12.5 (2.8)
wh-predicate, match	0	11.1 (2.6)	15.4 (2.5)	15.7 (2.9)
wh-predicate, mismatch	0	14.7 (3.3)	20.0 (2.7)	11.4 (2.5)
Second pass reading time				
wh-argument, match	1367 (121)	378 (74)	380 (76)	352 (58)
wh-argument, mismatch	1477 (129)	365 (40)	392 (59)	309 (53)
wh-predicate, match	1433 (120)	389 (43)	348 (40)	337 (36)
wh-predicate, mismatch	1406 (116)	375 (48)	300 (40)	250 (39)

On the first fixation duration on the pre-reflexive region, there was no significant effect (Fs < 2.1, p > .1). On the first-pass measure on the pre-verbal region, there was a main effect of wh-type due to slower reading times in wh-argument conditions than in wh-predicate conditions (980ms vs. 890ms), $F1(1, 37) = 12.59$, $p < .005$, $F2(1, 23) = 7.99$, $p < .05$, and there was a marginal effect of gender type in by-participants analysis,

F1(1, 37) = 3.30, $p = .078$, *F2*(1, 23) < 1, but there was no significant interaction (*F*s < 1). The slower reading time for wh-argument conditions was observed in the wh-phrase region in the self-paced reading experiment (Experiment 9), and this is likely to be the reason why we observed a contrast in the first-pass reading times in this region. On the second pass reading time on the pre-reflexive region, there was no significant effect (*F*s < 1).

In the critical reflexive region, on the first fixation duration there was a main effect of gender in by-participants analysis due to slower reading times in gender mismatch conditions (250ms vs. 267ms), although this effect did not persist in by-items analysis, *F1*(1, 37) = 8.08, , $p < .01$, *F2*(1, 23) = 2.12, $p > .1$. There was no main effect of wh-type nor a significant interaction (*F*s < 1). On the first pass reading time in the reflexive region, there was a marginal effect of wh-type only in by-participants analysis, *F1*(1, 37) = 3.46, , $p = .071$, *F2*(1, 23) < 1. Critically, we found a main effect of gender in both by-participants and by-items analyses, *F1*(1, 37) = 12.30, $p < .005$, *F2*(1, 23) = 7.40, $p < .05$, and there was no significant interaction (*F*s < 1). A planned comparison revealed that the gender agreement effect was present in wh-argument conditions due to slower reading times in the gender mismatch condition (349ms vs. 398ms), *t*1(37) = 2.65, $p < .05$, *t*2(23) = 2.26, $p < .05$, and there was a significant difference in participants analysis and a marginally significant difference in items analysis between the two wh-predicate conditions, also due to slower reading times in the gender mismatch condition (376ms vs. 421ms), *t*1(37) = 2.49, $p < .05$, *t*2(23) = 1.90, $p = .07$.

The same pattern of reading time data was found in the regression path time on the critical reflexive region. There was a main effect of gender in both by-participants

and by-items analyses, $F1(1, 37) = 12.43$, $p < .005$, $F2(1, 23) = 9.28$, $p < .01$, and there was no main effect of wh-type nor a significant interaction (Fs < 1). A planned comparison revealed a marginally significant effect of gender agreement effect in wh-argument conditions due to the slower reading time in gender mismatch condition (424ms vs. 510ms), $t1(37) = 1.93$, $p = .06$, $t2(23) = 2.01$, $p = .056$, and we found a significant difference between the two wh-predicate conditions, also due to slower reading times in the gender mismatch condition (435ms vs. 565ms), $t1(37) = 2.42$, $p < .05$, $t2(23) = 2.45$, $p < .05$. The effects in percent regressions showed the same pattern as first pass time and regression path time. There was a main effect of gender in both by-participants and by-items analyses, $F1(1, 37) = 8.85$, $p < .01$, $F2(1, 23) = 9.25$, $p < .01$, and there was no main effect of wh-type nor a significant interaction (Fs < 1). In a planned comparison, we found a significant difference between the two wh-argument conditions due to larger percent regressions in the gender mismatch condition (9.0% vs. 17.1%), $t1(37) = 2.17$, $p < .05$, $t2(23) = 2.62$, $p < .05$, but there was no significant difference in percent regressions between the two wh-predicate conditions ($p > .2$). The second pass time on this region revealed no significant effects (Fs < 1).

In the post-reflexive region, there was no significant effect in the first fixation duration measure ($p > .1$). The first pass time on the post-reflexive region revealed a main effect of gender in by-participants analysis and a marginally significant effect of gender in by-items analysis, $F1(1, 37) = 4.81$, $p < .05$, $F2(1, 23) = 2.92$, $p = .10$, and we also observed a significant interaction of wh-type and gender in by-items analysis, though the effect was marginal in by-participants analysis, $F1(1, 37) = 3.60$, $p = .065$, $F2(1, 23) = 6.55$, $p < .05$. A planned comparison revealed that there was a significant difference

between wh-argument conditions due to slower reading times in the gender mismatch condition (352ms vs. 403ms), $t1(37) = 2.76$, $p < .01$, $t2(23) = 2.69$, $p < .05$, but there was no significant difference between the two wh-predicate conditions (ts < 1). The same pattern of reading time was found in regression path time on the post-reflexive region, though the effects were generally stronger on this measure than in the first pass reading time. We found a main effect of gender in both by-participants and by-items analysis, $F1(1, 37) = 12.22$, $p < .005$, $F2(1, 23) = 16.81$, $p < .001$, and we also observed a significant interaction of wh-type and gender in both by-participants and by-items analysis, $F1(1, 37) = 5.31$, $p < .05$, $F2(1, 23) = 9.43$, $p < .01$. A planned comparison revealed that there was a significant difference between wh-argument conditions due to slower reading times in the gender mismatch condition (435ms vs. 645ms), $t1(37) = 3.87$, $p < .001$, $t2(23) = 6.84$, $p < .001$, but there was no significant difference between the two wh-predicate conditions (ts < 1.3, $p > .1$). In percent regressions, there was no longer a significant interaction (Fs < 1), but there was a main effect of gender in both by-participants and by-items analysis, $F1(1, 37) = 7.99$, $p < .01$, $F2(1, 23) = 5.72$, $p < .05$. In a planned comparison, we found a significant difference between the two wh-argument conditions due to larger percent regressions in the gender mismatch condition (12.8% vs. 22.5%), $t1(37) = 2.35$, $p < .05$, $t2(23) = 2.97$, $p < .01$, but there was no significant difference in percent regressions between the two wh-predicate conditions ($p > .1$). The second pass time on this region revealed no significant effects ($p > .1$).

On the embedded clause subject region, there was no significant effect on the first fixation duration, first pass time, regression path time or percent regressions ($p > .1$). In the second pass time, there was a marginal effect of gender in by-participants analysis

due to slower reading times in gender match conditions (345ms vs. 280ms), although it did not reach significance in by-items analysis, $F1(1, 37) = 3.51$, $p = .069$, $F2(1, 23) = 1.24$, $p > .1$.

5.3.3 Discussion

Summary and implications for adults' processing of reflexives and null subjects

The main results of this experiment can be summarized as follows. First, the wh-argument and wh-predicate conditions showed very similar reading time data in the critical reflexive region, in that the first pass reading time and regression path time in the reflexive region was longer, and the percent regressions larger, in the gender mismatch conditions. In the post-reflexive region, on the other hand, the first pass reading time and regression path time measures revealed a gender mismatch effect when the fronted wh-phrase was a wh-argument, but there was no such difference when the fronted wh-phrase was a wh-predicate. Finally, the second pass reading time in the embedded clause subject region revealed a marginal slow down in the gender match conditions.

The fact that early measures revealed a gender mismatch effect in both wh-argument and wh-predicate conditions on the reflexive region suggests that the parser immediately attempted to establish a binding relation regardless of the type of structure in which the reflexive was embedded. This is an unexpected result for the domain extension hypothesis, because this hypothesis predicts that the gender mismatch effect should occur later when the reflexive is inside a wh-predicate. The domain extension hypothesis further predicts that the second pass reading time for the wh-predicate gender mismatch condition should be longer, but there was no evidence that readers spent more time re-reading the gender mismatching reflexive in the wh-predicate condition.

The present results lend support instead to the active co-indexation hypothesis, which predicts that a gender mismatch effect should manifest in an equally early measure in both wh-predicate and wh-argument environments. Furthermore, the fact that the gender mismatch effect persisted into the following spill-over region in the wh-argument condition but not in the wh-predicate condition is consistent with the active co-indexation hypothesis, which predicts that the hypothesized co-indexation of the main clause subject and the trace can be retracted as soon as the parser encounters the embedded clause subject NP, which must serve as the actual antecedent of the trace. Note also that the marginal slow down in the embedded subject region of gender match conditions could be seen as reflecting the cost of retracting the index that was actively assigned to the null subject of the fronted predicate, though we leave open the question of why this effect did not reach significance or why this cost only emerged in the relatively late, second pass reading time.

Evidence for active co-indexation of the null subject of a fronted predicate highlights that the parser incrementally attempts to construct grammatical and interpretive commitments despite the fact that there is no bottom-up evidence for such active index assignment. This active co-indexation mechanism resembles active gap creation in filler-gap dependency processing, where the parser postulates a gap in the absence of bottom-up evidence for the structural position of a gap. In fact, a similar mechanism may be at play in the processing of backward anaphora in sentences like (7). It has been observed that a gender mismatch effect arises at the main clause subject region *John/Mary*, which suggests that the parser attempts to establish a co-reference

relation between the pronoun *he* and the main clause subject (e.g., Kazanina et al., 2007; van Gompel & Liversedge, 2003).

(7) While he was watching the game, {John | Mary} was drinking beer.

Here, the first position at which a potential antecedent can be presented is the main clause subject region, and this is where the parser attempts to locate an antecedent for the pronoun. In our target sentences, there is uncertainty as to the antecedent of the null subject inside the fronted predicate. It is possible that this null subject could be co-referential with the main clause subject, if the embedded clause subject is a pronoun; but it is also possible that the null subject could be co-referential with an entirely different referent, if the embedded clause subject is not a pronoun. The parser apparently relies on the former possibility, incrementally assigning to the null subject co-reference with the main clause subject, rather than waiting for the embedded clause subject to disambiguate these two possible antecedents. This account also predicts that the parser should incrementally relate a potential antecedent to an unpronounced subject of a fronted predicate in general. For example, in a sentence like (8a) where the fronted adjunct contains a *pro*, the same backward anaphora processing mechanism as (7) could relate the null pronoun to the main clause subject. In (8b), however, the parser could wait until the later-arriving subject to assign an antecedent for *pro*, or it could actively analyze the *pro* as co-referential with the main clause subject *John*, as the active co-indexation hypothesis predicts.

(8) a. While *pro* watching the game, {John | Mary} was drinking beer.

b. John said that while *pro* watching the game, {he | she} was drinking beer.

Thus, by examining the presence of a gender mismatch effect in (8b), we could further test the active co-indexation hypothesis.

Revising Leddon and Lidz (2006)

Given our argument that the active co-indexation mechanism plays a role in the processing of reflexives inside fronted wh-phrases, let us revisit the child behavior reported in Leddon and Lidz (2006), who presented ambiguous sentences like (1).

(1) a. Wh-argument fronting condition

Miss Cruella$_1$ knew which painting of herself$_{1/2}$ Janie$_2$ put up ____ .

b. Wh-predicate fronting condition

Mr. Monkey$_1$ figured out how proud of himself$_{*1/2}$ Andy$_2$ was ____ .

The observation was that in a Truth Value Judgment Task, adults accepted both the surface interpretation (Miss Cruella=herself) and the reconstruction interpretation (Janie=herself) in (1a) (although with a slightly higher acceptance rate for the surface interpretation), and only accepted the reconstruction interpretation in (1b). Children were adult-like in one respect, in that they only accepted the reconstruction interpretation in (1b); but they differed from adults in that they only accepted the surface interpretation in (1a). Leddon and Lidz argued that this reflects the timing at which the two interpretations become available: In (1a), the surface interpretation becomes available as soon as the

reflexive is processed; children adopt this interpretation, and this commitment persists in their ultimate interpretation. In (2b), the reflexive cannot be grammatically bound by the main clause subject, and therefore children do not entertain this interpretive possibility, successfully adopting the reconstruction interpretation.

One the one hand, the time course data presented in the two reading experiments with adults suggested that the parser immediately attempts to establish a binding relation between the main clause subject and reflexive in sentences like (1a), lending plausibility to Leddon and Lidz's account for children's comprehension of (1a). On the other hand, our time course data indicated that the null subject inside the fronted predicate is immediately analyzed as being co-referential with the main clause subject. This, however, should allow surface interpretation to be available in sentences like (1b), which seems to cast doubt on Leddon and Lidz's account.

Let us now consider two possible explanations for why children only entertained the reconstruction interpretation in (1b). One is that children are much more conservative in assigning an antecedent to the null subject of a fronted predicate. If we further assume that children's antecedent search processes are strictly constrained by the locality domain for the reflexive, then it is expected that the surface interpretation is never made available for children. Alternatively, it may be the case that children actively analyze the null subject of the fronted predicate as co-referential as the main clause subject, but they are able to later retract this incremental commitment. In this case, the error signal that could plausibly lead children to succeed in reanalysis is the overt embedded clause subject, which must be the actual subject of the fronted predicate. This is a clear syntactic and semantic error signal that is based on the argument-predicate relation in the embedded

clause: when the embedded clause subject is encountered and the wh-phrase is reconstructed to the base position, this causes the external thematic role to be assigned to the actual subject (Andy in (1b)). However, grammatically it is impossible for the same thematic role to be assigned to two different NPs, and this verb-based error signal may serve as an effective interpretation revision cue.

Although the current evidence does not distinguish these two accounts, the second account that relates the finding to children's reanalysis mechanism seems more plausible. The first explanation that children are conservative in establishing binding relations seems incompatible with observations in Chapter 4 that children actively hypothesize syntactic structures without relying on critical bottom-up evidence. On the other hand, the second explanation that children can use thematic information based on the argument-predicate relation in the embedded clause is consistent with the conclusion in Chapter 4 that children are able to use verb-based information to retract their active commitments. Future work with children is obviously needed to further investigate this possibility.

5.4 Conclusion

This chapter examined the processing of anaphor reconstruction, and demonstrated that the parser not only incrementally establishes a binding relation between the reflexive and the grammatically accessible subject in the local domain, but also actively assigns an antecedent to a null subject inside a fronted predicate. Considering this evidence together with the child comprehension behaviors reported in Leddon and Lidz (2006), I argued that children are able to recover from active syntactic commitments that are disconfirmed by later arriving verb information.

Chapter 6: Where did parents tell their children that wh-phrases should be attached?

This chapter examines potential consequences of the child parser for language acquisition.[15] The studies reported in Chapters 4 and 5 suggested that children actively construct syntactic representations without relying on critical bottom-up evidence, but that they can in principle retract such active syntactic commitments in cases where the actively constructed syntactic representations are disconfirmed by later arriving verb-based error signals. On the other hand, it was also observed that in the absence of verb-based error signals, actively constructed syntactic representations persist in the syntactic analysis and interpretation of the sentence. For example, Chapter 5 discussed the observation by Leddon and Lidz (2006) that in processing a sentence like *Miss Cruella wondered which picture of herself Janie put up,* children only accept the surface interpretation of the reflexive (i.e., *Miss Cruella = herself*) and fail to accept the grammatically permissible reconstruction interpretation (i.e., *Janie = herself*). Moreover, the Truth Value Judgment Study in Chapter 4 revealed that even adults fail to accept the association of *where* with the second VP *hurt herself* in a sentence like (1); in this environment, adults actively associate *where* with the first VP *said to someone,* and even when they encounter an NP that would make the embedded clause interpretation true (e.g., tree), this interpretation remains unavailable to them.

[15] Imogen Davidson White and Angela Stanley made a large contribution to the work presented in this chapter.

(1) The place where Emily said to someone that she hurt herself was the {pool | tree}.

These observations raise the following question: Given that children actively process long distance dependencies like adults do, if children fail to represent a certain analysis due to their pervasive parsing biases and immature revision capacities, how might it affect their language acquisition processes? For example, the studies reported in Chapter 4 show that English-speaking children are generally biased towards the main clause interpretation in processing a globally ambiguous wh-question like (2), which essentially contains the same wh-dependency as (1).

(2) Where did Emily tell someone that she hurt herself?

However, the Russian counterpart of (2) is unambiguous (3), in that only the main clause interpretation is grammatically permissible (Stepanov, 2000; Stepanov & Stateva, 2006).

(3) Gdje Emili skazala komu-to chto ona ushiblas'?
Where Emily said some-person that she hurt+REFL
"Where did Emily tell someone that she hurt herself?"

This is due to the fact that Russian is a strict scope marking language, in which overt wh-dependencies never cross a clausal boundary, unlike in such languages as English where cross-clausal wh-dependencies can be overtly formed (McDaniel, 1989; Riemsdijk, 1982; for a review of the grammar of wh-scope marking across languages, see Lutz, Müller, &

von Stechow, 2000). Instead, Russian inserts a wh-scope marker in the clause where the actual wh-phrase should take scope, as shown in (4). This sentence is unambiguous in the opposite way from (3) in that the question only allows the embedded clause interpretation.[16]

(4) Kak Emili dumaete, Gdje ona ushiblas' ?
how Emily think where she hurt-self
"(Lit.) What did she think where she hurt herself?"

Here, English-learning children must detect the availability of the embedded clause interpretation in (2) in order to distinguish their target grammar from that of Russian,[17] and it is reasonable to expect that the actual English input should contain sufficient

[16] The example in (4) has *dumaete* 'think' as the main clause predicate, because the Russian wh-scope marking construction does not allow verbs like *skazala* 'say' (Stepanov, 2000). This is a language-specific constraint in Russian and not a general property of wh-scope marking, because in other languages with wh-scope marking such as German or Hindi, any bridge verbs including 'say' or 'think' can be used in wh-scope marking constructions. The verb restriction to 'think' may lead one to think that the Russian wh-scope marking constructions should be analyzed as involving a special parenthetical clause, but this analysis is untenable because in a sentence like (i), a bound variable reading of the pronoun in the embedded clause is possible, suggesting that the clause that contains the verb 'think' c-commands the embedded clause.

(i) Kak sčitaet [každyj iz studentov]$_1$ kuda ego$_1$ mogut otpravit' ?
how thinks every from students where him can send-they
"Where does every student think that they can send him"

Note also that the embedded clause interpretation in (4) was constructed to be a close match with the sentence in (3), but it may sound somewhat infelicitous because usually a person who has been hurt should remember where it happened and thus should not have to think about it. The embedded clause interpretation in this sentence becomes perfectly natural however if, for example, Emily lost her memory in an accident and is trying to remember where she got hurt.

