ABSTRACT

Title of Thesis: THE TEMPORAL DIMENSION OF LINGUISTIC PREDICTION

Wing Yee Chow

Doctor of Philosophy 2013

Directed by: Professor Colin Phillips

Department of Linguistics

This thesis explores how predictions about upcoming language inputs are computed during real-time language comprehension. Previous research has demonstrated humans' ability to use rich contextual information to compute linguistic prediction during real-time language comprehension, and it has been widely assumed that contextual information can impact linguistic prediction as soon as it arises in the input. This thesis questions this key assumption and explores how linguistic predictions develop in real-time. I provide event-related potential (ERP) and reading eye-movement (EM) evidence from studies in Mandarin Chinese and English that even prominent and unambiguous information about preverbal arguments' structural roles cannot immediately impact comprehenders' verb prediction. I demonstrate that the N400, an ERP response that is modulated by a word's predictability, becomes sensitive to argument role-reversals only when the time interval for prediction is widened. Further, I provide initial evidence that different sources of contextual information, namely, information about preverbal arguments' lexical identity vs. their structural roles, may impact linguistic prediction on different time scales. I put forth a research framework that aims to characterize the mental computations underlying linguistic prediction along a temporal dimension.

THE TEMPORAL DIMENSION OF LINGUISTIC PREDICTION

by

Wing Yee Chow

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
2013

Advisory Committee:

Professor Colin Phillips, Chair
Professor Ellen Lau
Professor Jeffrey Lidz
Professor Yi-Ting Huang
Professor Nan Jiang

© Copyright by
Wing Yee Chow
2013

Dedication

To my parents

Acknowledgements

I have been extremely fortunate to have the privileges to do what I do. This dissertation is a product of all the encouragements and support (intellectual as well as emotional) that I have been blessed with, and I have many people to thank for that.

First and foremost I would like to express my deepest gratitude to Colin Phillips. It's been a real blessing to have Colin as my mentor, who continually pushes me to go beyond doing well-designed experiments to think about questions that I thought I had no chance of ever answering. Our weekly meetings have provided the best training for my perseverance – I consider it one of my greatest triumphs in grad school to have finally convinced Colin (after months of trying!) of the idea that turned into this dissertation. I am extremely grateful for all the support Colin has given me – I cannot even begin to describe how much it meant to me when he took time off his holidays in the UK to support me and my career in person. I really hope one day I can become as good a mentor to my students as Colin has been to me.

Ellen Lau has been a great influence to my growth as a researcher. Ellen is like a big sister to me and has been my role model ever since I visited Maryland as a prospective student when she was still a student. I am particularly grateful to the Department for bringing her back. Discussions with Ellen have continued to make research ever more exciting and fun. They have also been a great source of inspirations – many ideas in this dissertation can be traced back to the discussions we have had in the past couple of years. I really wish our time at Maryland had overlapped more.

I would like to thank Suiping Wang, my mentor and host at South China Normal University. For two consecutive summers she welcomed me into her lab and provided me with all the support and resources that I could ask for. Much of my dissertation research would not have been realized without her support.

I'd like to thank Jeff Lidz, Yi-Ting Huang, and Nan Jiang for serving on my dissertation committee. They have been engaged in various stages and aspects of my graduate career, and Jeff and Yi-Ting deserved special credits for always being there for my lab meeting presentations and for pushing me to articulate my ideas more clearly.

A big thank you goes to Naomi Feldman for her enthusiasm and tirelessness in introducing me to computational psycholinguistics. It's with her encouragement that I managed to do what I would have been too intimated to even try otherwise. I am very grateful for the time she took to mentor me despite the little overlap between our research areas.

Thank you to Norbert Hornstein, for his generosity and for showing me how wonderful it is to stay curious and keep learning about new things every day; to Howard Lasnik, for his stories and for being the teacher that I aspire to become; to Bill Idsardi, for his insights and for his generous help during my computer crisis; to Alexander Williams, for saving me from a potential rice shortage and for taking me under his wing at my first NACCL conference.

Thank you to my bigger brothers and sisters who have taken care of me both inside and outside of Maryland: Brian Dillon, Ming Xiang, Akira Omaki, Masaya Yoshida, Matt Wagers, Robert Fiorentino, So-One Hwang, Phil Monahan and Eri Takahashi. I am grateful to Brian, who took me under his wing when I had just started graduate school and for his continued generosity in sharing his knowledge through the years. Thank you to Ming, who introduced me to Suiping and encouraged me to do research abroad. Thank you to Akira, Masaya, Matt, and Rob for the wisdom they have shared with me and for encouraging me to keep pursuing what I am interested in. Thank you to So-One, for her friendship and for showing me the power of determination and devotion. Thank you to Phil and Eri, whose help and support have been invaluable in my preparation for the next step in my career.

Thank you to my classmates: Shevaun Lewis, Brad Larson, Ewan Dunbar, Dave Kush, Terje Lohndal, Michaël Gagnon, and Chris LaTerza, and our recently adopted classmate Shannon Barrios, all of whom I have learned a lot from. A special thank goes to Shevaun, who has been an invaluable friend and an impeccable colleague, for dragging me out of College Park on many occasions and for inspiring me to be a better cook. Thank you to Shannon, who has been a good friend as well as a big sister to me. Thank you to Ewan, for generously sharing his knowledge about Statistics and many other things. Thank you to Brad, who is one of the most easy-going people I know, and Dave, who was writing his dissertation at the same time as me, for showing me it is possible to do good research and other interesting things at the same time.

Sol Lago has been a constant source of encouragements, and I cannot imagine what grad school would have been like without her companionship through the ups and downs. Our countless marathon discussions and debates on various topics have continued to fuel my enthusiasm for research. I am extremely lucky to have known Naho Orita, my perfect roommate, even though we were two cohorts apart. Together Sol and Naho have made living in College Park a colorful, multilingual and multicultural experience. My heartfelt thanks to Sun Yuxia and her lovely family, who have welcomed me into their home and have kept me company when I was away from home for the past five Chinese New Years. Very special thanks to Carmen Gallar, who I've had the great pleasure to share one semester with, and whose appearance at my door I will never forget.

A great team of dedicated people at SCNU, Beijing Normal University as well as at UMD have helped to carry out the research reported here. Brian Dillon kick-started my research when he included my experimental materials in his study as fillers. The 'pilot' study became my first paper and the rest is history. Li Nan has taught me most of the things I know about eye-tracking and has made sure I stayed sane through the days of

intensive data collection. I am extremely grateful for all the students at SCNU who volunteered to help: Li Jie, Sun Qiong, Chen Lan, Zhang Wenjia, Yu Luodi, Guan Shaowei, Cui Yanling, Huang Jian, Deng Zhizhou, Sun Dongxia, Zhu Xiayan, Feng Gangyi and many more. Cybelle Smith, Glynis MacMillan, Shefali Shah and Erin Mahoney all deserve credits for their help in making the English part of this research happen as smoothly as it did.

Many students and faculty, both within the Linguistics department and in the Maryland Language Science community, have also enriched my graduate experience, including Valentine Hacquard, Tonia Bleam, Andrea Zukowski, Peggy Antonisse, Jordan Boyd-Graber, Hal Daumé III, Sunyoung Lee-Ellis, Valerie van Mulukom, Shota Momma, Dan Parker, Lyn Tieu, Mahayana Godoy and many more. Thank you to Kathi Faulkingham, Kim Kwok and Csilla Kajtar for their help with all the important things that I am still learning to keep track of. I would also like to acknowledge the Graduate School. During the writing of the dissertation, I was supported by a Wylie Dissertation Fellowship.

This research has benefitted greatly from the discussions I have had with Patrick Sturt, Roger Levy, Peter Gordon, Adrian Staub, Kara Federmeier, Gina Kuperberg, Jakub Szewczyk, Wonil Choi, Nathaniel Smith, Jinmian Yang and many others. I'd also like to thank the all the speakers and the organizing committee for MayFest 2013 for the wonderful prediction themed weekend.

My interests have grown from the seeds that were planted during my undergraduate years at McGill. I am grateful to Karsten Steinhauer, Kris Onishi, Fred Genesee, Erin White, Nino Grillo and Andrea Santi for their guidance and support. They had given me the best training to prepare me for grad school.

Although I dreamt of studying abroad when I was small, I never thought I'd have the courage to pursue a research career thousands of miles away from home. Thank you, Ivan, for always being there for me. Your love and support meant the world to me. I cannot imagine doing this without you. My big brother has taught me to pursue what I love and not just what I'm good at. My mom and dad have given me the best preparation for life wherever it takes me. Their optimism and love of learning are infectious, and I am eternally grateful for their unconditional support to pursue my dream.

Table of Contents

Dedication ..... **ii**
Acknowledgements ..... **iii**
Table of Contents ..... **vi**
List of Tables ..... **viii**
List of Figures ..... **x**
1 Introduction **1**
1.1 Overview 1
1.2 What is prediction? 3
1.3 Evidence for prediction in language comprehension 5
1.4 Representations 6
1.5 The temporal dimension of prediction 8
1.6 Outline of the dissertation 10
2 A surprising case of prediction failure **14**
2.1 Introduction 14
2.2 When do semantic anomalies elicit a P600 effect? 18
2.3 When do semantic anomalies fail to elicit an N400 effect? 22
2.4 Experiments 1 and 2 28
2.4.1 Methods 34
2.4.2 Results 40
2.5 Discussion 50
2.6 Conclusion 60
3 Prediction as a race against time **61**
3.1 Introduction 61
3.2 Why is the N400's insensitivity to argument role-reversals surprising? 61
3.3 Experiment 3: Can offline predictability measures capture online predictions? 63
3.3.1 Methods 65
3.3.2 Results 70
3.3.3 Discussion 76
3.4 Experiment 4: How do linguistic predictions develop over time? 78
3.4.1 Methods 80
3.4.2 Results 82
3.4.3 Discussion 88
3.5 Experiment 5: Prediction vs. Integration over the extended time interval 90
3.5.1 Methods 91
3.5.2 Results 92
3.5.3 Discussion 98
3.6 General Discussion 100
4 Predictive computations on different time scales **108**
4.1 Introduction 108

4.2 Garnsey, Tanenhaus & Chapman (1989) 109
4.3 Reinterpretation of Garnsey et al.'s (1989) findings 109
4.4 Experiment 6: Impact of argument role vs. identity on verb prediction 111
4.4.1 Methods....113
4.4.2 Results....117
4.4.3 Discussion....124
4.5 General Discussion....129

5 Aligning the eyes and the brain 137

5.1 Introduction 137
5.2 Basic considerations for studying EMs during reading Chinese....140
5.3 Comparison between standard EM and ERP paradigms....141
5.4 Experiment 7: A striking mis-alignment 143
5.4.1 Methods....144
5.4.2 Results....150
5.4.3 Discussion....153
5.5 Experiment 8: Alignment upon further considerations 155
5.5.1 Methods....156
5.5.2 Results....158
5.5.3 Discussion....165
5.6 Experiment 9: Contribution of parafoveal preview in reading....170
5.6.1 Methods....173
5.6.2 Results....176
5.6.3 Discussion....179
5.7 General Discussion....182

6 Conclusion 190

6.1 Overview 190
6.2 Synthesis of empirical findings 190
6.3 Theoretical implications 192
6.4 Recommendations for future research....194
6.5 General conclusion 201

Appendix (Experimental materials)....202

Experiment 1 202
Experiment 2 207
Experiment 3 209
Experiment 4 215
Experiment 5 225
Experiment 6 230
Experiment 7 234
Experiment 8 237
Experiment 9 240

References....243

List of Tables

Table 2-1. Experimental conditions and example sentences in Experiment 1. The target word is underlined. 31

Table 2-2. Experimental conditions and example sentences in Experiment 2. The target word is underlined. 32

Table 2-3. Target response and accuracy on acceptability judgment task. 41

Table 2-4. Repeated measures ANOVA F values at the target word in Experiment 1 47

Table 2-5. Repeated measures ANOVA F values at the target word in Experiment 2 49

Table 2-6. Summary of some of the factors that have been proposed to account for the Semantic P600 phenomenon. 59

Table 3-1. Experimental conditions and sample materials in Experiment 3. 67

Table 3-2. Target response and accuracy on plausibility judgment task in Experiment 3. 70

Table 3-3. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 3 72

Table 3-4. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 3. 75

Table 3-5. Experimental conditions and sample materials in Experiment 4. 81

Table 3-6. Target response and accuracy on plausibility judgment task in Experiment 4. 82

Table 3-7. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 4 84

Table 3-8. ANOVA F-values and ROI analysis t-values at the target verb for the short- and long-distance comparisons in Experiment 4. 87

Table 3-9. Experimental conditions and sample materials in Experiment 5. 91

Table 3-10. Target response and accuracy on plausibility judgment task in Experiment 5. 92

Table 3-11. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 5 94

Table 3-12. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 5. 97

Table 4-1. A comparison between the methods in Experiments 2-4 and Garnsey et al. (1989) 111

Table 4-2. Experimental conditions and sample materials in Experiment 6. 114

Table 4-3. Average plausibility judgment in Experiment 6. 118

Table 4-4. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 6.....119

Table 4-5. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 6. 123

Table 5-1. Experimental conditions and regions of interest in sample materials in Experiment 7.....146

Table 5-2. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 7. Standard error of the mean is in parentheses. 152

Table 5-3. Linear mixed-effect and logistic regression model estimates in Experiment 7.....152

Table 5-4. Experimental conditions and regions of interest in sample materials in Experiment 8.....157

Table 5-5. Target response and accuracy on plausibility judgment task in Experiment 8. 158

Table 5-6. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 8 (original region definition). Standard error of the mean is in parentheses. 161

Table 5-7. Linear mixed-effect and logistic regression model estimates in Experiment 8 (original region definition).....161

Table 5-8. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 8 (alternate region definition). Standard error of the mean is in parentheses. 164

Table 5-9. Linear mixed-effect and logistic regression model estimates in Experiment 8 (alternate region definition). 164

Table 5-10. Experimental conditions and regions of interest in sample materials in Experiment 9. The pseudoword in the invalid preview condition is presented on the left of the target. 175

Table 5-11. Target response and accuracy on plausibility judgment task in Experiment 9.....176

Table 5-12. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 9. Standard error of the mean is in parentheses. 178

Table 5-13. Linear mixed-effect and logistic regression model estimates in Experiment 9.....178

List of Figures

Figure 2-1. Grand average ERPs in six regions of interests in Experiment 1. 42

Figure 2-2. Grand average ERPs in midline posterior region in (A) 300-500 ms, (B) 600-800 ms, and (C) 800-1000 ms time intervals in Experiment 1.....46

Figure 2-3. Grand average ERPs in six regions of interest in Experiment 2. 49

Figure 3-1. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 3. 71

Figure 3-2. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the low-predictability sentences in Experiment 3. 73

Figure 3-3. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the high-predictability sentences in Experiment 3. 74

Figure 3-4. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 3. 74

Figure 3-5. Illustration of the distance manipulation in Experiment 4. 79

Figure 3-6. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 4. 83

Figure 3-7. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the short-distance sentences in Experiment 4. ...85

Figure 3-8. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the long-distance sentences in Experiment 4. 86

Figure 3-9. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 4. 86

Figure 3-10. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 5. 93

Figure 3-11. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the low-predictability sentences in Experiment 5. 95

Figure 3-12. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the high-predictability sentences in Experiment 5. 96

Figure 3-13. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 5. 96

Figure 3-14. Summary of results in the experimental conditions in Experiments 3-5. *Left and center*: Grand average ERP waveforms at CPZ. *Right*: Topographic map of the effect of argument role-reversals (role-reversed–canonical) in the 350-450ms time interval. 102

Figure 3-15. Summary of results in the control conditions in Experiments 3-5. *Left:* Grand average ERP waveforms at CPZ. *Center and right*: Topographic map of effect of cloze probability (unexpected – expected) in 350-450ms and 600-800ms time intervals.....104

Figure 3-16. An experimental paradigm for studying the time course of prediction. 105

Figure 4-1. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 6.....119

Figure 4-2. Grand average ERPs at the target verb in expected (black) and unexpected (red) conditions in the argument role-reversal sentences in Experiment 6.....121

Figure 4-3. Grand average ERPs at the target verb in expected (black) and unexpected (red) conditions in the argument substitution sentences in Experiment 6.....121

Figure 4-4. Topographic distribution of ERP effects in 300-500 ms (top) and 700-900ms (bottom) across experimental and control comparisons in Experiment 6.....122

Figure 4-5. Average ERP amplitude in each experimental condition in Experiment 6. Left: 300-500ms in central midline region. Right: 700-900ms in posterior midline region.....127

1 Introduction

1.1 Overview

The goal of this dissertation is to provide evidence regarding the temporal dimension of predictive computations in language comprehension and to outline a framework for studying the ingredients and steps involved in predictive computations. A key goal of psycholinguistics is to understand the mental representations and computations that underlie humans' capacity to process language in real-time. Although language comprehension generally comes automatically and effortlessly to us as readers and listeners, it involves many challenges that even state-of-the-art technology cannot deal with adequately. In order to recover structure and meaning from physical inputs that are sequential in nature (e.g., speech and written texts), comprehenders must continuously coordinate representations that are computed at different points in time. Further, since language inputs unfold rapidly in time (e.g., 3-5 words per second in speech) and are often noisy (e.g., due to speaker/listener errors or noise in the environment), the computations underlying language comprehension must be both fast and robust to noise.

A key to the speed and robustness of real-time language comprehension likely lies in our ability to predict upcoming input. Much like anticipating the trajectory of a ball helps soccer goalkeepers to position their bodies to block it, the ability to anticipate upcoming input (e.g., words, grammatical categories) can help comprehenders to process incoming language more efficiently. In fact, much recent research in cognitive and

computational neuroscience suggests that generating predictions about the future is a fundamental principle underlying the brain's operations (Bar, 2011; Hawkins, 2004; Llinás, 2002). After an explosion of work on this topic across the last decade, predictive processes have now been repeatedly demonstrated across domains such as visual and auditory perception (e.g., Bar, 2007; Bendixen, Schröger & Winkler, 2009; Houde, Nagarajan, Sekihara & Merzenich, 2002; Summerfield, Trittschuh, Monti, Mesulam & Egner, 2008), motor planning (e.g., Davidson & Wolpert, 2005; Wolpert, 1997) as well as language comprehension (e.g., DeLong, Urbach & Kutas, 2005; Wicha, Moreno & Kutas, 2004; Van Berkum, Brown, Zwitserlood, Kooijman & Hagoort, 2005).

Linguistic prediction can facilitate comprehension in several non-mutually exclusive ways. For example, contextual information (e.g., "Amy went by the pond at the petting zoo to feed the... ") may be used to pre-select likely grammatical categories (e.g., a noun). It may also be used to predict certain semantic or other features (e.g., animate, can swim) of an upcoming word. Such pre-selection may allow early disambiguation of incoming words without activating and inhibiting contextually inappropriate meanings. For example, comprehenders may use the context presented above to pre-select the waterbird meaning of the string of letters "duck" without activating other irrelevant meanings associated with the same string of letters (e.g., the action of lowering one's body). Further, contextual information may also be used to pre-activate stored lexical representations (e.g., "goose"). With prediction, comprehension need not be compromised even when the input is noisy or imperfect (e.g., /ʌk/). This also frees up cognitive resources (e.g., attention) for comprehenders to process other information simultaneously.

The extent to which prediction can facilitate processing critically depends on its timing relative to the predicted event. To catch a flying ball we not only need to estimate its trajectory, but also compute such predictions quickly enough in order to position ourselves and move our arms to a predicted position before the ball gets there. In the domain of language comprehension, even highly accurate predictions can fail to benefit language comprehension if they are not generated quickly enough. Therefore, in order to study how prediction contributes to the speed and robustness of the language processing system it is key to study how predictions are computed *in real time.* In this dissertation I focus on the processing of thematic relations and explore how different sources of information (in particular, the identity and the structural roles of event-participants) impact the predictive computations involved in real-time.

1.2 What is prediction?

Here I define prediction as mental operations that occur in anticipation of upcoming inputs or events. Conceptualizing the human brain as a computing machine, there are three key components to prediction – predictive computations and their inputs and outputs. From a mathematical perspective, these mappings are functions. A function is a deterministic mapping from elements of one set of distinct entities (domain) to elements from another set of distinct entities (co-domain). An input is an element of the domain and an output is the element of the domain that it gets mapped to. Therefore, an important goal of the study of linguistic prediction concerns characterizing the inputs to and outputs of linguistic prediction and how they are represented in the human brain.

Further, since linguistic prediction requires complex mappings that involve putting many simpler functions together, an important goal of a model of linguistic prediction is to specify the functions (sub-processes) involved in computing predictions and how these sub-processes combine to achieve such complex mappings. To begin I will define what makes computations predictive and explore what each of the three components of linguistic prediction might look like.

One defining characteristic of prediction is that it involves mappings between existing contextual information (input) and upcoming information (output). In other words, prediction involves using contextual information to form hypotheses about upcoming events. This contrasts with computations that serve to interpret incoming information, for instance, by mapping information from one level of representation to another (e.g., from sequences of phonemes to words and vice versa).

This definition of prediction posits a distinction between predictive and top-down computations. Although predictive computations often involve top-down processes (mappings from higher to lower levels of representation), top-down computations need not be predictive in nature. Top-down computations can be realized non-predictively, for instance, through using higher-level contextual information to select among lower-level representations that are computed bottom-up. One account that acknowledges the role of top-down computations but explicitly argues against prediction is Marslen-Wilson (1987)'s cohort model of word recognition:

> *"A lexical unit is assumed to become active when the sensory input matches the acoustic-phonetic pattern specified for that unit. The model prohibits top-down activation of these units in normal word-recognition, so that only the sensory*

> *input can activate a unit. There is no contextually driven pre-selection of candidates, so that words cannot become active as potential percepts without some bottom-up (sensory) input to the structures representing these words. ... Once the word-initial cohort has been accessed, and the model has entered into the selection phase, then top-down factors begin to affect its behavior."* (Marslen-Wilson, 1987, p. 78)

1.3 Evidence for prediction in language comprehension

Although evidence for top-down computations does not by itself constitute evidence for predictive computations, psycholinguistic research in the past fifteen years or so has provided a growing body of evidence that comprehenders do in fact compute predictions for upcoming linguistic inputs on the fly. Studies using different experimental and computational techniques have found that comprehenders are highly sensitive to the predictability of the inputs. For example, Altmann and Kamide (1999) showed that listeners to an utterance like '*The boy will eat…*' (compared to '*The boy will move…*') showed more anticipatory eye-movements to the picture of an edible object (e.g., a cake) even before the direct object was named.

Meanwhile, comprehenders' electrical brain responses have also been shown to be sensitive to the syntactic, semantic, and even phonological properties of likely upcoming words that are not yet present in the input (e.g., DeLong et al., 2005; Wicha et al., 2004; Van Berkum et al., 2005). For example, Van Berkum and colleagues (2005) observed listeners' ERPs to Dutch sentences such as '*The burglar had no trouble locating the secret family safe. Of course, it was situated behind a* big_{NEU} */* big_{COM} *but unobtrusive* $painting_{NEU}$ */* $bookcase_{COM}$*'* and found that, even prior to the onset of the noun, an early positivity was observed at the adjective when its grammatical gender is inconsistent with that of the predicted noun.

Further, along with research that uses computational methods to estimate a word's predictability (*surprisal*; e.g., Hale, 2001; Levy, 2008), many have shown that word predictability affects fixation durations and regression probabilities in large reading eye-movement corpora (e.g., Boston, Hale, Kliegl, Patil, & Vasishth, 2008; Demberg & Keller, 2008; Roark, Bachrach, Cardenas & Pallier, 2009; Smith & Levy, 2013). Therefore, in this dissertation I will build on existing evidence for the presence of predictive mechanisms in language comprehension and study the properties of the computations that, by hypothesis, are involved in real-time linguistic prediction.

1.4 Representations

What might predictive computations and their inputs and outputs look like? Let's start with a (relatively) simple and much simplified example – upon hearing the utterance *'The gardener talked as the barber trimmed the …'* a listener expects the word 'mustache' to appear next. What is the nature of the listener's predictions and what gives rise to such predictions? These are questions about mental representations: (i) How is contextual information mentally represented and what are the representations of contextual information that feed into predictive computations? (ii) What representations do predictive computations operate over? (iii) How are the outputs of predictive computations represented?

Questions about the outputs of predictive computations have received the most attention in existing research, which has primarily studied prediction by examining its consequences on the ease of processing. By examining what aspects of processing are facilitated (or disrupted) when bottom-up inputs are more (or less) compatible with the

predictions licensed by context, we can draw inferences about the representation of the outputs of prediction. For example, Dikker and colleagues (Dikker, Rabagliati & Pylkkänen, 2009) observed that mismatch with a syntactic category prediction affects activity in visual cortex at 130ms and suggested that predictions about upcoming syntactic categories are translated into form-based estimates, which activate representations in the visual cortex. Meanwhile, DeLong et al. (2005) observed that brain potentials are sensitive to the mismatch between a determiner (e.g., "a", "an") and the onset of a likely upcoming noun (e.g., "kite") and suggested that predictive computations can activate the phonological representation of predicted words.

However, much less is known about the representation of the inputs to prediction as well as the representations that predictions are computed on. Existing research has often abstracted away from these questions, and computational psycholinguistic models have commonly operated under the assumption that the inputs and outputs of predictions share the same level of representation. For example, in some recent probabilistic models of syntactic prediction (e.g., Hale, 2001; Levy, 2008), information about the syntactic category of previous words is fed into predictive computations, which output predictions about the syntactic category of an upcoming word. While these models are compatible with proposals that syntactic predictions are translated into form-based estimates that pre-activate phonological and/or orthographic representations, they remain largely agnostic about whether and how other sources of information (e.g., semantic and pragmatic knowledge) might impact predictions about upcoming syntactic categories.

In this dissertation I will discuss how careful considerations about the relevant sources of contextual information and their availability may help us get at questions about mental representations that underlie linguistic prediction. I will use the processing of thematic relations to illustrate how psycholinguistic experimentation can begin to address questions about the representations that feed into predictive computations.

1.5 The temporal dimension of prediction

Existing evidence for the impact of prediction on real-time language comprehension has given rise to the common assumption that all contextual information impacts comprehenders' predictions as soon as it becomes available in the input stream. In fact, this assumption is implicit in all studies that use offline cloze task or language corpora to estimate a word's predictability during real-time comprehension (e.g., Ehrlich & Rayner, 1981; Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Staub, 2011). Similarly, probabilistic models of linguistic prediction commonly assume that comprehenders take into account all contextual information for purposes of estimating the likelihood of upcoming input (e.g., Boston et al., 2008; Demberg & Keller, 2008; Roark et al., 2009; Hale, 2001; Levy, 2008; Smith & Levy, 2013).

Here I question this key assumption and claim that even prominent and unambiguous contextual information can fail to immediately impact linguistic predictions. I reason that, given the complexity of language input and the speed at which it becomes available to a reader/listener, the mental computations that are involved in generating linguistic predictions might still be incomplete when predictable input arises. Under this view, prediction is a race against time. To hit a flying ball in a game of tennis we not only

need to estimate its trajectory and spin, but also we need to compute those predictions quickly enough to position ourselves and move the racquet to the predicted position before the ball gets there. The extent to which prediction can facilitate processing critically depends on its timing relative to the predicted event.

Further, I argue that questions about the temporal properties of predictive computations (the 'when' questions) can help address questions about how the brain computes predictions during real-time comprehension (the 'how' questions). Based on timing estimates about when a piece of contextual information can feed into predictive computations and when an output is generated, we can set boundary conditions on the amount of time available for the relevant computations. Combined with considerations about cognitive theories about information flow (e.g., what representations must be computed before others) and neuroanatomical evidence (e.g., where certain processes are carried out in the brain and how quickly information can travel between relevant brain regions), such boundary conditions can further constrain our hypotheses about the sub-processes involved.

Although existing research has examined how the outputs of predictive computations impact bottom-up processes at different time points (e.g., sensory processes, lexical semantic access, structure building), much less is known about how the computations responsible for such predictions unfold in real-time. To study this, we need to determine when an output is generated by predictive computations relative to when the relevant inputs feed into such computations. However, empirically, these time points cannot be measured directly and can only be estimated indirectly. For contextual

information that is tied to a specific piece of input (e.g., a specific word), we may assume that such information cannot feed into predictive computations any earlier than when it appears in the input stream[1]. For example, in a discourse like '*Carol listened to the radio. The storm last night had only left a dusting of snow. Nonetheless, the schools were…',* we can examine how the connective 'nonetheless' impacts comprehenders' prediction for the upcoming word (e.g., 'closed') across time by manipulating when it appears in the sentence. Similarly, we can estimate an upper limit on when an output is generated by the computations that take into account the connective by examining the earliest point in time when bottom-up processing (e.g., at the word 'closed') is impacted by the supposed outputs of predictive computations. I will use these estimates to delimit a time window during which the predictive computations of interest occur. Through examining the 'when' questions more rigorously and systematically, this dissertation aims to provide a key component for future research on the 'how' questions.

1.6 Outline of the dissertation

In Chapter 2 I propose that the "Semantic P600" phenomenon in the ERP literature presents an interesting puzzle to the study of linguistic prediction. This phenomenon centers on the observations that, contrary to the long-held generalization that semantic anomalies elicit an N400 effect and grammatical anomalies elicit a P600 effect, thematic role-reversals (e.g., in Dutch: *De vos die op de stroper joeg* ... English

[1] Strictly speaking, this may not be true if the contextual information of interest is predicted and can impact further predictive computations before it appears in the input stream. For ease of exposition, however, I will assume that contextual information that is tied to specific bottom-inputs can impact predictive computations only after such inputs appear.

word-by-word translation: The fox that at the poacher hunted…) elicit a P600 effect but not an N400 effect despite being semantically/ pragmatically anomalous and fully grammatical. These findings have attracted much attention in the past decade and have often been taken as evidence for a cognitive architecture in which semantic interpretation can proceed independently from surface syntax. Based on the results from two ERP experiments in Mandarin Chinese (Experiments 1 and 2), I argue that (i) the P600's robust sensitivity to argument role-reversals and implausibility in general show that comprehenders' compute an accurate semantic interpretation using surface syntax, and (ii) the N400's insensitivity is not attributable to 'semantic illusions', but instead shows that comprehenders fail to use information about the arguments' structural roles to compute predictions for an upcoming verb.

In Chapter 3 I explore potential causes for this apparent prediction failure in three ERP experiments in Mandarin Chinese (Experiments 3-5). I introduce a new experimental paradigm for studying the time course of linguistic prediction and investigate whether the prediction failure reported in Chapter 2 is attributable to a delayed impact of structural role information on comprehenders' verb prediction and show that it can be remedied by extending the time interval for predictive computations. These results provide the first evidence that even unambiguous and prominent contextual information may have a delayed impact on comprehenders' predictions and highlight the significance of the temporal dimension for the study of linguistic prediction.

In Chapter 4 I turn to ask whether certain sources of contextual information can impact predictive computations more quickly than others. In an ERP experiment in

English (Experiment 6) I compare the impact of two different types of inputs on comprehenders' predictions for a verb given its arguments. I find that, despite matching in terms of off-line cloze probability measures, information about the arguments' lexical identity has a more immediate impact on comprehenders' predictions than information about their structural roles. These findings provide the first direct evidence that different sources of contextual information can impact predictive computations on different time scales. I outline future studies and new experimental paradigms that aim to establish a clearer time line of some of the predictive computations involved in processing thematic relations.

In Chapter 5 I evaluate the proposed interpretation of the ERP evidence by examining the extent to which ERP and eye-movement (EM) evidence align. In three EM experiments in Mandarin Chinese (Experiments 7-9), I explore various methodological considerations that may give rise to mis-alignments between EM and ERP results. By taking into considerations differences in stimulus presentation methods and information flow in EM and ERP experiments, I show that the effect of argument role-reversals on readers' EMs aligned with the ERP evidence reported in Chapter 3. I argue that EM evidence provides further support for the proposal that information about arguments' structural roles cannot immediately impact comprehenders' verb prediction.

In the final chapter, I synthesize the empirical findings reported in this dissertation and discuss their implications for the study of real-time linguistic prediction more broadly. I discuss several issues pertaining to developing more explicit models of real-time linguistic prediction and outline recommendations for future research.

2 A surprising case of prediction failure

This chapter has been published as Chow, W.Y. & Phillips, C. (2013). No semantic illusion in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research, 1506, 76–93.*

2.1 Introduction

Surface syntax is critical in determining the meaning of a sentence. Two sentences with the same words ordered differently (e.g., (1) and (2)) can have drastically different meanings.

1. The rebels killed the king.
2. The king killed the rebels.

Given the ease with which we detect the difference in meanings in sentences like (1) and (2), it can perhaps be taken for granted that we use surface syntax to compute the meaning of a sentence. In fact, most models of human sentence processing (e.g., Ferreira & Clifton, 1986; MacDonald, Pearlmutter & Seidenberg, 1994; Trueswell, Tanenhaus & Garnsey, 1994) have assumed that surface syntax is always used to guide online semantic composition.

The assumption that semantic composition relies on surface syntax should not be confused with the "syntax-first" position in the debate over online syntactic analysis in the study of structural ambiguity resolution. Although there are disagreements over whether syntactic information has priority over other sources of information, such as lexical bias, in online syntactic analysis (Ford, Bresnan, & Kaplan, 1982; Frazier, 1987;

Pickering, Traxler, & Crocker, 2000; Trueswell, Tanenhaus, & Kello, 1993), it is commonly assumed that only analyses that are compatible with the surface syntax are ever considered. Similarly, the view that semantic interpretation combines word meanings in accordance with syntactic constraints is independent of claims that syntactic anomalies are more rapidly detected than semantic anomalies (Friederici, 1995; McElree & Griffith, 1995). The assumption that semantic interpretation is based on the syntactic structure of the sentence is related to the claim that syntactic anomalies block the detection of semantic anomalies (e.g., Friederici, Steinhauer & Frisch, 1999; Hahne & Friederici, 2002), but these are logically distinct claims.

However, this assumption has not gone unchallenged (e.g., Bever, 1970; Caramazza & Zurif, 1976; Ferreira, Bailey & Ferraro, 2002; Slobin, 1966; Townsend & Bever, 2001; Jackendoff, 2002). In fact, many have argued that the recent discovery of the "Semantic P600" phenomenon in the electrophysiological literature directly challenges this assumption (e.g., Kim & Osterhout, 2005; Kolk, Chwilla, van Herten & Oor, 2003). These studies used event-related potentials (ERPs) to examine brain responses to fully grammatical sentences that contradict stereotypical thematic relationships ("role-reversed sentences", e.g., a criminal arresting a policeman, as opposed to being arrested by a policeman). The amplitude of the N400, a centro-parietal negative-going waveform peaking at around 400ms after stimulus onset, is generally modulated by the cloze probability and semantic/pragmatic congruity of the word in a given context (e.g., Kutas & Hillyard, 1980; Kutas & Hillyard, 1984; van Berkum, 2009). The P600, on the other hand, is a late posterior positive-going ERP waveform that has been associated with the presence of grammatical anomalies and syntactic processing

difficulty (e.g., Osterhout & Holcomb, 1992; Hagoort, Brown & Groothusen, 1993). Interestingly, although role-reversed sentences are clearly semantically anomalous, they typically fail to elicit a larger N400 than their canonical control (e.g., Hoeks, Stowe & Doedens, 2004; Kim & Osterhout, 2005; Kim & Sikos, 2011; Kolk et al., 2003; Kuperberg, Sitnikova, Caplan & Holcomb, 2003; Kuperberg, Caplan, Sitnikova, Eddy & Holcomb, 2006; Kuperberg, Kreher, Sitnikova, Caplan & Holcomb, 2007; Stroud & Phillips, 2012; van Herten, Kolk & Chwilla, 2005; van Herten, Chwilla & Kolk, 2006). Further, despite being fully grammatical and structurally unambiguous, role-reversed sentences consistently elicit a larger P600 compared to the canonical control condition.

In this chapter, we will refer to the phenomenon that grammatically well-formed role-reversed sentences elicit (i) only a P600 effect, and (ii) no N400 effects as the "Semantic P600" phenomenon. Various accounts of the phenomenon have proposed processing architectures that assume a semantic interpretation mechanism that is independent of surface syntax, i.e., an *independent semantic composition* mechanism, and thereby challenge the assumption that online semantic composition relies on surface syntax (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Hagoort, Baggio & Willems, 2009; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007; van Herten et al., 2005, 2006; van de Meerendonk, Kolk, Chwilla & Vissers, 2009). An influential study by Kim and Osterhout (2005) examined ERP responses to unambiguous, grammatically well-formed sentences that depict an anomalous thematic relation (e.g., (3) and (4)). They reported that semantically anomalous sentences with a "semantically attractive" predicate-argument combination (e.g., (3), in which *meal* is a likely Theme argument for *devour*) elicited only a P600 effect and no N400 effect. In contrast,

semantic anomalies such as (4), where the predicate and its argument are not semantically attractive, elicited only an N400 effect and no P600 effects.

3. Semantic anomaly with a plausible non-surface interpretation:

   The hearty meal was devouring… (control: *the hearty meal was devoured*)

4. Semantic anomaly (no plausible non-surface interpretation):

   The dusty tabletops were devouring… (control: *the hearty meal was devoured*)

Kim and Osterhout (2005) present a two-part argument that online semantic composition can be independent of surface syntax. First, when the subject and the verb are semantically attractive, as in (3), the processor constructs a plausible semantic representation, i.e., *the hearty meal* as the Theme of *devour*, even if it contradicts what is unambiguously dictated by surface syntax, i.e., *the hearty meal* as the Agent of *devour*; henceforth a "non-surface interpretation". Therefore, the processor is blind to the semantic anomaly in the input (a 'semantic illusion', see also Hoeks et al., 2004) and hence no N400 effects are elicited. Meanwhile, since the surface syntax of the input conflicts with that of the semantic representation computed, the processor in turn perceives the sentence as ungrammatical, resulting in a P600 effect. Second, when the subject and the verb are not semantically attractive, as in (4), and therefore no plausible semantic interpretation can be constructed, even by altering the structure or word order of the sentence, the processor perceives the sentence as semantically anomalous and generates an N400 effect and no P600 effect. Taken together, Kim and Osterhout argued that these results show that the processing system uses the meaning of individual words

to compute a plausible interpretation, even when surface syntax unambiguously conflicts with that interpretation.

In sum, both the presence of a P600 effect and the absence of N400 effects have been taken as evidence for an independent semantic composition mechanism. Below, we use evidence from Mandarin Chinese to evaluate these two key pieces of evidence in turn, and propose that (i) the presence of a P600 effect in role-reversed sentences may be attributed to factors that are independent from, but often confounded with, the presence of plausible non-surface interpretations; and (ii) the absence of N400 effects in role-reversed sentences is attributable to a combination of lexical priming and weak contextual constraints. Most of the comparisons presented in this study build upon previous studies, and our conclusions have precursors in the literature. The primary contribution of the current study is that it takes advantage of the properties of Mandarin Chinese to better assess proposals for syntax-independent semantic composition and the impact of factors such as animacy and implausibility.

2.2 When do semantic anomalies elicit a P600 effect?

Among the accounts that assume a processing architecture with a syntax-independent interpretation mechanism, several of them maintain that certain semantic anomalies elicit a P600 effect because the processor computes plausible interpretations that are incompatible with the surface syntax (e.g., Kim and Osterhout, 2005; Kolk et al., 2003; van Herten et al., 2005, 2006). The strongest evidence for this account involves arguments that the P600 response to semantic anomalies is selective. If semantic anomalies elicit a P600 response only if a plausible non-surface interpretation is available,

then this suggests that the non-surface interpretation plays a role in the processing of the sentence. On the other hand, if the P600 effect is elicited by semantic anomalies regardless of the availability of a plausible non-surface interpretation, then the observation of P600 effects in role-reversed sentences is compatible with accounts that assign no role to computation of non-surface interpretations.

To date, however, evidence for such selectivity is rather limited. Many studies have shown that semantic anomalies can elicit a P600 effect regardless of the availability of a plausible non-surface interpretation (e.g., Hoeks et al., 2004; Kuperberg et al., 2006, 2007; Paczynski & Kuperberg, 2011; Stroud & Phillips, 2012; van Herten et al., 2006;). For example, Hoeks et al. (2004) found that, along with role-reversed sentences such as *The javelin has the athletes thrown* (Dutch: De speer heeft de atleten geworpen.), semantically anomalous sentences that lack a plausible non-surface interpretation, such as "*The javelin has the athletes summarized*." (Dutch: De speer heeft de atleten opgesomd.), also elicited a significant P600 effect. Similar findings have been reported in studies across different languages, consistently showing that the presence of P600 effects to semantic anomalies is not restricted to cases in which a plausible non-surface interpretation is available (e.g., English: Kuperberg et al., 2006, 2007; Paczynski & Kuperberg, 2011; Stroud, 2008; Dutch: van Herten et al., 2006; Spanish: Stroud & Phillips, 2012; Japanese: Oishi & Sakamoto, 2010).

In light of the finding that the P600 is not selectively elicited by role-reversals, some authors have proposed that other factors can elicit semantic P600s. Some of these proposals still assume some form of syntax-independent semantic interpretation

mechanism (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg, 2007; van de Meerendonk, Kolk, Vissers & Chwilla, 2010). For example, Kuperberg (2007) emphasized that P600 effects to semantic anomalies are not solely modulated by thematic role-reversals. She identified that implausibility, along with the presence of animacy violations, played a key role in evoking a P600 effect in semantically anomalous sentences (see also Kuperberg & Paczynski, 2011). Meanwhile, van de Meerendonk et al. (2010) proposed that the P600 is modulated by the severity of the conflict between what is expected (i.e., likely to be true) and what is observed, and found evidence that deeply implausible sentences such as *The eye consisting of among other things a pupil, iris, sticker...* elicit a larger P600 response than mildly implausible sentences such as *The eye consisting of among other things a pupil, iris, eyebrow....*

On the other hand, others proposals do not assume a syntax-independent semantic composition mechanism and have argued that the P600's sensitivity to role-reversals can be fully attributed to surface properties of the materials (e.g., Brouwer, Fitz & Hoeks, 2012; Stroud, 2008; Stroud & Phillips, 2012). For example, Stroud (2008) observed that much existing evidence of P600 effects to role-reversals comes from studies that have confounded role-reversals with animacy violations. For instance, the role-reversal anomaly in (3) also involves a violation of the verb's requirement for an animate Agent. Stroud (2008) suggested that such P600 effects are attributable to the detection of animacy violations and therefore should not be taken as evidence for independent semantic composition. Meanwhile, van Petten and Luka (2012) suggested that the P600 reflects reanalysis processes that are triggered by the detection of implausibility, whereas

Brouwer et al. (2012) proposed that the P600 reflects the process of integrating the lexical information activated by a word into the current mental representation.

In order to evaluate whether factors such as the availability of plausible non-surface interpretations, animacy violations and implausibility make a unique contribution to the P600, comparisons need to be made between ERP responses to independent manipulations of non-surface plausibility and animacy congruity. However, only two studies to date (one in Dutch: van Herten et al., 2005; one in Mandarin Chinese: Ye & Zhou, 2008) have examined the effects of thematic role-reversals using fully grammatical and animacy-congruous sentences. Both of these studies used clauses with a subject-object-verb (SOV) word order, e.g., (5a) vs. (5b), and reported that role-reversal anomalies elicit a P600 effect and no N400 effect.

5. Role-reversal Anomaly in Animacy-congruous Sentences (Dutch)

a. *De stroper die op de vos joeg slopen door het bos.*

   The poacher$_{[singular]}$ that at the fox$_{[singular]}$ hunted$_{[singular]}$ stalked through the woods.

   "The poacher that hunted the fox stalked through the woods."

b. *De vos die op de stroper joeg sloop door het bos.*

   The fox$_{[singular]}$ that at the poacher$_{[singular]}$ hunted$_{[singular]}$ stalked through the woods.

   "The fox that hunted the poacher stalked through the woods."

Since these studies differed from those that examined animacy-violated role-reversals in many respects (e.g., language, word order of the sentence, the grammatical category of the target word), it remains difficult to compare across studies to determine to whether the availability of plausible non-surface interpretations, animacy violations,

and/or mere implausibility contribute uniquely to the P600 effects observed in role-reversed sentences. Therefore, in the present study we aim to provide a more rigorous test by comparing ERP responses to manipulations of animacy congruity and non-surface plausibility.

2.3 When do semantic anomalies fail to elicit an N400 effect?

Although it has attracted less attention than the P600 effects elicited by semantic anomalies, the N400's insensitivity to role-reversal anomalies is also surprising and central to arguments for independent semantic composition. Based on the functional interpretation of the N400 as reflecting the process of computing a coherent semantic representation by incorporating each new word into its context (e.g., Brown & Hagoort, 1993; Hagoort, Hald, Bastiaansen & Petersson, 2004), several existing accounts have interpreted the lack of N400 effects in role-reversed sentences as evidence that the parser temporarily fails to detect the semantic anomaly in role-reversed sentences, i.e., a 'semantic illusion' (e.g., Kolk et al., 2003; van Herten et al., 2005, 2006; Kim & Osterhout, 2005; Hagoort et al., 2009). For example, van Herten et al. (2005) proposed that the lack of N400 effects shows that comprehenders initially consider the interpretation that fits their world knowledge best. According to this hypothesis, a role-reversed phrase such as *the cat that fled from the mice* is initially interpreted as the assertion that the mice are fleeing from the cat, since "this describes a far more plausible real life event than the situation that the cat is fleeing from the mice" (p. 252). Meanwhile, Kuperberg (2007) proposed that the attenuation of the N400 in semantic P600 cases is driven by a "non-combinatorial semantic memory-based mechanism (that) computes the

semantic features, associative relationships and other types of semantic relationships between content words (including verbs and arguments) within a sentence, and compares these relationships with those that are pre-stored within lexical semantic memory" (p. 37). Taken together, these accounts posit that the processor can ignore surface syntax to compute a plausible interpretation in role-reversed sentences and therefore is effectively (temporarily) blind to the semantic anomaly and thus experiences no difficulty in semantic interpretation.

It has also been proposed that animacy information makes a unique contribution to the N400's sensitivity to semantic anomalies (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg et al., 2007; Paczynski & Kuperberg, 2011). For instance, Bornkessel-Schlesewsky & Schlesewsky (2008) noted that arguments with a dispreferred animacy feature (e.g., an inanimate subject, or an animate object) elicited larger N400 responses and proposed that the N400 reflects core argument interpretation based on prominence information such as animacy in addition to syntax-independent computation of plausible interpretations. Meanwhile, Kuperberg and colleagues observed that animacy-violated semantically incongruous sentences do not elicit an N400 effect and proposed that full semantic analysis, as indexed by the N400, can be 'switched off' when a reader's animacy-based expectations are violated (e.g., Kuperberg et al., 2007).

In this chapter, we test the hypothesis that neither non-surface plausibility nor animacy violations makes a unique contribution to the N400. We adopt a lexical access account of the N400, according to which N400 amplitude reflects the cost of access to a lexical entry in the lexicon (Deacon, Hewitt, Yang & Nagata, 2000; Kutas & Federmeier,

2000; Lau, Phillips, & Poeppel, 2008). We propose that the absence of N400 effects in role-reversed sentences indicates that the cost of accessing the target verb in the lexicon does not differ between the canonical and role-reversed conditions due to a combination of weak contextual constraint and strong lexical semantic association, and not due to the plausibility of a non-surface interpretation or to the presence of animacy violations. Based on previous findings regarding the effects of contextual constraint and lexical association on the N400, we aim to relate evidence of the N400's insensitivity in role reversals to other cases in which the N400 has been found to be insensitive to semantic anomalies.

A number of previous studies have found evidence of the N400's insensitivity to the compositional semantic meaning of a sentence. But these findings have previously been analyzed as independent phenomena. For example, Fischler and colleagues examined ERP responses to semantic anomaly in affirmative and negated sentences (Fischler, Bloom, Childers, Roucos and Perry, 1983). They observed that, for affirmative sentences like (6), false sentences elicited a larger N400 compared to true sentences. However, in negated sentences like (7) it was the true sentences that elicited a larger N400. Based on the assumption that the N400 reflects sentence meaning computation, the authors suggested that their results support a two-step theory of negation (e.g., Carpenter and Just, 1975), according to which the meaning of a proposition such as *A robin is not a bird* is hypothesized to be computed initially without the negation as *A robin is a bird*, and the semantic effect of negation is only computed in a second step. Under this account the N400 reflects only the first of these two steps.

6. Affirmative sentences

   A robin is a bird/tree.

7. Negated sentences

   A robin is not a tree/bird.

More recently Urbach and Kutas (2010) reported that the N400 is insensitive to semantic incongruity in sentences with certain types of quantifiers. They examined ERP responses to sentences such as (8) and (9) and observed that the atypical object (e.g., *worms*) elicited a larger N400 than the typical object (e.g., *crops*) in all cases, despite the fact that the relative semantic congruity in the *most/often* sentences is reversed in the *few/rarely* sentences. That is, in the *most*/*often* sentences the N400 amplitude was larger in the semantically incongruous conditions than in the congruous conditions, but in the *few*/*rarely* sentences the N400 amplitude was in fact smaller in the semantically incongruous conditions than in the congruous conditions. Based on this pattern of results, the authors suggested that semantic processing of quantifiers such as *most* and *often* occurs rapidly and incrementally, whereas quantifiers such as *few* and *rarely* are processed more slowly.

8. Sentences with noun phrase quantifiers

   a) *Most farmers grow crops/worms*

   b) *Few farmers grow crops/worms*

9. Sentences with adverbial quantifiers

   a) Farmers often grow crops/worms

   b) Farmers rarely grow crops/worms

One important similarity between these studies and previous studies on role-reversals may be the relatively low predictability of the target word in congruous and incongruous conditions alike, given that their sentence contexts are often minimally predictive. For example, in the case of negated sentences, given a context like "*A robin is not a ...*", the range of possible continuations is very broad, and hence an incremental processor might not expect the congruous target word *tree* any more than the incongruous target word *bird*. The sentence contexts in these studies do not provide sufficient information to facilitate access to the congruous target word relative to the incongruous target word. Under these circumstances it should not be surprising that the amplitude of the N400 is not reduced in the congruous condition relative to the incongruous condition.

In fact, a recent study by Nieuwland and Kuperberg (2008) contrasted ERP response profiles for sentences in which negation was pragmatically licensed (e.g., "*With proper equipment, scuba-diving isn't very dangerous /safe...*") vs. those in which negation was pragmatically unlicensed (e.g., "*Bulletproof vests aren't very dangerous / safe...*"). They found that in the conditions with pragmatically unlicensed negation, which were compatible with many possible continuations, N400 amplitudes were not reduced in the congruous condition. But in the conditions with pragmatically licensed negation, which more tightly constrains the likely continuations, the N400 was reduced in the congruous condition relative to the incongruous condition.

Further, a recent study by Bornkessel-Schlesewsky, Kretzschmar, Tune, Wang, Genç, Philipp, Roehm & Schlesewsky (2011) examined the effects of role-reversals by swapping the case marker or word order of an animate and an inanimate argument in

verb-final sentences in Turkish and Mandarin Chinese. They found that the verb sometimes elicited a larger N400 in the role-reversed condition than in the canonical control condition. Although the authors attributed the contrast between the presence of an N400 effect in their studies and the absence of N400 effects in previous studies to whether the language studied has rigid or flexible word order, it is plausible that the N400 effect showed that the processor uses the animacy feature of the arguments to predict different verbs in the canonical vs. role-reversed sentences, since the canonical sentences in these studies always had an animate Agent and an inanimate Theme and the opposite is true for the role-reversed sentences.

However, the low predictability of the target words alone does not explain why N400 amplitude was in fact larger in the congruous condition than in the incongruous condition in the studies by Fischler et al. (1983) and Urbach & Kutas (2010). Both of these studies compared ERP responses to lexical items that differed in terms of their semantic relatedness to the words in the preceding sentence context. For example, in sentences such as (8) and (9), the typical object "*crops*" is more closely associated to the context words "*farmers"* and "*grow"* than the atypical object "*worms*" is. The N400 amplitude is known to be reduced by semantic priming in word lists (e.g., Rugg, 1985) as well as in sentences (e.g., Camblin, Gordon & Swaab, 2007; Ditman, Holcomb & Kuperberg, 2006). Therefore, in a situation where the compositional meaning of the sentence context does not make one target word more expected than the other, it is unsurprising that the N400 amplitudes are modulated by effects of lexical relatedness (Nieuwland & Kuperberg, 2008). In previous studies of role-reversals, on the other hand, the canonical and role-reversed sentences differed only in either voice (active vs. passive)

or word order, and so the lexical items were perfectly matched between conditions. The fact that the target words were therefore lexically associated to the same degree across conditions is consistent with the absence of N400 effects in these studies.

This brief survey of different cases in which the N400 is insensitive to semantic incongruity highlights the commonalities among them and suggests the following generalization: The amplitude of an N400 response to a word is modulated by the processor's expectation for that word, which in turn is mediated by the compositional meaning of the sentence context as well as by semantic association among words in the sentence. Therefore, in the present study we aimed to examine how the N400's sensitivity to semantic anomalies is modulated by lexical semantic association and whether non-surface plausibility and animacy congruity makes any unique contribution to the N400.

2.4 Experiments 1 and 2

The present study aimed to clarify the implications of the Semantic P600 phenomenon for architectural questions about the relations between syntax and online semantic interpretation. To this end, we devised two ERP experiments in Mandarin Chinese in tandem to examine the contributions of plausible non-surface interpretations, animacy violations, lexical association and mere implausibility to the ERP responses to role-reversals. We first explain the design of both experiments and then discuss the predictions of different hypotheses for the two experiments.

Both experiments examined the ERP responses to role-reversals. The role-reversals in Experiment 1 co-occurred with an animacy violation (e.g., *the student baffled*

the math problem). The role-reversals in Experiment 2 were fully animacy-congruous (e.g., *the suspect arrested the inspector*). Due to practical constraints on generating fully animacy-congruous role-reversed sentences in sentences with a SVO word order, and in order to allow comparisons between the current study and previous studies on both kinds of role-reversals, all of our experimental sentences had a SOV word order. Despite having a SVO basic word order, Mandarin Chinese has a highly frequent SOV *Ba*(把)-construction. This construction requires a transitive verb, and the coverb *Ba* always follows the Actor argument and immediately precedes the Patient argument. Therefore, in this construction unambiguous and reliable cues about the arguments' syntactic roles are present in advance of the verb. Further, the fact that a clear role-reversal manipulation can be achieved by simply reversing the order of the arguments allowed us to avoid the ambiguity that occurs when role-reversed sentences are also morpho-syntactically anomalous in sentences such as "*The meal was devoured/devouring...*". Lastly, in order to maximize comparability among conditions across the two experiments, sentences in both experiments were intermixed and presented within a single experimental session.