[17] Some wh-scope marking languages (e.g., certain dialects of German; see Höhle, 2000) allow English-like overt long distance wh-dependencies in addition to the wh-scope marking construction. Thus, there are actually at least three possible grammars with respect to long distance wh-dependency formation: a) only overt long distance dependency formation is allowed (e.g., English), b) only wh-scope marking is used (e.g., Russian), and c) a mixed language in which both strategies are available (e.g., dialects of German). Examining the acquisition process for all three types of grammars goes beyond the scope of this dissertation, and therefore we will focus on the clear-cut cases like English and Russian and leave open the question about the input distribution and acquisition of wh-dependency formation rules in mixed languages.

evidence for the availability of the embedded clause interpretations. However, given the findings in Chapter 4 that children actively create gaps and are thus generally biased towards the main clause analysis for wh-dependencies exemplified in (2), the distribution of the actual input to children may be skewed by the child's perceptual mechanism in such a way that the embedded clause interpretation in the *effective* input could seem rarer than it actually is, and this may affect the rate of learning the English grammar for long-distance wh-dependencies.

In fact, it has been observed that English-speaking children behave as if Russian-like wh-scope marking sentences were possible in their grammars. For example, Thornton (1990) observed that English-speaking 3-year-old and 4-year-old children produced sentences like (5a), and the acceptability judgment data reported in McDaniel, Chiu, and Maxfield (1995) also suggested that 3-year-old and 4-year-old children judged sentences like (5a) as acceptable. Moreover, de Villiers and colleagues (de Villiers et al., 1990, 2008) as well as de Villiers and Roeper (1995) conducted Question-after-Story experiments with sentences like (5b), and reported that 4-year old children answer the question as if the lower wh-phrase had matrix scope, answering what the boy caught as a response to (5b).

(5) a. What do you think who ate this?

b. How did the boy say what he caught?

It is still questionable whether these production and comprehension behaviors truly demonstrate that English-speaking children have adopted wh-scope marking strategies in

their developing English or whether these data reflect children's performance constraints (for discussions, see Jakubowicz & Strik, 2008; Schulz, 2006). However, this type of behavior could plausibly reflect that children are at least temporarily considering the wh-scope marking construction as a viable grammatical option, which is a reasonable consequence if children analyzed the preponderance of main clause interpretations in sentences like (2) as evidence that wh-dependencies in English do not cross a clausal boundary.[18]

Taking wh-dependency processing as a case study, this chapter discusses how the actual input distribution would be distorted in a child's mind if we take the child's parsing constraints seriously. Section 6.1 presents a brief review of recent work on computational modeling of language acquisition that assumes that parser successes and failures are integral components of the language acquisition mechanisms. We illustrate that these studies have generally made the assumption that children's parsers are perfect, in that so long as there is a grammatically possible parse for a sentence, children will be able to assign that parse to the input. Section 6.2 presents a distributional analysis of long distance wh-dependencies in child-directed speech in CHILDES (MacWhinney, 2000), and demonstrates how the distributional characteristics of the input can vary when we take into consideration children's active gap creation processes and abilities to use verb-based error signals to retract their active commitments. Section 6.3 discusses the implications for models of language acquisition.

[18] Alternatively, the presence of wh-scope marking in child English could indicate that children are going through the phase of *grammar competition*, in which children initially consider all relevant grammars as potential candidates for the target grammar, and probabilistically select one of the competing grammars to test their parsing performance against the input (Yang, 2002, 2004). See section 6.1 for more discussions on Yang's model.

6.1 Parsing and distributional learning

The idea that the parser plays a critical role in language learning processes has been an active assumption since the time of early studies on formal language learnability (Berwick, 1985; Clark, 1992; Gold, 1967; Pinker, 1984; Wexler & Culicover, 1980). This is an attractive notion in that learners can internally create negative evidence for their hypothesis about the target grammar. Moreover, given that learners in reality must parse the input for comprehension purposes anyway, this approach does not need to assume any special learning device that is dedicated to language learning. This approach remains popular in more recent computational models of language acquisition, in which it is generally assumed that the parser's successes and failures are critical in determining the evidence for or against grammatical hypotheses. (Berwick & Niyogi, 1996; Gibson & Wexler, 1994; Fodor & Sakas, 2004; Yang, 2002, 2004).

For example, Yang (2002, 2004) proposed a variational model of language acquisition that essentially characterizes acquisition as a process of grammar competition. In this model, the learner is assumed to have access to all the possible sets of grammars within the limits of Universal Grammar, and for each input sentence, the learner probabilistically chooses one of the possible grammars to evaluate how successful that grammar is in parsing the target language. If the input can be assigned a parse, then the selected grammar can be rewarded by increasing its probability weight, but if the parse fails, then the selected grammar is punished by decreasing its probability weight. The Structural Triggers Learners model proposed by Fodor (1998) and Fodor and Sakas (2004) also assumes that learners have access to all possible grammars of human language (*supergrammar*), and in this model, the parser uses the supergrammar to assign

any of the possible syntactic analyses. This procedure is used as a grammar selection procedure: The model adopts a new grammatical setting when an input sentence could be parsed only by that grammar. Decisions based on this type of evidence are considered to be most accurate because the fact that parsing of the input requires that unique grammar presents unambiguous evidence that this grammatical rule is necessary for the target language.

The goal of this chapter, however, is not to test these computational models of language acquisition, but rather to examine the consequences of immature child parsing on language acquisition processes. In all the models mentioned above, the parsing success or failure is considered a critical component of language acquisition, but the parser is assumed to be *perfect* in that it can assign any structural analysis allowed by the current grammatical hypothesis (which is probabilistically selected in Yang's model, or is assumed to be the entire Universal Grammar by Fodor and Sakas); moreover, the grammatical hypothesis itself is also considered to be adult-like. In other words, in these models the success and failure of a parse is determined by whether the selected grammar can generate a structure that can accommodate the input string, and not by whether a real learner can assign structures and interpretations to the input. As discussed by Valian (1990) and as discussed in the previous chapters of this dissertation, however, this assumption ignores the fact that children have psychological constraints on parsing that often prevent them from assigning adult-like structures to the input even when they have the proper grammatical knowledge for that construction.

Let us hypothetically consider how the input distribution could be altered when we incorporate psychological constraints on the parser into the account of how long

distance wh-dependency formation is learned. The relevant input for learning long distance wh-dependencies is sentences like (2), which is repeated here for convenience.

(2) Where did Emily tell someone that she hurt herself?

First, suppose that a learner's parser is 'perfect' as assumed in the computational modeling and formal learnability research discussed above, such that a learner can assign any structural analysis that can be generated by the hypothesized grammar. With this assumption, the sentence in (2) is compatible with either an English-type grammar or a Russian-type grammar, because both grammars can assign an analysis, even though the Russian grammar is limited to the analysis in which the wh-phrase is attached to the main clause VP.

On the other hand, incorporating properties of the actual child parser could turn ambiguous sentences like (2) into evidence in favor of one of the possible grammars. For example, one relevant factor is children's active dependency completion bias that we observed in Chapter 4. If we assume that this processing bias plays a role in naturalistic language learning, then ambiguous sentences like (2) that could allow either a main clause interpretation or an embedded clause interpretation are no longer perceived to be ambiguous, because the children's parsing biases should generally favor the main clause interpretation in the absence of signals that would require them to retract such active commitments. This could have a significant consequence on English-speaking children's acquisition of long distance dependencies, when we take into account the role of indirect negative evidence (Braine, 1971; Chomsky, 1981; Pinker, 1989) or the size principle

(Tenenbaum & Griffiths, 2001; Xu & Tenenbaum, 2007) in language acquisition. In the English-Russian comparison, English is a superset of Russian in that English allows both the main clause interpretation and the embedded clause interpretation, whereas Russian only allows the main clause interpretation. If the active dependency completion bias would only yield the main clause interpretation for children, then this is compatible with both the subset grammar (Russian) and the superset grammar (English). However, if the superset language is the correct target grammar, it would seem reasonable to expect that there should be some evidence for the presence of the embedded clause interpretation. If the data that would uniquely favor the superset grammar is absent, then this could serve as potential indirect negative evidence, and lead children to conclude that the Russian grammar is more likely to be the target grammar. The same problem arises even if a learner operates under a subset principle (Berwick, 1985; Fodor & Sakas, 2005; Manzini & Wexler, 1987) that forces them to start out with the subset grammar as the default grammar, which in this case would be Russian, since that only allows the main clause interpretation. Here again, the active dependency completion bias would essentially eliminate the embedded clause interpretation, and consequently the unique evidence for the superset grammar would not be available in the input. In these ways, the parsing bias that skews the input distribution could potentially lead children to adopt a Russian-type wh-scope marking grammar.

Note, however, that once we consider the second property of the child parser established in this dissertation, namely that children can retract their active dependency completion when the first VP (i.e., the main clause VP in English) is semantically or pragmatically incompatible with the fronted wh-phrase, we see that children have a

mechanism that allows them to (re)associate the wh-phrase with the second VP (i.e., the embedded clause VP in English). This would not only alter the distribution and decrease the frequency of the main clause interpretation, but also increase the proportion of unambiguous evidence for the English grammar (Fodor, 1998; Fodor & Sakas, 2004), because the embedded clause interpretation is unavailable in the Russian-type wh-scope marking grammar.

Moreover, even if a learner's parser was perfect as assumed in the computational modeling research, it is possible that children's immature grammatical knowledge could skew the input distribution. For example, it has been argued that 4-year-old children do not respect factive island constraints in sentences like (6). For adults, (6) only allows the main clause interpretation, suggesting that extraction out of the clausal complement of a factive verb like *know* is ungrammatical (Cattell, 1978). However, Roeper and de Villiers (1992) demonstrated in a Question-after-Story experiment that children allowed either the main clause interpretation or the embedded clause interpretation for (6).

(6) When did the boy know that he hurt himself?

Thus, even though sentences like (6) are supposed to be grammatically unambiguous and hence irrelevant to deciding whether the target grammar in principle allows a long distance wh-dependency, children's developing grammars could turn such sentences into evidence for the presence of long distance wh-dependencies.

Now, let us summarize in what ways children's developing grammars and parsers could contribute to skewing the input distribution. There are four possible ways in which

the child grammar and parser can interact with the input distribution. First, as assumed in the formal learnability research and computational modeling studies, the child parser could be perfect, such that the child is always able to find the appropriate parse of the input based on the current grammar. Moreover, when children focus on a certain language as a possible target grammar, then it is assumed in the computational modeling research that the grammar itself is supposed to be adult-like. On these assumptions, the input distribution from language learners' perspectives would be identical to the input distribution from the adults' perspectives. This position is summarized in (7a). On the other hand, even if the child parser is indeed perfect, children may not have access to the same set of grammatical constraints as adults do. For example, children may not always respect factive islands (Roeper & de Villiers, 1992), such that the sentences that are grammatically unambiguous for adults might nevertheless be considered ambiguous for children. This scenario is summarized in (7b).

(7) a. The child parser and grammar are perfect, and parsing biases and limitations do not skew the input distribution. The input distribution for children will be the same as the one from the adult's perspective.

b. Children's parsers are perfect, but children do not respect factive islands (Roeper & de Villiers, 1992). This increases the number of ambiguous sentences.

Under (7a) or (7b), ambiguous sentences as well as unambiguous sentences that only allow an embedded clause interpretation would be considered as evidence for the English-type grammar that allows overt long distance wh-dependencies.

Alternatively, biases and limitations in children's parsers could actually skew the input distribution, such that the effective distributional information that would be used for learning purposes may be different from the actual input distribution in the environment. One relevant factor is children's active dependency completion bias, which causes the parser to initially attach the wh-phrase to the main clause VP. Given that children are generally not as competent as adults in retracting their syntactic commitments (Trueswell et al., 1999), this active wh-processing bias could lead to an increase of main clause interpretation data in the input. This position is summarized in (8a). Another relevant factor is the observation in Chapter 4 that children are able to retract their active syntactic commitments in wh-dependency processing when the wh-phrase and the first verb are not compatible with each other for semantic and pragmatic reasons. This could allow children to recognize evidence for the presence of embedded clause interpretation in the input. This scenario is summarized in (8b).

(8) a. Children's parsers actively complete wh-dependencies, and generally fail to revise their initial syntactic commitments.

b. Children's parsers actively complete wh-dependencies and the revision capacities are limited, but children are able to use verb-based, semantic and pragmatic incompatibility information to retract their active commitments.

In these ways, it seems feasible that properties of the child parser and grammar could skew the input distribution, but it is unknown whether long distance wh-dependencies with adjuncts in child-directed speech actually contain features that would

render the sentences amenable to children's parsing biases and limited reanalysis capacities. In order to investigate this question, the next section examines the distribution of English wh-dependencies (2) in child-directed speech in CHILDES (MacWhinney, 2000).

6.2 Analysis of wh-dependency distribution in child-directed speech

6.2.1 Corpus information and analysis procedure

We examined child-directed speech in transcripts from Brown (child names: Adam, Eve, Sarah), Sachs (Naomi), Snow (Nathaniel), Suppes (Nina), and Warren-Leubecker (20 children who visited the lab, each of whom was recorded once; names not included). This produced a total of 146363 lines of child-directed speech. The age range across the recording sessions, the number of recording sessions, as well as the number of child-directed utterance lines in each corpus is summarized in Table 5.

Table 5. Details of corpora used in the current study

corpus	child	age range	number of sessions	number of lines with child-directed speech
Brown (1973)	Adam	2;3-4;10	55	26698
	Eve	1;6-2;3	20	11548
	Sarah	2;3-5;1	139	34780
Sachs (1983)	Naomi	1;1-5;1	93	12024
MacWhinney and Snow (1990)	Nathaniel	2;3-3;9	30	19841
Suppes (1974)	Nina	1;11-3;11	56	35616
Warren-Leubecker (1982)	20 children	1;6-3;1 4;6-6;2	20 (1 per child)	5856

In each corpus, we used a python script to extract all sentences with wh-dependencies that include a wh-phrase.[19] Note that this underestimates the actual occurrences of wh-dependencies, given that English wh-dependencies may not necessarily include a wh-phrase, as represented by *that* relative clauses, infinitival relatives, or cleft constructions. This decision was made for a practical reason, which is that including *that* as a search criterion extracts too many sentences without wh-dependencies, and it would thus be impractical to sort the output and extract *that* relatives and clefts. *That* relatives, infinitival relatives and clefts were still counted as including wh-dependencies when they were found in the sentences that also included a wh-phrase.

For each sentence with a wh-phrase, we evaluated whether it included a wh-dependency with a predicate or a preposition that hosted a gap position. Thus, we excluded the following types of sentences from further analyses: a) echo questions (e.g., *Mommy forgot what?*), which do not exhibit an overt wh-dependency; b) fragment questions (e.g., *Which friend?*); and questions with sluicing or VP ellipses (e.g., *I don't know how*, or *Why can't you?*) that make it unclear where the gap position is, c) expressions that include a wh-phrase but are pragmatically used as a suggestion rather than a question (e.g., *why don't you/we…*, *how/what about…*), d) formulaic greetings (e.g., *how are you*?), and e) questions with *how come*, with which long distance wh-dependencies cannot be formed (Collins, 1991). Next, the remaining wh-dependencies were categorized as involving wh-argument fronting or wh-adjunct fronting. Wh-arguments included wh-phrases like *what, what NP, who, which NP, whose NP, how many/much NP*, *how AdjP* as well as *where,* which serves as the complement of directional verbs (e.g., *go, arrive*) or prepositions. Subject wh-questions were also

[19] We thank Yakov Kronrod for writing this python script.

counted as involving wh-argument fronting, because the presence of dislocation is visible when the wh-subject is moved long distance (e.g., *Who did you think __ ate the cake?*). Complements of copulas (e.g., *Where is it ___?*) and complex predication constructions like resultatives (e.g., *how big did you blow it __ ?*) were also treated as involving wh-argument fronting. Wh-adjuncts included wh-phrases like *when, how, where, how AdvP, what/which NP* used as an adverbal (e.g., which way), and *why*. In cases where the argumenthood was not clear, a native speaker informant was consulted for constituency tests to examine whether the fronted wh-phrase should be analyzed as an argument or not. The temporal adjunct clause marker *when* was also included in the analyses, because it shows similar properties to regular wh-dependencies in wh-questions or relative clauses; e.g., interpretive ambiguities arise when the temporal adjunct marker selects a bi-clausal complement, and long distance dependency formation is blocked across island domains (e.g., Geis, 1970; Haegeman, 2009; Larson, 1987). For the same reason, free relatives were included in the analysis as they behave like regular wh-dependencies (Bresnan & Grimshaw, 1978; for a review, see van Riemsdijk, 2006). If a single utterance line contained more than one wh-dependency, then each dependency contributed to the total count of wh-dependencies.

Our primary concern is in how frequently wh-dependencies in sentences like (2) appear in the input, and whether the child parser could treat such sentences as an unambiguous sentence that only allows the main clause interpretation. For this reason, wh-dependencies in the child-directed utterances were further coded as (potentially) crossing one predicate or two predicates in order to identify the frequency of overt long distance wh-dependencies and short distance wh-dependencies. In coding the long

distance dependencies, we assumed that children would respect wh-island and relative clause island constraints based on previous findings (de Villiers & Roeper, 1990; de Villiers et al., 1990; de Villiers et al., 2008). Thus, sentences that could theoretically contain a long distance dependency if children did not respect island constraints (e.g., *how do you know what we find at the carnival?*) were coded as involving a clause-internal, short wh-dependency.

It is important to note that in languages that allow wh-scope marking, the wh-scope marking strategy is available only for wh-dependencies that originate in finite sentential complements and cross a finite complement clause boundary (McDaniel, 1989). For this reason, we separated counts of wh-dependencies that potentially cross finite clausal complements (e.g., *Why do you think he has hands on his tummy?*) from counts of wh-dependencies that potentially cross infinitival clausal complements such as raising and control constructions (e.g., *Why did you ask me to do that?*). Moreover, predicates like *going to, have to, be supposed to, used to, get to, ought to* appear to involve infinitival clause complementation, but these expressions were treated as auxiliaries without infinitival complementation, because these expressions differ from raising or control predicates in that the predicate itself does not retain the original meaning of the verbs and is therefore likely to be lexicalized as an auxiliary. Therefore, wh-dependencies crossing these expressions were only categorized as a short dependency within a clause.