In Experiment 1 we orthogonally manipulated animacy-congruity and the 'combinability' of the verb and its arguments (see Table 2-1 for a sample set of experimental materials). Using test sentences that had an animate subject and an inanimate direct object, animacy-congruity was manipulated by using verbs that can or cannot take an inanimate object. For purposes of the current study a verb was considered 'combinable' with its arguments if they can be combined to form a plausible sentence. For example, in the example in Table 2-1, the verb "hang" cannot be combined with the NPs "student" and "math problem" in a simple sentence to describe a plausible scenario,

and therefore this verb-argument triplet was classified as non-combinable. We considered 'combinability' as a more restrictive criterion than mere lexical association, since verbs that are combinable with their arguments are likely also lexically related to the arguments (e.g. doctor - patient - cure), but lexically related verb-argument triplets might not be combinable (e.g., doctor - nurse - cure). Since accounts that assume independent semantic composition mechanisms predict that the N400 is modulated by the presence of a plausible non-surface interpretation (combinability) and not just lexical association, we manipulated combinability in the current study (see Kuperberg et al., 2006 for a discussion about the relative contribution of these factors). This approach allowed us to evaluate these accounts and our proposal at the same time, because combinability and lexical association are correlated (see Methods). The manipulations of animacy-congruity and combinability resulted in a fully crossed 2 × 2 within-subjects design. Importantly, all sentences in the animacy-violated and combinable condition were role-reversed (i.e., they had a plausible non-surface interpretation), but the design of the experiment was such that the role reversal was simply a consequence of the two independent factors. This design made it possible to assess whether the presence of a plausible non-surface interpretation made any unique contribution to the observed ERP effects, as predicted by accounts that assume independent semantic composition mechanisms.

As shown in Table 2-1, the four conditions in each item set had the same subject and object arguments and only differed in the target verb. Further, verbs were shuffled among item sets to appear in different experimental conditions, thereby minimizing lexical confounds (see Methods). Therefore, all comparisons were made between sentences with the same pre-target context and different target verbs. A related

experimental design was used by Kuperberg et al. (2007), who also manipulated animacy congruity and lexical association. However, due to the constraints of SVO word order in English, comparisons in that study had to be made between sentences that differed in multiple ways. Animacy congruity was manipulated by varying the subject noun while holding the verb constant (e.g., *For breakfast the boys/the eggs would eat...*). But animacy-congruous violations ('pragmatic violations' in Kuperberg's terminology) were created by combining a plausible subject-verb pair with an incongruous adverbial (e.g., *For breakfast the boys would plant ...*). The lexically unrelated animacy violations were created by combining a lexically associated adverb-subject sequence with an unassociated verb (e.g., *For breakfast the eggs would plant...*). The SOV word order of the Chinese BA construction made it possible to tighten the manipulations, and also to provide closer comparisons with previous findings from languages with SOV order.

Experimental condition	Sample materials
1. Animacy-congruous, Combinable (Control)	gaocaisheng ba shuxueti jieda-le student BA math problem *solve-ASP* "The student *solved* the math problem"
2. Animacy-violated, Combinable (Role-reversed)	gaocaisheng ba shuxueti nandao-le student BA math problem *baffle-ASP* "The student *baffled* the math problem"
3. Animacy-congruous, Non-combinable	gaocaisheng ba shuxueti guaqi-le student BA math problem *hang-ASP* "The student *hung* the math problem"
4. Animacy-violated, Non-combinable	gaocaisheng ba shuxueti kunzhu-le student BA math problem *restrain-ASP* "The student *restrained* the math problem"

Table 2-1. Experimental conditions and example sentences in Experiment 1. The target word is underlined.

In Experiment 2 we manipulated the structural role of the arguments in simple BA-construction sentences (see Table 2-2 for a sample set of experimental materials). Unlike Experiment 1, both pre-verbal arguments in these sentences were animate NPs

and therefore this role-reversal manipulation never co-occurred with an animacy violation. This manipulation was related to experimental designs used by Kolk and colleagues (e.g., Kolk et al., 2003; van Herten et al., 2005) in Dutch, and Ye and Zhou (2008) in Mandarin Chinese. However, unlike the materials used by Kolk and colleagues, where the second arguments are prepositional phrases (e.g., *at the fox;* Dutch: *op de vos*), both arguments are noun phrases in the BA-construction in Mandarin Chinese. Further, unlike the materials used by Ye and Zhou (2008), where half of the sentences were in the active BA-construction while the other half were in the passive BEI-construction, the current study only used the active BA-construction in the experimental materials to ensure that the structural roles of the arguments were unequivocal to comprehenders.

Experimental condition	Sample materials
5. Canonical control	chen-tanzhang ba zhege-yifan jubu-le Inspector Chen BA the suspect *arrest-ASP* "Inspector Chen arrested the suspect"
6. Role-reversed (Animacy-congruous)	zhege-yifan ba chen-tanzhang jubu-le Inspector Chen BA the suspect *arrest-ASP* "The suspect arrested Inspector Chen"

Table 2-2. Experimental conditions and example sentences in Experiment 2. The target word is underlined.

Based on previous results, role-reversed sentences in both experiments were expected to elicit a P600 effect and no N400 effects relative to their canonical counterparts. However, competing theoretical accounts make different predictions in the other conditions. Specifically, if non-surface plausibility makes a unique contribution to the P600 (e.g., Kim & Osterhout, 2005), then the P600 effect should be largest in the role-reversed conditions in both experiments. If animacy violations make a unique contribution to the P600 (Stroud & Phillips, 2012), then the P600 effect should be larger

in the animacy-violated conditions than in the animacy-congruous conditions, including the role-reversed condition in Experiment 2. If both of these factors uniquely contribute to the P600 and their effects are independent, then a P600 effect should be observed in all role-reversed and animacy-violated conditions, but the effect should be largest in the animacy-violated role-reversed condition in Experiment 1. However, if the P600 is fully attributable to the general implausibility of the surface meaning (e.g., Brouwer et al., 2012; Kuperberg, 2007; van de Meerendonk et al., 2009, 2010), then we should merely expect a significant P600 effect in all implausible conditions relative to the plausible control in both experiments.

Predictions for the N400 effects also differ in competing accounts, although the N400 effects alone are less theoretically decisive. Accounts that assume independent semantic composition mechanisms predict an N400 effect in the two non-combinable conditions in Experiment 1, but no N400 effects in role-reversed sentences in both experiments because of the presence of a plausible non-surface interpretation. An account that attributes N400 effects to lexical association differences makes very similar predictions, since the lexical association between the target verb and its preceding words was much stronger in the role-reversed condition in both experiments than in the non-combinable conditions in Experiment 1. Although these two rather different hypotheses cannot be distinguished based on the N400 results alone, only the former predicts that non-surface plausibility uniquely contribute to the P600. Meanwhile, Bornkessel-Schlesewsky and Schlesewsky's (2008) proposal predicts that the N400 should be larger in the animacy-violated conditions than in the animacy-congruous conditions, whereas

Kuperberg et al.'s hypothesis (2007) predicts that the N400 should be attenuated or eliminated in the presence of an animacy violation.

2.4.1 Methods

Participants

Nineteen students (11 female, mean age = 22 years, range 18-25 years) from Beijing Normal University participated in the current study. All participants were native speakers of Mandarin Chinese, were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. Data from five additional participants were excluded due to excessive artifacts (≥50% epochs rejected in one or more conditions). All participants gave informed consent and were paid 50 RMB/hour for their participation.

Materials

Each item set in Experiment 1 contained four sentence types (see Table 2-1). The materials preceding the verb in all sentences within each item set were identical, consisting of an animate subject, the coverb *Ba* and an inanimate object. Each condition had a different target verb, followed by materials that were identical within each item set. All of the verbs were relatively common and three-characters long. Animacy-congruity was manipulated by using verbs that do or do not allow inanimate direct objects. Around 10% of verbs in the animacy-incongruous conditions showed a strong animacy bias rather than a strict animacy requirement. It should be noted that animacy-congruity does not entail the fulfillment of all of the verb's selectional restrictions. For example, *The student*

hung the math problem is considered animacy congruous because the verb 'hang' can take inanimate objects, even though 'math problem' does not fulfill the verb's requirement for a concrete object noun. Ninety percent of the sentences in the non-combinable animacy-congruous condition involved a violation of the verb's selectional restrictions. Combinability was manipulated by using verbs that either do or do not yield a plausible interpretation when combined with the two arguments. A lexical association rating study was conducted to obtain objective measures of the lexical association between the verbs and their arguments (see below for more details). Role-reversal anomalies in this experiment were created by using a verb that does not allow inanimate direct objects but is combinable with the two arguments, i.e., the sentence would have a plausible interpretation had the two arguments been reversed. Therefore, within the current experimental design role-reversed sentences can be characterized as *animacy-violated* and *combinable*.

In order to avoid lexical confounds in the ERP data, each verb that was used in a combinable condition in one item set was used in a non-combinable condition in another item set. Specifically, the verbs in the canonical sentences were shuffled across item sets to create the *animacy-congruous non-combinable conditions*, and the verbs in the role-reversed sentences were shuffled to create the *animacy-violated non-combinable* sentences. Therefore, the two animacy congruous conditions and the two animacy-violated conditions used an identical set of verbs. Care was taken in the verb-shuffling procedure to ensure non-combinability in the resulting sentences.

Assessing the lexical frequency of the target verbs was made difficult by the fact that the target verbs often consisted of a main verb and an adjectival resultative, e.g., du-si, meaning 'poison-dead', which are considered separate lexical items in some Chinese corpora. We were able to obtain lexical frequency estimates for 66 of the 120 target verbs (after removing the aspectual marker -*le*) using SUBTLEX-CH (Cai & Brysbaert, 2010), which suggested that lexical frequencies were well-matched between the animacy-congruous verbs (mean=2.34; SE=0.11) and the animacy-violated verbs (mean=2.32; SE=0.14). Meanwhile, the average log character frequency of the first two characters of the target verbs (i.e., without the common aspectual marker -le) was numerically slightly higher for animacy-congruous verbs (mean=4.30; SE=0.09) than for animacy-violated verbs (mean=4.04; SE=0.08).

We asked 44 native Mandarin Chinese speakers who did not participate in the ERP study to rate, on a 7-point scale, the degree to which the verbs are considered 'related' to the corresponding pair of noun phrases used in the ERP study. Each participant only saw one of the target verbs for any given item. The results showed that the verbs were judged to be closely related to the arguments in the combinable conditions (animacy-congruous: $M = 6.55$, $SE = 0.042$; animacy-violated: $M = 5.15$, $SE = 0.083$), but not in the non-combinable conditions (animacy-congruous: $M = 1.76$, $SE = 0.056$; animacy-violated: $M = 2.17$, $SE = 0.068$).

As illustrated in Table 2-2, each item set in Experiment 2 contained two conditions: a canonical condition and a role-reversed condition. Role-reversed sentences were created by reversing the structural position of the arguments in the canonical

sentences. Therefore, the two sentences within each set used an identical verb-argument triplet, and the two conditions differed only in the order of the arguments. Further, unlike Experiment 1 both preverbal arguments were animate and therefore the role-reversals in Experiment 2 never co-occurred with an animacy-violation.

All experimental sentences consisted of an adverbial phrase followed by a main clause. In order to avoid sentence-final wrap-up effects at the critical clause-final verb the SOV BA-construction was embedded in the adverbial phrase (*Zai… zhihou*, After…), followed by a grammatical main clause that was held constant across conditions within each item set. In all experimental sentences, no anomaly was evident before the critical verb.

Sixty sets of items were generated for each of the experiments and the sentences were distributed in 2 presentation lists, such that half of the participants read sentences from one presentation list and the remaining participants read sentences from the other list. Each list contained 180 experimental sentences (120 for Experiment 1 and 60 for Experiment 2) along with 180 unrelated fillers of similar length and structural complexity. Each list contained one sentence from each item set in Experiment 2, and two sentences from each item set in Experiment 1 (one combinable and one non-combinable). The sentences were presented in 6 blocks of 60 sentences each, and the order of the blocks was randomized across participants. The two conditions from the same item set never appeared within the same presentation block. Care was taken to ensure that the overall congruous-to-anomalous ratio in each presentation list was 1:1.

Procedure

Participants were comfortably seated in a testing room around 100cm in front of a computer screen. Sentences were presented one word at a time in a white font (30 pt simplified Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500ms. Each word appeared on the screen for 400ms, followed by 200ms of blank screen. The last word of each sentence was marked with a period "。", followed 1000ms later by a response cue "?". Participants were instructed to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively and to indicate whether the sentence was an acceptable sentence of Mandarin Chinese by pressing one of two buttons. The current study used this task because the phenomenon of interest has been observed in previous studies that used the same task. Prior to the experimental session, participants were presented with 12 practice trials to familiarize themselves with the task. The experimental session was divided into six blocks of 60 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 2.5 hours on average.

EEG Recording

EEG was recorded continuously from 30 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. Recordings were referenced online to the left mastoid and re-referenced to linked mastoids offline. The electro-oculogram (EOG) was recorded at four electrode sites; vertical EOG was recorded from electrodes placed above and below the left eye and the horizontal EOG

was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5kΩ. The EEG and EOG recordings were amplified (bandpass = 0.5-100Hz) and digitized online at 1kHz with a bandpass filter of 0.1-70 Hz.

ERP Data Analysis

All trials were evaluated individually for EOG or other artifacts. Trials contaminated by eye blinks, excessive muscle artifact, or amplifier blocking were excluded from the averaging procedure. This affected 10.6% of experimental trials, equally distributed across conditions (ranging between 9.6 and 11.6% across conditions). Event-related potentials were computed separately for each participant and each condition for the 1000ms after the onset of the critical verb relative to a 100 ms baseline preceding the critical verb. Averaged waveforms were filtered offline using a 10 Hz low-pass filter for presentation purposes only. All statistical analyses were performed using the original data.

Statistical analyses on average voltage amplitudes were conducted separately for four time windows chosen based on previous literature and on visual inspection of the data: 0-300 ms for possible early differences, 300–500 ms for the N400, and 600-800 ms and 800-1000 ms for the P600. Separate analyses were conducted for mean amplitudes in each time window. While the 600-800 ms interval is most commonly used to analyze P600 effects in previous studies, the additional 800-1000 ms interval was included in the current study to examine possible component overlap between the N400 and the P600 effects. We reasoned that, if the P600 partially overlapped in time with the N400, the amplitude of the P600 might be affected by the N400 more strongly in an earlier interval

(e.g., 600-800 ms) than in a later interval (e.g., 800-1000 ms). Although similarity between the ERPs in these intervals does not exclude the possibility that the P600 did overlap with the N400, systematic differences in these P600 intervals could be informative about the extent to which our findings about the P600 is attributed to potential overlap with the N400.

Data from the two experiments were analyzed separately and in two ways. A traditional omnibus repeated measures ANOVA was conducted using anteriority (anterior vs. posterior) and laterality (left vs. midline vs. right) as topographic factors. Since the current study was designed to modulate two ERP components with well established topographic distributions, a repeated measures ANOVA was conducted in each of the six regions of interest (ROI) to test for the predicted differences (left-anterior: F3, FC3, C3; midline-anterior: FZ, FCZ, CZ; right-anterior: F4, FC4, C4; left-posterior: CP3, P3, O1; midline-posterior: CPZ, PZ, OZ; right-posterior: CP4, P4, O2). Data from Experiment 1 were analyzed using combinability (combinable vs. non-combinable) and animacy-congruity (animacy-congruous vs. animacy-violated) as within-subjects factors. Follow-up comparisons were carried out only when the interaction between animacy-congruity and combinability reached statistical significance. Data from Experiment 2 were analyzed using role-reversal (control vs. role-reversed) as a within-subjects factor.

2.4.2 Results

Acceptability Judgments

Participants' average acceptability judgment accuracy and the target response in each condition is shown in Table 2-3. With an overall accuracy of 86.7%, participants

reliably accepted canonical control sentences and rejected the semantically anomalous sentences, regardless of the presence or absence of thematic role-reversals. In Experiment 1 a repeated measures ANOVA revealed a marginal effect of animacy-congruity ($F(1,19) = 3.03$, $p < .10$), due to more accurate responses for animacy-violated sentences than animacy-congruous sentences. In Experiment 2 mean accuracy did not differ significantly between the canonical and role-reversed conditions ($t(18) = 0.71$, $p = 0.49$).

	Target Response	Percent accurate (sd)
Experiment 1		
Animacy-congruous, Combinable	Yes	83.2 (11.6)
Animacy-violated, Combinable (Role-reversed)	No	89.5 (8.3)
Animacy-congruous, Non-combinable	No	83.5 (12.2)
Animacy-violated, Non-combinable	No	91.8 (7.5)
Experiment 2		
Canonical control	Yes	87.2 (7.4)
Role-reversed (Animacy-congruous)	No	84.9 (10.2)

Table 2-3. Target response and accuracy on acceptability judgment task.

Event-related Potentials (Experiment 1)

Figure 2-1 shows the grand average ERPs (n=19) at the target word in all four conditions in Experiment 1. The target words in all conditions elicited the pattern characteristic of ERPs to visual stimuli. These components include an initial positivity (P1) peaking at about 80 ms, followed by a negativity (N1) at 170 ms, and a positivity (P2) around 275 ms. These responses were followed by a centro-posterior negativity between about 300 and 500 ms (N400). In the conditions involving animacy violations, the N400 was followed by a large late positive-going wave starting from approximately 550 ms (P600).

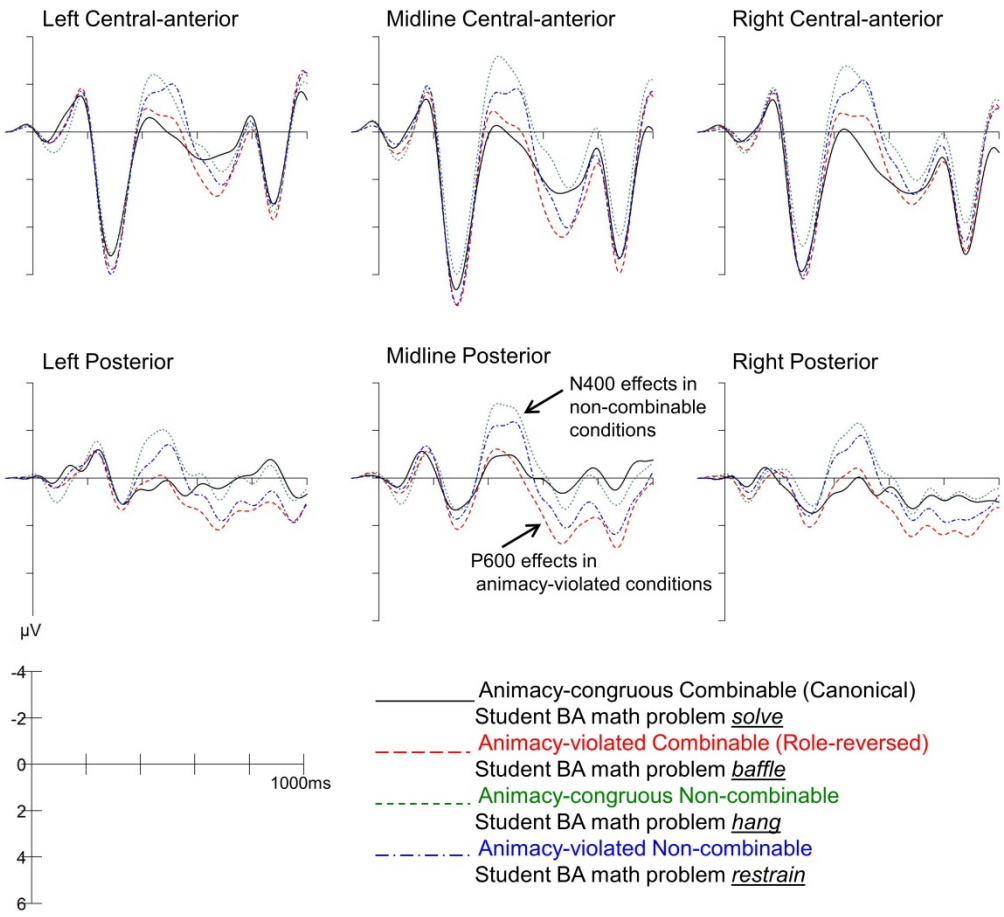

Figure 2-1. Grand average ERPs in six regions of interests in Experiment 1.

Inspection of Figure 2-1 reveals clear effects of both experimental factors. Combinability affected N400 amplitude and animacy-congruity affected P600 amplitude. Non-combinable target verbs elicited a larger N400 response compared to combinable target verbs. Starting at about 550 ms, animacy-incongruous target verbs elicited a larger posterior positivity (P600) than the animacy-congruous target verbs. These observations were confirmed by the statistical analyses. Results from the overall ANOVA and region of interest (ROI) analyses are presented in Table 2-4. The mean ERP values in the N400 and P600 intervals in the midline posterior region are presented in Figure 2-2.

No significant differences were observed in the 0-300 ms interval. In the 300-500 ms interval the overall ANOVA revealed a significant main effect of combinability, showing that ERPs in the non-combinable conditions were more negative than in the combinable conditions across the entire scalp. The interaction between combinability and animacy, and a four-way interaction between combinability, animacy, anteriority and showing that the effect of combinability was slightly larger and more broadly distributed in the animacy-congruous condition than in the animacy-violated condition. ROI analyses revealed a significant main effect of combinability in all ROIs, and a significant interaction between animacy-congruity and combinability in three ROIs (midline central-anterior, right central-anterior, and right posterior regions). Follow-up comparisons revealed that the amplitude of the N400 was not different between the animacy-violated and animacy-congruent combinable conditions in any of these regions, whereas the N400 was less negative in the animacy-violated non-combinable condition than in the animacy-congruent non-combinable condition and this difference reached statistical significance in the midline central-anterior region ($t(1,18) = 2.41$, $p < .05$).

These results were also corroborated by pair-wise comparisons of the amplitude of the N400 between each of the anomalous conditions and the canonical control condition in each ROI. These comparisons revealed that the N400 never differed between the animacy-violated combinable condition and the control condition, and that the N400 was more negative in both non-combinable conditions compared to the control condition across the scalp. The N400 in the animacy-congruent non-combinable condition was significantly more negative than that in the canonical control condition across all ROIs (all ps $< .02$); the effect in the animacy-violated non-combinable condition was

marginally significant in the left central-anterior region (p <.06) and significant in all other regions (ps < .02).

Starting at around 550 ms ERPs in the animacy-violated condition were more positive than in the animacy-congruous condition and this effect persisted throughout the entire epoch. The effect was present across the entire scalp, but was largest at midline posterior sites. In the 600-800 ms interval the overall ANOVA revealed a marginally significant main effect of animacy-congruity (p = .054) and significant interactions between animacy and anteriority and between animacy-congruity and laterality. Consistent with the typical distribution of P600 effects, ROI analyses confirmed a main effect of animacy-congruity that was significant in three regions (left posterior, midline posterior, and midline anterior) and marginally significant in the right posterior region (p = .06). No significant interaction effects between animacy-congruity and combinability were observed. In the 800-1000 ms interval the overall ANOVA revealed a significant interaction between animacy-congruity and anteriority, along with a significant four-way interaction between combinability, animacy-congruity, anteriority, and laterality. ROI analyses revealed a main effect of animacy-congruity that was statistically significant in the midline posterior region and marginally significant in the left posterior region (p = .08). No significant interaction effects between animacy-congruity and combinability were observed.

In summary, ERPs in Experiment 1 were significantly more negative in the non-combinable conditions than in the combinable conditions in the N400 interval, and significantly more positive in the animacy-violated conditions than in the animacy-

congruous conditions in the P600 intervals. But with the exception of one ROI in the N400 analyses, the ERP effects of animacy-congruity and combinability were independent of one another: the N400 amplitude was modulated by combinability but not by animacy-congruity, and the P600 amplitude was modulated by animacy-congruity but not by combinability.

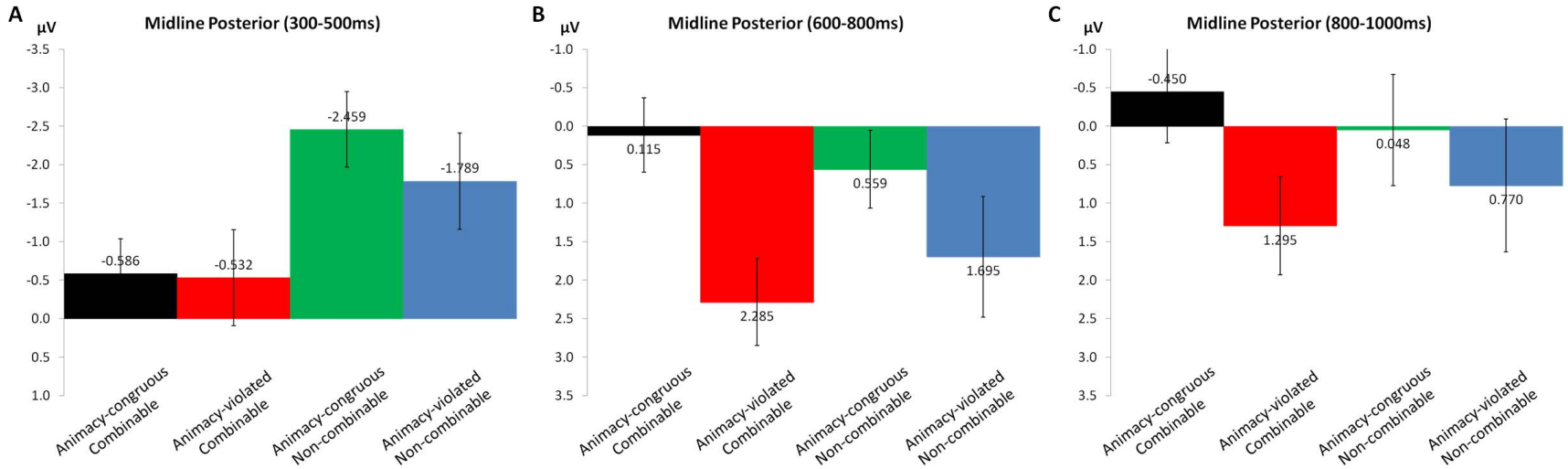

Figure 2-2. Grand average ERPs in midline posterior region in (A) 300-500 ms, (B) 600-800 ms, and (C) 800-1000 ms time intervals in Experiment 1.

	df	0-300ms	300-500ms	600-800ms	800-100ms
Omnibus ANOVA					
comb	1,18	<1	26.04**	<1	<1
anim	1,18	<1	<1	4.24^	<1
comb * anim	1,18	<1	11.04**	1.63	<1
comb * ant	1,18	<1	<1	<1	2.02
anim * ant	1,18	<1	1.23	7.54*	14.49**
comb * anim * ant	1,18	2.06	1.31	3.44^	5.66*
comb * lat	2,36	2.24	2.33	2.56	3.27^
anim * lat	2,36	<1	2.19	4.34*	2.02
comb * anim * lat	2,36	<1	<1	<1	1.57
comb * ant * lat	2,36	3.02^	2.4	2.33	3.37^
anim * ant * lat	2,36	<1	<1	<1	<1
comb * anim * ant * lat	2,36	<1	3.49*	3.12^	4.16*
ROI analyses					
Left central-anterior					
comb	1,18	<1	7.55*	<1	<1
anim	1,18	<1	<1	<1	<1
comb * anim	1,18	<1	2.04	1.77	<1
Midline central-anterior					
comb	1,18	<1	24.93**	<1	2.38
anim	1,18	1.09	<1	4.86*	<1
comb * anim	1,18	<1	7.56*	<1	2.78
Right central-anterior					
comb	1,18	2.84	25.61**	2.27	4.31^
anim	1,18	<1	<1	<1	<1
comb * anim	1,18	<1	14.18**	1.94	3.83^
Left posterior					
comb	1,18	<1	23.79**	<1	<1
anim	1,18	<1	<1	7.53*	3.43^
comb * anim	1,18	<1	2.43	1.18	<1
Midline posterior					
comb	1,18	<1	19.88**	<1	<1
anim	1,18	<1	1.37	9.77**	5.85*
comb * anim	1,18	<1	1.96	4.36^	2.37
Right posterior					
comb	1,18	1.28	18.81**	<1	<1
anim	1,18	<1	<1	3.87^	1.94
comb * anim	1,18	<1	6.69*	3.03^	<1

Factors: *comb* = combinability; *anim* = animacy; *ant* = anteriority; *lat* = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

Table 2-4. Repeated measures ANOVA F values at the target word in Experiment 1.

Event-related Potentials (Experiment 2)

Figure 2-3 shows the grand average ERPs (n=19) at the target verb in Experiment 2. Results from the overall ANOVA and ROI analyses are presented in Table 2-5. As shown in Figure 2-3, the target words in both conditions elicited the pattern characteristic of ERPs to visual stimuli. The ERPs did not diverge early on, hence no significant effects of role-reversal were observed in the 0-300 ms and 300-500 ms intervals. Starting at around 550 ms the ERPs became more positive in the role-reversed condition than in the canonical condition and the effect persisted throughout the entire epoch. The effect was present across the entire scalp, but was most pronounced at posterior sites, showing a topographic distribution that is typical of P600 effects.

In the 600-800 ms interval the omnibus ANOVA revealed no significant main effect or interactions involving Role-reversal (p = 0.13), but ROI analyses confirmed a significant or near-significant effect of Role-reversal in all posterior regions, consistent with the characteristically posterior distribution of P600 effects. In the 800-1000 ms interval the omnibus ANOVA revealed a marginal interaction between role-reversal and anteriority (p = .07), and ROI analyses revealed a marginal effect of role-reversal in the right posterior region (p = .06).

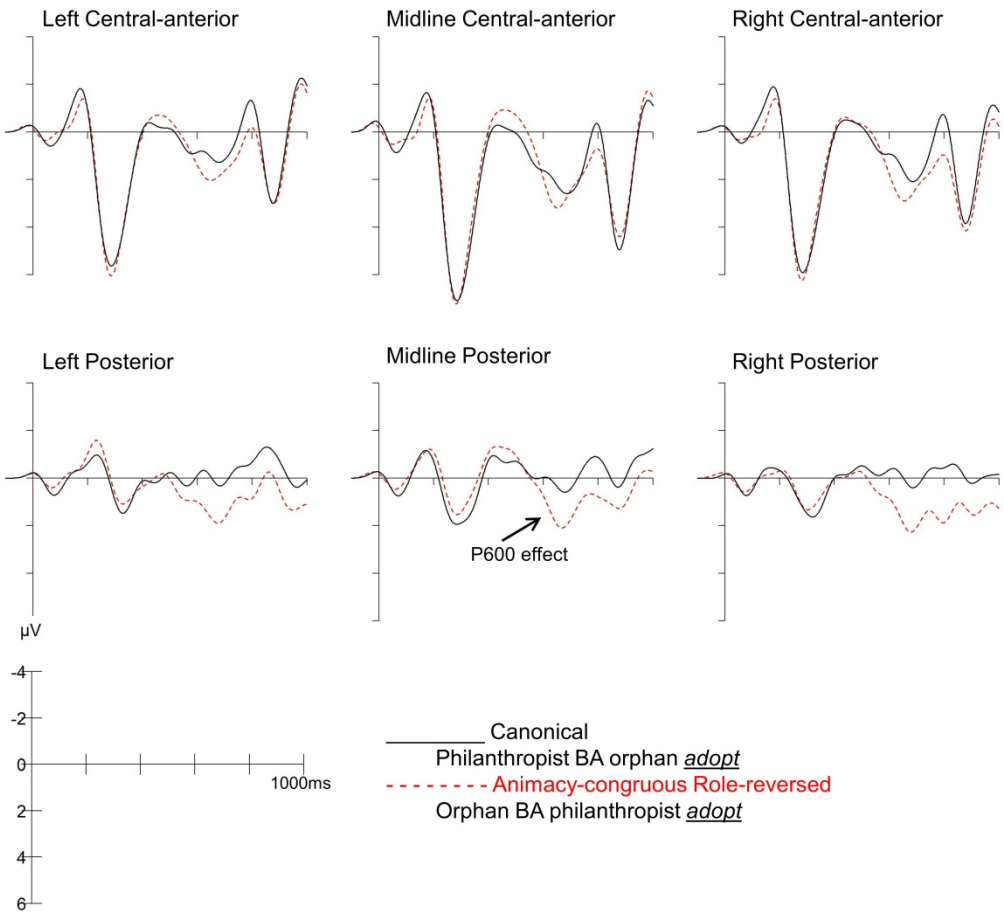

Figure 2-3. Grand average ERPs in six regions of interest in Experiment 2.

	df	0-300ms	300-500ms	600-800ms	800-100ms
Omnibus ANOVA					
rev	1,18	<1	<1	2.48	<1
rev x ant	1,18	1.11	<1	2.73	3.81^
rev x lat	2,36	<1	2.13	2.2	2.11
rev x ant x lat	2,36	<1	<1	<1	<1
ROI analysis					
Left central-anterior	1,18	<1	<1	<1	<1
Midline central-anterior	1,18	<1	<1	<1	<1
Right central-anterior	1,18	<1	<1	1.13	<1
Left posterior	1,18	1.35	<1	4.57*	2.65
Midline posterior	1,18	<1	<1	3.12^	1.51
Right posterior	1,18	<1	<1	10.06**	3.86^

Factors: *rev* = reversal; *ant* = anteriority; *lat* = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

Table 2-5. Repeated measures ANOVA F values at the target word in Experiment 2.

2.5 Discussion

The aim of the present study was to clarify the role of plausible non-surface interpretations ('semantic attraction'), lexical association/combinability, and animacy congruity in the ERP responses to role-reversal anomalies. Each of the individual results in the present study has precedents in previous studies, but the way in which they are combined here makes it possible to address architectural questions that were not so easily addressed before. First, the presence of a P600 effect and the absence of N400 effects in the role-reversed conditions of Experiments 1 and 2 are consistent with previous reports that role-reversed sentences, despite being syntactically well-formed and semantically incongruous, elicit a P600 effect and no N400 effects (e.g., Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003, Kuperberg et al., 2003, 2007; van Herten et al., 2005, 2006; Ye & Zhou, 2008). Further, the presence of a highly similar P600 effect across different semantically anomalous conditions in Experiments 1 and 2, independent of non-surface plausibility, suggests that the P600 is sensitive to the implausibility of the surface form of the sentence, but not to the availability of plausible non-surface interpretations. The current results suggest that the N400's disappearance in role-reversed sentences is likely due to strong lexical associations. Meanwhile, although the apparent reduction of the N400 in the animacy-violated non-combinable condition relative to the animacy-congruous non-combinable condition suggests that animacy-violations might attenuate the N400, we argue that such reduction is also attributable to component overlap.

P600 is not selectively sensitive to non-surface interpretations

Table 2-6 summarizes some of the factors that have been proposed to account for the Semantic P600 phenomenon. Current accounts for the observation of P600 effects to semantic anomalies consider factors such as surface plausibility, non-surface plausibility, animacy congruity and competing representations.

The current results, along with some previous results discussed in the Introduction, are not compatible with accounts that assume independent semantic composition (Hagoort et al., 2009; Hoeks et al., 2004; Kim & Osterhout, 2005; Kolk et al., 2003; van de Meerendonk et al. 2009, 2010; van Herten et al., 2005, 2006). These accounts predict that semantic anomalies that have a (partially or wholly) plausible non-surface interpretation should make a unique contribution to the P600, and that a semantic P600 response should be conditioned by the absence of an N400 effect. In the current study, however, the animacy-violated non-combinable condition in Experiment 1 nonetheless elicited a P600 effect despite the absence of a plausible non-surface interpretation, and the size of this effect was almost identical to that elicited in the role-reversed condition. Further, the current findings are also not compatible with an account based on 'partial plausibility' (van Herten et al., 2006), since the two non-combinable conditions had identical degrees of partial plausibility and only one of them elicited a P600 effect.

In fact, evidence for the P600's selective sensitivity to plausible non-surface interpretations has only been reported in the original study by Kim and Osterhout (2005). Other studies, including Stroud's (2008) replication study using Kim and Osterhout's (2005) experimental materials, have consistently found that semantic anomalies that have

no plausible non-surface interpretations nonetheless elicit a P600 effect (see also Kuperberg et al., 2006; Kuperberg 2007; Oishi et al., 2010; Paczynski & Kuperberg, 2011, 2012; Stroud & Phillips, 2012). And in cases where there is a direct comparison between conditions with and without a plausible non-surface interpretation ('semantic attraction'), the P600 effects typically show identical amplitude (e.g., Stroud, 2008; Stroud & Phillips, 2012). Meanwhile, our observation that an N400 effect preceded this P600 effect shows that the presence of a P600 response to semantic anomalies is not conditioned by the absence of an N400 response. This is consistent with the observation that semantic anomalies frequently elicit both an N400 and a late positivity (e.g., Friederici, Hahne & von Cramon, 1998; Kolk et al., 2003; Curran, Tucker, Kutas & Posner, 1993; van den Brink, Hagoort & Brown, 2001; van Herten et al., 2005). In fact, van Petten & Luka (2012) noted that the N400 elicited by semantically incongruous words is followed by a posterior positivity in about one third of the 64 published comparisons they reviewed. Taken together, our results provide convergent evidence that the P600 response to semantic anomalies is not modulated by the availability of (partially) plausible non-surface interpretations and therefore they undermine the original argument for independent semantic composition.

Meanwhile, however, we believe no existing accounts can straightforwardly capture the current results. A surface anomaly account that attributes the P600 to grammatical and animacy violations (e.g., Stroud, 2008) cannot capture the observations that fully grammatical and animacy-congruous role-reversed sentences (Experiment 2) nonetheless elicited a P600 effect. Meanwhile, accounts that attribute the P600 to the implausibility of the sentence (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008;

Brouwer et al., 2012; Kuperberg, 2007; van de Meerendonk et al., 2009, 2010) have yet to capture why certain kinds of implausible sentences fail to elicit a P600 effect. In the current study, even though the same group of participants had judged the sentences in all of the anomalous conditions as unacceptable, only the animacy-congruous non-combinable condition failed to elicit a significant P600 effect. As 90% of the sentences in the animacy-congruous non-combinable condition (as opposed to 100% in the animacy-violated conditions and only 55% in the animacy-congruous role-reversed condition) involved a violation of the verb's selectional restriction (see Methods), sentences in the animacy-congruous non-combinable condition should not be less implausible than those in the other conditions. Therefore, we propose that neither the presence of animacy violations nor surface implausibility can straightforwardly account for the full set of current findings.

Previous studies have discussed the possibility that a P600 effect might be attenuated if it temporally overlaps with a large N400 effect, which has opposite polarity (e.g., Hagoort, 2003). One possibility is that a P600 effect was elicited in all semantically anomalous conditions, but that the response was fully masked in the non-combinable conditions due to an overlapping N400 effect. However, we regard this possibility as rather unlikely. Given that no apparent effects were observed in the animacy-congruous non-combinable condition in either of the P600 time intervals (600-800 ms; 800-1000 ms) in Experiment 1, the supposedly masked P600 effect would have to have been completely overlapping in time with the N400 effect. The N400 effects elicited by visually presented stimuli are typically confined to a well-defined time interval (e.g., 300-500 ms), during which the divergence between the conditions peaks at around 400ms and gradually

returns to baseline afterwards (for review see Kutas & Federmeier, 2011). P600 effects, meanwhile, tend to extend over a longer time interval (e.g., 500-1000 ms). Therefore, if a P600 effect were present in all of the anomalous conditions in the present study, then its apparent absence in the animacy-congruous non-combinable condition must be attributable to (i) an N400 effect that extended well into the P600 time-window in that condition; and/or (ii) a particularly short-lasting P600 effect in that condition.

Note, however, that even if the presence of an N400 effect did obscure a potential P600 effect in the current study, our results would still be incompatible with semantic illusion accounts of the P600. Since an N400 effect was present in the non-combinable conditions but not in the combinable conditions, resolving this overlap would yield larger P600 effects in the non-combinable conditions than in the combinable conditions. In particular, the P600 in the animacy-violated non-combinable condition would be larger than that in the role-reversed conditions. This is the opposite of the predictions of semantic illusion accounts, according to which the role-reversed condition should elicit a larger P600 than conditions in which no plausible non-surface interpretations are available, since a plausible non-surface interpretation is present in the former but not in the latter.

In sum, the presence of non-surface plausibility makes no unique contribution to the P600, and thus the P600's sensitivity to role-reversal anomalies does not constitute evidence for syntax-independent semantic composition. The present results suggest that the P600 is sensitive to both animacy violations and surface implausibility, but neither of these factors can fully account for the current findings in isolation. The current results

add to the growing body of evidence that fully grammatical sentences with semantic anomalies do at times elicit a P600 effect. Future work is required to specify testable hypotheses about how these factors combine and/or interact in modulating the P600.

Lexical relations, not non-surface plausibility or animacy-congruity, modulate the N400

Current accounts for the absence of N400 effects to semantic anomalies consider factors such as lexical semantic association, non-surface plausibility and animacy congruity. Accounts that assume independent semantic composition (e.g., Kim & Osterhout, 2005) attribute the lack of N400 in role-reversed sentences to the presence of a plausible non-surface interpretation ('semantic attraction'). Meanwhile, others have proposed that the presence of animacy-violations makes a unique contribute to the modulation of the N400 (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg, 2007). Alternatively, without positing any unique contribution of non-surface plausibility and/or animacy congruity, we outlined in the Introduction that the lack of N400 effects can be fully attributed to strong lexical association between the target verb and its arguments in role-reversed sentences. Our discussion about lexical association can be extended to include the semantic features, associative relationships and other types of semantic relationships between content words (e.g., the semantic memory-based mechanism in Kuperberg, 2007). It may also be extended to include event schemas (e.g. Bicknell, Elman, Hare, McRae and Kutas, 2010; Paczynski & Kuperberg, 2012).

In the current study, the non-combinable conditions elicited a significantly larger N400 effect than the combinable conditions. Since the current study operationalized the manipulation of lexical association as the combinability between the verb and its

arguments, these N400 findings alone do not allow us to determine whether the presence of a plausible non-surface interpretation makes a unique contribution to the N400 beyond that of lexical association. Meanwhile, we can better evaluate the merits of these competing accounts in the context of the broader array of findings. Firstly, Kuperberg and colleagues (2006) observed that, when semantic relatedness is held constant, semantically anomalous words elicited the same N400 and P600 responses regardless of the presence of plausible non-surface interpretations ('thematic fit', in the authors' terms) and suggested that non-surface plausibility makes no unique contribution to their findings. Secondly, accounts that assume independent semantic composition predict that both the lack of an N400 effect and the presence of a P600 effect in in role-reversed sentences are selectively conditioned by the presence of plausible non-surface interpretations. As we discussed above, however, most existing evidence shows that the P600 is less selective. Further, as discussed in the Introduction, instances of the N400's blindness to semantic incongruity have been reported outside of the "Semantic P600" literature and in different previous studies (e.g., Fischler et al., 1983; Nieuwland & Kuperberg, 2008; Urbach & Kutas, 2010). Even though accounts that assume independent semantic composition are compatible with the current N400 findings, such accounts require that other instances of N400 blindness be given different interpretations. In contrast, our proposal attributes the N400's insensitivity to role-reversals to the (roughly) equal accessibility of the target word across conditions, due to a combination of strong lexical association and weak contextual constraints. This account does not give a special status to sentences with a plausible non-surface interpretation, and it can potentially provide a unified explanation for other instances of the N400's blindness to semantic anomalies.

Further, the current findings are consistent with the claim that animacy congruity makes no unique contribution to the N400. Specifically, even though role-reversals co-occurred with animacy-violations in Experiment 1 but not in Experiment 2, they did not elicit an N400 effect in either case. Meanwhile, the small reduction of the N400 in the animacy-violated non-combinable condition relative to the animacy-congruous non-combinable condition is compatible with the hypothesis that animacy-violations attenuate the N400 (Kuperberg et al., 2007). However, such N400 reduction can also be attributed to (i) potential component overlap, and/or (ii) semantic relatedness differences between the items used in the two non-combinable conditions. Since the animacy-violated condition elicited a much larger late positivity than the animacy-congruous condition, it is plausible that the N400 was reduced as a result of its overlap with the P600. Further, in the semantic relatedness judgment study the target verbs were judged to be slightly more related to the argument NPs in the animacy-violated non-combinable condition than in the animacy-congruous non-combinable condition (2.17 vs. 1.76 on a 7-point scale). This difference in semantic relatedness might have led to the N400 difference between the two non-combinable conditions. Therefore, we argue that the current results do not provide clear evidence that lexical relations and animacy-congruity interact to modulate the N400 effect.

We propose that in simple sentence contexts like "*The student BA the math problem...*", the processor might fail to differentially expect upcoming information, i.e., a congruous vs. an incongruous verb, based only on information about the subject and the object noun phrases. Under such circumstances the ease of lexical access, and by hypothesis the N400 amplitude, should only be modulated by lexical association between

the target word and prior context. This, however, raises questions regarding the nature of expectation-generation mechanisms. Specifically, is the presumed difficulty in predicting an appropriate verb given its arguments reflective of a general property of the processor? Could the processor's apparent difficulty in predicting a verb be rectified? How do different word orders modulate the processor's success in predicting plausible thematic relations? Future work will need to address these questions by examining the effects of manipulations that are believed to facilitate predictions and by making carefully controlled comparisons across sentences with different word orders.

	N400				P600*			
Proposed contributors	Stored semantic relationships	Non-surface plausibility	Animacy (verb-dependent selectional restriction)	Animacy (linear order heuristic)	Surface plausibility	Non-surface plausibility	Competing representations	Animacy (verb-dependent selectional
Proposed generalizations / mechanisms	The N400 is reduced for words that are semantically/ associatively related.	Semantic anomalies do not elicit an N400 effect when a plausible non-surface interpretation is computed.	Violations of a verb's animacy selectional restriction can "switch off" semantic processing and thus reduce the N400.	The N400 is modulated by the match between the animacy of a noun and its linear position.	Implausible interpretations elicit a larger P600.	The presence of a plausible non-surface interpretation elicits a larger P600.	Mismatches between a non-compositional representation and a syntax-based representation elicit a larger P600.	Animacy violations lead to a larger P600.
Accounts for the Semantic P600 phenomenon								
Kolk et al. (2003); van Herten et al. (2005, 2006); van de Meerendonk et al. (2009, 2010)	X	X			X	X	X	
Kim & Osterhout (2005)	X	X				X		
Hoeks et al. (2004)	X	X				X		
Kuperberg et al. (2007)	X	X	X		X	X	X	X
Paczynski & Kuperberg (2011)	X			X^	X		X	X
Bornkessel-Schlesewsky & Schlesewsky (2008); Bornkessel-Schlesewsky et al. (2011)	X	X		X^	X		X	
Hagoort et al. (2009)	X	X				X		
Stroud & Phillips (2012)	X							X
Brouwer et al. (2012)	X				X			

^ Bornkessel-Scheleswsky & Schlesewsky (2008) proposed that violations of linear order animacy heuristics increase the N400, whereas Paczynski & Kuperberg (2011) proposed they reduce the N400.

* The role of context and task have also been discussed, although the generalization was not specified.

Table 2-6. Summary of some of the factors that have been proposed to account for the Semantic P600 phenomenon.

2.6 Conclusion

In this chapter we investigated the theoretical implications of the 'Semantic P600' phenomenon. In previous studies both the presence of a P600 effect and the absence of an N400 effect in role-reversed sentences have been regarded as two central pieces of evidence for a syntax-independent semantic composition mechanism. We presented two ERP studies that tested competing explanations for these two pieces of evidence. We found that the P600's sensitivity to semantic anomalies is not restricted to cases in which plausible non-surface interpretations are available, and argued that the presence of a P600 effect in role-reversed sentences does not constitute evidence for independent semantic composition. We also showed that the N400's insensitivity to role-reversals cannot be attributed to the presence or absence of animacy violations, and can instead be attributed to the lexical association between a verb and its arguments. We outlined a proposal in which the N400 reflects the ease of lexical access, and interpreted the lack of N400 effects in role-reversed sentences as reflecting the processor's temporary failure to generate specific lexical expectations in canonical vs. role-reversed sentences.

3 Prediction as a race against time

3.1 Introduction

In the previous chapter I argued that comprehenders rely on surface syntax to compute an accurate interpretation of a sentence and proposed that the N400's insensitivity to argument role-reversals may reflect comprehenders' temporary failure to generate differential predictions regarding an upcoming verb based on the word order of the preverbal arguments. In this chapter I situate this proposal in a larger context of previous research on linguistic prediction and explore potential causes for this apparent prediction failure. I present results from three ERP experiments in Mandarin Chinese that show that comprehenders' predictions can change depending on the amount of time available for predictive computations. I discuss the significance of the temporal dimension for our understanding of predictive computations.

3.2 Why is the N400's insensitivity to argument role-reversals surprising?

The results from Experiments 1 and 2 showed that the N400 to a verb is not modulated by the structural role of its arguments in the verb-final *Ba*-construction in Mandarin Chinese. These results appeared to be a striking exception to the well-established generalization that the N400 is sensitive to a word's predictability and have led to a proposal that challenges widely held assumptions about linguistic prediction.

As discussed in Chapter 2, the amplitude of the N400 response to a word can be taken to reflect the cost of accessing that word in long-term semantic memory (the lexicon; Deacon et al., 2000; Kutas & Federmeier, 2000; Lau et al., 2008). Under this view, the N400's sensitivity to a word's predictability is attributed to linguistic prediction – more expected words can be accessed from long-term semantic memory more easily because they are pre-activated ahead of time. This interpretation is supported by the fact that the N400's sensitivity to a word's predictability has been demonstrated in various contexts. For example, the effect of semantic priming on the N400 to words in isolation is modulated by the proportion of semantically related items in the experimental context (Lau, Holcomb & Kuperberg, 2013). In sentence comprehension, the amplitude of the N400 response to a word is inversely related to that word's cloze probability in a given sentence context (e.g., Federmeier & Kutas, 1999; Gunter, Friederici & Schriefers, 2000; Kutas & Hillyard, 1984). A word's cloze probability is the proportion of trials on which speakers continue the sentence context with that word in an untimed sentence fragment completion task.

In fact, many factors that are known to affect a word's cloze probability (e.g., sentence structure, event schemas, world knowledge) have also been shown to modulate the N400's amplitude during real-time comprehension (Bicknell et al., 2010; Hagoort et al., 2004; Kos et al., 2011; Paczynski & Kuperberg, 2012). Previous studies have shown that a word's predictability can modulate the N400 even when it appears in sentences with the same lexical contents. For example, Nieuwland & Kuperberg (2008) showed that, when the use of negation was pragmatically licensed, e.g., '*With proper equipment, scuba diving is/isn't ...* ', the N400 response to a target word (e.g., 'dangerous' or 'safe') was

modulated by its predictability. Meanwhile, Van Berkum and colleagues (Van Berkum, Van den Brink, Tesink, Kos & Hagoort, 2008) showed that, in an utterance like '*Every evening I drink some…*', listeners' N400 response to the word 'wine' is smaller when it was more predictable given the voice of the speaker (e.g., an adult vs. a child).

In this context, the observation that argument role-reversals do not modulate the N400 is rather surprising. This is because (i) by the structural roles of the arguments (e.g., whether the cop is the subject and the thief is the object or vice versa) should greatly affect the predictability of a verb (e.g., 'arrest') and (ii) information about the arguments' structural roles is prominent and unambiguously marked before the verb appears in the *Ba*-construction in Mandarin Chinese. Therefore, in this chapter I explore the potential cause for these surprising findings in three ERP experiments and discuss their implications for our understanding of linguistic prediction. By extending the basic paradigm used in Experiment 2, I examined the effects of argument role-reversals on comprehenders' ERPs at the verb in the *Ba*-construction in Mandarin Chinese in Experiments 3-5. In each experiment a further manipulation was included to examine what underlies the N400's (in)sensitivity to argument role-reversals.

3.3 Experiment 3: Can offline predictability measures capture online predictions?

In Experiment 3 I explored the possibility the N400 insensitivity to argument role-reversals in previous studies indicates that argument role-reversals did not impact the verb's predictability. In particular, I examined whether reversing the arguments' structural roles (by reversing their word order) can elicit an N400 effect when it has a greater impact on the verb's cloze probability. I compared the effects of argument role-

reversals in sentences in which the critical verb in the canonical control condition has high vs. low cloze probability (hereafter *high- vs. low-predictability sentences*; Table 3-1). In the role-reversed condition the verbs commonly have zero cloze probability. Therefore, argument role-reversals have a greater impact on the verb's cloze probability in high-predictability sentences than in low-predictability sentences.

If the word order (and thereby the structural roles) of the arguments can immediately impact comprehenders' predictions for the verb, then argument role-reversals should have a bigger effect on the N400 when they have a greater impact on the verb's cloze probability (i.e., in high-predictability sentences). Alternatively, if the word order of the arguments cannot impact comprehenders' predictions before the verb is presented, then the N400 should remain insensitive to thematic role-reversals in both high- and low-predictability sentences.

On the other hand, since lexical semantic association between the target verb and its arguments is much weaker in low predictability sentences than in high predictability sentences, a semantic illusion account, or an account that attributes the N400 insensitivity to a floor effect due to strong lexical association, would predict a significant N400 effect to role-reversals in low predictability sentences but not in high predictability sentences.

As a control comparison, I also examined comprehenders' sensitivity to a standard cloze probability manipulation across Experiments 3 to 5 (e.g., *From the sheep the herdsman collected a lot of wool …* vs. *Last year the boss bought a lot of wool…*).

3.3.1 Methods

Participants

Twenty-four students (19 female, mean age = 21 years, range 18-24 years) from South China Normal University participated in the current study. All participants were native speakers of Mandarin Chinese, were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. Data from one additional participant were excluded due to experimenter error. All participants gave informed consent and were paid 20 RMB/hour for their participation.

Cloze probability norming

Norming was done with 60 student volunteers at South China Normal University on a total of 190 pairs of subject-*BA*-object sentence frames, e.g., *cop BA thief* and its role-reversed counterpart *thief BA cop*. Participants were asked to provide the best continuations for the sentence frames. Individual participants saw only one version, with each normed by 30 participants.

Materials

The stimuli consisted of 120 pairs of sentences, each with a canonical and reversed argument order. All experimental sentences used the highly frequent SOV *ba*-construction in Mandarin Chinese. This construction requires a transitive verb, and the arguments' identity and their syntactic roles are evident before the critical verb because the particle *Ba* always follows the subject and immediately precedes the direct object. Role-reversed sentences were created by reversing the order of the pre-verbal arguments

in the canonical sentences, both of which were animate. Within the same item set the canonical and role-reversed sentences had an identical verb-argument triplet and differed only in the order of the arguments. No anomaly was evident before the critical verb. Sentences were extended beyond the target verb with words that were held constant across conditions within each item set.

Argument order was manipulated in sentences in which the verb had high vs. low predictability (Table 3-1). The verb had high predictability (average cloze = 63.6%, range 41.4 - 96.6%) in half of the canonical sentences and low predictability (average cloze = 6.6%, range 3.3 - 20.7%) in the other half. The verb was not predictable (0% cloze) in the role-reversed sentences. In order to avoid sentence-final wrap-up effects, the sentences were extended beyond the critical verb with words that were held constant across conditions within each item set.