Furthermore, for sentences with wh-adjunct fronting that (potentially) crosses a finite clause boundary, properties of the main clause verb type and its relation to interpretive possibilities were examined. This is relevant since these sentences could

potentially be treated by children as positive evidence for wh-scope marking. The experimental findings in Chapter 4 demonstrated that children are sensitive to the semantic relation between the wh-phrase and the verb (Experiments 5 and 8) in such a way that they are able to overcome their bias to associate the wh-phrase with the first verb in the sentence. Experiment 5 demonstrated that children generally adopt an embedded clause interpretation when the main clause verb is bare *say*, but not when the main clause VP is more complex, as in *tell someone* or *say to someone*, and that this type of verb bias effect can plausibly affect children's interpretations in naturalistic settings as well. There are other wh-verb combinations that are likely to affect children's interpretive possibilities. For example, a main clause interpretation is unavailable in a question like *How do you think you have to do it?*, because it is infelicitous to ask about the manner of thinking in English. Even though the experiments reported in Chapter 4 did not examine this type of combination, it is feasible that children might be sensitive to this wh-verb incompatibility and only accept the embedded clause interpretation.

Finally, children's answers to wh-questions could potentially be useful in understanding how they parsed the wh-dependencies, but we decided not to analyze these responses for two reasons. First, children do not reliably answer wh-questions in the corpora. Second, wh-dependencies often appear as indirect questions or relative clauses, such that they do not require answers from children.

6.2.2 Results

Of the 146363 lines of child-directed speech, there were 14427 overt wh-dependencies, which accounts for approximately 10% of the child-directed utterance line. Table 6 summarizes the distribution of short vs. long distance wh-dependencies and

finiteness of the complement clause in the long-distance wh-dependencies for wh-arguments, and Table 7 summarizes the same data for wh-adjuncts.

Table 6. Distribution of argument wh-dependencies

corpus	child	long distance		short distance	total
		finite	non-finite		
Brown	Adam	80	52	2064	2196
	Eve	2	17	535	554
	Sarah	16	44	1581	1641
Sachs	Naomi	12	35	747	794
MacWhinney and Snow	Nathaniel	20	57	1013	1090
Suppes	Nina	83	173	4401	4657
Warren-Leubecker	20 children	6	24	677	707
	Total	219	402	11018	11639

Table 7. Distribution of adjunct wh-dependencies

corpus	child	long distance		short distance	total
		finite	non-finite		
Brown	Adam	34	25	612	671
	Eve	0	1	103	104
	Sarah	2	10	561	573
Sachs	Naomi	9	3	204	216
MacWhinney and Snow	Nathaniel	11	26	237	274
Suppes	Nina	28	31	765	824
Warren-Leubecker	20 children	2	10	113	125
	Total	86	106	2595	2787

As Tables 6 and 7 show, overall there were more instances of wh-argument fronting than wh-adjunct fronting: Approximately 81% (11639/14427) of the wh-dependencies involved argument fronting and approximately 19% (2787/14427) involved adjunct fronting. Long distance wh-dependencies that crossed a finite or non-finite clause

boundary were generally very rare both for arguments or adjuncts: Long distance dependencies accounted for just 5.4% (621/11639) of the argument wh-dependencies and 6.9% (192/2787) of the adjunct wh-dependencies, and long distance dependencies that crossed a finite clause boundary accounted for only 1.9% (219/11639) of the argument wh-dependencies, and 3.1% (86/2787) of the adjunct wh-dependencies. This illustrates that the samples that are critical for learning the long distance wh-dependency formation rule are sparse (for related observations, see Dabrowska, Rowland, & Theakston, 2009; Yang, 2004).

Next, the critical 86 long distance wh-adjunct dependencies were examined further to investigate how the actual input distribution from an adult's perspective could be skewed when the input is represented in a child's mind. As summarized in (7) and (8), which are repeated below for convenience, there are four ways in which the child parser and grammar could interact with the input distribution.

(7) a. The child parser and grammar are perfect, and parsing biases and limitations do not skew the input distribution. The input distribution for children will be the same as the one from the adult's perspective.

b. Children's parsers are perfect, but children do not respect factive islands (Roeper & de Villiers, 1992). This increases the number of ambiguous sentences.

(8) a. Children's parsers actively complete wh-dependencies, and generally fail to revise their initial syntactic commitments.

b. Children's parsers actively complete wh-dependencies and the revision capacities are limited, but children are able to use verb-based, semantic and pragmatic incompatibility information to retract their active commitments.

Let us start with (7a), in which the child parser and grammar are assumed to be perfect, such that all the grammatical parses in the input would be recognized. In other words, this can be considered to reflect the input distribution from an adult's perspective. The target sentences with long distance wh-adjunct dependencies could be categorized into three types in terms of interpretive possibilities: a) ambiguous, allowing both main clause and embedded clause interpretations; b) main clause interpretation only, when the main clause verb is a factive verb, which blocks long distance dependencies; and c) embedded clause interpretation only, when the association of the wh-phrase with the main clause predicate is blocked as it would yield an infelicitous interpretation. In order to assess how the assumptions made in (7a) interact with the input distribution, it is critical to investigate the distribution of main clause verb and wh-phrase type, such that we can identify what proportion of the 86 sentences involve a factive verb, as well as a wh+verb combination that would be infelicitous.

Out of the 86 sentences with long distance dependencies, only four verbs were used as a main clause predicate, namely, *think, say, tell,* and *know*. The distribution of these verbs for each corpus is summarized in Table 8. Two example sentences for each verb type are shown in (9), and all 86 sentences are listed in Appendix I.

Table 8. Distribution of main clause verb type in adjunct wh-dependencies

corpus	child	think	say	tell	know	total
Brown	Adam	18	2	2	12	24
	Eve					0
	Sarah	2				2
Sachs	Naomi	6			3	9
MacWhinney and Snow	Nathaniel	8			3	11
Suppes	Nina	23		4	1	28
Warren-Leubecker	20 children	1		1		2
	Total	59	2	7	19	86

(9) a. Two example sentences with *think* as a main clause verb (child, speaker)

How do you think you have to do it? (Adam, Mother)

Why do you think he has an umbrella? (Nina, Mother)

b. Two example sentences with *say* as a main clause verb (child, speaker)

Why do you say he has wheels? (Adam, Mother)

Why do you say he's a bad boy? (Adam, Mother)

c. Two example sentences with *tell* as a main clause verb (child, speaker)

How can you tell that you're strong? (Adam, Ursula)

How can you tell that it's nighttime? (Nina, Mother)

d. Two example sentences with *know* as a main clause verb (child, speaker)

That's how you know it's winter. (Nathaniel, Mother)

How did you know that those are mother elephants? (Naomi, Mother)

As Table 8 shows, most of the sentences with potential long distance wh-dependencies with adjuncts use the main verb *think*, which accounts for 68.6% (59/86) of such wh-dependencies. The next most frequent verb is *know* (22.1%, 19/86), followed by *tell*

(8.1%, 7/86) and *say* (2.3%, 2/86).[20] It is important to note that *know* is a factive verb that generally blocks long-distance extraction for adults (Cattell, 1978), though, as noted above, children may potentially be insensitive to factive islands (Roeper & de Villiers, 1992). Also, the verb *tell* in the extracted utterances was not used as an utterance verb, but rather functions as a factive verb. The main clause of the examples in (9c) (*how can you tell…*) can roughly be translated as *How do you know…*, and for this reason, this use of *tell* serves to create a factive island according to our native speaker informants. Moreover, the two instances of *say* as shown in (9b) are not used in the usual sense of *say*, which generally marks the presence of an utterance event. In both of the two examples in (9b), taken together with the fact that the main clause is present tense, the verb *say* appears to be used to indicate an opinion of the speaker, as the main clause can be approximately considered as equivalent to asking *Why do you think*…. For these reasons, even though the experimental data reported in Chapter 4 indicated that *say* and *tell* affect children's interpretation biases, those findings are not applicable to these instances of *say* and *tell* because they are used with a different sense from the one that was used in the experiments.

Next, in order to evaluate the semantic and pragmatic compatibility between the wh-phrase and the verb, we examined the distribution of wh-verb combinations in the child directed speech. Table 9 combines data from different corpora and summarizes the wh × main verb distribution in child directed speech.

[20] Note that for sentences with long distance wh-argument fronting, the proportion of *think* was even more dominant: *think* 91.8% (201/219), *sa*y 4.6% (10/219), *tell* 2.3% (5/219), *know* 0.5% (1/219), *suppose* 0.5% (1/219), and *bet* 0.5% (1/219).

Table 9. Distribution of wh-phrase types and main verb types

Main verb type	Wh-phrase type				Total
	how	where	which way	why	
know	19				19
say				2	2
tell	7				7
think	8	11	1	38	58
Total	34	11	1	40	86

Table 9 shows that the verb *think* is distributed across various adjunct wh-phrases, but that other verbs are used selectively with certain wh-phrases. For example, factive verbs like *know* and *tell* only occur with *how*, and the verb *say* (which can be understood as *think* in the extracted sentences) only occurs with *why*, although the apparent distributional bias with *say* may be due to the small sample size (n=2).

Native speaker informants were consulted to assess the distribution of three interpretive possibilities (ambiguous, main clause only, embedded clause only). First, there were 40 sentences that can be truly ambiguous (46.5%, 40/86), and these sentences include sentences with *why* as a wh-phrase and *say* or *think* as a main clause verb. Next, as discussed above, 26 sentences (30.2%, 26/86) contained factive verbs like *know* and *tell* (given the way these verbs are used in the extracted child directed speech), and therefore adult informants only allowed the main clause interpretation.[21] The remaining 20 sentences (23.3%, 20/86) were judged to allow only the embedded clause interpretation. These sentences include *how* (n=8)*, where* (n=11)*,* or *which way* (n=1) as a

[21] It was also found that in all of these 26 sentences with factive verbs and the wh-phrase *how*, the embedded clause predicate was pragmatically incompatible with *how* because the embedded clause predicates are stative or individual-level predicates, which are unlikely to be modified by a manner wh-adverbial like *how*. Thus, even if children were insensitive to factive islands (Roeper & de Villiers, 1992), children may be able to use the incompatibility between the wh-phrase and the embedded clause predicate to only accept the main clause interpretation.

wh-phrase and *think* as the main clause verb in present tense (e.g., *how/where/which way do you think…*). In all of these sentences, the wh-association with the main clause verb was unacceptable because it is infelicitous to ask about the manner, location, or direction of thinking.[22]

Let us revisit the possible assumptions one could make about the child parser. In the first assumptions described in (7a), the learner has a perfect parser which allows the learner to recognize all the parses in the input that are acceptable under the current grammar. In other words, the input distribution for this learner should directly correspond to the input distribution that was determined based on our native speaker informants. The input distribution for (7a) is summarized in Table 10, which also presents the input distribution estimate for other hypothetical learners to be explained below.

[22] Note that it is not impossible to construct a scenario in which asking for a location for thinking can be made felicitous. For example, if a person habitually entertains a thought that he wants to visit the beach whenever he goes to a swimming pool, for a sentence like *Where do you think that you want to visit the beach?,* the main clause interpretation "at the swimming pool" is felicitous. However, these are highly constrained contexts and the context for the child-directed utterances was checked to ensure that this type of contextual information was not present.

Table 10. The input distribution and its interaction with the child grammar and parser

	Interpretation assigned by a child		
	ambiguous	main clause only	embedded clause only
The child parser is perfect and recognizes all grammatical parses			
(7a): Children respect factive islands (= distribution from an adult's perspective)	46.5% (40/86)	30.2% (26/86)	23.3% (20/86)
(7b): Children do not respect factive islands	76.7% (66/86)	0	23.3% (20/86)
Properties of the child parser skew the input distribution			
(8a): Active wh-attachment, no reanalysis	0	100% (86/86)	0
(8b): Active wh-attachment, revision based on verb information	0	76.7% (66/86)	23.3% (20/86)

Now let us consider the assumptions in (7b), namely that the learner has a perfect parser but does not respect factive islands. For this learner, all the 26 sentences that contained a factive verb as the main clause verb could be reclassified as belonging to the ambiguous category, because these sentences would now allow either the main clause interpretation or the embedded clause interpretation. Under this assumption, 76.7% (66/86) of the sentences with long distance wh-adjunct dependencies become ambiguous. For both (7a) and (7b), a large proportion of the input contains sentences that make the embedded clause interpretation available (69.8% (60/86) for (7a), and 100% for (7b)), while availability of the main clause interpretation is also recognized (76.7% (66/86) for both (7a) and (7b)), and in this sense, learners with assumptions in (7a) or (7b) should have no difficulty in arriving at the target English grammar.[23]

[23] Given that factive islands syntactically block long distance wh-dependencies, it is also reasonable to simply eliminate the 26 sentences from analysis for the learner operating under the

On the other hand, the input distribution can drastically change once we adopt the assumptions in (8a) and (8b) that properties of the child parser could skew the input distribution. If we assume as in (8a) that children actively associate the wh-phrase with the main clause verb and fail to retract this syntactic commitment, then for all the 86 sentences the main clause interpretation will be perceived as the only available interpretation. The fact that the embedded clause interpretation is completely absent could serve as indirect negative evidence and lead children to conclude that the target language is not English, because English should generate at least some evidence for the availability of the embedded clause interpretation.

However, the picture changes once we consider the fact that children are able to utilize information about semantic or pragmatic incompatibility between the wh-phrase and the first verb in the sentence. This error signal would allow children to retract their active wh-attachment to the main clause VP, and arrive at the embedded clause interpretation 23.3% (20/86) of the time. In this way, the child revision mechanism can create evidence for an English-type grammar that allows overt long distance wh-dependencies.

6.2.3 Discussion

The descriptive results of the distributional analysis can be summarized as follows. First, wh-dependencies that could potentially cross a finite clausal boundary were extremely rare in the child-directed speech (less than 3%), since most of the wh-dependencies found in the child-directed speech consisted of a clause-internal, short

---

assumptions in (7a). Note that changing the denominator from 86 to 66 would not affect the point that there is sufficient input for a learner under the assumptions in (7a), to whom both the main clause and embedded clause interpretations are available.

distance dislocation. Second, the main clause verb type in sentences with potential long distance wh-dependencies was restricted to four types (*know, say, tell, think*), and the combination of certain wh-phrases with these verbs eliminated ambiguities and forced the main clause interpretation (factive verbs like *know* or *tell*) or the embedded clause interpretation (the combination of *how, where, which way* with the verb *think*); but some sentences were ambiguous and allowed either main clause or embedded clause interpretation.

The distributional analysis of long distance wh-dependencies in child-directed speech has interesting implications for the role of the parser in language acquisition. If we assume that a learner's parser only considers whether the hypothesized grammar can assign a syntactic analysis to the input string, then the distributional pattern observed for wh-adjunct long distance dependencies would be compatible with either an overt long distance dependency formation grammar as in English, or a wh-scope marking grammar like Russian. All of the 86 sentences with (potentially) long distance wh-dependencies could syntactically accommodate a wh-association with the main clause predicate or an embedded clause predicate, and therefore these sentences are uninformative with respect to which grammar should be favored.

Once we examine the learning consequence of the actual psychological constraints on child parsing, the effective input distribution could be different. For example, if we assume that the active dependency completion bias plays a role in naturalistic language learning, sentences that could allow either the main clause interpretation or the embedded clause interpretation cease to be ambiguous, due to children's parsing biases that cause them to generally favor the main clause

interpretation. If this is the case, then all the 86 sentences with a potentially long distance wh-adjunct dependencies would be analyzed as unambiguous, allowing only a main clause interpretation.

On the other hand, if we assume that children are sensitive to the semantic and pragmatic compatibility between the wh-phrase and the verb in making wh-attachment decisions, then children can (re)associate the wh-phrase with the second VP (i.e., the embedded clause VP in English). This would not only alter the distribution and decrease the frequency of the main clause interpretation from 100% to 76.7%, but also increase the proportion of clear evidence for the English grammar, because the embedded clause interpretation is unavailable in the Russian-type wh-scope marking grammar. In other words, although the majority of the input is in favor of the main clause analysis, learners are led to perceive the input as containing unambiguous support for the English-type grammar. It is likely that children would be able to use this evidence for the presence of embedded clause interpretation to rule out the Russian-like strict wh-scope marking grammar. It is important to note, however, that the observation that children can use verb-based error signals to retract their active wh-attachment decisions was based on the Japanese data in Experiment 9, where the wh-phrase was semantically incompatible with the embedded clause VP rather than the main clause VP. Future work is needed to determine whether English-speaking children can also retract their active wh-attachment to the main clause VP when the wh-phrase is semantically incompatible with the main clause VP.

So far, we have concentrated on children's developing grammar, active dependency completion bias and their reanalysis capacity to investigate how these

properties of the developing child parser might skew the input distribution. However, there are other potential sources of information that may skew the input distribution and affect the learning process for long distance wh-dependency formation. For example, it is possible that the bi-clausal sentences with *think* and *say* (used in the sense of *think* and not as an utterance verb) that we have treated as ambiguous may not actually be perceived as such, if the apparent main clause behaves as a parenthetical clause that is inserted as an epistemic marker and does not contain its own assertion (Diessel & Tomasello, 2001; Thompson & Mulac, 1991). This is particularly plausible in the present case, because the majority of the sentences with *think* have the form *WH do you think/say…*, where the *do you think* clause could be used only to indicate the source of the main assertion of the sentence. [24] If this is the case, one possible outcome is that when

[24] In a wh-scope marking language like German, sentences that involve true complementation and sentences that involve parentheticals demonstrate overtly different syntactic behaviors. When there is true complementation with a wh-dependency, then the sentence takes a wh-scope marking construction as shown in (ia); but when the first clause with *think* is a parenthetical, then the verb in the second clause shows a verb-second phenomenon (ib), suggesting that this second clause is treated as a root clause in (ib) and that the first clause *glaubt de Student* does not select the second clause. The availability of a bound variable reading in (iia) and the lack of the bound variable reading in (10b) further demonstrates that the first clause does not c-command the second clause in (iib).

(i) a. Was glaubt der Student$_1$, was der Professor beibringen sollte?
what thinks the student what the professor teach should
b. Was glaubt der Student, sollte der Professor beibringen?
what thinks the student should the professor teach
"What does the student think that the professor should teach?"