Low-predictability

Canonical	老刘 把 鹦鹉 Mr. Liu BA parrot	**训练了** 好一段时间 。 **train** quite some time.
Role-reversed	鹦鹉 把 老刘 Parrot BA Mr. Liu	

"Mr. Liu trained the parrot for quite some time."
vs. "The parrot trained Mr. Liu for quite some time."

High-predictability

Canonical	警察 把 小偷 cop BA thief	**抓了** 回警局。 **arrest** (and bring back) to police station.
Role-reversed	小偷 把 警察 thief BA cop	

"The cop arrested the thief (and brought him back) to the station."
vs. "The thief arrested the cop (and brought him back) to the station."

Table 3-1. Experimental conditions and sample materials in Experiment 3.

Experimental sentences were distributed in two presentation lists, such that exactly one version of each item appearing in each list. Each list contained 120 experimental sentences (30 per condition), 60 plausible filler sentences from a previous experiment examining the effects of cloze probability (high: 88% vs. low: 27%) and 60 unrelated implausible filler sentences of similar length and structural complexity, so that the overall plausible-to-implausible ratio in each presentation list was 1:1. The sentences were presented in four blocks of 60 sentences each, and the order of the blocks was randomized across participants.

Procedure

Participants were comfortably seated about 100cm in front of a computer screen in a testing room. Sentences were presented one word at a time in a white font (30 pt simplified Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500ms. Each word appeared

on the screen for 400ms, followed by 200ms of blank screen. The last word of each sentence was marked with a period "。", followed 1000ms later by a response cue "?". Participants were instructed to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively and to indicate whether the sentence meaning was plausible by pressing one of two buttons. Prior to the experimental session, participants were presented with 6 practice trials to familiarize themselves with the task. The experimental session was divided into four blocks of 60 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 1.5 hours on average.

EEG Recording

EEG was recorded continuously from 30 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. The electro-oculogram (EOG) was recorded at four electrode sites; vertical EOG was recorded from electrodes placed above and below the left eye and the horizontal EOG was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5kΩ. The EEG and EOG recordings were amplified (band-pass filtered at DC - 200 Hz) and digitized online at 1kHz. Online recordings were referenced to the left mastoid. They were re-referenced to the average of the left and right mastoids and filtered using a 0.1 - 40 Hz band-pass filter offline.

ERP Data Analysis

Event-related potentials were computed separately for each participant and each condition for the 1000ms after the onset of the critical verb relative to a 100ms baseline preceding the critical verb. ERP data were evaluated for EOG or other artifacts. Trials contaminated by artifacts were excluded from the averaging procedure. This affected 7.8% of experimental trials, roughly equally distributed across conditions (ranging between 6.3 and 9.6% across conditions).

Statistical analyses on average voltage amplitudes were conducted separately for two time windows chosen based on previous literature and on visual inspection of the data: 350-450 ms for the N400, and 600-800ms for the P600. Data from high- and low-predictability sentences were analyzed separately. We conducted repeated measures ANOVAs that fully crossed Role-reversal (congruous vs. role-reversed) with Anteriority (anterior vs. central vs. posterior) and Laterality (Left vs. Midline vs. Right). The topographic factors effectively defined nine regions of interest (ROIs): left-anterior: F3, FC3; midline-anterior: FZ, FCZ; right-anterior: F4, FC4; left-central: C3, CP3; midline-central: CZ, CPZ; right-central: C4, CP4; left-posterior: P3, O1; midline-posterior: PZ, OZ; right-posterior: P4, O2. Univariate F-tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser correction. Since the current study was designed to modulate two ERP components with well-established topographic distributions, a paired-sample t-test was conducted to test for the predicted differences in each of the ROIs.

For the control comparison, the same procedures were used to analyze data from filler sentences to examine the effect of predictability (high vs. low cloze probability) instead of argument role-reversal.

3.3.2 Results

Plausibility Judgments

Participants' average plausibility judgment accuracy in each condition is shown in Table 3-2. Across both predictable and non-predictable sentences, participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 90.6%.

	Target Response	Percent accurate (SE)
Low predictability, Canonical	Yes	88.3 (1.8)
Low predictability, Role-reversed	No	90.4 (1.2)
High predictability, Canonical	Yes	90.3 (1.4)
High predictability, Role-reversed	No	93.3 (0.9)

Table 3-2. Target response and accuracy on plausibility judgment task in Experiment 3.

Event-related Potentials (Control comparison)

A control predictability (cloze probability) manipulation in the filler sentences elicited a clear N400 effect and a frontally distributed late positivity (Figure 3-1). Statistical analyses in the 350-450 ms time interval showed that expected words elicit a smaller N400 response than unexpected words (Table 3-3). A significant predictability × anteriority × laterality interaction showed that the effect was largest at midline central and posterior sites. In the 600-800 ms time interval the ERPs were more positive to unexpected than expected words. A significant predictability × anteriority interaction showed that, unlike the P600 effect which has a posterior distribution, this late positivity was largest at frontal sites. This establishes that the participants showed standard electrophysiological responses to cloze manipulations.

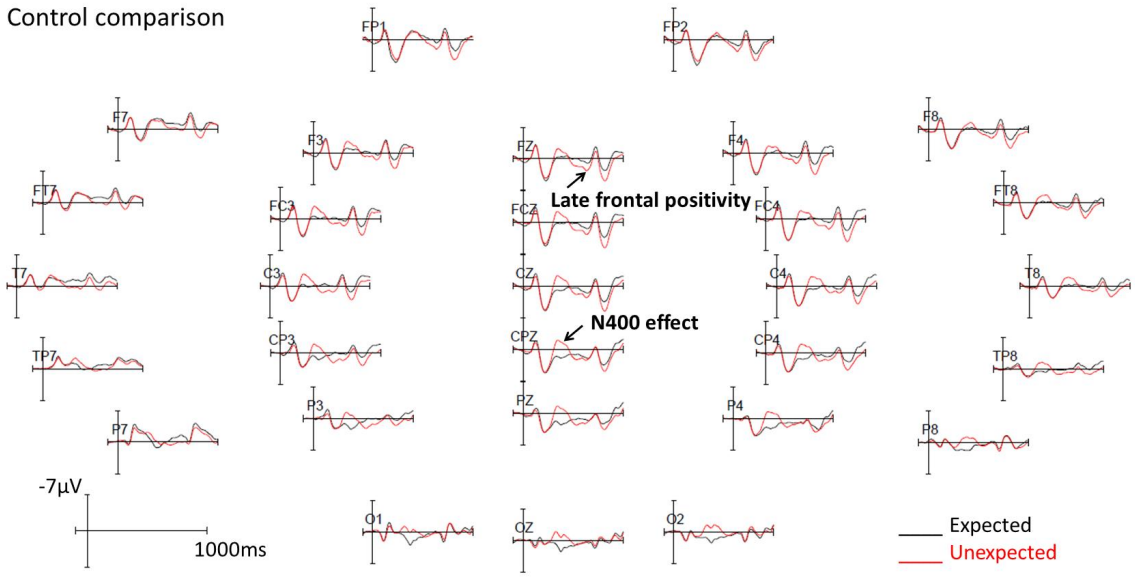

Figure 3-1. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 3.

	df	350-450ms	600-800ms
Omnibus ANOVA			
pred	1,23	20.3**	2.25
pred * ant	2,46	7.21**	5.18*
pred * lat	2,46	3.56*	< 1
pred * ant * lat	4,92	2.91*	1.67
ROI analyses			
Left anterior	23	2.27*	1.43
Midline anterior	23	2.43*	2.06^
Right anterior	23	2.41*	2.08*
Left central	23	3.53**	2^
Midline central	23	4.15**	< 1
Right central	23	4.61**	1.74^
Left posterior	23	4.88**	< 1
Midline posterior	23	4.83**	< 1
Right posterior	23	5.85**	< 1

Factors: *pred* = predictability; *ant* = anteriority; *lat* = laterality.
** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-3. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 3.

Event-related Potentials (Experimental comparisons)

Figure 3-2 and Figure 3-3 show the grand average ERPs at the target word in the low- and high- predictability sentences respectively. Figure 3-4 shows the grand average ERPs at CPZ and the topographic distribution of the effects at the target word in the 350-450ms and 600-800ms time intervals across the control and experimental comparisons.

Argument role-reversals elicited a clear P600 effect in both high- and low-predictability sentences. Further, although the amplitude of the N400 at the target verb was generally higher in the low-predictability sentences than in the high-predictability sentences, it did not differ between the role-reversed and canonical sentences in either case. These observations were confirmed by the statistical analyses, which were

conducted on high- and low-predictability sentences separately (Table 3-4). Repeated measures ANOVAs in the 350-450 ms interval revealed no significant effects involving reversal in either high- or low-predictability sentences. In the 600-800 ms interval a significant main effect of reversal was observed in both high- and low-predictability conditions. Paired-sample t-tests in individual regions of interest (ROIs) revealed that the P600 effect was statistically significant at central and posterior sites in both cases. This experiment established that, contrary to its widely-demonstrated sensitivity to a word's cloze probability, the N400 can be completely insensitive to differences in a verb's cloze probability when those differences are due to the word order of preverbal arguments.

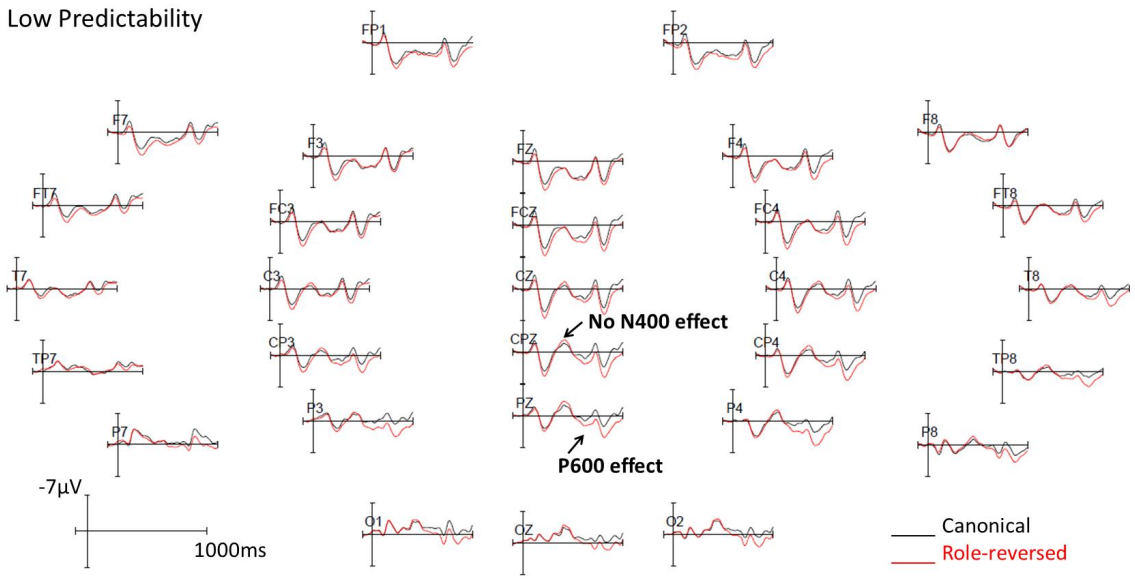

Figure 3-2. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the low-predictability sentences in Experiment 3.

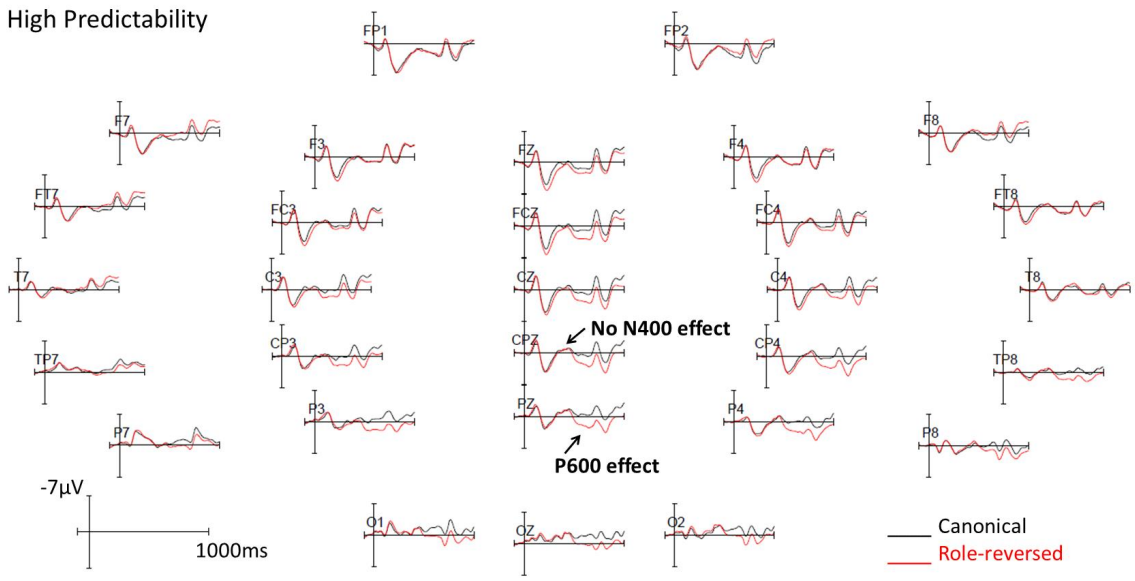

Figure 3-3. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the high-predictability sentences in Experiment 3.

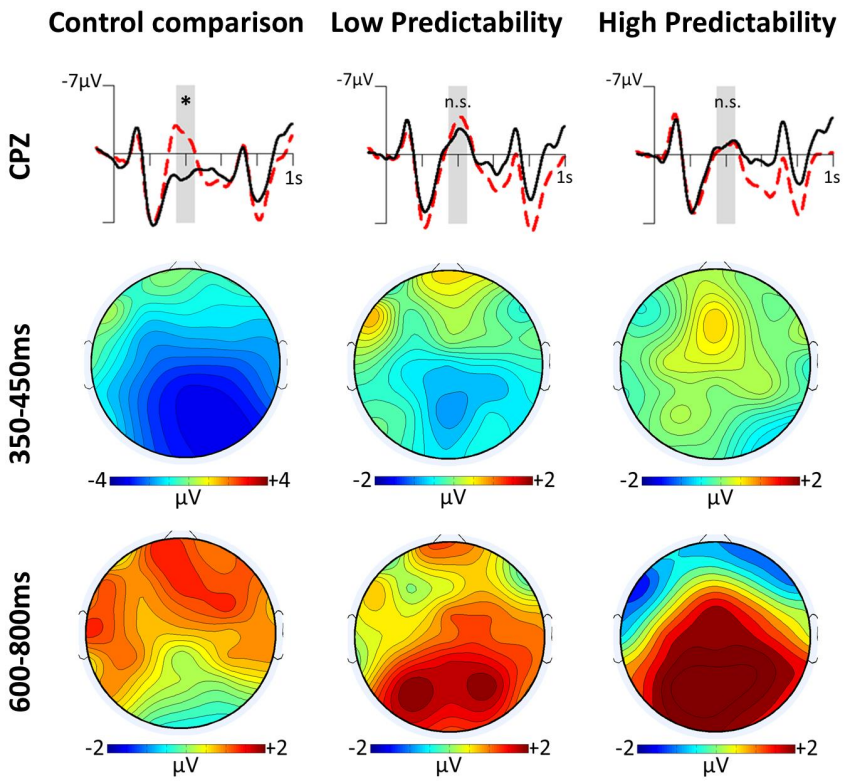

Figure 3-4. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 3.

	df	350-450ms	600-800ms
Low Predictability			
rev	1,23	< 1	5.81*
rev * ant	2,46	< 1	2.14
rev * lat	2,46	< 1	1.54
rev * ant * lat	4,92	1.64	2.1
ROI analyses			
Left anterior	23	< 1	< 1
Midline anterior	23	< 1	1.29
Right anterior	23	< 1	1.52
Left central	23	< 1	1.55
Midline central	23	1.26	2.08*
Right central	23	1.18	2.13*
Left posterior	23	< 1	3.09**
Midline posterior	23	1.46	3**
Right posterior	23	< 1	3.73**
High Predictability			
rev	1,23	< 1	18.11**
rev * ant	2,46	< 1	7.86**
rev * lat	2,46	< 1	4.24*
rev * ant * lat	4,92	2.33^	3.92**
ROI analyses			
Left anterior	23	< 1	< 1
Midline anterior	23	< 1	2.73*
Right anterior	23	< 1	1.2
Left central	23	< 1	3.59**
Midline central	23	< 1	3.68**
Right central	23	< 1	4.62**
Left posterior	23	< 1	6.07**
Midline posterior	23	< 1	6.6**
Right posterior	23	< 1	5.43**

Factors: *rev* = reversal; *ant* = anteriority; lat = laterality.

** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-4. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 3.

3.3.3 Discussion

The current results showed that (i) differences in target word's cloze probability elicited a clear N400 effect and a late frontal positivity in the control comparison, (ii) role-reversal anomalies in the experimental sentences were readily detected and elicited a clear P600 effect at the verb, but (iii) the N400 remained insensitive to argument role-reversals regardless of the predictability of the target verb.

First, the presence of a clear N400 effect in the control comparison established that the participants in the current experiment displayed a standard N400 effect to a common cloze probability manipulation in which the same target word has high vs. low cloze probability in different sentence contexts. Further, the N400 effect was followed by a frontally distributed late positivity. As shown in Figure 3-4, the topographic distribution of this late positivity is distinct from the posteriorly distributed P600 effect in the experimental sentences. As the unexpected target words in the filler sentences were nonetheless plausible, the observation of a late frontal positivity in the control comparison is consistent with Van Petten and Luka's (2012) generalization that plausible but unexpected words tend to elicit a larger late frontal positivity than expected words, while implausible words tend to elicit a posteriorly distributed late positivity (a P600 effect). Further, the sentence contexts in the filler sentences were all strongly constraining (maximum cloze value ranges from 0.53 to 1). Similarly, Federmeier et al. (2007) observed a late frontal positivity in strongly constraining sentence contexts not in weakly constraining contexts. This suggests that Van Petten and Luka's (2012) observation that plausible but unexpected words elicited a late frontal positivity only about two thirds of

the time may be attributed to differences in contextual constraints in the materials used across studies.

In the experimental comparisons, the presence of a clear P600 effect and high plausibility judgment accuracy in both high- and low- predictability sentences showed that comprehenders computed an accurate interpretation of the sentences using structural role information and readily detected the implausibility resulted from argument role-reversals. A comparison of the waveforms in Figure 3-2 and Figure 3-3 shows that the time course of the P600 effect was roughly the same across the high- and low-predictability sentences (see also Figure 3-15). Figure 3-4 shows that the effects also had the same topographic distribution. Taken together, these findings suggest that the P600 is modulated by a common variable across these sentences, namely, the plausibility of the sentence as it is being interpreted incrementally.

On the other hand, no significant N400 effect was observed in either high- or low-predictability sentences. This was not predicted by a semantic illusion account (e.g., Kim & Osterhout, 2005). Given that the verb and its arguments were much less lexically associated (or "semantically attractive" to one another) in the low-predictability sentences than in the high-predictability sentences, an account that attributes N400 insensitivity to semantic illusion would predict a significant N400 effect to role-reversal anomalies in the low-predictability sentences. Also, had strong lexical associations between the verb and its arguments in previous studies given rise to a floor effect and thus a lack of N400 effect, we should expect to see a significant N400 effect in the low predictability sentences since a floor effect is much less likely to occur in that condition. However, this

prediction was not borne out by the current results. In fact, although the amplitude of the N400 response to the verb was higher in the low-predictability sentences than in the high-predictability sentences, argument role-reversals did not impact the N400 in either case.

Meanwhile, the current N400 results were also not expected by a prediction account if we assume that the structural roles of the preverbal arguments can impact comprehenders' predictions about the verb. The observation that role-reversals did not elicit an N400 effect even when the verb is highly predictable based on offline cloze measures suggests information about preverbal arguments fails to impact comprehenders' verb predictions when the verb was presented immediately after its arguments. This constitutes a clear exception to the generalization that the N400 is modulated by a word's cloze probability and raises many new questions – Did comprehenders simply not use structural role information for computing predictions, or did they fail to incorporate such information in time?

3.4 Experiment 4: How do linguistic predictions develop over time?

In Experiment 4 I examined the possibility that structural role information has a delayed impact on comprehenders' predictions. In order to examine whether the N400 becomes sensitive to cloze probability differences that result from argument role-reversals when comprehenders have slightly more time to compute verb predictions, I manipulated the linear distance between the verb and its arguments by varying the position of an adverbial time expression (e.g., "*zai shangxingqi*", last week).

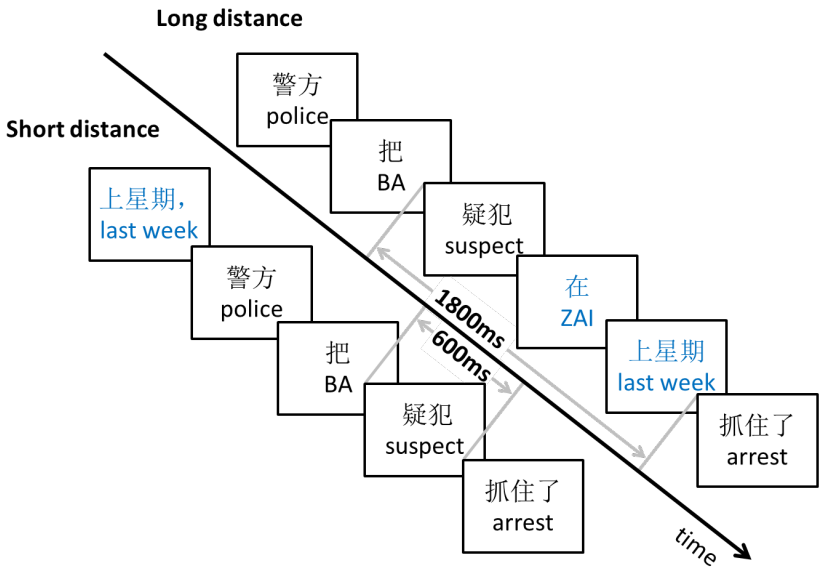

Figure 3-5. Illustration of the distance manipulation in Experiment 4.

As illustrated in Figure 3-5, placing the temporal phrase after the second argument effectively delays the presentation of the verb relative to its arguments by 1200ms. Since temporal phrases appear in sentence-initial position (as in the short-distance condition) more often than between the arguments and the verb (as in the long-distance condition) in the *ba*-construction in Mandarin Chinese, comparisons were made between canonical and role-reversed sentences with the same configuration only (i.e., within the same level of distance).

If structural role information can impact comprehenders' predictions for the verb within the extended time interval between the arguments and the verb, then the N400 is expected to become sensitive to role-reversals when the verb is further away from its arguments (in the long-distance condition). Alternatively, if the N400's insensitivity to role-reversals reflects a genuine failure for structural role information to impact linguistic

predictions, then the N400 should remain insensitive to role-reversals regardless of the amount of time elapsed between the arguments and the verb.

3.4.1 Methods

Participants

Twenty-four students (22 female, mean age = 22 years, range 19-28 years) from South China Normal University participated in the current study. Informed consent was obtained from all participants. All were right-handed, native Mandarin Chinese speakers with normal or corrected-to-normal vision and were paid for their participation.

Materials

Table 3-5 shows a sample set of experimental materials. As in Experiment 3, the stimuli in consisted of 120 pairs of sentences, each with a canonical and reversed argument order and used the highly frequent SOV Ba(把)-construction in Mandarin Chinese. Argument order was manipulated in sentences in which the verb was closer vs. further away from the arguments (Figure 3-5). The linear distance between the arguments and the verb was manipulated by varying the position of an adverbial time expression (e.g., "*zai shangxingqi*", *last week*). The time expression appeared either at the start of the sentence, where it did not interrupt the subject-object-verb sequence (the short-distance condition), or between the direct object and the verb, where it created a delay between the arguments and the verb (the long-distance condition). Therefore, the lexical material in the sentence before the verb was the same across all four conditions, and the conditions differed only in the order of presentation. The verb was predictable in the canonical

sentences (average cloze = 37.4%, range 16.7 - 69.0%) but not in the role-reversed sentences (0% cloze). The verb was not always the most likely continuation in the canonical sentences.

Short distance			
Canonical	上星期，	警方 把 疑犯 Police BA suspect	抓住了…
Role-reversed	Last week,	疑犯 把 警方 Suspect BA police	**arrest** ….
	"The cop arrested the thief last week..." *vs. "The thief arrested the cop last week…"*		
Long distance			
Canonical		警方 把 疑犯 Police BA suspect	在 上星期 抓住了…
Role-reversed		疑犯 把 警方 Suspect BA police	ZAI last week **arrest** …
	"The cop arrested the thief last week..." *vs. "The thief arrested the cop last week…"*		

Table 3-5. Experimental conditions and sample materials in Experiment 4.

Experimental sentences were distributed in four presentation lists, such that exactly one version of each item appearing in each list. Each list contained 120 experimental sentences (30 per condition) and the same set of 120 filler sentences used in Experiment 3. The sentences were presented in four blocks of 60 sentences each, and the order of the blocks was randomized across participants.

Procedure

The experimental procedures were identical to those in Experiment 3.

EEG Recording

The EEG recording procedures were identical to those in Experiment 3.

ERP Data Analysis

The procedures for data analysis were identical to those in Experiment 3. Data from short- and long-distance sentences were analyzed separately. A total of 6.6% of experimental trials, roughly equally distributed across conditions (5% - 8.3%), were excluded from the averaging procedure due to artifacts.

3.4.2 Results

Plausibility Judgments

Table 3-6 shows participants' average plausibility judgment accuracy in each condition. Regardless of the distance between the arguments the verb, participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 89.5%.

	Target Response	Percent accurate (SE)
Short distance, Canonical	Yes	93.4 (1.1)
Short distance, Role-reversed	No	87.5 (2.2)
Long distance, Canonical	Yes	89.9 (1.4)
Long distance, Role-reversed	No	86.7 (1.7)

Table 3-6. Target response and accuracy on plausibility judgment task in Experiment 4.

Event-related Potentials (Control comparison)

As in Experiment 3, participants showed a clear N400 effect and a frontally distributed late positivity to the standard cloze probability manipulation in the filler sentences (Figure 3-6). Statistical analyses (Table 3-7) showed that expected words elicit a smaller N400 response than unexpected words in the 350-450 ms time interval, but the ERPs were more positive to unexpected than expected words in the 600-800 ms time

interval. As in Experiment 3, the N400 effect has a central-posterior distribution while the late positivity was more frontally distributed.

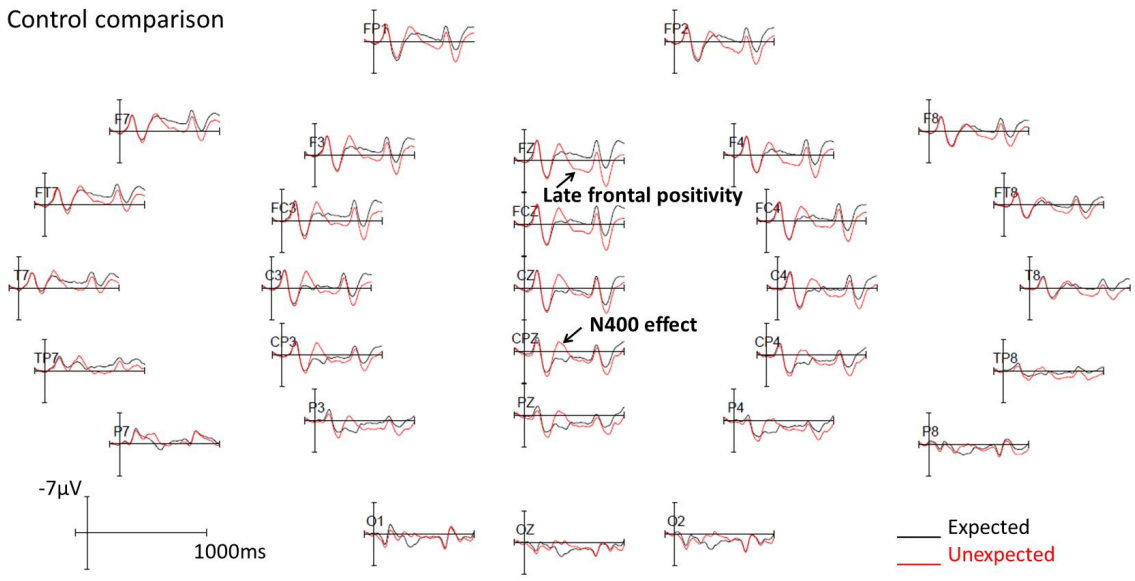

Figure 3-6. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 4.

	df	350-450ms	600-800ms
Omnibus ANOVA			
pred	1,23	6.93*	3.11^
pred * ant	2,46	10.44**	8.69**
pred * lat	2,46	< 1	< 1
pred * ant * lat	4,92	3.07*	5**
ROI analyses			
Left anterior	23	< 1	2.39*
Midline anterior	23	< 1	2.98**
Right anterior	23	< 1	2.11*
Left central	23	3.04**	1.36
Midline central	23	3.15**	< 1
Right central	23	2.66*	1.65
Left posterior	23	3.81**	1.1
Midline posterior	23	4.3**	< 1
Right posterior	23	4.4**	< 1

Factors: *pred* = predictability; *ant* = anteriority; *lat* = laterality.

** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-7. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 4.

Event-related Potentials (Experimental comparisons)

Figure 3-7 and Figure 3-8 show the grand average ERPs at the target word in the short and long distance conditions respectively. Figure 3-9 shows the grand average ERPs at CPZ and the topographic distribution of the effects at the target word in the 350-450ms and 600-800ms time intervals across the control and experimental comparisons.

A P600 effect was present in both short- and long-distance conditions in Experiment 4. Meanwhile, argument role-reversals did elicit an N400 effect, but only in the long-distance condition (Figure 3-8). These observations were confirmed by the statistical analyses, which were conducted on the short- and long-distance conditions separately (Table 3-8). Repeated measures ANOVAs in the 350-450 ms interval revealed

no significant effects involving reversal in the short-distance condition; a significant reversal × laterality interaction was observed in the long-distance condition. ROI analyses revealed that the N400 effect was statistically significant in several central and posterior sites. Repeated measures ANOVAs in the 600-800 ms interval revealed a significant main effect of reversal in both short- and long-distance conditions. ROI analyses revealed that the P600 effect was statistically significant across central and posterior sites in both cases. This experiment showed that the N400 at the verb became sensitive to argument role-reversals when the time-interval between the presentation of the arguments and the verb was widened.

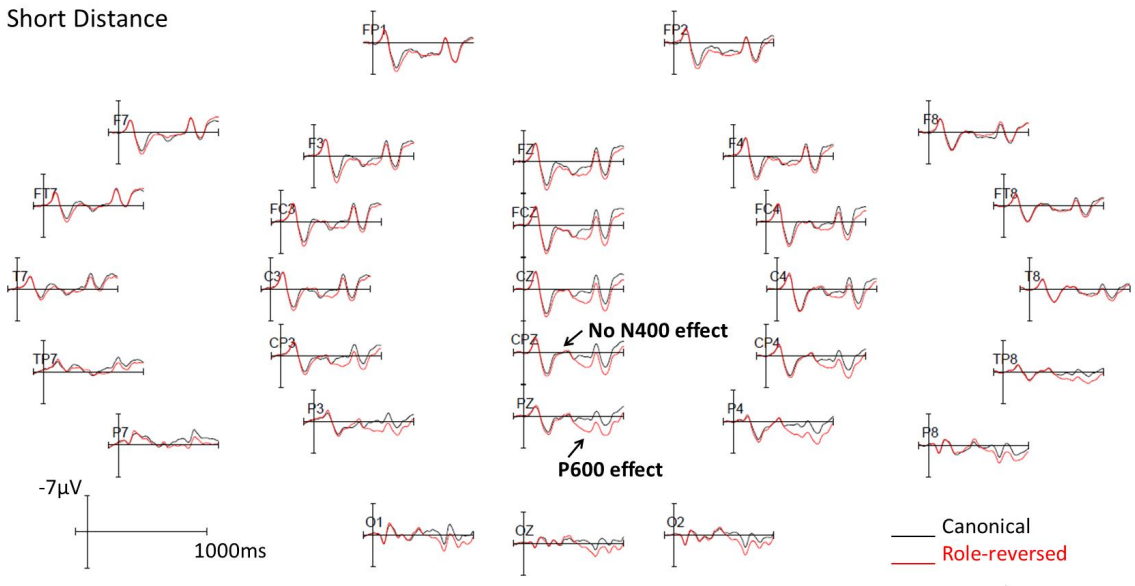

Figure 3-7. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the short-distance sentences in Experiment 4.

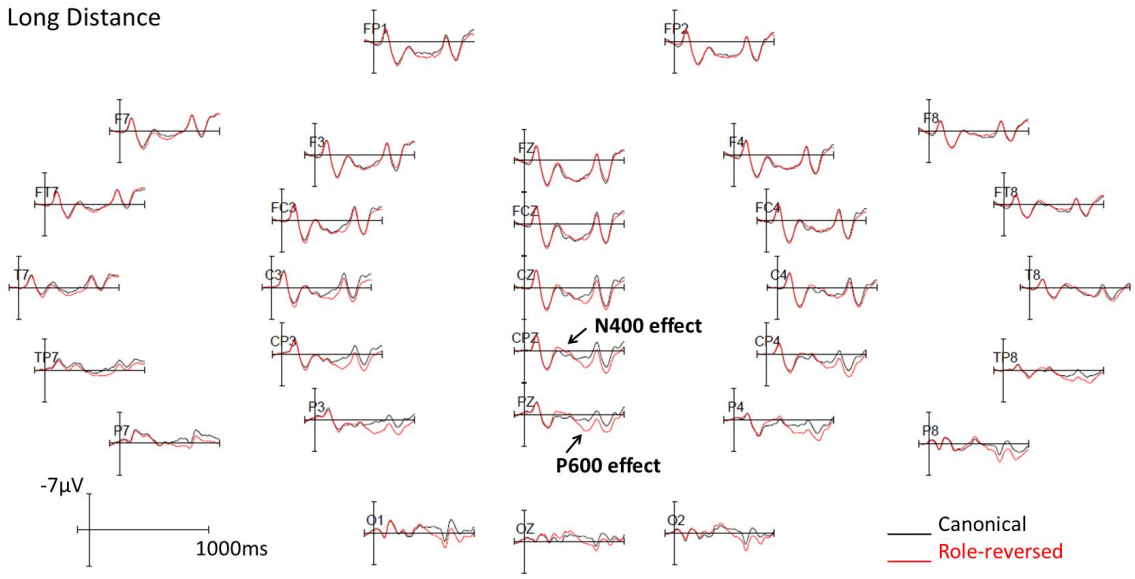

Figure 3-8. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the long-distance sentences in Experiment 4.

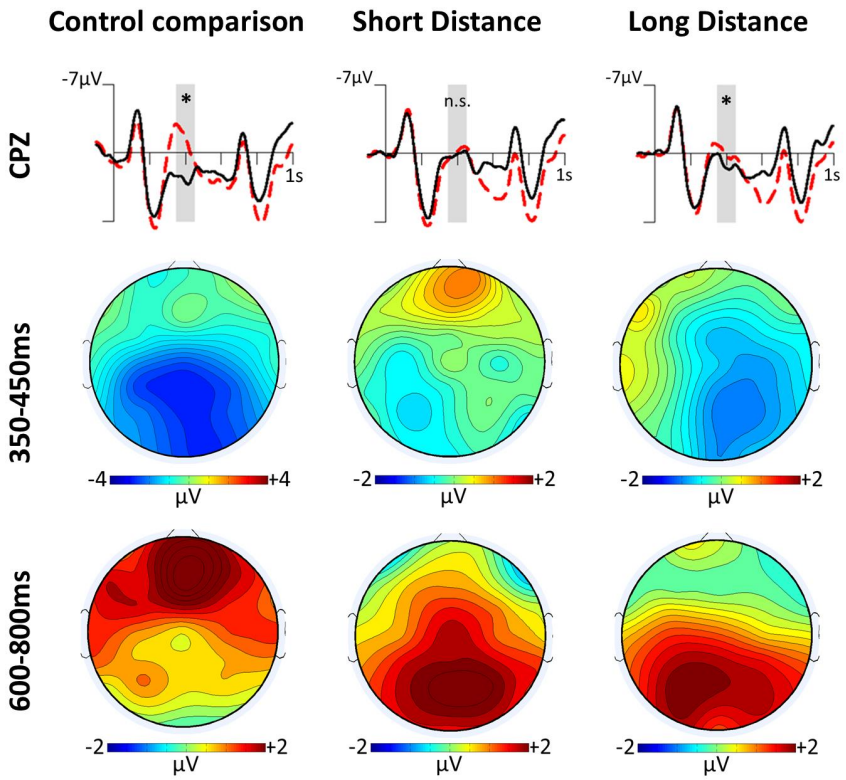

Figure 3-9. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 4.

	df	350-450ms	600-800ms
Short Distance			
rev	1,23	< 1	8.86**
rev * ant	2,46	< 1	2.49
rev * lat	2,46	< 1	2.36
rev * ant * lat	4,92	< 1	< 1
ROI analyses			
Left anterior	23	< 1	1.33
Midline anterior	23	< 1	1.82^
Right anterior	23	< 1	1.56
Left central	23	< 1	2.26*
Midline central	23	< 1	3.15**
Right central	23	< 1	2.64*
Left posterior	23	1.04	3.67**
Midline posterior	23	< 1	3.42**
Right posterior	23	< 1	3.47**
Long Distance			
rev	1,23	1.85	6.77*
rev * ant	2,46	< 1	9.67**
rev * lat	2,46	6.46**	< 1
rev * ant * lat	4,92	1.14	< 1
ROI analyses			
Left anterior	23	< 1	< 1
Midline anterior	23	< 1	< 1
Right anterior	23	< 1	< 1
Left central	23	< 1	2.99**
Midline central	23	1.41	2.92**
Right central	23	2.1*	2.03^
Left posterior	23	1.04	5.6**
Midline posterior	23	2.52*	3.44**
Right posterior	23	2.47*	3.44**

Factors: *rev* = reversal; *ant* = anteriority; lat = laterality.

** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-8. ANOVA F-values and ROI analysis t-values at the target verb for the short- and long-distance comparisons in Experiment 4.

3.4.3 Discussion

The most important finding from Experiment 4 is that the N400 became sensitive to role-reversals when the presentation of the target verb was delayed (in the long-distance condition). Other aspects of the results were consistent with those from previous experiments: (i) differences in target word's cloze probability elicited a clear N400 effect and a late frontal positivity in the control comparison, (ii) role-reversal anomalies in the experimental sentences were readily detected and elicited a clear P600 effect at the verb, and (iii) the N400 was insensitive to argument role-reversals when the verb immediately followed its arguments (in the short-distance condition). Below I will focus on the impact of the distance manipulation on the N400's sensitivity to argument role-reversals and its theoretical implications. For a discussion of the significance of other aspects of the current results please refer to page 76.

As I reviewed in Chapter 2, previous studies have commonly observed that the implausibility resulting from reversals of animate preverbal arguments elicited only a P600 effect and no N400 effect. Although this phenomenon has been reported across studies using different languages, a common characteristic among the materials used in these studies is the close proximity between the arguments and the target verb. For example, van Herten et al. (2005, 2006) used experimental materials in which the target verb followed the second argument with a 645ms SOA. Similarly, in Experiments 1-3 reported above, the target verb followed the second argument with a 600ms SOA. The N400's insensitivity to the verb's cloze probability in these cases suggests that

information about preverbal arguments' structural roles cannot immediately impact comprehenders' prediction about the verb.

By manipulating the linear distance between the arguments and the verb, the current experiment showed that the N400 became sensitive to argument role-reversals when the presentation of the target verb was delayed. The N400's sensitivity to argument role-reversals in the long distance condition (1800ms SOA) stands in sharp contrast with its insensitivity in the short distance condition (600ms SOA) and in all previous studies. Since the sentences in the short- and long-distance conditions contained identical words and differed only by the location of the temporal phrase, they provide qualitatively the same information for computing verb predictions. Instead, the contrast between N400's sensitivity in these cases is likely attributable to the amount of time elapsed between the arguments and the verb.

Specifically, I propose that the N400's reemerged sensitivity to argument role-reversals reveals that the widened time interval allowed information about the arguments' structural roles to impact comprehenders' verb prediction before the verb appeared. Under this account, the N400 effect emerged because the verb became more pre-activated in the canonical condition than the role-reversed condition when comprehenders were given slightly more time to incorporate structural role information into their predictive computations. This implies that the N400 did not become sensitive to the implausibility of role-reversed sentences per se – the effect of role-reversals on the N400 should be tightly linked to the verb's predictability given its arguments.

3.5 Experiment 5: Prediction vs. Integration over the extended time interval

In this experiment I investigated what gave rise to the N400's reemerged sensitivity to argument role-reversals in the long-distance condition in Experiment 4. In order to test the hypothesis that role-reversals elicited a significant N400 effect because the widened time interval allowed comprehenders to incorporate structural role information into their verb predictions, I combined the low- vs. high-predictability manipulation of Experiment 3 with the long-distance structure introduced in Experiment 4. The materials used in this experiment were directly adapted from those in Experiment 3 by placing an adverbial time expression between the direct object and the verb.

Following the logic outlined in Experiment 3, I examined whether argument role-reversals have a bigger effect on the N400 when they have a greater impact on comprehenders' predictions for the verb. If the effect of distance observed in Experiment 4 reflects limitations on how quickly structural role information impacts predictions, then the effect of role-reversals on the N400 should be tightly linked to the verb's predictability. That is, argument role-reversals should elicit a significant N400 effect only when they have a clear impact on the verb's predictability, i.e., in the high-predictability conditions.

Alternatively, if the N400 became sensitive to role-reversals in the long-distance condition simply because the widened time interval allowed comprehenders to build a more robust semantic representation of the sentence context and thus detect the implausibility at the verb more quickly, then the N400 should be modulated by

plausibility. That is, role-reversals should elicit a significant N400 effect in the current experiment regardless of the verb's predictability.

3.5.1 Methods

Participants

Twenty-four students (19 female, mean age = 19.9 years, range 18-24 years) from South China Normal University participated in the current study. Informed consent was obtained from all participants. All were right-handed, native Mandarin Chinese speakers with normal or corrected-to-normal vision and were paid for their participation.

Materials

Low-predictability		
Canonical	老刘 把 鹦鹉 Mr. Liu BA parrot	在 那年夏天 **训练了** 好一段时间 。 ZAI that summer **train** quite some time.
Role-reversed	鹦鹉 把 老刘 Parrot BA Mr. Liu	
	"Mr. Liu trained the parrot for quite some time that summer." *vs. "The parrot trained Mr. Liu for quite some time that summer."*	
High-predictability		
Canonical	警察 把 小偷 cop BA thief	在 那天傍晚 **抓了** 回警局。 ZAI that evening **arrest** (and bring back) to police station.
Role-reversed	小偷 把 警察 thief BA cop	
	"The cop arrested the thief (and brought him back) to the station that evening." *vs. "The thief arrested the cop (and brought him back) to the station that evening."*	

Table 3-9. Experimental conditions and sample materials in Experiment 5.

As shown in Table 3-9, the design of Experiment 5 was identical to that of Experiment 3 with the exception that an adverbial time expression always appeared

between the direct object and the verb, as in the long-distance condition in Experiment 4. The materials of Experiment 5 were adapted directly from those in Experiment 3.

Procedure

The experimental procedures were identical to those in Experiments 3 and 4.

EEG Recording

The EEG recording procedures were identical to those in Experiments 3 and 4.

ERP Data Analysis

The procedures for data analysis were identical to those in Experiment 3. A total of 11.5% of experimental trials, roughly equally distributed across conditions (9% - 13.6%), were excluded from the averaging procedure due to artifacts.

3.5.2 Results

Plausibility Judgments

Table 3-10 shows participants' average plausibility judgment accuracy in each condition. Participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 90.9%.

	Target Response	Percent accurate (SE)
Low predictability, Canonical	Yes	88.0 (2.6)
Low predictability, Role-reversed	No	89.5 (1.5)
High predictability, Canonical	Yes	90.7 (2.0)
High predictability, Role-reversed	No	95.1 (1.1)

Table 3-10. Target response and accuracy on plausibility judgment task in Experiment 5.

Event-related Potentials (Control comparison)

As in Experiments 3 and 4, a cloze probability manipulation in the filler sentences elicited a clear N400 effect and a frontally distributed late positivity (Figure 3-10). Statistical analyses (Table 3-11) showed that expected words elicit a smaller N400 response than unexpected words in the 350-450 ms time interval, but the ERPs were more positive to unexpected than expected words in the 600-800 ms time interval. As in the control comparison in Experiments 3 and 4, the N400 effect was broadly distributed while the late positivity was largest at frontal sites.

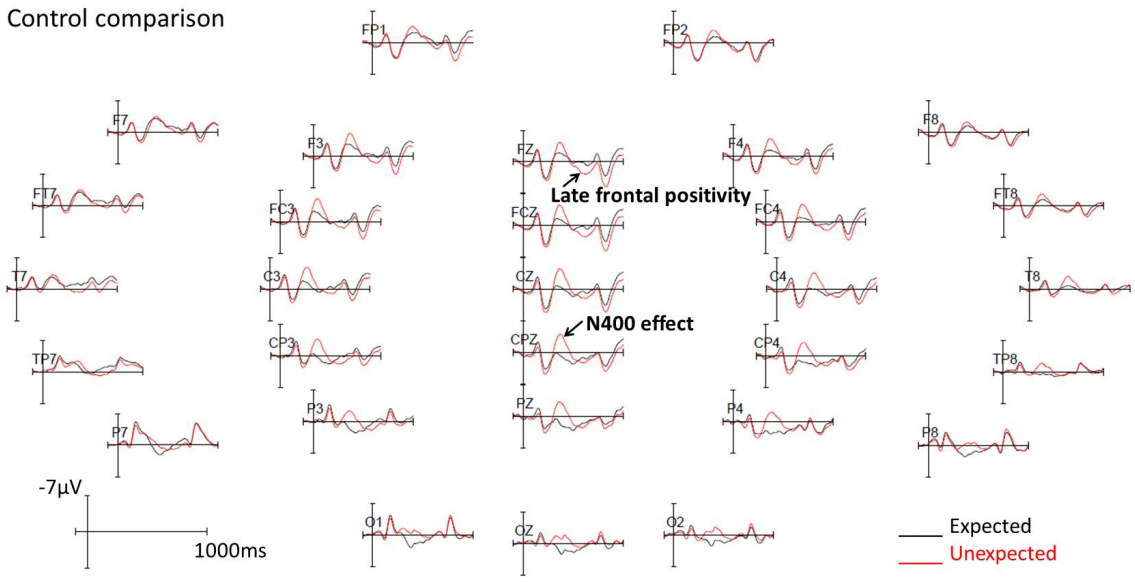

Figure 3-10. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 5.

	df	350-450ms	600-800ms
Omnibus ANOVA			
pred	1,23	45.8**	2.09
pred * ant	2,46	9.02**	4.26*
pred * lat	2,46	< 1	1.68
pred * ant * lat	4,92	3.51*	4.13**
ROI analyses			
Left anterior	23	4.08**	1.89^
Midline anterior	23	3.2**	2.77*
Right anterior	23	3.69**	1.42
Left central	23	6.67**	2.1*
Midline central	23	6.26**	1.16
Right central	23	6.92**	< 1
Left posterior	23	8.09**	< 1
Midline posterior	23	7.7**	< 1
Right posterior	23	7.82**	< 1

Factors: *pred* = predictability; *ant* = anteriority; *lat* = laterality.

** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-11. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 5.

Event-related Potentials (Experimental comparisons)

Figure 3-11 and Figure 3-12 show the grand average ERPs at the target word in the low- and high- predictability sentences respectively. Figure 3-13 shows the grand average ERPs at CPZ and the topographic distribution of the effects at the target word in the 350-450ms and 600-800ms time intervals across the control and experimental comparisons.

As in Experiment 3, a P600 effect was present in both high- and low-predictability conditions. Meanwhile, argument role-reversals elicited an N400 effect in the high-predictability condition only. These observations were confirmed by the statistical analyses (Table 3-12). Repeated measures ANOVAs in the 350-450 ms interval revealed no significant effects involving reversal in the low-predictability condition. In

the high-predictability condition, a marginally significant main effect of reversal and a significant reversal × anteriority × laterality interaction were observed. ROI analyses revealed that the N400 effect was statistically significant across several central and posterior sites. Repeated measures ANOVAs in the 600-800 ms interval revealed a significant main effect of reversal and reversal × anteriority interaction in both high- and low-predictability conditions. ROI analyses revealed that the P600 effect was statistically significant across all central and posterior sites in both cases. This experiment showed that the benefit of the added time interval is found only in situations where the argument roles are predictive of the verb.

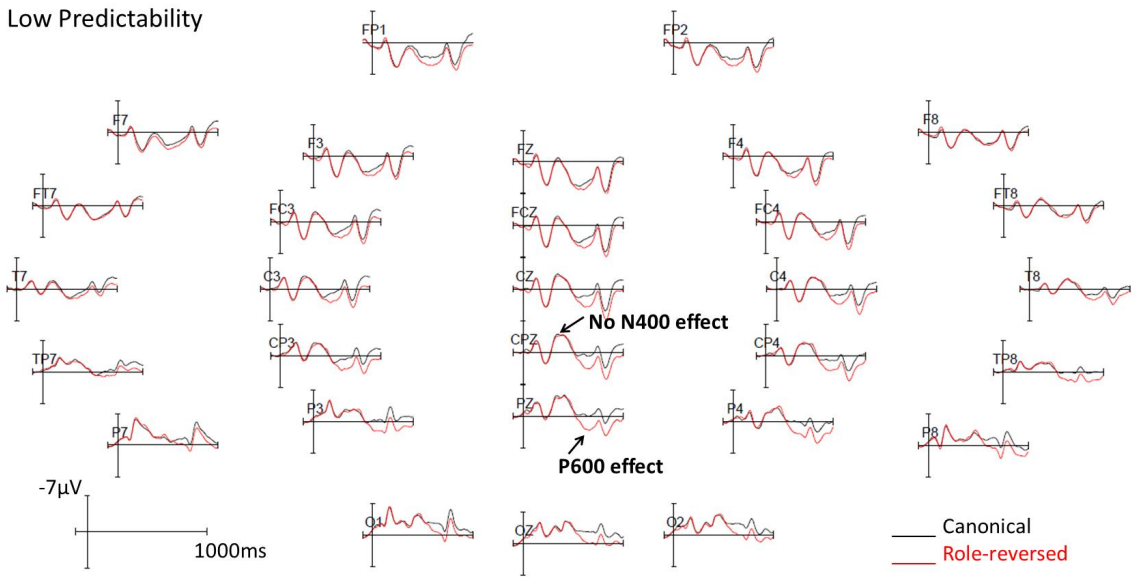

Figure 3-11. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the low-predictability sentences in Experiment 5.

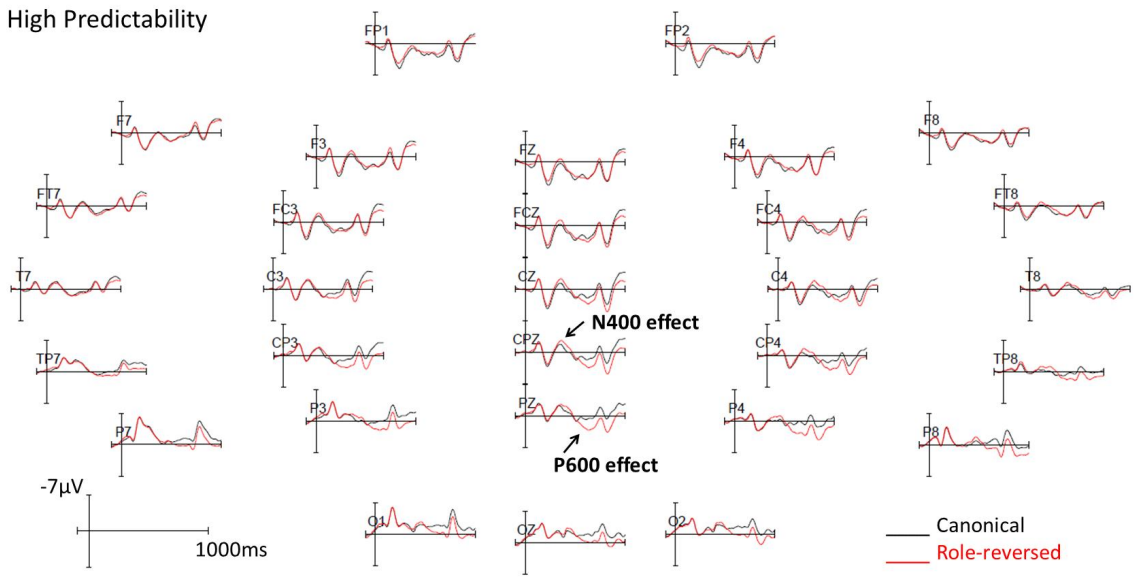

Figure 3-12. Grand average ERPs at the target verb in canonical (black) and role-reversed (red) conditions in the high-predictability sentences in Experiment 5.