(ii) a. Was glaubt jeder Student$_1$, was der Professor ihm$_1$ beibringen sollte?
what thinks every student what the professor him teach should
b. Was glaubt jeder Student$_1$, sollte der Professor ihm$_{*1 / OK2}$ beibringen?
what thinks every student should the professor him teach
"What does every student think that the professor should teach?"

children perceive these sentences as including a parenthetical, these sentences no longer serve as informative data for the purpose of acquiring long distance dependency formation rules. In this case, 69.8% (60/86) of the potentially relevant data would be discarded, leaving very few data points for children to use to decide on the long distance wh-dependency formation rule in the target language. Moreover, these 60 sentences were critical for observing the presence of the embedded clause interpretation; the decision to exclude sentences with *think/say* discards all of the 40 sentences that were potentially ambiguous (which mainly consisted of *why-think/say* combinations) as well as the 20 sentences in which the embedded clause interpretation was forced due to the incompatibility between the wh-phrase and the verb *think*. As a result, the remaining 26 sentences are all biased toward the main clause interpretation due to the factive island. Given that this effectively eliminates the clearest evidence for the embedded clause interpretation, this could drastically slow down the learning rate for this construction.

Another source of information that may affect children's acquisition of long distance dependency formation is the distribution of long distance wh-dependencies with arguments. In order to examine the effects of the child parser on the process of language acquisition, we have so far concentrated on the distribution of long distance wh-dependencies with adjuncts that show structural ambiguities. However, long distance wh-dependencies with arguments are unambiguous in that the wh-phrase must be associated with a gap position in the embedded clause. If we assume that long distance dependency data on arguments (there were 219 sentences of this type) can be generalized to adjuncts

---

English is unique in that even if the first clause is treated as a parenthetical, the second clause does not show a root clause phenomenon like subject-auxiliary inversion, unlike the German sentence in (ib) where a root phenomenon like verb-second is observed.

(there were 86 sentences of this type), then the evidence for English-type overt long distance dependency formation dramatically increases. Thus, even if we assume that children cannot retract their active syntactic commitments in processing long distance adjunct wh-dependencies (i.e., 86 sentences are all analyzed as allowing only the main clause interpretation), the proportion of sentences that make the embedded clause wh-association becomes larger (71.8%, 219/305).

It is also important to note, however, that not all the syntactic properties of wh-arguments should be generalized to wh-adjuncts. For example, it is well known that arguments and adjuncts show differences in terms of constraints on long distance dependency formation. For example, extraction out of a wh-island as in (10) is only slightly degraded with argument fronting (10a), but it is significantly worse with adjunct fronting (10b) (Cinque, 1990; Rizzi, 1990).

(10) a. ? Who do you wonder whether John talked to ___ ?

b. * How do you wonder whether John talked to Mary ___ ?

This suggests that at least the syntactic constraints on long distance dependency formation cannot be generalized straightforwardly from arguments to adjuncts. Thus, even if a generalization across distinct syntactic representational classes is possible, theories of language acquisition must place appropriate constraints on what kind of syntactic properties can be generalized.

We have so far assumed that English-speaking children's processing of local dependencies and long distance dependencies in the input is critical for learning the long

distance dependency formation rule, which is a fair assumption to make in models of language acquisition in which the parsing experience is assumed to play a critical role in language acquisition (e.g., Fodor & Sakas, 2004; Yang, 2002, 2004). However, it is also possible that the lack of wh-scope marking constructions in the English input can serve as indirect negative evidence and lead children to acquire the target English grammar. Under a rational Bayesian model of language acquisition (e.g., Pearl & Lidz, 2009; Tenenbaum & Griffiths, 2001; Tenenbaum & Xu, 2007), for example, the learner searches through the input for strings that are predicted by the competing grammars, i.e., the English-type grammar and the Russian-type grammar. According to this model, the rational learner expects the input to contain some occurrences of wh-scope marking construction if the Russian-type grammar is the correct grammar. Since English-speaking adults do not produce wh-scope marking constructions, however, the learner would be led to reject the Russian-type wh-scope marking grammar. Thus, in addition to children's revision capacities based on verb-based error signals, a rational model of language acquisition could provide another solution to the skewed input problem in the acquisition of long distance dependencies.

6.3 Conclusion

This chapter investigated the distribution of long distance wh-dependencies in child directed speech, and how the properties of the child parser that we uncovered in Chapter 4 would affect the input distribution. It was found that the active dependency completion bias in children could potentially skew the input distribution to the extent that the input may not clearly indicate the availability of the embedded clause interpretation, but that children's ability to use verb information to retract such active dependency

completion commitments would help to create clear evidence for the English-type long distance dependency formation rule. These results illustrate that the input distributional pattern could vary depending on properties of the children's perceptual mechanisms, and that it is important to consider these perceptual mechanisms in understanding the role of input distribution in language acquisition.

Chapter 7: Conclusion

The goal of this dissertation was to understand how sentence interpretation mechanisms in adults and children incrementally compute syntactic and interpretive processes, as well as how reanalysis mechanisms function to retract incremental syntactic and interpretive commitments when those commitments are disconfirmed by later arriving information. As a case study, we concentrated on adults and children's active completion of long distance dependencies, where the sentence processing mechanism was shown to construct syntactic and semantic representations that were not (yet) supported by critical bottom-up evidence for the gap position.

This dissertation took advantage of the characteristics of active syntactic processing to shed light on how the parser uses linguistic information to actively construct syntactic dependencies and execute subsequent interpretive processes in real time. Furthermore, we extended the investigation to child sentence processing, asking whether children actively process long distance dependencies, and if so, whether such active commitments can be retracted based on later arriving information. These experimental investigations used a variety of on-line and off-line measures in English and Japanese to better understand what kind of information is critically used to trigger active processing and revisions of active commitments. The next section summarizes the major empirical findings and specific conclusions of this dissertation.

7.1 Summary of empirical findings and specific conclusions

Active filler-gap dependency computation in adults (Chapters 2 and 3)

The starting point of our investigation was to examine to what extent active dependency completion in sentence interpretation proceeds without relying on critical bottom-up information about the gap position, which only becomes available at or after the tail of long distance dependencies. Previous work on long distance dependency processing in verb-final languages suggested that active dependency completion is executed without relying on bottom-up information like verb transitivity information, but there has not been conclusive evidence for this view in English filler-gap dependency completion. A self-paced reading study (Experiment 1) and an eye-tracking during reading study (Experiment 2) examined whether active object gap creation processes are dependent on the verb transitivity information in English. In both experiments, the presentation of intransitive verbs in a potential gap creation environment yielded evidence for reading disruption. This suggests that the syntactic representation for filler-gap dependencies was constructed before the parser had access to transitivity information of the verb.

In order to further our understanding of how the comprehension mechanism maps the actively constructed syntactic representations to semantic and discourse representations, Experiments 3 and 4 used a visual world eye-tracking experiment with a Question-after-Story design. Here, the pattern of fixations on objects that were described in the story were measured during the presentation of wh-questions like *Can you tell me what Emily was eating the cake with ___?* Fixation patterns from Experiment 4 suggest that the interpretation of filler-gap dependencies can be completed within 600-700ms

after the onset of the verb, when there is no pre-verbal information that would allow listeners to predict which verb they are about to hear. On the other hand, this process appears to be completed earlier (within 200-300ms after the onset of the verb) when the combination of contextual and linguistic information in the sentence allows listeners to predict the upcoming verb.

These results led us to propose the following procedure for the real time computation of filler-gap dependencies (1).

(1) a. Actively construct the syntactic dependency between the filler and the gap.
b. Access the lexical information that becomes available at the tail of the dependency, and compute the semantic and discourse representations.

This illustrates the sequence in which the input is mapped onto different levels of linguistic representations in real time sentence interpretation. Moreover, it was argued that some of these processes could be executed predictively. For example, the active syntactic dependency completion process that did not rely on the verb information (Experiments 1 and 2) effectively led the parser to predict that the verb was going to be transitive. Moreover, the visual world eye-tracking studies (Experiments 3 and 4) suggested that the specific semantic content of the verb can be predicted based on a combination of contextual and linguistic information, such that interpretive processes can be executed early. In summary, the findings reported in this dissertation lend support to psycholinguistic models that accommodate predictive computation of linguistic representations (Aoshima, Phillips, & Weinberg, 2004; DeLong, Urbach, & Kutas, 2005;

Federmeier, 2007; Gibson, 1998; Hale, 2003; Kamide, Altmann, & Hayward, 2003; Kimball, 1975; Lau, Phillips, & Poeppel, 2008; Levy, 2008).

Active processing and reanalysis mechanisms in children (Chapters 4 to 6)

Active syntactic processing that has not been validated by bottom-up evidence provides an extremely efficient way to encode language input in linguistic representations, but it also creates the risk that the processing burden will be increased, since active commitments are sometimes disconfirmed by later arriving bottom-up information and must be retracted in order for sentence interpretation to be successful. Previous work on child sentence processing suggests that children incrementally incorporate bottom-up information to compute syntactic and interpretive processes, and that when such incrementally constructed representations are disconfirmed by later arriving information, they perseverate and fail to retract their incremental commitments (producing a 'kindergarten-path effect'). This dissertation presented a novel investigation of whether children's revision difficulties extend to contexts in which children make active, hypothesis-driven commitments that are to be tested against bottom-up information.

The first step in this investigation was to establish that children actively process long distance dependencies. In order to probe this, Experiments 5 through 8 (Chapter 4) used story-based comprehension tasks that investigated how adults and children comprehend ambiguous wh-questions like (2) in English and Japanese.

(2) a. Where did Emily tell someone that she hurt herself?

b. Doko-de Emily-chan-wa *pro* ashi-o kegashita to itteta-no?

where-at Emily-Dim-Top she foot-Acc hurt Comp was telling-Q

The comprehension results from Experiments 5, 6 and 7 showed that adults and children consistently associate the fronted wh-phrase with the first VP in both English and Japanese versions of sentences like (2), such that English-speaking adults and children systematically provided an answer that reflects the main clause interpretation in (2a), while Japanese adults and children systematically provided an answer that reflects the embedded clause interpretation in (2b). These cross-linguistic data attesting a systematic preference for the wh-association with the first VP in the sentence accords with the adult time course evidence from previous work that wh-dependencies are completed at the first potential thematic positions (Aoshima et al., 2004; Nakano et al., 2002). This strongly suggests that the interpretation preferences observed in our experiments demonstrate evidence for active filler-gap dependency completion in children.

Based on this conclusion, Experiment 8 investigated whether children can retract their active syntactic commitments. This experiment manipulated the revision cues to compare the effectiveness of verb-based error signals versus syntactic error signals. The results demonstrated that children were successful at retracting their active syntactic commitments when the error signals were based on verb information, whereas they failed to retract the active commitment when the error signal consisted of syntactic information.

When this finding is taken together with previous demonstrations of kindergarten-path effects, a new generalization emerges. Previous demonstrations of these effects were

primarily based on garden-path sentences involving a temporary PP attachment ambiguity, such as *Put the frog on the napkin in the box.* In such sentences, the argument structure of the ditransitive verb *put* is completed when the structurally ambiguous PP is processed, and this in turn leads the child sentence processing mechanism to construct a series of semantic and discourse representations that are confirmed by the bottom-up information. Thus, based on these findings, I propose the following generalization to account for the distribution of kindergarten-path effects (3).

(3) a. Kindergarten-path effects arise when

(i) children make interpretive commitments that are fully confirmed by bottom-up evidence, and

(ii) active commitments are disconfirmed by syntactic information.

b. Kindergarten-path effects are not observed when active commitments are disconfirmed by semantic or pragmatic information coming from the verb

Experiments 9 and 10 (Chapter 5) extended the domain of investigation to the incremental processing of binding relations. The reading time evidence reported here for the incremental processing of sentences similar to (4), combined with children's comprehension of sentences like (4) as reported in Leddon and Lidz (2006), is consistent with the generalization about children's reanalysis mechanism in (3).

(4) a. Ms. Cruella wanted to know which picture of herself Janie put up on the wall.

b. Mr. Monkey wondered how proud of himself Andy was.

Leddon and Lidz demonstrated that for (4a), children accept only one of the two grammatically permissible interpretations. For (4a), they accept the reading in which *herself* is bound by *Ms. Cruella* (the 'surface' interpretation), but fail to accept the reading in which *herself* is bound by *Janie* ('reconstruction' interpretation). In (4b), on the other hand, children only accepted the reconstruction interpretation, which is indeed the only grammatically permissible interpretive option. Leddon and Lidz argued that children incrementally adopt the grammatically permissible interpretation that becomes available first in the sentence, which was assumed to be the surface interpretation in (4a) and the reconstruction interpretation in (4b). Adult reading time data from a self-paced reading study (Experiment 9) and an eye-tracking during reading study (Experiment 10) supported the argument that the surface interpretation becomes available first in (4a), but also suggested that in real-time processing, the sentence processing mechanism actively hypothesizes that the main clause subject and the null subject of the fronted wh-predicate are co-referential in (4b), which effectively leads readers to transiently entertain the surface interpretation. This pattern turns out to support the child reanalysis generalization in (3): The incrementally constructed binding relation in (4a) is based on the bottom-up information that has become available in the input, and therefore this is expected to be difficult to retract (category (3a(i)). On the other hand, the actively hypothesized co-reference relation between the main clause subject and the reflexive in (4b) is disconfirmed by the fact that the actual external argument of the fronted predicate (i.e.,

the embedded clause subject) is not co-referential with the main clause subject. Thus, this thematic information serves as an effective bottom-up error signal that allows children to retract the actively hypothesized co-index relation (the category (3b)).

In Chapter 6, I examined potential consequences of children's revision difficulties for language acquisition processes. Experiments 6 to 7 revealed that children and adults actively associate the wh-phrase with the main clause verb in sentences like (5) and do not retract this active analysis.

(5) Where did Emily tell someone that she hurt herself?

This experimental evidence for such a strong bias in children's comprehension of (5) in turn raises the possibility that the input distribution for long distance dependencies in sentences like (5) may be skewed in such a way as to affect children's learning of the long-distance wh-movement rule. For rexample, if the active dependency completion bias leads sentences like (5) to be always perceived as yielding only the main clause interpretation, then the absence of embedded clause interpretation could be taken as indirect negative evidence and English-speaking children might adopt the Russian grammar. Alternatively, if sentences like (5) in child-directed speech contain features that allow children to retract such active wh-association with the main clause VP, then the input distribution could be corrected to provide evidence for the availability of overt long distance wh-dependencies.

In order to explore this question, I examined the distribution of English wh-questions like (5) in child-directed speech in CHILDES (MacWhinney, 2000). The

distributional analysis showed that the active dependency completion bias in children could potentially skew the input distribution in such a way that the input may not clearly indicate the availability of the embedded clause interpretation, but that children's ability to use verb information to retract such active dependency completion commitments would help to create clear evidence for the permissibility of overt long distance wh-dependencies in English. The observation that the input distributional pattern could vary depending on properties of children's perceptual mechanisms calls for a re-evaluation of the assumptions made in learnability and computational modeling research on language acquisition, where it is generally assumed that the learner has a perfect parser that is capable of assigning any parse that is possible within the realm of the hypothesized grammar.

7.2 Broader implications for psycholinguistic research

The psycholinguistic studies reported in this dissertation drew on a variety of experimental methods as well as cross-linguistic and developmental data to shed light on the real-time procedures for computing linguistic representations. The fact that we were able to shed light on the nature of real time computation of filler-gap dependencies in adults and children as well as on children's reanalysis mechanisms highlights the usefulness of such diverse approaches to psycholinguistic problems.

As discussed in Chapters 2 and 3, different measures of real time processing are inherently suitable for understanding different kinds of linguistic representations and computations. For example, the visual world eye-tracking method is particularly useful for understanding interpretive processes and how listeners rapidly establish reference to the world, but eye-tracking during reading techniques may be more informative with

regards to how syntactic and semantic representations can be rapidly constructed. On the other hand, I showed that a careful articulation of linking assumptions between the hypothesized cognitive processes and dependent measures across different methodologies can help us establish what kinds of global linguistic computations take place during sentence interpretation.

Cross-linguistic sentence processing data are not only necessary for understanding how speakers of languages with different grammatical features compute linguistic representations in real time, but also extremely useful for revealing underlying principles of sentence processing mechanisms. The off-line comprehension data reported in Chapter 4 took advantage of the verb order differences between Japanese and English to construct an argument for active dependency completion. Moreover, the insight that English object gap creation processes may be executed independently of lexical information at the tail of the dependency was inspired by previous work on long distance dependency processing in verb final languages (Aoshima et al., 2004; Nakano et al., 2002). Thus, increasing the empirical coverage of cross-linguistic sentence processing is likely to help further progress in sentence processing research.

Finally, our developmental research on reanalysis mechanisms gained important insights from adults' interpretation persistence and revision difficulties. Of particular importance was the observation that incremental commitments are difficult to retract in the context of bottom-up information that appears to temporarily support these commitments. On the other hand, given that children's sensitivities to various error signals are pronounced and easily detectable, further investigations into when children succeed and fail in reanalysis may shed light on the factors that affect adults' reanalysis

processes but may be masked by adults' efficient recovery capacities. For example, in research on adults' active syntactic processing, the focus has been on what causes adults to construct representations actively, rather than why it is so easy for them to recover when the active commitments are disconfirmed by later arriving information. Thus, we invite further cross-population psycholinguistic research, which is likely to provide new insights on sentence processing mechanisms in both adults and children.

7.3 Broader implications for language acquisition research

7.3.1 Incremental processing and input distribution for long distance reflexives

Chapter 6 investigated whether children's incremental syntactic commitments may in principle skew the input distribution and affect the course of language acquisition, but it is unknown how widely such problems actually occur in child language acquisition. One case that may have similar characteristics is the acquisition of long distance reflexives. In languages like Chinese, the reflexive pronoun *ziji* can be bound by a c-commanding noun in the local clause, as in English; but unlike English reflexives, *ziji* can also be bound by a c-commanding noun in a non-local clause (for a review of *ziji* binding, see Huang & Liu, 2001). Thus, a sentence like (6) is a globally ambiguous sentence, because *ziji* can be bound by the subject of the local clause (Da-xingxing 'Big Gorilla') or the subject of the non-local clause (Milaoshu 'Mickey Mouse').

(6) Milaoshu$_1$ mengjian Da-xingxing$_2$ bei-zhe ziji$_{1/2}$ -de didi
Mickey Mouse dream Big-Gorilla carry self's baby-brother

There is little research on the processing of long distance reflexives, but recently Dillon, Chow, Wagers, Guo, Liu, and Phillips (2010) used a speed-accuracy-tradeoff task to examine the time course of *ziji* binding in Chinese and found a time course advantage for local subject binding. This suggests the possibility that children may also adopt local subject binding first in sentences like (6). Moreover, given that this commitment is based on bottom-up information that has become available in the input (see the generalization in (3)), children may predominantly commit to the local binding and may not be able to consider the long distance binding possibilities as often as they should. In fact, there are some experimental reports showing that children do prefer local subject binding in sentences like (6), even when the context biases the long distance binding interpretation (e.g., Chien & Lust, 2006; Su, 2003). This suggests the possibility that locality biases in children's reflexive processing mechanism may potentially skew the input and affect the developmental time course of long distance reflexive learning.