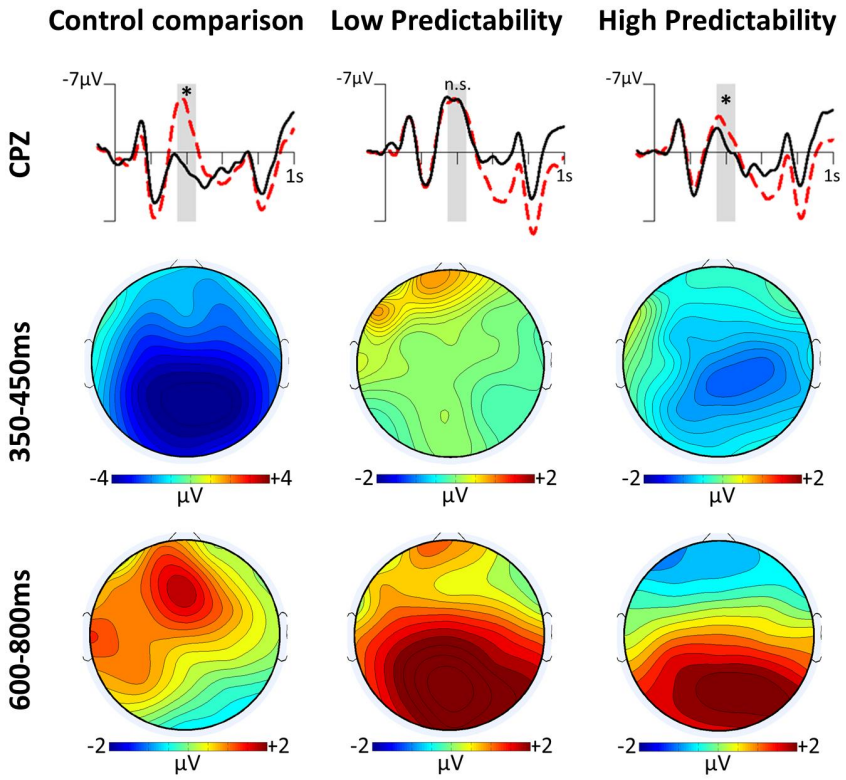

Figure 3-13. Grand average ERPs at CPZ (top) and topographic distribution of ERP effects in 350-450 ms (middle) and 600-800ms (bottom) across experimental and control comparisons in Experiment 5.

	df	350-450ms	600-800ms
Low Predictability			
rev	1,23	< 1	10.76**
rev * ant	2,46	< 1	8.58**
rev * lat	2,46	< 1	1
rev * ant * lat	4,92	< 1	1.64
ROI analyses			
Left anterior	23	< 1	1.58
Midline anterior	23	< 1	1.33
Right anterior	23	< 1	< 1
Left central	23	< 1	3.45**
Midline central	23	< 1	3.47**
Right central	23	< 1	2.96**
Left posterior	23	< 1	3.97**
Midline posterior	23	< 1	4.72**
Right posterior	23	< 1	4.45**
High Predictability			
rev	1,23	3.85^	6.7*
rev * ant	2,46	< 1	20.44**
rev * lat	2,46	< 1	< 1
rev * ant * lat	4,92	3.25*	< 1
ROI analyses			
Left anterior	23	1.12	< 1
Midline anterior	23	1.13	< 1
Right anterior	23	1.07	< 1
Left central	23	1.16	2.95**
Midline central	23	2.41*	2.42*
Right central	23	2.38*	2.34*
Left posterior	23	2.68*	4.3**
Midline posterior	23	2.27*	4.52**
Right posterior	23	1.58	4.09**

Factors: *rev* = reversal; *ant* = anteriority; lat = laterality.

** $p < .01$ * $p < .05$ ^ $.05 < p < .1$

Table 3-12. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 5.

3.5.3 Discussion

The most important finding from Experiment 5 is that, even when the presentation of the verb was delayed by 1200ms, the N400 became sensitive to role-reversals only when the target verb's cloze probability differed greatly depending on the arguments' structural roles (i.e., in the high-predictability sentences). Other aspects of the results were consistent with those from previous experiments: (i) differences in target word's cloze probability elicited a clear N400 effect and a late frontal positivity in the control comparison, (ii) role-reversal anomalies in both high- and low- predictability sentences were readily detected and elicited a clear P600 effect at the verb. Below I will focus on the N400 results in the experimental sentences and their theoretical implications. For a discussion of the significance of other aspects of these results please refer to page 76.

The current study extended beyond the findings from Experiment 4 and showed that predictive computations played a central role in the N400's reemerged sensitivity to argument role-reversals. First, the current results are not predicted by a view that links the N400 to implausibility (or semantic anomaly) detection. If the N400 effect observed in Experiment 4 simply indicated that comprehenders detected the implausibility more quickly in the long-distance condition because the extended time interval allowed them to build a more robust representation of the preceding context, then role-reversals should have elicited an N400 effect as long as they rendered a sentence implausible. However, although role-reversed sentences were overwhelmingly judged as 'implausible' in both high- and low- predictability sentences, role-reversals elicited a significant N400 effect

only in the high-predictability sentences. Therefore, these results showed that the presence of an N400 effect cannot be simply attributed to implausibility detection.

Further, the observation that argument role-reversals elicited an N400 effect only if they clearly impacted the target verb's cloze probability (i.e., in high-predictability sentences) provides support for the hypothesis that the extended time window allowed comprehenders to incorporate information about the arguments' structural roles into their verb predictions. According to the lexical (semantic memory) access view of the N400, the N400 response to a word is reduced when access to this word in the lexicon is facilitated. In the current experiment, the reduction of the N400 in the canonical condition relative to the role-reversed condition in the high-predictability sentences suggests that access to the target verb was facilitated because information about the arguments' structural roles led comprehenders to pre-activate the target verb in the canonical condition. This observation is complemented by the fact that role-reversals do not modulate the N400 when the target verb was not predictable even in the canonical condition (i.e., in the low-predictability sentences). Therefore, the current results suggest that the N400 became sensitive to role-reversals in the long-distance condition in Experiment 4 because the extended time interval between the arguments and the verb allowed comprehenders to use information about the arguments' structural role to *pre-activate* the target verb before it appeared in the input.

3.6 General Discussion

In this chapter I investigated the cause for the N400's surprising insensitivity to argument role-reversals in previous studies. The results from three ERP experiments (Experiments 3 – 5) were summarized in Figure 3-14 (experimental comparisons) and Figure 3-15 (control comparisons). Firstly, the implausibility created by reversal of the arguments in verb-final sentences was readily detected and it elicited a P600 effect at the verb and. Further, I showed that the N400 became sensitive to argument role-reversals if and only if (i) there was additional time between the arguments and the verb, and (ii) the arguments' structural roles had a bit impact on the verb's cloze probability. Meanwhile, the results from the filler sentences showed that plausible but unexpected words elicited an N400 effect and a late positivity that has a frontal distribution. These findings license three conclusions.

Word order impacts comprehension, but it may fail to impact prediction.

In the *Ba*-construction in Mandarin Chinese, the structural roles of the pre-verbal arguments are unambiguously marked by their relative word order. Comprehenders reliably use the word order of the arguments to compute an accurate interpretation of a sentence and readily detect the implausibility resulted from argument role-reversals, as evidenced by their accurate behavioral judgments and the presence of a P600 effect at the verb. This is seen in Experiments 1 through 5 as well as in previous studies (e.g., Kolk et al., 2003). In contrast, the same piece of information may fail to impact comprehenders' predictions. The N400's insensitivity to argument role-reversals in Experiment 3 and in previous studies suggested that comprehenders fail to take into account the arguments'

structural roles when computing predictions for an upcoming verb. The contrast between comprehenders' behavior and the P600's sensitivity on the one hand and the N400's insensitivity on the other illustrates a dissociation between bottom-up comprehension processes and top-down predictive processes. The same piece of information can impact comprehension without informing prediction. As long as comprehenders can process language inputs using bottom-up mechanisms, their failure to predict upcoming input accurately in time does not preclude successful comprehension.

Word order can impact linguistic predictions, but its effect is not immediate.

Experiment 4 showed that the N400's sensitivity reemerged when the time interval for predictive computations (i.e., time elapsed between the arguments and the verb) was widened. The results from Experiment 5 show that the N400 effect is specifically associated with the expectancy for a word. The N400's selective sensitivity to cloze probability suggests that it is not simply modulated by implausibility. Combined with the effect of distance, these findings show that word order information about the arguments' structural roles does not have an immediate effect on comprehenders' verb predictions. Further, the reemerged sensitivity in the widened time interval (1800ms between the onset of the object and the verb) places an upper limit on the time required for word order information to impact predictive computations. This demonstrates that even unambiguous and prominent contextual information might have a delayed impact on comprehenders' predictions. More broadly, these results undermine widely held assumptions about the immediacy of linguistic prediction.

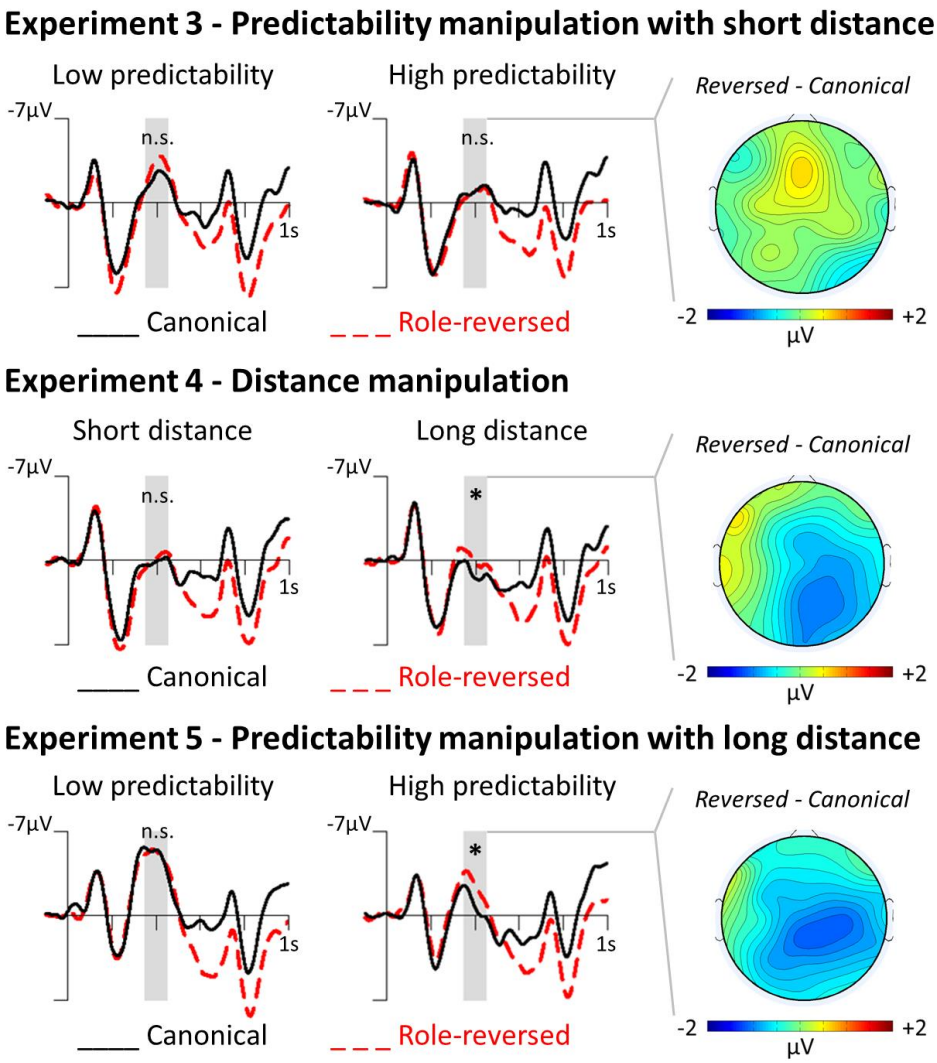

Figure 3-14. Summary of results in the experimental conditions in Experiments 3-5. *Left and center*: Grand average ERP waveforms at CPZ. *Right*: Topographic map of the effect of argument role-reversals (role-reversed– canonical) in the 350-450ms time interval.

Unexpected (but plausible) words and implausible words elicit distinct neural responses

Across Experiments 3 to 5, a simple cloze probability manipulation in the filler sentences elicited a clear N400 effect and a late frontal positivity. Figure 3-14 and Figure 3-15 showed that the N400 effect had a central-posterior distribution across experimental as well as control comparisons, which suggests that they reflect common underlying processes. However, the effect was bigger in the control comparisons (3-4μV) than in the

experimental conditions (~1μV). This difference in effect size is likely due to the fact that the expected and unexpected target words for the control comparison appeared in vastly different sentence contexts (e.g., *From the sheep the herdsman collected a lot of wool ...* vs. *Last year the boss bought a lot of wool ...*), while the contexts in the experimental sentences only differed by the order of the preverbal arguments.

Meanwhile, although a late positivity was observed in both experimental as well as control comparisons, their topographic distributions were clearly different (Figure 3-4; Figure 3-9; Figure 3-13). Role-reversals elicited a posterior late positivity (a P600 effect); unexpected but plausible target words in the filler sentences elicited a positivity in the same time interval that has a frontal distribution. This suggests that the posterior P600 and the frontal positivity reflect distinct underlying processes. The P600 is likely modulated by processes that are triggered in case of comprehension failures (e.g., upon detecting implausibility / grammatical anomalies in the input). These processes may involve re-analyses (e.g., Friederici, 1995), context updating (e.g., Coulson, King & Kutas, 1998), and/or error corrections in a noisy channel model (e.g., Gibson, Stearns, Bergen, Eddy & Fedorenko, 2013). Meanwhile, the frontal positivity may be modulated by processes that are triggered by inputs that disconfirm strong predictions. Federmeier et al. (2007) linked the frontal positivity to "semantic revision" processes and proposed that it may reflect "increased resource demands entailed by the need to override or suppress a strong prediction for a different word or concept" (p.81) and/or a "learning signal" from which to update future predictions (see also Federmeier, Kutas, Schul, 2010). Although future work is needed to better understand the functional significance of the

P600 and the late frontal positivity, the current results show that these components are modulated by different variables and likely reflect distinct underlying cognitive processes.

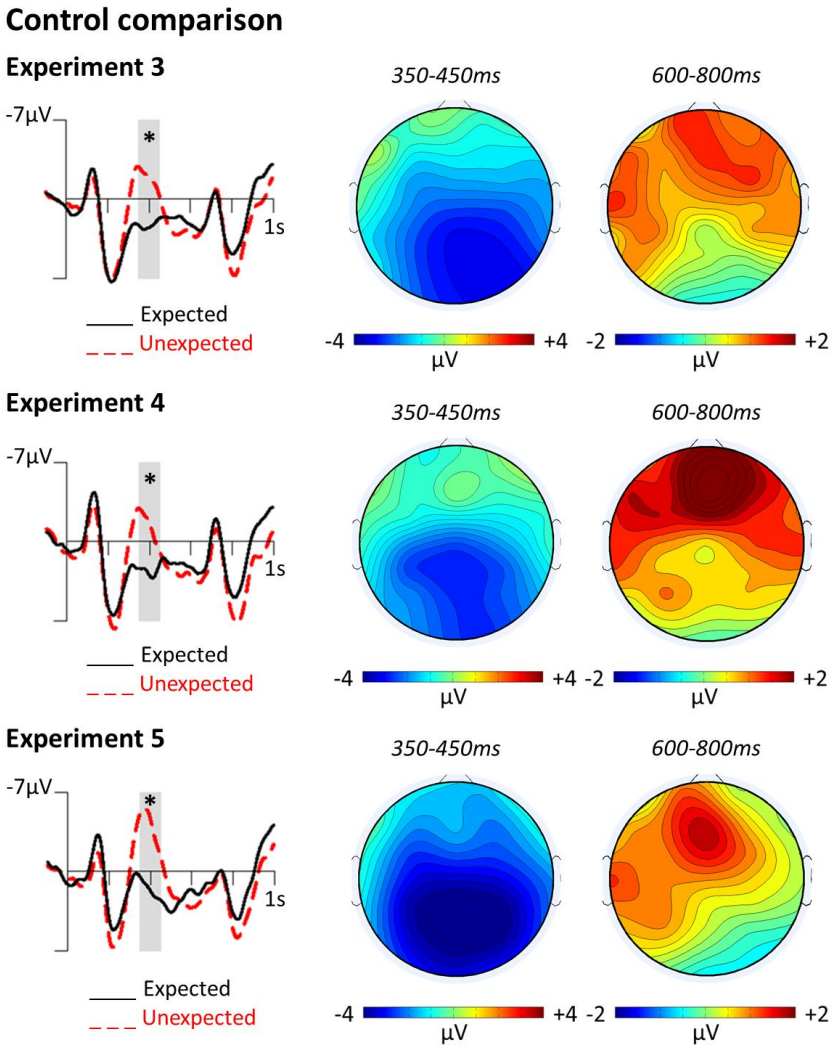

Figure 3-15. Summary of results in the control conditions in Experiments 3-5. *Left:* Grand average ERP waveforms at CPZ. *Center and right*: Topographic map of effect of cloze probability (unexpected – expected) in 350-450ms and 600-800ms time intervals.

A new experimental paradigm for studying the time course of prediction

In order to examine the time course of predictive computations, we must (i) identify critical pieces of information that inform such computations, (ii) delimit the time period during which predictive computations might occur, and (iii) be able to measure the output of predictive computations. Under the guiding hypothesis that the amplitude of the N400 is modulated by comprehenders' online expectations for specific words (condition *iii*), I introduced a novel experimental paradigm that satisfies the other two conditions for studying the time course of predictive computations (Figure 3-16).

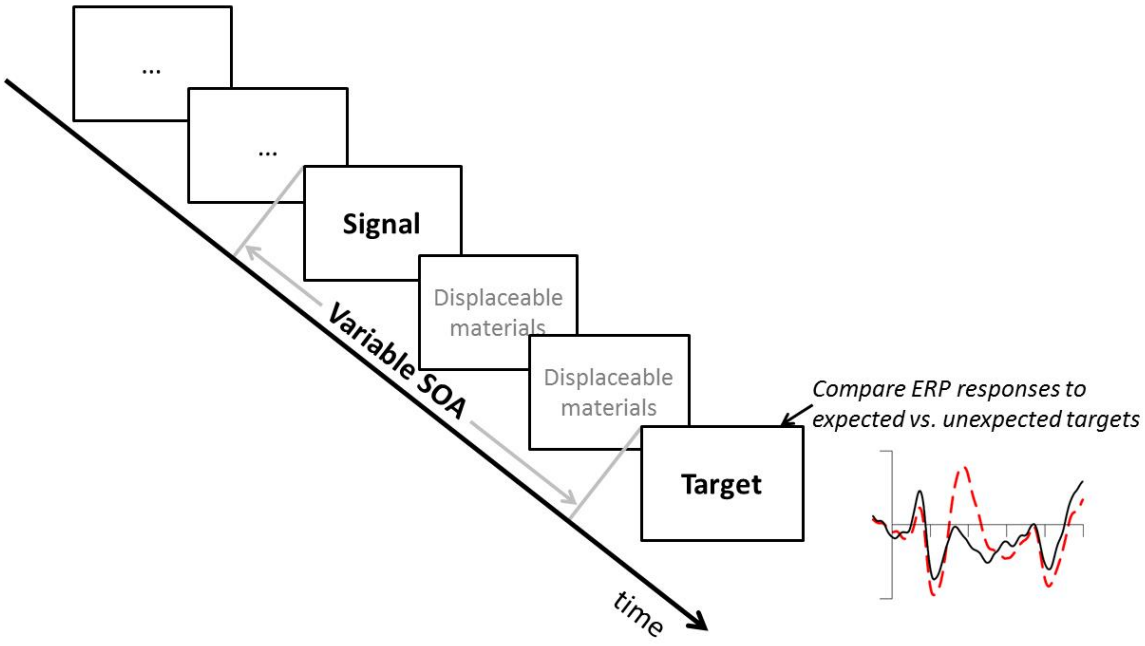

Figure 3-16. An experimental paradigm for studying the time course of prediction.

Following the violation-of-expectation paradigm commonly used in ERP research, this paradigm also presents sentences to comprehenders word-by-word at a fixed rate and compares their ERP response to an unexpected vs. expected word (the "target"). Unique to this paradigm is a manipulation of the relative timing of the target and a critical piece

of linguistic information (the "signal") that should render the target expected or unexpected. This manipulation allows comparisons of the effect of unexpected targets on the N400 depending on the amount of time comprehenders can incorporate the signal to update expectations for the target. Through such comparisons we can infer how the signal's impact on comprehenders' expectations has changed over the time period examined. A specific implementation of this general paradigm can be seen in Experiment 4 (Figure 3-5; page 79).

Broader implications

The results reported in this chapter undermine the common assumption about the immediacy of the linguistic predictions. I demonstrated that even prominent and unambiguous contextual information such as word order may fail to impact healthy adult native speakers' linguistic predictions immediately. These results have three main implications.

First, the observation that offline measures of predictability systematically failed to capture the linguistic predictions computed during real-time comprehension warns against drawing conclusions about online linguistic predictions directly from offline predictability measures. Since offline predictability measures such as those obtained in language corpora do not take into consideration how contextual information impacts linguistic prediction over time, models of linguistic prediction that are based solely on these measures (e.g., Hale, 2001; Levy, 2008; Smith & Levy, 2013) would fail to capture cases in which predictive computations are not yet complete when the relevant bottom-up input arises.

To build an accurate model of linguistic prediction, future research should consider the relevant inputs for predictions and the time point at which they can potentially impact comprehenders' predictions. Just as language inputs unfold in time, comprehenders' access to and use of difference sources of contextual information also develop over time. These considerations are crucial for differentiating between genuine failures to engage predictive mechanisms and cases in which certain sources of contextual information cannot impact predictions quickly enough to facilitate bottom-up processing, and they will likely be particularly important for understanding the development of fast and robust language comprehension abilities in children (e.g., Trueswell, Sekerina, Hill & Logrip, 1999) and second language learners (e.g., Moreno & Kutas, 2005) as well as the challenges faced by elderly individuals (e.g., Federmeier et al., 2002) and people with language impairments (e.g., Kuperberg, 2010).

More generally, the current results highlight the significance of the temporal dimension in the study of predictive computations. As in the language domain, prediction in other complex systems also involves mental computations that must race against rapid bottom-up input. Considerations about the time at which relevant contextual information becomes available and the time at which prediction starts to facilitate bottom-up processes are likely to have critical consequences for understanding the mental computations that underlie prediction across different domains.

4 Predictive computations on different time scales

4.1 Introduction

In Chapter 3 I presented evidence that comprehenders' real-time linguistic predictions might not be immediately sensitive to even prominent and unambiguous contextual information. In particular, I argued that word order information about the structural roles of preverbal arguments has a delayed impact on comprehenders' predictions for an upcoming verb. However, these results should not be taken to indicate that verb-prediction is slow in general. Instead, I argued that it is important to first identify the relevant inputs to predictive computations and carefully consider how each of them might impact linguistic predictions along the temporal dimension. In this chapter I put forth a framework for studying how different sources of contextual information impact linguistic prediction in real time. As a first step, I expand on the work reported in Chapter 3 and compare how information about pre-verbal arguments' lexical identity and their structural roles impacts comprehenders' verb-prediction across time. I discuss the theoretical and methodological considerations behind this approach and explain how it could be applied more broadly.

Here I explore the possibility that different sources of information may impact linguistic predictions on different time scales. In particular, I ask whether comprehenders can use information about the arguments' lexical identity to compute verb predictions even when information about their structural roles is not yet available for predictive

processes. Of particular relevance are the findings in a class study by Garnsey and colleagues (Garnsey, Tanenhaus & Chapman, 1989).

4.2 Garnsey, Tanenhaus & Chapman (1989)

In the original study, Garnsey and colleagues set out to test whether comprehenders compute a filler-gap dependency as soon as a potential gap site arises (the 'active gap-filling' parsing strategy; Fodor, 1978). They examined comprehenders' ERP responses when they read sentences such as '*The businessman knew which customer/article the secretary will call ....*' They found that the N400 response to the embedded verb was larger when the extracted object was inanimate and formed an implausible filler-gap dependency with the verb (e.g., *'... which article the secretary will call __ '*). The authors took the N400 effect to indicate that comprehenders detect an implausible dependency at the verb 'call' when the extracted object is 'which article' as opposed to 'which customer' because they assign the filler (the wh- phrase) to the first possible gap without waiting for disambiguating information.

4.3 Reinterpretation of Garnsey et al.'s (1989) findings

While these results indicate that comprehenders are sensitive to the potential dependency between the extracted object and the embedded verb, I propose that the N400 effect at the embedded verb may instead show that the verb was more or less predicted depending on the identity and/or animacy of the extracted object. Under the view that the N400 is modulated by ease of lexical semantic access (Deacon et al., 2000; Kutas & Federmeier, 2000; Lau et al., 2008), the N400 effect at the embedded verb may show that

access to the verb in long-term semantic memory is easier in the plausible condition than in the implausible condition because comprehenders have stronger predictions for the target verb 'call' when the extracted object is animate (e.g., 'customer') compared to when it is inanimate (e.g., 'article'). Further, this interpretation may be preferred as more recent findings have suggested that the process of constructing a filler-gap dependency itself is reflected by a positivity (the P600) that is elicited at a later time point than the N400 (Kaan, Harris, Gibson & Holcomb, 2000; Phillips, Kazanina & Abada, 2005). Therefore, I propose that the original findings by Garnsey et al. (1989) suggest that comprehenders can use information about the animacy and/or lexical identity of preverbal arguments to compute predictions about an upcoming verb.

Taken together, Garnsey et al.'s (1989) findings suggest that information about preverbal arguments' animacy and/or lexical identity can impact verb predictions quite quickly, while results from the experiments on Mandarin Chinese in Chapters 2 and 3 suggest that structural role information has a delayed impact on verb-predictions. These findings are compatible with the hypothesis that comprehenders can use information about the arguments' lexical identity to compute verb predictions even when information about their structural roles is not yet available. However, it remains very difficult to directly compare the results from these studies, as they differed in many important aspects (see Table 4-1). Therefore, in order to test whether different sources of input can impact predictive computations on different time scales, we must examine their contribution in minimally different environments. Here I describe an ERP experiment designed with this goal in mind.

	Exp 2–4$_{\text{short distance}}$	Garnsey et al. (1989)
Input of interest / Target manipulation	Structural roles of preverbal arguments	Animacy / lexical identity of filler (extracted object)
Language	Mandarin Chinese	English
Syntactic structure of verb-final clause	*Ba*-construction	Object-extracted embedded question
Word order of verb-final clause	SOV	OSV
Frequency of verb-final clause structure	High	Low
SOA between second argument and verb	600ms	1000ms
Animacy of preverbal arguments	Always animate	The extracted object was animate in plausible condition and inanimate in implausible condition; The subject was always animate.

Table 4-1. A comparison between the methods in Experiments 2-4 and Garnsey et al. (1989).

4.4 Experiment 6: Impact of argument role vs. identity on verb prediction

In this experiment I explored the possibility that different sources of contextual information impact predictive computations on different time scales. In particular, I hypothesized that information about arguments' lexical identity can impact comprehenders' predictions for an upcoming verb more quickly than information about their structural roles. Under this view, comprehenders' predictions for a verb should show sensitivity to the arguments' lexical identity before their structural roles, even when both sources of information become available in the input stream at the same time. In other words, there should be a time interval during which comprehenders' verb-predictions have been impacted by the arguments' lexical identity but not their structural roles.

Assuming the guiding hypothesis that the N400 is a reflection of comprehenders' real-time lexico-semantic prediction, I examined comprehenders' verb prediction by measuring the N400 response to a target verb when it appears shortly after the arguments. The target verb's cloze probability was manipulated in two different ways: (i) by

reversing the order of the preverbal arguments ('*Argument role-reversal*'), or (ii) by replacing one of the preverbal arguments with a different but discourse-compatible noun phrase ('*Argument substitution*'). A noun phrase is considered discourse-compatible if its occurrence in the give context is semantically and pragmatically congruous up to the point when the target word appears. Critically, these manipulations were done in sentences with identical structures and the distance between the arguments (the input) and the verb (the target) was held constant. Care was taken to match the cloze probability difference (by quartile) across argument role-reversal and argument substitution sentences. Further, in order to minimize potential confounds, all preverbal arguments were animate[2].

Based on the results of Experiments 1-5 as well as previous findings on the effects of role-reversals, I expected the N400 at the verb to be insensitive to cloze probability differences that result from argument role-reversals. If information about the arguments' lexical identity also fails to impact comprehenders' prediction immediately, then the N400 should also be insensitive to cloze probability differences that result from argument substitution. However, if information about the arguments' lexical identity impacts comprehenders' verb prediction by the time the target verb appears, then cloze probability differences that result from argument substitution should elicit a clear N400 effect at the verb.

---

[2] This differed from Garnsey et al.'s (1989) original study, in which the plausibility of the filler-gap dependency was confounded by the animacy of the filler (extracted object NP).