In order to properly examine this question, as shown with regard to the distributional analysis of wh-dependency in Chapter 6, it is critical to understand whether children are able to use linguistic information to retract such syntactic and interpretive commitments to short distance binding. Animacy information might constitute a reasonable revision cue: Chinese reflexive *ziji* is only compatible with a human antecedent, and therefore it cannot be bound by an inanimate subject. Thus, even if children prefer local binding, if the local subject turns out to be inanimate, they may be able to overcome the locality bias and retract an initial commitment to a local binding relation to obtain the long distance binding interpretation. This question needs to be addressed in future research.

7.3.2 Active processing may assist grammar acquisition

The findings reported in this dissertation reveal that children actively hypothesize syntactic structures, and that when later arriving bottom-up information disconfirms the actively constructed analyses, children are able to retract such active commitments. This 'hypothesize and evaluate' process could possibly generate a useful error signal for children's developing grammars as well (Phillips, 2009). For example, it is well known that children's grammars often 'over-generate', allowing more grammatical options than are actually licensed in the target grammar. This is most prominent in the case of argument structure, where children have been shown to utter non-adult-like sentences such as *you giggled me* or *fill the water into the glass* (Bowerman, 1982; Kim, Landau, & Phillips, 1999; Pinker, 1989). Now, suppose that children actively hypothesize upcoming syntactic structures using their developing grammar. Then, at the point of processing a verb like *fill*, children may expect to hear a figure (e.g., water) rather than a ground (e.g., glass) as the upcoming object NP. However, if children actually hear a ground in the input, this generates local negative evidence that they could use to revise their learning hypothesis about the argument structure of verbs like *fill*. In fact, it has been observed that children as young as 19-months old actively predict an upcoming object NP upon hearing a potential transitive verb (Lidz & Baier, 2010), and this could be taken as evidence that children are using the active processing mechanism in order to test their hypotheses about the target grammar. (For similar ideas about how children use empirical data to test hypotheses relating to their grammar, see Valian, 1990). Questions about the extent to which this type of self-generated negative evidence actually helps children to correct their grammatical hypotheses, as well as what other domain of grammar

acquisition might potentially be assisted by this mechanism, are left open for future research.

Appendix A: List of target items used in Experiment 1

1a/c. The house that the woman wrote/grinned slyly about was inspected by the board.
1b/d. The house that the woman who wrote/grinned slyly appreciated was inspected by the board.

2a/c. The city that the author wrote/chatted regularly about was named for an explorer.
2b/d. The city that the author who wrote/chatted regularly saw was named for an explorer.

3a/c. The injury that the victim called/sighed repeatedly about was treated improperly by doctors.
3b/d. The injury that the victim who called/sighed repeatedly saw was treated improperly by doctors.

4a/c. The jewelry that the sheriff questioned/joked sharply about was recovered after the robbery.
4b/d. The jewelry that the sheriff who questioned/joked sharply watched was recovered after the robbery.

5a/c. The billboard that the girl threw/snorted angrily at was located near the beach.
5b/d. The billboard that the girl who threw/snorted angrily saw was located near the beach.

6a/c. The wall that the boy threw/grinned mischievously at was painted fire engine red.
6b/d. The wall that the boy who threw/grinned mischievously noticed was painted fire engine red.

7a/c. The artist that the architect designed/competed passionately with was nationally well known.
7b/d. The artist that the architect who designed/competed passionately admired was nationally well known.

8a/c. The design that the professor lectured/sighed resignedly about was discussed in the seminar.
8b/d. The design that the professor who lectured/sighed resignedly saw was discussed in the seminar.

9a/c. The equipment that the employee phoned/frowned disapprovingly about was mentioned by the President.
9b/d. The equipment that the employee who phoned/frowned disapprovingly saw was mentioned by the President.

10a/c. The drugs that the principal threatened/frowned sternly about were discussed during the meeting.
10b/d. The drugs that the principal who threatened/frowned sternly confiscated were discussed during the meeting.

11a/c. The bridge that the tourist read/napped peacefully under was photographed by the group.

11b/d. The bridge that the tourist who read/napped peacefully missed was photographed by the group.

12a/c. The clock that the collector read/smiled fondly about was found while shopping for antiques.

12b/d. The clock that the collector who read/smiled fondly discovered was found while shopping for antiques.

13a/c. The bacteria that the biologist instructed/struggled intensely about turned out to be highly productive.

13b/d. The bacteria that the biologist who instructed/struggled intensely studied turned out to be highly productive.

14a/c. The match that the athlete trained/struggled endlessly for was ended by the authorities.

14b/d. The match that the athlete who trained/struggled endlessly played was ended by the authorities.

15a/c. The manager that the custodian cleaned/smiled obediently for was ruined by a financial crisis.

15b/d. The manager that the custodian who cleaned/smiled obediently liked was ruined by a financial crisis.

16a/c. The recording that the instructor taught/beamed enthusiastically about was heard throughout the auditorium.
16b/d. The recording that the instructor who taught/beamed enthusiastically presented was heard throughout the auditorium.

17a/c. The princess that the clown sang/danced cheerfully with was adored by the media.
17b/d. The princess that the clown who sang/danced cheerfully admired was adored by the media.

18a/c. The accident that the governor scolded/yelled angrily about was seen on the news.
18b/d. The accident that the governor who scolded/yelled angrily witnessed was shown on the news.

19a/c. The theories that the teacher scolded/corresponded vigorously about were taught throughout the term.
19b/d. The theories that the teacher who scolded/corresponded vigorously criticized were taught throughout the term.

20a/c. The jobs that the instructor taught/sulked grudgingly about were all in food service.

20b/d. The jobs that the instructor who taught/sulked grudgingly knew were all in food service.

21a/c. The country that the man killed/prayed endlessly for was destroyed by Mongol military.
21b/d. The country that the man who killed/prayed endlessly hated was destroyed by Mongol military.

22a/c. The party that the designer dressed/laughed obnoxiously for was thought to be very important.
22b/d. The party that the designer who dressed/laughed obnoxiously enjoyed was thought to be very important.

23a/c. The poster that the manager paid/appeared unexpectedly for was sent to the office.
23b/d. The poster that the manager that paid/appeared unexpectedly saw was sent to the office.

24a/c. The magazine that the children fought/giggled persistently about could not be found anywhere.
24b/d. The magazine that the children who fought/giggled persistently read could not be found anywhere.

25a/c. The sculpture that the critics lectured/quarreled seriously about was seen in the

park.

25b/d. The sculpture that the critics who lectured/quarreled seriously denounced was seen in the park.

26a/c. The money that the criminal kidnapped/chuckled cruelly for was missing during the investigation.

26b/d. The money that the criminal who kidnapped/chuckled cruelly despised was missing during the investigation.

27a/c. The game that the journalist bet/shrieked compulsively on was discussed at the pub.

27b/d. The game that the journalist who bet/shrieked compulsively watched was discussed at the pub.

28a/c. The client that the cook prepared/flirted skillfully for was disliked by the waiters.

28b/d. The client that the cook who prepared/flirted skillfully favored was disliked by the waiters.

Appendix B: List of target items used in Experiment 2

1a/c. The studio that the students designed/remained peacefully in while the professors conferred was small and ugly.

1b/d. The studio that the students who designed/remained peacefully rested in while the professors conferred was small and ugly.

2a/c. The warehouse that the trucker phoned/departed nervously from last week was very old.

2b/d. The warehouse that the trucker who phoned/departed nervously visited last week was very old.

3a/c. The opponent that the veteran tennis player played/prevailed skillfully with/over during the game was very gracious.

3b/d. The opponent that the veteran tennis player who played/prevailed skillfully beat during the game was very gracious.

4a/c. The newsroom that the reporter called/emerged moodily from was full of problems.

4b/d. The newsroom that the reporter who called/emerged moodily oversaw was full of problems.

5a/c. The group that the speaker lectured/appeared reluctantly with at the conference was very dogmatic.

5b/d. The group that the speaker who lectured/appeared reluctantly spoke to at the conference was very dogmatic.

6a/c. The impaired plane that the pilot landed/arose wearily behind/from was a mass of twisted metal.
6b/d. The impaired plane that the pilot who landed/arose wearily had flown/crashed did not meet safety standards.

7a/c. The knight that the warrior killed/died nobly for went on to save the princess.
7b/d. The knight that the warrior who killed/died nobly admired/succumbed to went on to save the princess.

8a/c. The quarrel that the girl heard/persisted reluctantly about/in was incomprehensible and pointless.
8b/d. The quarrel that the girl who heard/persisted reluctantly resolved/won was incomprehensible and pointless.

9a/c. The building that the thief climbed/disappeared quickly over/behind during the chase was a nondescript warehouse.
9b/d. The building that the thief who climbed/disappeared quickly entered during the chase was a nondescript warehouse.

10a/c. The manager that the customer fought/erupted angrily with/at was near the

register.

10b/d. The manager that the customer who fought/erupted angrily shouted at was near the register.

11a/c. The rave that the teenage boy planned/remained obligingly for/in at the underground club was hopping.

11b/d. The rave that the teenage boy who planned/remained obligingly watched at the underground club was hopping.

12 a/c. The old bed that the cat scratched/arose lazily at/in was really worn out but comfortable.

12b/d. The old bed that the cat who scratched/arose lazily sharpened her claws on was really worn out but comfortable.

13a/c. The table that the chemical burned/disappeared quietly on/from was very old and dirty.

13b/d. The table that the chemical that burned/disappeared quietly had left a stain on was very old and dirty.

14a/c. The people that the conqueror killed/prevailed fiercely for/over were [grateful to have their city back]/[upset to lose their city].

14b/d. The people that the conqueror who killed/prevailed fiercely defended were grateful to have their city back.

15a/c. The athlete that the coach taught/appeared proudly about/with before the game was nominated for a big award.

15b/d. The athlete that the coach who taught/appeared proudly trained before the game was nominated for a big award.

16a/c. The accident that the lady escaped/died mysteriously from/in last night was thoroughly investigated.

16b/d. The accident that the lady who escaped/died mysteriously had photographed last night was thoroughly investigated.

17a/c. The research that the scientist prepared/persisted determinedly for/in during his whole career was finally completed.

17b/d. The research that the scientist who prepared/persisted determinedly pursued during his whole career was finally completed.

18a/c. The gate that the limousine passed/emerged slowly through/from as it left the house was closed shortly thereafter.

18b/d. The gate that the limousine which passed/emerged slowly crossed as it left the house was closed shortly thereafter.

19a/c. The frat boy that the woman fought/erupted aggressively with/at in the department store was very obnoxious.

19b/d. The frat boy that the woman who fought/erupted aggressively disciplined in the department store was very obnoxious.

20a/c. The assistant that the magician trained/vanished skillfully with was good at her job.

20b/d. The assistant that the magician who trained/vanished skillfully complimented after the show was good at her job.

21a/c. The fugitive that the mobster hid/appeared abruptly from/with was rumored to be very dangerous.

21b/d. The fugitive that the mobster who hid/appeared abruptly feared/shot was rumored to be very dangerous.

22a/c. The airport that the ambassador left/departed rapidly for/from during the unrest was closed to most traffic.

22b/d. The airport that the ambassador who left/departed rapidly had visited during the unrest was closed to most traffic.

23a/c. The computer lab that the IT technician phoned/arrived tardily from/at was full of college kids studying for finals.

23b/d. The computer lab that the IT technician who phoned/arrived tardily despised was full of college kids studying for finals.

24a/c. The costume party that the student planned/arrived eagerly for at the fraternity house was pretty lame.

24b/d. The costume party that the student who planned/arrived eagerly attended/threw at the fraternity house was pretty lame.

Appendix C: List of target items used in Experiment 3

1. Jody was eating breakfast one morning when she saw a big hairy spider creeping across the table towards her. Jody, whose terrible arachnaphobia had caused her to seek therapy a few years ago, drew on the techniques of relaxation and anxiety management that her therapist had taught her. Instead of screaming or freaking out, she calmly took off her shoe and slammed it down on top of the spider. She ate the rest of her Froot Loops in peace.

Wh-question: What did Jody squash a spider with?

Yes/No question: Did Jody squash a spider with her shoe?

2. Courtney was bored out of her mind during fifth period algebra. To occupy herself, she resorted to doodling, and ended up drawing a fairly elaborate rose on the cover of her math book. Pleased with her work, Courtney entered the picture in the school's drawing competition.

WH-question: What did Courtney draw a rose on?

Yes/No question: Did Courtney draw a rose on her math book?

3. Shannon had planned a surprise party for her roommate Marion's birthday. She had just finished wrapping the final present when she heard Marion come in the front door. She barely had time to shove the present under the bed before Marion walked in the room, asking if Shannon had seen her car keys. Shannon told her they were on the kitchen table. As she left the room, Marion gave Shannon a funny look. Shannon couldn't figure out why until she reached up and realised that she was still wearing the party hat she had tried on earlier when going through supplies. She was very disgusted with herself. All that hurrying with the present, and she had given away the secret by her own carelessness

WH-question: What did Shannon hide the present under?

Yes/No question: did Shannon hide the present under the bed?

4. Leslie was thoroughly enjoying her summer at the lake. On cool summer evenings, she would sit out on the deck relaxing and reading the newspaper. On this particular evening, however, Leslie could not enjoy herself properly because of one mosquito that kept circling her head and buzzing in her ear. When it finally landed, Leslie rolled up the newspaper and swatted the mosquito. Now she could read undisturbed.

WH-question: What did Leslie swat the mosquito with?

Yes/No question: Did Leslie swat the mosquito with a newspaper?

5. Cody and his brother Josh were preparing Thanksgiving dinner for their family. Since Cody was only six years old, Josh had been doing most of the work, and Cody was beginning to feel a little left out. Suddenly, Josh turned and handed him a pot of boiled potatoes and a fork and said, ‘‘Here you go, buddy. You’re in charge of the potatoes. Mash ’em up good so not even Grandma can find a lump.’’ Cody proudly took the fork and proceeded to mash the living daylights out of the unsuspecting potatoes.

WH-question: What did Cody mash the potatoes with?

Yes/No question: Did Cody mash the potatoes with a fork?

6. While playing outside one day, Karen chanced across a baby bird that had fallen out of its nest. The bird wasn’t moving, and Karen wasn’t sure if it was alive or not. Knowing the mother would abandon it if she smelled that a human had touched her baby, Karen gently nudged the bird with a twig. It didn’t move. She poked a little harder. Still no reaction. Realising that it was dead, Karen got a shoebox, put the baby bird in, and buried the package in the back yard.

WH-question: What did Karen bury the bird in?

Yes/No question: Did Karen bury the bird in a shoebox?

7. While cleaning out his attic, Otto began emptying his old suitcase. In addition to old clothes and musty towels, he was surprised to discover his old scrapbook inside. Otto began thumbing through the pages looking at old pictures of himself and his childhood friends. Soon Otto had forgotten all about cleaning the attic, and was sitting in the middle of the dust and clutter, examining every page of is scrapbook and reliving happy memories.

WH-question: What did Otto find his scrapbook in?

Yes/No question: Did Otto find his scrapbook in his suitcase?

8. It was Robbie's night to make dinner for the members of the fraternity house and he was a little perplexed. He had never booked for twenty-five people all at once before, and he wondered how he was going to manage. He decided he would keep it as simple as possible, and that meant spaghetti for everyone. Robbie found a huge stew pot, filled it with water and began boiling the spaghetti. This was going to be easier than he thought.

WH-question: What did Robbie cook the spaghetti in?

Yes/No question: Did Robbie cook the spaghetti in a stew pot?

9. For years, Darryl had been saving his allowance in a little pink ceramic piggy bank. One day on the way home from school, Darryl saw a shiny new bike in the window of the toy store. He ran home, found a hammer and smashed his little bank to bits. Inside, there was forty-three dollars and seventy-four cents in change. That would surely be enough. Excitedly, Darryl scooped the money into a plastic bag and heaved back to the toy store. He was horrified to discover that the bike he had wanted was ninety-nine, ninety-nine, plus tax.

WH-question: What did Darryl smash the piggy bank with?

Yes/No question: Did Darryl smash the piggy bank with a hammer?

10. Harvey was taking a cooking course at the local community college. The recipe he was currently working on called for the meat to be pounded flat. Harvey was surprised—he had never seen that kind of thing before. After removing all bones and fat, he dutifully got out his little wooden mallet and hammered his slab of meat until it was only a centimetre thick. Harvey was delighted when he tasted how tender and delicious the dish turned out. He also received an A for the project.

WH-question: What did Harvey flatten the meat with?

Yes/No question: Did Harvey flatten the meat with a mallet?

Appendix D: List of target items used in Experiment 4

Below is a list of the two events that occurred in each story, and the the target questions for wh- and yes-no conditions. Note that all the questions follow "Can you tell me…".

	1st event	2nd event
1	eat cake w/ fork	wash dishes with sponge
2	peel potatoes w/ knife	wipe table w/ sponge
3	Build sandcastle w/ bucket	blow up beach ball with air pump
4	light candle with match	dust trunk w/ rag
5	draw picture w/ crayon	attach bows with glue
6	mow lawn w/ lawn mower	chop wood with axe (for the
7	gather leaves w/ rake	sweep porch w/ broom
8	Shine saddle w/ cloth	groom his horse Clipclop w/ brush
9	wash car w/ hose	polish vase with towel
10	fix bike w/ wrench	paint mailbox with a spray can

	wh-questions	if questions
1	…what Emily was eating the cake with?	...if Emily was eating the cake with the fork?
2	...what Sammy was peeling the potatoes with?	... if Sammy was peeling the potatoes with the knife?
3	...what Ethan was building the sandcastle with?	... if Ethan was building the sandcastle with the bucket?
4	…what Robbie was lighting the candle with?	... if Robbie was lighting the candle with the match?
5	...what Lizzie was drawing the picture with?	... if Lizzie was drawing the picture with the crayon?
6	...what Jimmy was chopping the wood with?	... if Jimmy was chopping the wood with the axe?
7	...what rosie was sweeping the porch with?	... if Rosie was sweeping the porch with the broom?
8	...what Oscar was grooming the horse with?	... if Oscar was grooming the horse with the brush?
9	...what Kelly was polishing the vase with?	... if Kelly was polishing the vase with the towel?
10	...what Sally was paintig the mailbox with?	... if Sally was painting the mailbox with the spray can?

Appendix E: List of target items used in Experiment 5

Below is the list of stories and questions in the target items. The questions are shown with the verb *tell someone* only, but this experiment had three main clause predicate conditions, namely, *say*, *tell someone*, and *say to someone*.

Item 1

Jeff is an astronaut and he is looking for aliens. Back when he was on earth, Jeff wanted to learn more about where he could find aliens. But he was really busy preparing for his trip, and by the time Jeff had made everything ready to go, he had to leave right away and he had no time to do research on aliens. Because he couldn't get any research done, Jeff thought he should ask his friend the scientist who lives in space about it. Jeff went to visit the scientist at his space station, and Jeff took all of the scientist's advice and decided where he should go. He said to the scientist, "I'm going to go find aliens on Mars!" The scientist agreed that that was an excellent plan, and Jeff headed off on his adventure. On his way to Mars, he dropped by the moon in case there were aliens there too. He looked everywhere on the moon, but he didn't find any aliens, so he continued towards Mars. When he arrived on Mars, Jeff finally found aliens! They were very friendly, and Jeff was happy that he got to meet nice aliens.

Question: Where did Jeff tell someone that he was going to see aliens?

Item 2

John had a long day at school and he was really hungry. When he got home, John found a piece of cake sitting in his room! He remembered that his Mom and grandma had made a cake together this morning. Perfect! But when he went to eat the cake, he tripped over a toy and dropped it on the floor! What a mess! John couldn't eat the cake now and he was still hungry, so he went to look for the rest of the cake to get another piece. He found the cake in the kitchen, and ate a piece there. The recipe that his Mom and grandma used turned out really well, and the cake was really delicious! Once he finished the cake, he went to the laundry room to tell his Mom about how great the cake tasted, but he could not talk to her because she was on the phone. John waited for a little while, but Mom kept talking on the phone and didn't pay any attention to him. He was disappointed that he couldn't talk to his Mom about how good the cake was. But John remembered that his grandma helped his Mom to make the cake, so he decided to go to his grandma's house next door to tell her how wonderful the cake was! John walked over to grandma's house and said to his grandma: "I dropped the piece of cake in my room, but then I ate a piece of the cake in the kitchen, and it tasted great!" John was really glad he could let her know how well the recipe worked, and John's grandma was very happy with herself.
Question: Where did John tell someone that he ate a piece of cake?

Item 3

It was a beautiful day in spring so Lizzie decided she was going to go catch some butterflies in the park. Her Mom and Dad weren't home, so Lizzie thought she should tell her brother or sister about going to the park, so that Mom and Dad will know where she

is when they get back. She first went to her brother's room, but he was taking a nap and she couldn't tell him about catching butterflies. Instead, Lizzie looked for her sister. She looked all over the house but didn't see her sister anywhere! When she was about to give up, Lizzie heard her sister's voice in the basement! She went to the basement and said to her sister: "I'm going to catch butterflies in the park!" Then, on her way to the park, Lizzie passed by a parking lot and saw a butterfly near it! She walked slowly towards the butterfly, but before Lizzie could get there, another girl came along and caught the butterfly! Lizzie didn't see any more butterflies there, so she kept walking towards the park. There were lots and lots of butterflies in the park, and she caught one in a jar and took it home with her. She liked the one that she caught, but she wished she could have caught more butterflies.

Question: Where did Lizzie tell someone that she was going to catch butterflies?

Item 4

Bill and his Mom went shopping yesterday and bought a new soccer ball, so Bill wanted to try it out after school today. At school, he wanted to tell his friends about playing soccer later, to see if some of them want to play with him. But he didn't want the teacher to get angry at him for talking in class, so he didn't say anything at school. After school, Bill knew that his Mom would be worried if he didn't get home till late, so he went home to tell her about his plans. He said to his mom: "I'm going to go play soccer on the soccer field!" His mom was happy that he was already using the new soccer ball. On his way to the soccer field, he passed by a playground and saw some friends there. He tried to play soccer with them, but they were already flying kites, so they didn't want to play. But

Bill’s best friend decided to leave the playground and go to the soccer field with Bill. They played soccer together in the soccer field, and Bill scored lots of goals with his new soccer ball.

Question: Where did Bill tell someone that he was going to play soccer?

Item 5

Annie was taking horseback riding lessons. Today she got to go on a long ride all by herself. First she tried to ride up a mountain, but the mountain trail was too steep for the horse so she couldn't ride the horse there at all. Instead, Annie and her horse went riding in the forest. It was so much fun to ride the horse in the forest - along the way, Annie and her horse even got to jump over a log that was blocking the path! Annie was really proud of herself for riding all alone in the forest, so she wanted to tell her riding teacher about her great ride. But her teacher was busy teaching another student, so she couldn’t talk to her. That night, Annie and her friend went camping at a campsite. Annie's friend was taking horseback riding lessons too, so he asked Annie about her ride. Annie said to him, “I rode a horse in the forest today, and we jumped over a big log!” Annie's friend was jealous that she's already allowed to go riding by herself.

Question: Where did Annie tell someone that she rode her horse?

Item 6

Emily likes to play outdoors. One day she was swinging on the swings, and she jumped off the swings from really high up! She balanced herself really well on the landing, so she didn't fall. When Emily got bored with the swings, she decided to go climb a really tall

tree. She got up pretty high, but suddenly one of the branches broke, so she fell off the tree and hurt herself! But Emily was a brave girl, so she got back up right away and didn't cry at all. Emily wanted to tell her friends how pleased she felt about not crying after hurting herself, so she went to find some of her friends at the library. She found a friend of hers there, but the librarian came out and told her that in the library they must be very quiet! So, Emily couldn't talk to her friend and felt disappointed. But she had a good idea: She could go to the swimming pool to see more friends, because at the pool she can talk as much as she wants! When she got to the pool and found her friends, she said to them, "I hurt myself falling out of the tree, but didn't cry at all!" Emily was happy she could finally tell someone about her day, and her friends were impressed by how brave she was. Question: Where did Emily tell someone that she hurt herself?

Item 7

Hannah watched a movie about treasure hunting one day, and she thought it would be fun to go in search of treasure herself. But first, she needed advice on where to find it. She went to ask a professor at the university about the best places to find treasure, but he had gone on vacation so she couldn't talk to him. Then she decided that the guide at the museum might know where to find treasure too, so she went to the museum and talked to him. The museum guide there told her all that he knew about where to find treasure, and Hannah said to him : "Thanks! I think I'll go find treasure in a shipwreck!" The guide agreed that that was a good place, and told her about a shipwreck nearby. On her way to the shipwreck, Hannah passed a cave. She thought there might be treasure in the cave too, so she decided to go in and have a look around. But the cave was really, really dark, and

Hannah couldn't see anything! Hannah got really scared so she left the cave right away and kept going towards the shipwreck. When she made it to the sea, she put on her goggles, dove underwater, and went inside the shipwreck. She looked around really carefully, and finally found a chest full of treasure! Hannah was really excited, and thought she should give some to the museum guide because he helped her find it.
Question: Where did Hannah tell someone that she was going to find treasure?

Item 8

Heidi really likes airplanes and always wanted to sit in a pilot's seat on an airplane. One day when she was at the airport with her family, she decided to sneak inside one of the airplanes to see the pilot's seat, but before she could get to the plane, a police officer stopped her, because it was too dangerous. Heidi was really sad that she didn't get to see the inside of an airplane, so she went to an airplane museum in the neighborhood. At the airplane museum, there was an old airplane and they let Heidi go sit in the pilot's seat! She felt like she was a real pilot, and Heidi liked it so much that she decided to be a pilot when she grows up. Heidi went home feeling really happy, and she wanted to tell her parents about her new dream of becoming a pilot. But they might worry that it's a dangerous job, so Heidi decided not to tell them about her dream. Instead, she went to school the next day, and told her favorite science teacher about her plan. She said to her teacher, "I sat in a pilot's seat at the museum, and now I want to be a pilot!" The teacher was really happy for Heidi, and taught her about things she needs to learn to become a pilot.
Question: Where did Heidi tell someone that she sat in a pilot's seat?

Appendix F: List of target items used in Experiment 7

Japanese target sentences in Experiment 7 are listed below. These were created by translating the English target questions listed in Appendix E.

1. Doko-de Takuya-kun-wa uchuujin-o mitsukeru-to itteta-no?
   where-at Takuya-Dim-Top he alien-Acc find Comp was telling-Q
   "Where was Takuya telling someone that he was going to find aliens?"

2. Doko-de Kenta-kun-wa keeki-o tabeta-to itteta-no?
   where-at Kenta-Dim-Top he cake-Acc ate-Comp was telling-Q
   "Where was Kenta telling someone that he ate a piece of cake?"

3. Doko-de Yukiko-chan-wa choucho-o tsukamaeru-to itteta-no?
   where-at Yukiko-Dim-Top she butterfly-Acc catch- Comp was telling-Q
   "Where was Yukiko telling someone that she will catch butterflies?"

4. Doko-de Hiroshi-kun-wa sakkaa-o suru-to itteta-no?
   where-at Hiroshi-Dim-Top he soccer-Acc do-Comp was telling-Q
   "Where was Hiroshi telling someone that he was going to play soccer?"

5. Doko-de Yuko-chan-wa uma-ni notta-to itteta-no?

where-at Yuko-Dim-Top she horse-Dat rode-Comp was telling-Q

"Where was Yuko telling someone that she rode a horse?"

6. Doko-de Emily-chan-wa ashi-o kegashita-to itteta-no?

where-at Emily-Dim-Top she foot-Acc hurt Comp was telling-Q

"Where was Emily telling someone that she hurt her foot?"

7. Doko-de Hana-chan-wa takara-o sagasu-to itteta-no?

where-at Hana-Dim-Top she treasure-Acc search-Comp was telling-Q

"Where was Hana telling someone that she was going to look for treasure?"

8. Doko-de Megumi-cha-wa pairottoseki-ni noreta-to itteta-no?

where-at Megumi-Dim-Top she pilot seat-Dat could ride-Comp was telling-Q

"Where was Megumi telling someone that she could sit in a pilot's seat?"

Appendix G: List of target items used in Experiment 8

This appendix presents the target materials used in Experiment 8. The target sentences in the embedded clause PP condition (1-8a) and the main clause PP condition (1-8b) are identical to the sentences used in Experiment 7 (Appendix F) except for the extra overt PP. Thus, the list below only provides the overt PP with its literal translation. Readers can reconstruct the target sentences by taking the ambiguous sentences in Appendix F, and insert the respective PP in the positions that are indicated by underlines in the example below (here, (a) indicates the position for embedded clause PP, (b) indicates the position for the main clause PP).

Doko-de Emily-chan-wa _____(a) ashi-o kegashita-to _____(b) itteta-no?
where-at Emily-Dim-Top foot-Acc hurt Comp was telling-Q

Note that the embedded clause PP always occurred between the subject and the object in the embedded clause, and the main clause PP always occurred between the embedded clause complementizer *-to* and the main clause verb.

1a. kasei-de "Mars at"

1b. uchuusen-de "spaceship at"

1c. Doko-de Takuya-kun-wa kasei-ni iku-to itteta-no?
where-at Takuya-Dim-Top he Mars-to go-Comp was telling-Q
"Where was Takuya telling someone that he was going to Mars?"

2a. daidokoro-de “kitchen at”

2b. obaachan-no-ie-de “grandmother’s house at”

2c. Doko-de Kenta-kun-wa keeki-ga oishikatta-to itteta-no?

where-at Kenta-Dim-Top he cake-Nom tasty-Comp was telling-Q

“Where was Kenta telling someone that the cake was tasty?”

3a. kouen-de “park at”

3b. monooki-de “storage room at”

3c. Doko-de Yukiko-chan-wa kouen-ni iku-to itteta-no?

where-at Yukiko-Dim-Top she park-to go- Comp was telling-Q

“Where was Yukiko telling someone that she was going to catch butterflies?”

4a. gurando-de “field at”

4b. ouchi-de “home at”

4c. Doko-de Hiroshi-kun-wa gurando-ni iku-to itteta-no?

where-at Hiroshi-Dim-Top he field-to go-Comp was telling-Q

“Where was Hiroshi telling someone that he was going to the field?”

5a. mori-no-naka-de “forest’s inside at”

5b. kyanpujou-de “campsite at”

5c. Doko-de Yuko-chan-wa mori-no-naka-ni itta-to itteta-no?

where-at Yuko-Dim-Top she forest-Gen-inside-to went-Comp was telling-Q

“Where was Yuko telling someone that she went inside the forest?”

6a. ki-no-shita-de “tree’s foot at”

6b. puuru-de “pool at”

6c. Doko-de Emily-chan-wa ki-kara ochita-to itteta-no?

where-at Emily-Dim-Top she tree-from fell-Comp was telling-Q

“Where was Emily telling someone that she fell off the tree?”

7a. chinbotsusen-de “shipwreck at”

7b. hakubutsukan-de “museum at”

7c. Doko-de Hana-chan-wa chinbotsusen-ni iku-to itteta-no?

where-at Hana-Dim-Top she shipwreck-to go-Comp was telling-Q

“Where was Hana telling someone that she was going to the shipwreck?”

8a. hikouki hakubutsukan-de “airplane museum at”

8b. gakkou-de “school at”

8c. Doko-de Megumi-cha-wa pairotto-ni nari-tai-to itteta-no?

where-at Megumi-Dim-Top she pilot-Dat become-want-Comp was telling-Q

“Where was Megumi telling someone that she wants to become a pilot?”

Appendix H: List of target items for Experiments 9 and 10

1a Robert wondered which story about himself the grumpy old alcoholic was trying to spread around at the community center.

1b Nicole wondered which story about himself the grumpy old alcoholic was trying to spread around at the community center.

1c Robert wondered how sure of himself the grumpy old alcoholic was trying to look when he saw the cop.

1d Nicole wondered how sure of himself the grumpy old alcoholic was trying to look when he saw the cop.

2a Patrick knew which rumor about himself the popular highschool quarterback was likely to spread after the game.

2b Rachel knew which rumor about himself the popular highschool quarterback was likely to spread after the game.

2c Patrick knew how ashamed of himself the popular highschool quarterback was likely to feel after the game.

2d Rachel knew how ashamed of himself the popular highschool quarterback was likely to feel after the game.

3a Nathan realized which video of himself the careless teenage babysitter had thrown away while cleaning out the cabinet.

3b Lauren realized which video of himself the careless teenage babysitter had thrown away while cleaning out the cabinet.

3c Nathan realized how angry with himself the careless teenage babysitter had looked while cleaning up the spill.

3d Lauren realized how angry with himself the careless teenage babysitter had looked while cleaning up the spill.

4a Kevin wondered which snapshot of himself the smart young nutritionist should upload to his professional website.

4b Laura wondered which snapshot of himself the smart young nutritionist should upload to his professional website.

4c Kevin wondered how delighted with himself the smart young nutritionist would feel if he won the big research grant.

4d Laura wondered how delighted with himself the smart young nutritionist would feel if he won the big research grant.

5a Justin forgot which description of himself the snobby British actor had talked about repeatedly during the interview.

5b Julie forgot which description of himself the snobby British actor had talked about repeatedly during the interview.

5c Justin forgot how mad at himself the snobby British actor had seemed during the interview.

5d Julie forgot how mad at himself the snobby British actor had seemed during the interview.

6a Brian was certain about which video of himself the talented ballet dancer would submit to the new reality show.
6b Laura was certain about which video of himself the talented ballet dancer would submit to the new reality show.
6c Brian was certain about how confident in himself the talented ballet dancer would pretend to be on the new reality show.
6d Laura was certain about how confident in himself the talented ballet dancer would pretend to be on the new reality show.

7a Sarah was worried which story about herself the spiteful tennis coach might slip to the journalists at the press conference.
7b Steven was worried which story about herself the spiteful tennis coach might slip to the journalists at the press conference.
7c Sarah was worried how in love with herself the spiteful tennis coach might look to the journalists at the press conference.
7d Steven was worried how in love with herself the spiteful tennis coach might look to the journalists at the press conference.

8a Heather figured out which photograph of herself the foolish police officer had ruined while searching the house.
8b Jeffrey figured out which photograph of herself the foolish police officer had ruined while searching the house.
8c Heather figured out how frustrated with herself the foolish police officer would feel after searching the house.
8d Jeffrey figured out how frustrated with herself the foolish police officer would feel after searching the house.

9a Alice recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.
9b Andrew recalled which drawing of herself the attractive young nanny had damaged during the summer vacation.
9c Alice recalled how pleased with herself the attractive young nanny had been during the summer vacation.
9d Andrew recalled how pleased with herself the attractive young nanny had been during the summer vacation.

10a Kathryn explained which rumor about herself the struggling stage actress had misrepresented in the talk show.
10b Raymond explained which rumor about herself the struggling stage actress had misrepresented in the talk show.
10c Kathryn explained how unhappy with herself the struggling stage actress had seemed during the talk show.
10d Raymond explained how unhappy with herself the struggling stage actress had

seemed during the talk show.

11a Karen noticed which picture of herself the famous French photographer had discussed in a recent magazine article.

11b Joseph noticed which picture of herself the famous French photographer had discussed in a recent magazine article.

11c Karen noticed how unsure of herself the famous French photographer had sounded in a recent magazine article.

11d Joseph noticed how unsure of herself the famous French photographer had sounded in a recent magazine article.

12a Hannah remembered which sketch of herself the sweet old grandmother had hung in the small living room.

12b Dennis remembered which sketch of herself the sweet old grandmother had hung in the small living room.

12c Hannah remembered how comfortable with herself the sweet old grandmother had looked in the small living room.

12d Dennis remembered how comfortable with herself the sweet old grandmother had looked in the small living room.

13a Josh knew which picture of himself the shrewd business executive would be likely to prefer since the two used to work together.

13b Jennifer knew which picture of himself the shrewd business executive would be likely to prefer since the two used to work together.

13c Josh knew how pleased with himself the shrewd business executive was likely to be after the merger went through.

13d Jennifer knew how pleased with himself the shrewd business executive was likely to be after the merger went through.

14a Dan wondered which description of himself the dignified British ambassador might find most believable.

14b Melissa wondered which description of himself the dignified British ambassador might find most believable.

14c Dan wondered how proud of himself the dignified British ambassador might feel after the truth about the prisoners came out.

14d Melissa wondered how proud of himself the dignified British ambassador might feel after the truth about the prisoners came out.

15a Cindy considered which recording of herself the quiet switchboard operator could have played for Bill.

15b Kevin considered which recording of himself the quiet switchboard operator could have played for Bill.

15c Cindy considered ho hw critical of herself the quiet switchboard operator was likely to be when she found out she had been fired.

15d Kevin considered how critical of himself the quiet switchboard operator was likely to be when she found out she had been fired.

16a Justin figured out which picture of himself the young makeup artist had secretly posted on the internet.

16b Jessica figured out which picture of himself the young makeup artist had secretly posted on the internet.

16c Justin figured out how disgusted with himself the young makeup artist had become when he found out the picture was posted on the internet.

16d Jessica figured out how disgusted with himself the young makeup artist had become when he found out the picture was posted on the internet.

17a Jason remembered which characterization of himself the brave American astronaut was likely to approve of most.

17b Michelle remembered which characterization of himself the brave American astronaut was likely to approve of most.

17c Jason remembered how angry with himself the brave American astronaut had seemed to be when he heard the news.

17d Michelle remembered how angry with himself the brave American astronaut had seemed to be when he heard the news.

18a Aaron wondered which story about himself the inexperienced substitute teacher was going to have to present in court.

18b Heather wondered which story about himself the inexperienced substitute teacher was going to have to present in court.

18c Aaron wondered how sure of himself the inexperienced substitute teacher was going to feel on the first day of school.

18d Heather wondered how sure of himself the inexperienced substitute teacher was going to feel on the first day of school.

19a Scott figured out which video clip of herself the popular guitar player had viewed on FaceBook.

19b Courtney figured out which video clip of herself the popular guitar player had viewed on FaceBook.

19c Scott figured out how unhappy with herself the popular guitar player had been when he viewed the video clip on FaceBook.

19d Courtney figured out how unhappy with herself the popular guitar player had been when he viewed the video clip on FaceBook.

20a Megan wanted to know which anecdote about herself the patriotic Army nurse would tell when addressing the students at the assembly.

20b Ryan wanted to know which anecdote about herself the patriotic Army nurse would tell when addressing the students at the assembly.

20c Megan wanted to know ho w upset with herself the patriotic Army nurse would feel when she heard the news about the tragedy back home.

20d Ryan wanted to know how upset with herself the patriotic Army nurse would feel when she heard the news about the tragedy back home.

21a Amy worried about what opinion of herself the grumpy bank president would have when she saw the story on the front page.

21b Andrew worried about what opinion of herself the grumpy bank president would have when she saw the story on the front page.

21c Amy worried about how frustrated with herself the grumpy bank president would be when she saw the story on the front page.

21d Andrew worried about how frustrated with herself the grumpy bank president would be when she saw the story on the front page.

22a Sarah forgot which rumor about herself the famous newspaper editor had tried to keep quiet.

22b Jake forgot which rumor about herself the famous newspaper editor had tried to keep quiet.

22c Sarah forgot how embarrassed with herself the famous newspaper editor had at first seemed to be when news of the project's cancellation was revealed.

22d Jake forgot how embarrassed with herself the famous newspaper editor had at first seemed to be when news of the project's cancellation was revealed.

23a Lisa recalled which image of herself the famous fashion model had seemed to prefer.

23b Adam recalled which image of herself the famous fashion model had seemed to prefer.

23c Lisa recalled ho w ashamed of herself the famous fashion model had seemed to be when she heard about the arrest.

23d Adam recalled how ashamed of herself the famous fashion model had seemed to be when she heard about the arrest.

24a Laura remembered which photograph of herself the crabby old librarian had been so happy to display when they finished the project.

24b Mike remembered which photograph of herself the crabby old librarian had been so happy to display when they finished the project.

24c Laura remembered how happy with herself the crabby old librarian had been when they finished the project.

24d Mike remembered how happy with herself the crabby old librarian had been when they finished the project.

Appendix I: Wh-dependencies in child directed speech

Below is a list of long distance dependencies with wh-adjuncts attested in child-directed speech in CHILDES (MacWhinney, 2000). The sentences are sorted by the main clause verb type.

Main Verb	Wh-phrase	Corpus	Speaker	Sentence
know	how	Adam	*MOT:	how do you know you don't like Arizona ?
know	how	Adam	*MOT:	how d(o) you know she has any toys ?
know	how	Adam	*MOT	Adam (.) how would I know that those are the wheels that go on here?
know	how	Adam	*URS	well (.) how did you know it was no good ?
know	how	Adam	*MOT:	how d(o) you know there are beans in there ?
know	how	Adam	*MOT:	how do you know that it's lunch time ?
know	how	Adam	*MOT:	how do you know that's the top truck ?
know	how	Adam	*MOT:	how do you know tha(t)'s your pencil ?
know	how	Adam	*MOT:	how do you know that's coffee ?
know	how	Adam	*MOT:	how d(o) you know it's big air ?
know	how	Adam	*MOT:	how do you know it's not ?
know	how	Adam	*MOT:	how d(o) you know that's the kind he likes ?
know	how	Nina	*MOT:	and how did he know there was a child here ?
know	how	Nathaniel	*MOT:	how did you know we need to cut that ?
know	how	Nathaniel	*MOT:	that's how you know it's winter .
know	how	Nathaniel	*MOT:	how do you know that's Nathaniel ?
know	how	Naomi	*MOT:	how do you know that those are mother elephants( )
know	how	Naomi	*FAT	how did you know this one is a lady ?
know	how	Naomi	*FAT	how do you know that's a man (.) Nomi ?
say	why	Adam	*MOT:	why do you say he has wheels ?
say	why	Adam	*MOT:	why do you say he's a bad boy ?
tell	how	Adam	*MOT:	how can you tell it's going to stop ?
tell	how	Adam	*URS	how can you tell that you're strong ?
tell	how	Warren-Leubecker	*MOT	how can you tell that it's Peter rabbit .
tell	how	Nina	*MOT:	how can you tell he's a baker ?
tell	how	Nina	*MOT:	how can you tell that it's nighttime ?
tell	how	Nina	*MOT:	how can you tell it's nighttime ?
tell	how	Nina	*MOT:	how can you tell it's Nina ?

think	how	Adam	*MOT:	how do you think you have to do it ?
think	how	Adam	*URS	look inside and see how you think it works .
think	how	Adam	*URS	how d(o) you think they make dollars ?
think	how	Adam	*MOT	how d(o) you think they go on the truck so they won't tip over .
think	how	Nina	*MOT:	how do you think he got in our house ?
think	how	Nina	*MOT:	how do you think I should take it out ?
think	how	Nina	*MOT:	how do you think I could do that ?
think	how	Nina	*MOT:	how do you think they'll cook them ?
think	where	Adam	*MOT:	where d(o) you think you'd find all those
think	where	Warren-Leubecker	*FAT	where you think we can get one ?
think	where	Nina	*MOT:	where do you think he's gonna land ?
think	where	Nina	*MOT:	where do you think she'll go dancing ?
think	where	Nina	*MOT:	where do you think we can find some ?
think	where	Nina	*MOT:	oh (.) where do you think he's feeding the elephant?
think	where	Nina	*MOT:	where do you think he has a mustache ?
think	where	Nina	*MOT:	where do you think it finished up ?
think	where	Nina	*MOT:	where do you think one opens it ?
think	where	Nathaniel	*MOT:	where you think he's gonna [: going to] fall ?
think	where	Naomi	*MOT:	where do you think you see the moon Nomi ?
think	which	Sarah	*MOT:	which way do you think they're growing ?
think	why	Adam	*MOT	why d(o) you think she looks like a raccoon with the glasses on ?
think	why	Adam	*MOT:	why d(o) you think he has his hands on his tummy?
think	why	Adam	*MOT:	why d(o) you think he's going to be the eye ?
think	why	Adam	*MOT:	why d(o) you think they they have windows ?
think	why	Adam	*URS	why d(o) you think it makes the bubbles ?
think	why	Adam	*MOT:	why d(o) you think Mommy hugged you ?
think	why	Adam	*MOT:	why d(o) you think that's called a scooter ?
think	why	Adam	*MOT:	why d(o) you think it's lunch time ?
think	why	Adam	*MOT:	why d(o) you think it sticks ?
think	why	Adam	*MOT:	why d(o) you think it does ?
think	why	Adam	*MOT:	why d(o) you think they don't ?
think	why	Adam	*MOT:	why d(o) you think that's dark ?
think	why	Adam	*MOT	why d(o) you think that's the wrong place in Italy when you're not in Italy ?
think	why	Nina	*MOT:	why do you think she has to eat soft food ?
think	why	Nina	*MOT:	why do you think he has to wear a bib ?
think	why	Nina	*MOT:	why do you think he's going to fall down ?
think	why	Nina	*MOT:	why do you think he's giving him some water ?
think	why	Nina	*MOT:	why do you think he has an umbrella ?

think	why	Nina	*MOT:	why do you think she wears gloves ?
think	why	Nina	*MOT:	why do you think she's carrying it ?
think	why	Nina	*MOT:	why do you think she's meowing ?
think	why	Nina	*MOT:	why do you think he's crying ?
think	why	Nina	*MOT:	why do you think it's a squirrel ?
think	why	Nina	*MOT:	why do you think it's bad ?
think	why	Nina	*MOT:	why do you think he's mad ?
think	why	Sarah	*MOT	that's why I thought that thing was a (.) y(ou) know (.) just makebelieve .
think	why	Nathaniel	*MOT:	why do you think mama has to put the food on top of the tv ?
think	why	Nathaniel	*FAT	why do you think sharing is good ?
think	why	Nathaniel	*FAT	why do you think it would be nice for her to have a horse to ride ?
think	why	Nathaniel	*MOT:	why do you think the water got on the floor ?
think	why	Nathaniel	*MOT:	and why do you think she buys baby food ?
think	why	Nathaniel	*MOT:	why do you think you have wavy hair ?
think	why	Nathaniel	*MOT:	why do you think she did ?
think	why	Naomi	*MOT	no (.) Nomi (.) I meant why do you think he's jumping like that ?
think	why	Naomi	*MOT:	why do you think all those people are there ?
think	why	Naomi	*MOT:	why do you think he's auburn honey ?
think	why	Naomi	*MOT:	why do you think Kristen's crying ?
think	why	Naomi	*FAT	why do you think the man and the lady are kissing?

References

Abney, S. P. & Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, *20*, 233-250.

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford: Oxford University Press.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Almeida, D., & Poeppel, D. (submitted). Word–specific repetition effects revealed by MEG and the implications for lexical access. *Journal of Cognitive Neuroscience.*

Altmann, G. T. (1998). Ambiguity in sentence processing. *Trends in Cognitive Sciences*, *2*(4), 146-152.

Altmann, G. T. (2004). Language-Mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79-87.

Altmann, G. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247-264.

Altmann, G. T. M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347–386). New York, NY: Psychology Press.

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*, 191-238.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23-54.

Aoshima, S., Yoshida, M., & Phillips, C. (2009). Incremental processing of coreference and binding in japanese. *Syntax*, *12*(2), 93-134.

Badecker, W., & Straub K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 748-769.

Barss, A. (1986). *Chains and Anaphoric Dependencies.* Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Berwick, R. C. (1985). The acquisition of syntactic knowledge. Cambridge, MA: MIT Press.

Berwick, R. C. & Niyogi, P. (1996). Learning from triggers. *Linguistic Inquiry, 27,* 605–622.

Boland, J. E. (2004). Linking eye movements to sentence comprehension in reading and listening. In M. Carreiras & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond* (pp. 51–76). New York, NY: Psychology Press.

Boland, J. E. (2005). Visual arguments. *Cognition*, *95*(3), 237-274.

Boland, J. E. & Cutler, A. (1996). Interaction with autonomy: Multiple output models and the inadequacy of the great divide. *Cognition*, *58*(3), 309 - 320.

Bowerman, M. (1982). Reorganizational processes in lexical and syntactic development. In E. Wanner and L.R. Gleitman (eds.), *Language acquisition: The state of the art*, Cambridge University Press, New York.

Braine, M. (1971). On two types of models of the internalization of grammars. In D. Slobin (ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–86). New York, NY: Academic Press.

Bresnan, J., & Grimshaw, J. (1978). The syntax of free relatives in English. *Linguistic Inquiry, 9,* 331-391.

Britt, M. A. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language, 33*, 251–283.

Brown, R. 1973. A first language: The early stages. Cambridge, MA: Harvard University Press.

Büring, D. (2005). *Binding Theory*. Cambridge: Cambridge University Press.

Carlson, G. N. & Tanenhaus, M. K. (1989). *Linguistic structure in language processing* (illustrated ed.). Dordrecht ; Boston: Springer.

Cattell, R. (1978). Constraints on movement rules. *Language, 52,* 18-50.

Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, *52*(1), 144-169.

Chien, Y-C., & Lust, B. (2006). Chinese children's knowledge of the Binding Principles. In L. H. Tan, E. Bates & O. J. L. Tzeng (Eds.), *The handbook of East Asian linguistics vol.1* (pp. 23-38). Cambridge: Cambridge University Press.

Choi, Y. (2010). The role of information distribution and inhibitory control in preschooler's sentence processing. A poster presented at CUNY 2010, New York University.

Choi, Y. & Trueswell, J. C. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106*(1), 41 - 61.

Chomsky, N. (1981). *Lectures on Government and Binding.* Mouton de Gruyter: Berlin.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368 - 407.

Clark, R. (1992). The selection of syntactic knowledge. *Language Acquisition, 2,* 83–149.

Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, *49*(3), 317 - 334.

Collins, C. (1991). Why and how come. *MIT Working Papers in Linguistics*, *15*.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107.

Corder, S.P. (1967). The significance of learners' errors. *International Review of Applied Linguistics, 5*, 161–170.

Crain, S., & Fodor, J. D. (1985). How can grammars help parsers? In D. Dowty, L. Kartunnen & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives*. Cambridge: Cambridge University Press.

Crain, S., Ni, W. and Conway, L. 1994. Learning, parsing, and modularity. In C. Clifton, L. Frazier & K. Rayner (Eds.) *Perspectives on Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Crain, S., & Steedman, M. (1985). On not being led up the garden path; the use of context by the psychological parser. In D.R. Dowty, L. Karttunen, and A. Zwicky (Eds.), *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives* (pp. 320-358). Cambridge: Cambridge University Press.

Crain, S., & Thornton, R. (1998). *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, MA: MIT Press.

Crocker, M. W. (1996). *Computational psycholinguistics: An interdisciplinary approach to the study of language.* Dordrecht: Springer.

Crocker, M. W. (1999). Mechanisms for sentence processing. In S. Garrod and M. J. Pickering (Eds.), *Language processing* (pp. 191-232). New York, NY: Psychology Press.

Dąbrowska, E., Rowland, C., & Theakston, A. (2009) The acquisition of questions with long distance dependencies. *Cognitive Linguistics, 20*, 571-596.

Dahan, D., Magnuson, J. S., & Tanenhaus,M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42,* 317–367.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletein & Review, 12,* 453-459.

Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and

predictability effects on event-related potentials during reading. *Brain Research*, *1084*(1), 89-103.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-78.

de Villiers, J. G., & Roeper, T. (1995). Relative clauses are barriers to wh-movement for young children. *Journal of Child Language, 22*, 389–404.

de Villiers, J. G., & Roeper, T. (1996). Questions after stories: On supplying context and removing it as a variable. In D. McDaniel, H. Cairns, & C. McKee (Eds.), *Methods for assessing children's syntax* (pp. 163–188). Cambridge, MA: MIT Press.

de Villiers, J., Roeper, T., & Vainikka, A. (1990). The acquisition of long-distance rules. In L. Frazier & J. de Villiers (Eds.), *Language processing and language acquisition* (pp. 257-297). Dordrecht: Kluwer.

de Villiers, J., Roeper, T., Bland-Stewart, L., & Pearson, B. (2008). Answering hard questions: Wh-movement across dialects and disorder. *Applied Psycholinguistics*, *29*(01), 67-103.

Deane, P. D. (1991). Limits to attention: A cognitive theory of island phenomena. *Cognitive Linguistics*, *2*(1), 1–63.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117-21.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children

can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352-62.

Diamond, A. (2006). Bootstrapping conceptual deduction using physical connection: Rethinking frontal cortex. *Trends in Cognitive Sciences*, *10*(5), 212-8.

Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics, 12,* 97-141.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(32), 12203-8.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.

Federmeier, K. D. & Kutas, M. (1999). A rose by any other name: Long-Term memory structure and sentence processing. *Journal of Memory and Language*, *41*, 469-495.

Federmeier, K. D., Segal, J. B., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain*, *123*(12), 2552.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing* (pp. 73-96). Amsterdam: John Benjamins.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistic Compass, 1/1-2,* 71-83.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic*

Inquiry, *9*(3), 427-473.

Fodor, J. D. (1998). Parsing to learn. *Journal of Psycholinguistic Research*, *27*(3), 339-374.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research*, *23*(5), 407-434.

Fodor, J. D., & Sakas, W. (2004). Evaluating models of parameter setting. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *Proceedings of the 28th annual Boston University Conference on Language Development* (pp. 1-27). Somerville, MA: Cascadilla Press.

Fodor, J. D. & Sakas, W. G. (2005). The subset principle in syntax: Costs of compliance. *Journal of Linguistics*, *41*(3), 513-569.

Fox, D., & Nissenbaum, J. (2004). Condition A and scope reconstruction. *Linguistic Inquiry, 35,* 475-485.

Frazier, L. (1987). Syntactic processing: Evidence from Dutch. *Natural Language and Linguistic Theory, 5*, 519-559.

Frazier, L., & Clifton, C. (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes, 4*(2), 93-126.

Frazier, L., & Clifton, C., Jr. (1998). Sentence reanalysis, and visibility. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 143-176). Dordrecht: Kluwer.

Frazier, L. & Flores d'Arcais, G. B. (1989). Filler driven parsing: A study of gap filling in dutch. *Journal of Memory and Language*, *28*(3), 331-344.

Frazier, L. & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291 - 325.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178-210.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78-84.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58-93.

Garnsey, S. M., Tanenhaus, M. K., & Chapman, R. M. (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 51-60.

Gathercole, V. G. M., Hoff, E. (2007). Input and the Acquisition of Language: Three Questions. In E. Hoff, & M. Shatz (Eds.), *Blackwell Handbook of Language Development* (pp. 107-127). Oxford: Blackwell Publishers.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177-190.

Geis, M. (1970). Adverbial subordinate clauses in English. PhD dissertation, MIT.

Gerken, L. A. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), 67-74.

Gibson, E. A. F. (1991). *A computational theory of human linguistic processing: Memory limitations and processing breakdown.* Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition,*

68, 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium.* (pp. 95-126). Cambridge, MA: MIT Press.

Gibson, E., & Pearlmutter, N. J. (1998). Constraints on sentence comprehension. *Trends in Cognitive Science, 2,* 262-268.

Gibson, E. & Warren, T. (2004). Reading-Time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax*, *7*(1), 55-78.

Gibson, E. & Wexler, K. (1994). Triggers. Linguistic Inquiry, *25,* 407–454.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*, 447–474.

Gómez, R. L. (2002). Variability and Detection of Invariant Structure. *Psychological Science*, *13*(5), 431-436.

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load interference in syntactic processing. *Psychological Science*, *13*(5), 425.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 1-40.

Gualmini, A. (2004). Some knowledge children don't lack. *Linguistics, 42*(5), 957-982.

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The Question-Answer requirement for scope assignment. *Natural Language Semantics, 16,* 205-237.

Haarmann, H. J. & Cameron, K. A. (2005). Active maintenance of sentence meaning in working memory: Evidence from EEG coherences. *International Journal of*

Psychophysiology : Official Journal of the International Organization of Psychophysiology, *57*(2), 115-28.

Haegeman, L. (2009). The movement analysis of temporal adverbial clauses. *English Language and Linguistics, 13,* 385-408.

Hagoort, P. (2003). How the brain solves the binding problem for language: A neurocomputational model of syntactic processing. *Neuroimage*, *20 Suppl 1*, S18-29.

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research*, *32*(2), 101-123.

Harada, S.-I. (1977). Nihongo-ni 'Henkei'-wa Hituyoo-da ["Transformation is necessary in Japanese"]. *Gengo,* 6, 11-12.

Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*, *30*(4), 1383-400.

Heim, I. (1982). The semantics of definite and indefinite noun phrases. Unpublished doctoral dissertation, MIT.

Henderson, J. M., & Ferreira, F. (2004). *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. Psychology Press.

Heycock, C. (1995). Asymmetries in reconstruction. *Linguistic Inquiry, 26,* 547-570

Hofmeister, P. & Sag, I. A. (2010). Cognitive constraints and island effects. *Language*, *86*(2), 366-415.

Höhle, T. N. (2000). The W-…W- construction: Appositive or scope indicating? In Lutz, U., Müller, G. and von Stechow, A. (Eds), *Wh-scope marking* (pp. 249-270), Amsterdam: John Benjamins.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking* (Version 1.5). [Computer Software]. West Lafayette: Purdue University.

Huang, C-T James. (1993). Reconstruction and the structure of VP: Some theoretical consequences. *Linguistic Inquiry, 24,* 103-138.

Huang, C.-T. J., & Liu, C.-S. L. (2001). Logophoricity, attitudes, and *ziji* at the interface. In *Syntax and semantics: Long distance reflexives.*

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*(2), 151-195.

Huettig, F. & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-32.

Hurewitz, F., Brown-Schmidt, S., Thorpe, K., Gleitman, L., & Trueswell, J. (2000). One frog, two frog, red frog, blue frog: Factors aVecting children's syntactic choices in production and comprehension. *Journal of Psycholinguistic Research, 29*, 597–626.

Inoue, A. & Fodor, J. D. (1995). Information-Paced parsing of japanese. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing.* (pp. 9-63). Hillsdale, NJ: Lawrence Erlbaum.

Jakubowicz, C., & Strik, N. (2008). Scope-marking strategies in the acquisition of long distance wh-questions in French and Dutch. *Language and Speech, 51,* 101-132.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 98,* 122–149.

Just, M. A., Carpenter, P., & Wooley, J. D. (1982). Paradigms and processes in reading

comprehension. *Journal of Experimental Psychology: General, 111*, 228-238.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159-.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistic Compass*, *2*(4), 647-670.

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.

Kamide, Y. & Mitchell, D. C. (1999). Incremental pre-head attachment in japanese parsing. *Language and Cognitive Processes*, *14*(5), 631-662.

Kazanina, N., Lau, E. F., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*(3), 384-409.

Keyser, S. J. & Roeper, T. (1984). On the middle and ergative constructions in english. *Linguistic Inquiry*, *15*(3), 381-416.

Kidd, E., & Bavin, E. L. (2007). Lexical and referential influences on on-line spoken language comprehension: A comparison of adults and primary-school-age children. *First Language, 27,* 29-52.

Kim, M., Landau, B., & Phillips, C. (1999). Cross-linguistic differences in children's syntax for locative verbs. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Annual Boston University Conference on Language Development* (pp. 337-348). Somerville, MA: Cascadilla Press.

Kimball, J. (1975). Predictive analysis and over-the-top parsing. *Syntax and Semantics*, *4*,

155-179.

King, J. & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, *30*(5), 580-602.

Kluender, R. (1998). On the distinction between strong and weak islands: A processing perspective. In P. Culicover & L. McNally (Eds.), *Syntax and semantics: The limits of syntax* (pp. 241–280). New York: Academic Press.

Kluender, R. (2004). Are subject islands subject to a processing account. In V. Chand, A. Kelleher, A. Rodriguez, & B. Schmeiser (Eds.), *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL 23)* (pp. 475–499). Somerville, MA: Cascadilla Press.

Kluender, R., & Kutas, M. (1993). Subjacency as a Processing Phenomenon. *Language and Cognitive Processes, 8*, 573-633.

Koh, S. (1997). The resolution of the dative NP ambiguity in korean. *Journal of Psycholinguistic Research*, *26*(2), 265-273.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627-45.

Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146,* 23-49.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463-470.

Larson, R. (1987). Missing prepositions and the analysis of English free relative clauses. *Linguistic Inquiry, 18*, 239–266.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:

(De)constructing the N400. *Nature Reviews. Neuroscience*, *9*(12), 920-33.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language*, *98*(1), 74-88.

Leddon, E. M., & Lidz, J. L. (2006). Reconstruction effects in child language. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *BUCLD 30 Proceedings* (pp. 328-339). Somerville, MA: Cascadilla Press.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research, 33*(1), 51-73.

Levin, B. & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax-lexical semantics interface.* Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-Based syntactic comprehension. *Cognition*, *106*(3), 1126-77.

Lewis, R. L. & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375-419.

Lidz, J., & Baier, R. (2010). Predictive parsing impedes word learning in 19-month-olds. A paper presented at LSA 2010, Baltimore.

Love, T. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: A real-time investigation. *Journal of Psycholinguistic Research, 36,* 191-206.

Lutz, U., Müller, G., & von Stechow, A. (2000). Introduction. In Lutz, U., Müller, G. and von Stechow, A. (Eds), Wh-*scope marking* (pp. 1-16), Amsterdam: John Benjamins.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676-703.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Erlbaum.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17,* 457-472.

Manzini, M. R. & Wexler, K. (1987). Parameters, binding theory, and learnability. *Linguistic Inquiry, 18,* 413–444.

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature*, *244*(5417), 522-523.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372-380.

Mauner, G. & Koenig, J. P. (2000). Linguistic vs. Conceptual sources of implicit agents in sentence comprehension. *Journal of Memory and Language*, *43*(1), 110-134.

Mazuka, R., Jincho, N., & Onishi, H. (2009). Development of executive control and language processing. *Language and Linguistics Compass, 3/1,* 59-89.

McDaniel, D. (1989). Partial and multiple *WH*-movement. *Natural Language and Linguistic Theory, 7*, 565-604.

McDaniel, D., Chiu, B., & Maxfield, T. L. (1995). Parameters for *wh*-movement types: Evidence from child English. *Natural Language and Linguistic Theory, 13*, 709-753.

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67-91.

McElree, B., & Griffith, T. (1998). Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Learning, Memory*, *24*, 432-460.

McKinnon, R. & Osterhout, L. (1996). Constraints on movement phenomena in sentence processing: Evidence from event-related brain potentials. *Language and Cognitive Processes*, *11*(5), 495-524.

Meroni, L., & Crain, S. (2003). On not being led down the kindergarten path. In *Proceedings of the 27th Boston University Conference on Language Development* (pp. 531–544). Somerville, MA: Cascadilla Press.

Miller, G. A. (1956). The magical number seven, plus or minus two. *Psychological review*, *63*, 81-97.

Miller, K., & C. Schmitt (2006). The effect of variable input on comprehension: Evidence from Spanish. In D. Bamman, T. Magnitskaia & C. Zaller (Eds.), *BUCLD 30* (pp. 365-376). Somerville, MA: Cascadilla Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of memory and language*, *59*(4), 475-494.

Musolino, J., & Lidz, J. (2006). Why children aren't universally successful with quantification. *Linguistics, 44,* 817–852.

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long-distance scrambling. *Journal of Psycholinguistic Research*, *31*(5), 531-71.

Nation, K., Marshall, C. M., & Altmann, G. T. M. (2003). Investigating individual differences in children's real-time sentence comprehension using language-mediated eye movements. *Journal of Experimental Child Psychology*, *86*(4), 314-329.

Nicol, J. L. (1993). Reconsidering reactivation. In G. Altmann & R. Schillcock (Eds.),

Cognitive models of speech processing: The second sperlonga meeting. (pp. 321-47). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nicol, J. & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research*, *18*(1), 5-19.

Nieuwland, M. S. & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science : A Journal of the American Psychological Society / APS*, *19*(12), 1213-8.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience, 5*, 263–281.

Omaki & Schulz. (in press). Filler-gap dependencies and island constraints in second language sentence processing. *Studies in Second Language Acquisition.*

Osterhout, L., Holcomb, P.J., Swinney, D.A., (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning ,Memory, and Cognition, 20*, 786–803.

Otten, M. & Van Berkum, J. J. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research*, *1153*, 166-77.

Pearl, L, and J. Lidz. (2009). When Domain General Learning Succeeds and When it Fails. *Language Learning and Development*, 5, 235-265.

Perlmutter, D. M. (1978). Impersonal passives and the Unaccusative Hypothesis. *Proceedings of the 4th Annual Meeting of the Berkeley Linguistics Society* (pp. 157-

189). UC Berkeley.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795-823.

Phillips, C. (2009). Real-time structure building and retreat from over-generation. An invited paper presented at GLOW Asia 2009, Hyderabad.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, *22*(3), 407-428.

Pickering, M., & Barry, G. (1991). Sentence processing without empty categories. *Language and Cognitive Processes*, *6*(3), 229–259.

Pickering, M. J. & Traxler, M. J. (2001). Strategies for processing unbounded dependencies: Lexical information and verb-argument assignment,. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *27*(6), 1401-1410.

Pickering, M. J. & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469-503.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts,. *Journal of Memory and Language*, *43*(3), 447 - 475.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1989). *Learnability and cognition : The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pollard, C., & Sag, I. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry, 23,* 261–303.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance.* Chicago, IL: University of Chicago Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372-422.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *The Quarterly Journal of Experimental Psychology, 39A,* 657-673.

Rayner, K. & Pollatsek, A. (2006). Eye-Movement control in reading. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics (second edition)*. (pp. 613-657). New York, NY: Academic Press.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z reader model of eye-movement control in reading: Comparisons to other models. *The Behavioral and Brain Sciences*, *26*(4), 445-76; discussion 477-526.

Resnik, P. (1992). Left-Corner parsing and psychological plausibility. In *Proceedings of the 14th conference on computational linguistics.*

Riemsdijk, Henk van, 1982. Correspondence effects and the empty category principle. Tilburg Papers in Language and Literature, vol. 12. University of Tilburg

Riemsdijk, Henk van. (2006). Free relatives. In M. Everaert & H. van. Riemsdijk (Eds.), *The Blackwell Companion to Syntax, vol. II* (pp. 338-382). Oxford: Blackwell.

Rizzi, L. (1982). *Issues in Italian Syntax*. Dordrecht: Foris.

Roberts, L., Marinis, T., Felser, C., & Clahsen, H. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research*, *36*(2), 175-88.

Roeper, T., & de Villiers, J. (1992). Ordered decisions in the acquisition of wh-questions.

In H. Goodluck, J. Weissenborm, & T. Roeper (Eds.), *Theoretical issues in language development*. Hilsdale, NJ: Erlbaum.

Ross, J. R. (1967). *Constraints on variables in syntax.* Unpublished doctoral dissertation, MIT, Cambridge.

Rowland, C. F., & Pine, J. M. (2000). Subject–auxiliary inversion mismatches and whquestion acquisition:'what children do know?'. *Journal of Child Language*, *27*, 157–181.

Runner, J.T., Sussman, R.S. & Tanenhaus, M.K. (2006). Assigning referents to reflexives and pronouns in picture noun phrases: Experimental tests of Binding Theory. *Cognitive Science, 30*, 1-49.

Sachs, J. (1983). Talking about the there and then: The emergence of displaced reference in parent–child discourse. In K. E. Nelson (Ed.), *Children's language, Vol. 4,* Hillsdale, NJ: Lawrence Erlbaum Associates.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57*, 1030–1033.

Saito, M. (1985). Some asymmetries in Japanese and their theoretical consequences. *Unpublished doctoral dissertation, Massachusetts Institute of Technology*.

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America, 57*, 1030–1033.

Schulz, B. (2006). *Wh*-scope marking in English interlanguage grammars: Transfer and processing effects on the second language acquisition of complex *wh*-questions. Unpublished PhD dissertation, University of Hawaii.

Sekerina, I., Fernandez, E. M., & Clahsen, H. (Eds.) (2008). *Developmental*

Psycholinguistics: On-Line Methods in Children's Language Processing. Amsterdam: John Benjamins.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences, 7*, 489-493.

Shieber, S. & Johnson, M. (1993). Variations on incremental interpretation. *Journal of Psycholinguistic Research*, *22*(2), 287-318.

Simons, M. (2007). Observations on embedding verbs, evidentiality, and presupposition. *Lingua, 117,* 1034-1056.

Snedeker, J. & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238-99.

Spivey-Knowlton, M. & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55*(3), 227-267.

Sportiche, D. (2006). Reconstruction, binding, and scope. In M. Everaert & H. van. Riemsdijk (Eds.), *The Blackwell Companion to Syntax*. Oxford: Blackwell.

Stabler, E. P. (1994). The finite connectivity of linguistic structure. In L. Frazier, C. Clifton, & K. Rayner (Eds.), *Perspectives on sentence processing.* (pp. 303-36). Hillsdale, NJ: Lawrence Erlbaum Associates.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 550-69.

Staub, A. & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...Or. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*(2), 425-36.

Staub, A., Grant, M., Clifton, C., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(3), 806-14.

Staub, A. & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The oxford handbook of psycholinguistics.* (pp. 327-42). Oxford: Oxford University Press.

Stepanov, Arthur, 2000. Wh-scope marking in Slavic. *Studia Linguistica, 54,* 1–40.

Stepanov, A., & Stateva, P. (2006). Successive cyclicity as residual wh-scope marking. *Lingua, 116,* 2107-2153.

Stowe, L. (1986). Evidence for on-line gap location. *Language and Cognitive Processes, 1*, 227-245.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542-562.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477-88.

Su, Y-C. (2003). Chinese children's *ziji*. Paper presented at the 28th Boston University Conference on Language Development.

Suppes, P. (1974). The semantics of children's language. *American Psychologist, 29,* 103–114.

Sussman, R. S., & Sedivy, J. C. (2003). The time-course of processing syntactic dependencies: Evidence from eye movements. *Language and Cognitive Processes, 18*(2), 143-163.

Szabolcsi, A. & den Dikken, M. (2003). Islands. In L. Cheng & R. Sybesma (Eds.), *The*

second glot international state-of-the article book. (pp. 213-240). Berlin: Mouton de Gruyter.

Takano, Y. (1995). Predicate Fronting and Internal Subjects. *Linguistic Inquiry, 26*, 327–340.

Tanenhaus, M. K. (2007). Spoken language comprehension: Insights from eye movements. In G. Gaskell (Ed.), *The oxford handbook of psycholinguistics.* (pp. 309-26). Oxford: Oxford University Press.

Tanenhaus, M. K. & Carlson, G. N. (1989). Lexical structure and language comprehension. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process.* (pp. 529-561). Cambridge, MA: MIT Press.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information is spoken-language comprehension. *Science, 268,* 1632–1634.

Tanenhaus, M. K., & Trueswell, J. C. (1995). Sentence comprehension. In J. Miller & P. Eimas (Eds.), *Speech, language, and communication* (Vol. 11, pp. 217-262). San Diego, CA: Academic Press.

Tanenhaus, M. K., & Trueswell, J. C. (2005). Eye movements as a tool for bridging the langauge-as-product and language-as-action traditions. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *Approaches to studying world-situated language use* (pp. 3-38). Cambridge, MA: MIT Press.

Tenenbaum, J. & Griffiths, T. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629-640.

Thompson, S. A., & Mulac, A. (1991). A quantitative perspective on the

grammaticalization of epistemic parentheticals in English. In B. Heine & E. Closs Traugott (eds.), Approaches to Grammaticalization, vol. 2 (pp. 313-329). Amsterdam: John Benjamins.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3(1),* 1-42.

Thornton, R. (1990). Adventures in long-distance moving: the acquisition of complex wh-questions. Doctoral dissertation, University of Connecticut, Storrs.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Child Language Acquisition*. Cambridge, MA: Harvard University Press.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*, 454-475.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, *47*(4), 530-547.

Trueswell, J. C. (2008). Using eye movements as a developmental measure within psycholinguistics. In I. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing* (pp. 73-96). Amsterdam: John Benjamins.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition, 73*(2), 89-134.

Valian, V. (1990). Logical and psychological constraints on the acquisition of syntax. In L. Frazier & J. de Villiers (Eds.), *Language processing and language acquisition* (pp. 119-146). Dordrecht: Kluwer Academic Publishers.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from erps and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(3), 443-67.

van Gompel, R. P. & Liversedge, S. P. (2003). The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*(1), 128-39.

Wagers, M. W. & Phillips, C. (2009). Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics*, *45*(2), 395-433.

Warren-Leubecker, A. (1982). *Sex differences in speech to children.* Unpublished doctoral dissertation. Georgia Institute of Technology.

Weighall, A. R. (2008). The kindergarten path effect revisited: Children's use of context in processing structural ambiguities. *Journal of Experimental Psychology, 99,* 75–95.

Wexler, K. & Culicover, P. W. (1980). Formal principles of language acquisition. Cambridge, MA: MIT Press.

Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain and Language*, *108*(1), 40 - 55.

Xu, F. & Tenenbaum, J. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245-272.

Yang, C. (2002). *Knowledge and learning in natural language*. New York: Oxford University Press.

Yang, C. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Science, 8*(10), 451-456.

Yee, E., & Sedivy, J. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1-14.

Yoshida, M. (2006). *Constraints and mechanisms in long-distance dependency formation.* Unpublished doctoral dissertation, University of Maryland, College Park.

Yoshida, M., Dickey, M. W., & Sturt, P. (in press). Predictive processing of syntactic structure: Sluicing and ellipsis in real-time sentence processing. *Language and Cognitive Processes.*