4.4.1 Methods

Participants

Twenty-four students (7 female, mean age = 21.9 years, range 18-29 years) from the University of Maryland College Park participated in the current study. All participants were native speakers of English, were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. All participants gave informed consent and were paid 10 USD/hour for their participation.

Materials

Following Garnsey et al. (1989), an object-extracted embedded question was used in all experimental materials. Unlike the canonical Subject-Verb-Object (SVO) word order in English, this construction has an Object-Subject-Verb (OSV) word order. Therefore, the arguments' identity and their syntactic roles are evident before the verb.

The experimental stimuli consisted of 120 English sentence pairs. Each sentence pair had the same target verb (the main verb in the embedded question), which always appeared immediately following the auxiliary "had". As shown in Table 4-2, the sentence context was manipulated in one of two ways to render the target verb predictable in one version and unpredictable in the other, as determined by cloze probability norming (see below). In the *Argument role-reversal* condition, the two versions within an item set had an identical verb-argument triplet and differed only in the order of the arguments. In the *Argument substitution* condition, the extracted object in one version was substituted by a different but discourse-compatible noun phrase in the other version, while the embedded

subject and target verb were held constant. No anomaly was evident before the target verb in either case. A sample set of argument role-reversal and argument substitution items is presented in Table 4-2.

Argument Role-reversal

The target verb's expectancy was manipulated by reversing the order of the arguments in the embedded question.

Sample Material	Target cloze % median (range)
The restaurant owner knew which customer the waitress had **served** during dinner yesterday.	23.3 (13.3- 53.5)
The restaurant owner knew which waitress the customer had **served** during dinner yesterday.	zero

Argument Substitution

The target verb's expectancy was manipulated by substituting the extracted object in the embedded question.

Sample Material	Target cloze % median (range)
The secretary confirmed which illustrator the author had **hired** for the new book.	23.3 (13.3 - 76.7)
The secretary confirmed which readers the author had **hired** for the new book.	zero

Table 4-2. Experimental conditions and sample materials in Experiment 6.

To determine the cloze probability of the target verbs in their sentence frames, a norming procedure was conducted with native English speakers using the Amazon Mechanical Turk (AMT) online marketplace. A total of 220 pairs of sentence frames, e.g., *The restaurant owner knew which customer the waitress had* ___ , and their argument role-reversed or argument-substituted counterparts, e.g., *The restaurant owner knew which waitress the customer had* ___, or *The restaurant owner knew which bartender the waitress had* ___, were divided into four lists of 110 items each. Each list was completed by 30 participants. In accordance with standard cloze norming procedures, participants were asked to read each sentence frame and to write down a word or phrase they expected to see next. Overwhelmingly participants responded with a verb. From the resulting database, 60 pairs of sentence fragments (in which the verb had a cloze value of 13% or greater in one version and 0% in the argument role-reversed version) were

selected as *Argument role-reversal* items; another 60 pairs (in which the verb had a cloze value of 13% or greater in one version and 0% in the argument-substituted version) were selected as *Argument substitution* items. In order to avoid sentence-final wrap-up effects, the sentences were extended beyond the critical verb with words that were held constant across conditions within each item set.

The sentences were distributed in two presentation lists, such that exactly one member of each pair appeared in each list. Each list contained 120 experimental sentences (30 per condition), 60 filler sentences adapted from a previous experiment examining the effects of cloze probability (each sentence context had an expected and an unexpected target word as a continuation; the expected target word in one context appeared as the unexpected target word in another context) and 80 unrelated filler sentences of similar length and structural complexity, so that the overall plausible-to-implausible ratio in each presentation list was 1:1. The sentences were presented in five blocks of 52 sentences each. The order of the blocks and the sentences within each block were randomized across participants.

Procedure

Participants were comfortably seated about 100cm in front of a computer screen in a testing room. Sentences were presented one word at a time in a white font on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500ms. Each word appeared on the screen for 300ms, followed by 230ms of blank screen (i.e., 530ms SOA). The last word of each sentence was marked with a period, followed 1000ms later by a response cue "?". Participants were instructed

to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively and to indicate whether the sentence meaning was plausible by pressing one of two buttons. Prior to the experimental session, participants were presented with 6 practice trials with feedback to familiarize themselves with the task. The experimental session was divided into five blocks of 52 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around two hours on average.

EEG Recording

EEG was recorded continuously from 29 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. Recordings were referenced online to the left mastoid and re-referenced to linked mastoids offline. The electro-oculogram (EOG) was recorded at four electrode sites; vertical EOG was recorded from electrodes placed above and below the left eye and the horizontal EOG was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5kΩ. The EEG and EOG recordings were amplified and digitized online at 1kHz with a bandpass filter of 0.1-100 Hz.

ERP Data Analysis

All trials were evaluated individually for EOG or other artifacts. Trials contaminated by artifacts were excluded from the averaging procedure. This affected 9.7% of experimental trials. Event-related potentials were computed separately for each participant and each condition for the 1000ms after the onset of the target word relative to

a 100ms pre-stimulus baseline. Statistical analyses on average voltage amplitudes were conducted separately for two time windows: 300–500 ms for the N400 and 700-900ms for the P600 / late positivity. Separate analyses were conducted for data from argument role-reversal and argument substitution sentences since neither the target verbs nor the arguments were matched across these sentences by design. Data from filler sentences for control comparison were also analyzed separately. We conducted repeated measures ANOVAs that fully crossed cloze probability (high vs. low) with Anteriority (anterior vs. central vs. posterior) and Laterality (Left vs. Midline vs. Right). The topographic factors effectively defined nine regions of interest (ROIs): left-anterior: F3, FC3; midline-anterior: FZ, FCZ; right-anterior: F4, FC4; left-central: C3, CP3; midline-central: CZ, CPZ; right-central: C4, CP4; left-posterior: P3, O1; midline-posterior: PZ, OZ; right-posterior: P4, O2. Univariate F-tests with more than one degree of freedom in the numerator were adjusted by means of the Greenhouse-Geisser correction. Since the current study was designed to modulate two ERP components with well-established topographic distributions, paired sample t-tests were conducted in individual ROI to examine the topographic distribution of the predicted effects.

4.4.2 Results

Plausibility Judgments

As shown in Table 4-3, participants judged sentences in the expected condition to be plausible at a much higher rate than those in the unexpected condition. Accuracy data were analyzed using mixed-effects logistic regression, which yielded significant main effects of both factors (expectancy: $\beta = -1.60$, p(Wald) < .0001; argument manipulation:

β = -0.22, *p*(Wald) < .0001) and no interaction between them: while argument substitution sentences were judged 'plausible' 5-6% more often than argument role-reversal sentences, the effect of cloze probability was much bigger, with expected sentences judged 'plausible' 60% more often than unexpected sentences.

Experimental Condition	% "Plausible" judgment Mean (SE)
Argument role-reversal, expected	85.4 (2.7)
Argument role-reversal, unexpected	23.8 (3.1)
Argument substitution, expected	90.3 (2.2)
Argument substitution, unexpected	31.1 (2.9)

Table 4-3. Average plausibility judgment in Experiment 6.

Event-related Potentials (control comparison)

The control comparison showed that comprehenders display clear N400 sensitivity to the standard cloze manipulation in the filler sentences. The grand average ERPs at the expected and unexpected target words for the control comparison are shown in Figure 4-1. The N400 was reduced for expected than unexpected words, and unexpected words elicited a larger late positivity than expected target verbs. These observations were confirmed by the statistical analyses. Results from the omnibus ANOVAs and ROI analyses are presented in Table 4-4.

In the 300-500 ms interval the omnibus ANOVA revealed a significant main effect of cloze probability, two-way cloze probability × anteriority and cloze probability × laterality interactions, and a three-way cloze probability × anteriority × laterality interaction. ROI analyses revealed that the N400 effect was present across the scalp. In the 700-900 ms interval the omnibus ANOVA revealed a significant main effect of cloze probability. ROI analyses revealed that the P600 effect was present across the scalp.

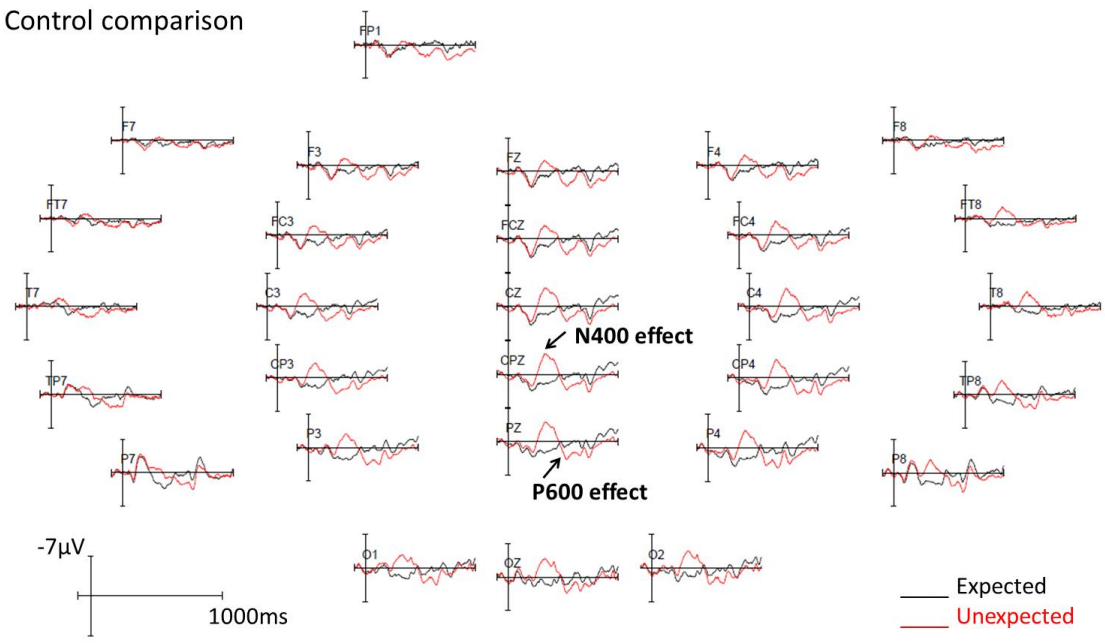

Figure 4-1. Grand average ERPs at expected (black) and unexpected (red) target words in the control comparison in Experiment 6.

	df	300-500ms	700-900ms
Control comparison			
cloze	1,23	21.84**	12.12**
cloze * ant	2,46	5.71*	< 1
cloze * lat	2,46	3.77*	1.97
cloze * ant * lat	4,92	3.78*	1.3
ROI analyses			
Left anterior	23	2.56*	2.95**
Midline anterior	23	2.22*	3.85**
Right anterior	23	3.35**	3.01**
Left central	23	4.17**	2.93**
Midline central	23	4.88**	3.36**
Right central	23	5.71**	3.54**
Left posterior	23	4.02**	1.79^
Midline posterior	23	4.88**	2.13*
Right posterior	23	5.18**	2.73*

Factors: cloze = cloze probability; ant = anteriority; lat = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

Table 4-4. ANOVA F-values and ROI analysis t-values at the target word in the control comparison in Experiment 6.

Event-related Potentials (experimental comparisons)

Figure 4-2 and Figure 4-3 show the grand average ERPs at the target word in the argument role-reversal and argument substitution conditions respectively. Figure 4-4 shows the topographic distribution of ERP effects in both experimental and control comparisons. An N400 effect was elicited only in the argument substitution sentences, in which the N400 was reduced for expected than unexpected target verbs. Meanwhile, unexpected target verbs elicited a larger late positivity than expected target verbs (a P600 effect) in both conditions, though the effect was bigger in the argument role-reversal condition than in the argument substitution condition. These observations were confirmed by the statistical analyses. Results from the omnibus ANOVAs and ROI analyses are presented in Table 4-5.

In the 300-500 ms interval, repeated measures ANOVAs revealed no significant effects involving cloze probability in the argument role-reversal condition; a significant main effect of cloze probability and a cloze probability × laterality interaction were observed in the argument substitution condition. ROI analyses revealed that the N400 effect was reliable across the scalp. A repeated measures ANOVA in the 700-900 ms interval revealed a significant main effect of cloze probability in the argument role-reversal condition; no significant effects involving cloze probability was observed in the argument substitution condition (all Fs < 2.2). ROI analyses revealed that the P600 effect was largest at central and posterior sites.

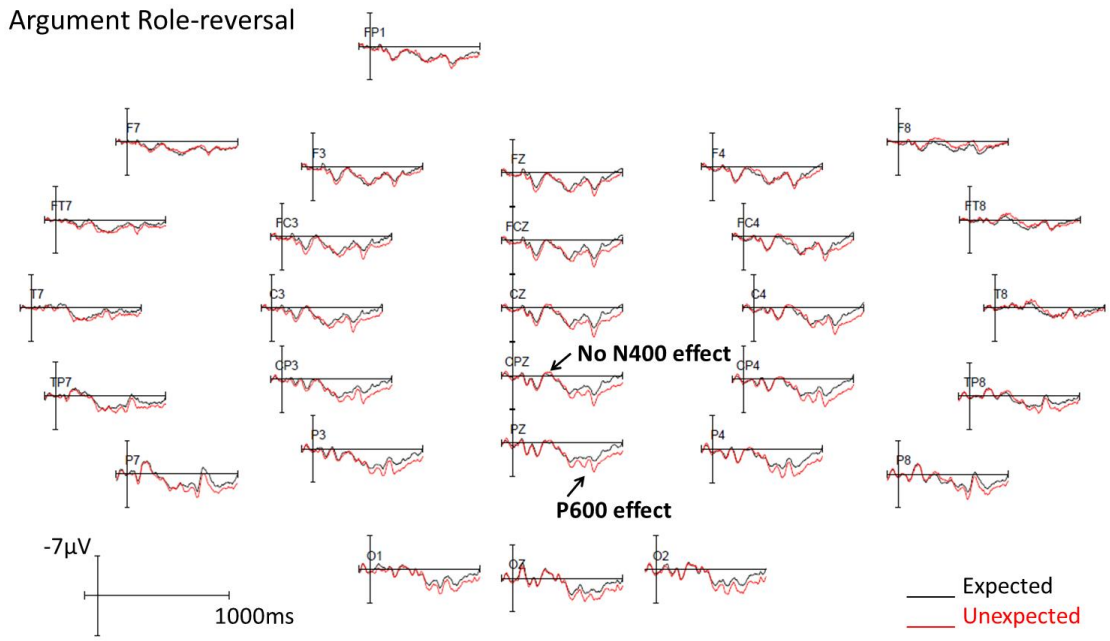

Figure 4-2. Grand average ERPs at the target verb in expected (black) and unexpected (red) conditions in the argument role-reversal sentences in Experiment 6.

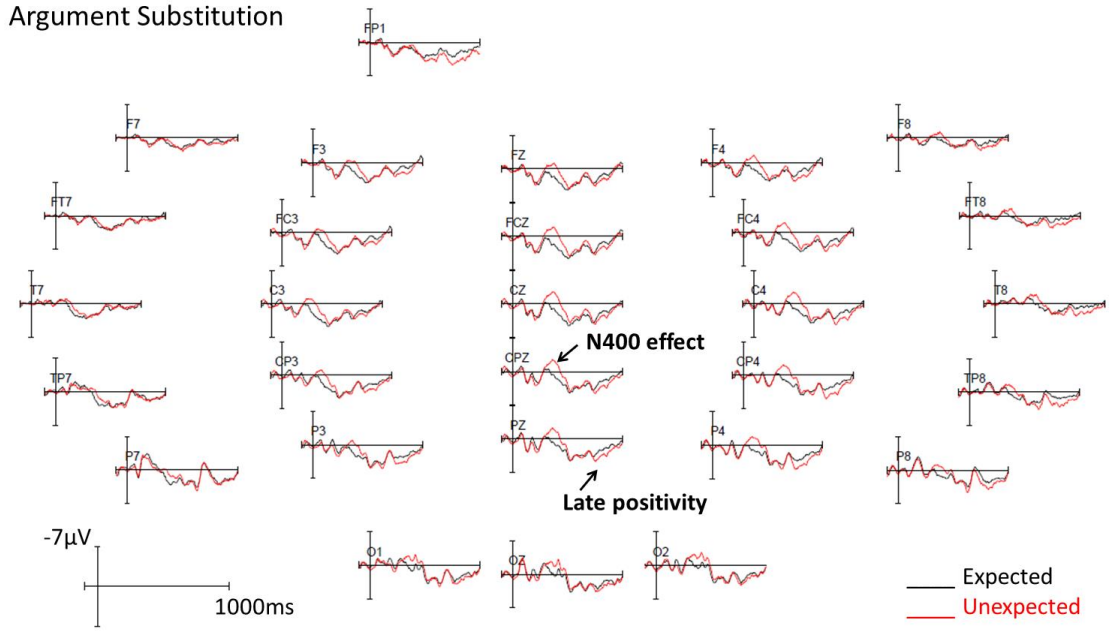

Figure 4-3. Grand average ERPs at the target verb in expected (black) and unexpected (red) conditions in the argument substitution sentences in Experiment 6.

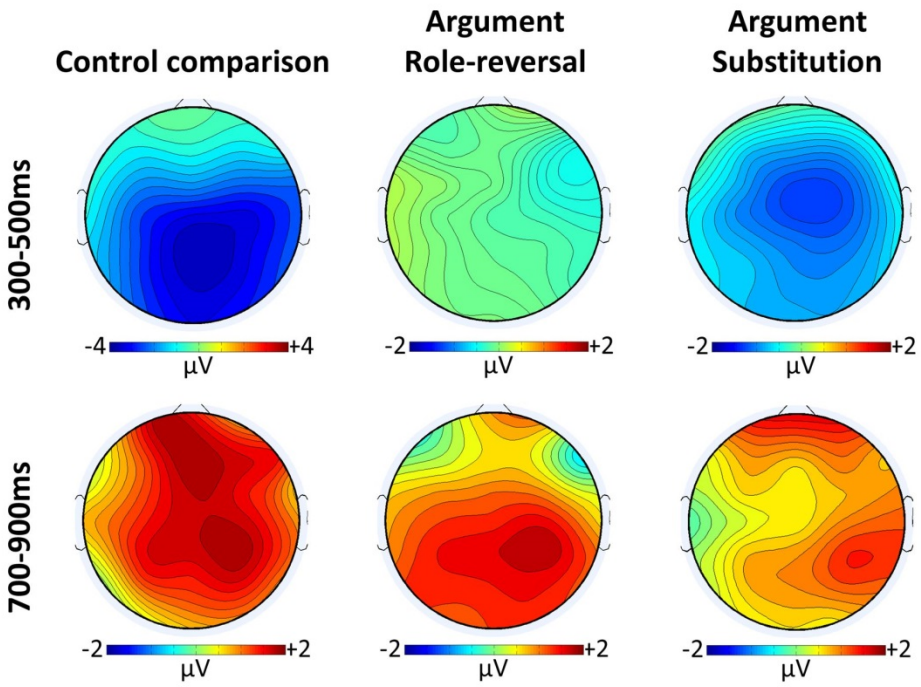

Figure 4-4. Topographic distribution of ERP effects in 300-500 ms (top) and 700-900ms (bottom) across experimental and control comparisons in Experiment 6.

	df	300-500ms	700-900ms
Argument Role-reversal			
cloze	1,23	< 1	5.69*
cloze * ant	2,46	< 1	2
cloze * lat	2,46	2.27	< 1
cloze * ant * lat	4,92	2.59^	1.11
ROI analyses			
Left anterior	23	< 1	1.34
Midline anterior	23	< 1	1.61
Right anterior	23	1.39	1.3
Left central	23	< 1	2.61*
Midline central	23	< 1	2.39*
Right central	23	< 1	3.21**
Left posterior	23	< 1	2.07*
Midline posterior	23	< 1	2.18*
Right posterior	23	< 1	2.56*
Argument Substitution			
cloze	1,23	8.22**	1.52
cloze * ant	2,46	1.09	< 1
cloze * lat	2,46	4.33*	1.93
cloze * ant * lat	4,92	1.56	2.15
ROI analyses			
Left anterior	23	2.47*	1.07
Midline anterior	23	2.92**	< 1
Right anterior	23	3**	1.19
Left central	23	2.68*	< 1
Midline central	23	2.97**	< 1
Right central	23	3**	1.63
Left posterior	23	2.26*	1.25
Midline posterior	23	2.51*	1.29
Right posterior	23	2.67*	1.4

Factors: cloze = cloze probability; ant = anteriority; lat = laterality.
** $p < .01$
* $p < .05$
^ $.05 < p < .1$

Table 4-5. ANOVA F-values and ROI analysis t-values at the target verb for the high- and low-predictability comparisons in Experiment 6.

4.4.3 Discussion

The current study showed that (i) comprehenders reliably detected the implausibility that resulted from both argument role-reversal and substitution, as reflected in explicit judgments; (ii) when the target verb appeared 1060ms following the onset of its second argument, argument role-reversals elicited only a significant P600 effect and no N400 effect; (iii) cloze probability differences that resulted from argument substitution elicited a significant N400 effect.

First, comprehenders were successful in detecting the implausibility in both argument role-reversal and argument substitution sentences. While sentences with an expected target verb were judged 'plausible' over 85% of the time, those with an unexpected target verb were deemed 'plausible' much less often (< 30% of the time). This shows that comprehenders computed an accurate representation of the meaning of the sentence in both cases. Further, although a plausible non-surface interpretation was available only in the unexpected argument role-reversed sentences, those sentences were not judged to be plausible any more often than the unexpected argument substituted sentences. In fact, unexpected argument role-reversed sentences were judged as implausible (76%) slightly more often than unexpected argument substituted sentences (69%). The observation that the acceptance rate for unexpected sentences was not at floor may be attributed to the nature of the task. Since participants were asked to perform a binary judgment to each sentence they read, their decision criterion may have been affected by the presence of clearly implausible filler sentences in the experiment. Taken together, the judgment data show that comprehenders were able to use both the

arguments' lexical identity and their structural roles to compute an accurate interpretation of the sentences.

Secondly, in line with previous studies and the results from Experiments 1 to 5, argument role-reversals in a verb-final configuration in English also elicited only a P600 effect and no N400 effect. Previous evidence regarding the effects of argument role-reversals has come from verb-final languages (e.g., Dutch, Japanese) and highly frequent verb-final constructions in a SVO language (the active *ba*- and passive *bei*- constructions in Mandarin Chinese). The current study showed the same pattern of results in object-extracted embedded questions in English, despite the fact that English has a SVO word order, and the experimental materials had a relatively complex sentence structure with an OSV word order. This highlights the generality of the previous ERP findings and suggests that the N400's insensitivity to argument role-reversals is not dependent on the language or particular sentence structures being studied. Meanwhile, the observation of a significant P600 effect in the current study shows that comprehenders can readily detect the implausibility that results from argument role-reversals, even when it is embedded in a complex and relatively infrequent sentence structure.

Further, a significant N400 effect was elicited by the cloze probability differences that resulted from argument substitution. This suggests that access to lexical semantic memory, as indexed by the N400, was facilitated for the expected verb compared to the unexpected verb in the argument substitution sentences. Such facilitation suggests that information about the preverbal arguments' lexical identity impacted comprehenders' predictions about the verb before it appeared in the input stream. Although the magnitude

of cloze probability differences was carefully matched between the argument role-reversal and argument substitution sentences, a significant N400 effect was observed only in the argument substitution sentences and not in the argument role-reversal sentences. This contrast suggests that, when the verb appeared around 1000ms following the onset of its second argument, only information about the arguments' lexical identity, but not their structural roles, could impact comprehenders' predictions about the verb. This provides the first evidence that different sources of contextual information may impact linguistic prediction on different time scales.

Characterization of the late positivity

First, a significant late positivity was observed in the control comparison. This effect had a frontal-central distribution (Figure 4-4) and bears resemblance to the late frontal positivity that has been associated with the processing costs that result from the violation of strong predictions (e.g., Federmeier et al., 2007; Federmeier et al., 2010; Van Petten & Luka, 2012; control comparisons in Experiments 3-5). Since the expected target words in the filler sentences in the current experiment had very high cloze probabilities (all > 70%), the unexpected target words likely violated rather strong predictions. However, unlike the materials used in these previous studies, the unexpected target words in the current experiment were implausible. Therefore, it is possible that the late frontal-central positivity in fact co-occurred with a P600 effect, which potentially explains why it was more broadly distributed than the frontal positivity observed in previous studies. Taken together with the lack of a frontal positivity in the experimental sentences, which had weaker contextual constraints compared to the filler sentences (average maximum

cloze value < 30%), these results provide support for proposals that associate the late frontal positivity to processes that are triggered when strong predictions are disconfirmed.

Meanwhile, a late posterior positivity was observed in both types of experimental sentences, in which unexpected verbs elicited a larger late positivity than expected verbs. Although this effect failed to reach statistical significance in the argument substitution sentences, inspection of the means (Figure 4-5) suggests that the effect was present in both argument substitution and argument role-reversal sentences in the 700-900ms time interval. In fact, these results contrast sharply with the results in the 300-500ms time interval, where the effect of cloze probability was observed only in the argument substitution condition and not in the argument role-reversal condition.

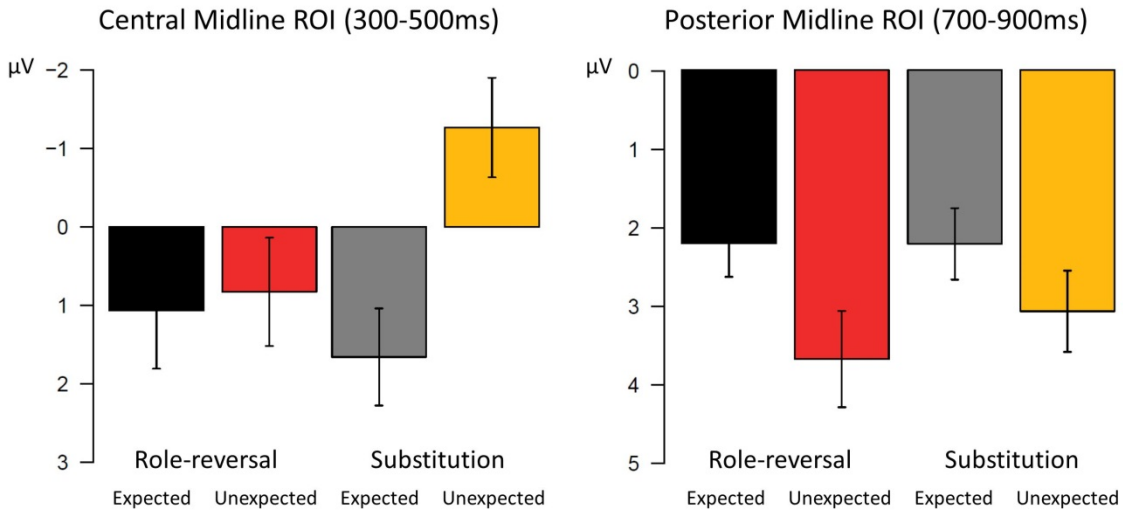

Figure 4-5. Average ERP amplitude in each experimental condition in Experiment 6. Left: 300-500ms in central midline region. Right: 700-900ms in posterior midline region.

To examine the cause for this quantitative difference between the size of the effect in these conditions, I explored the possibility that the P600 effect was reduced in the argument substitution condition because fewer unexpected argument substitution

trials (69%) were judged 'implausible' (compared to 76% in the argument role-reversal condition). This was motivated by previous observations that the P600 is sensitive to task requirements (e.g., Coulson et al., 1998; Kolk et al., 2003). If the P600 is associated with decision processes that are engaged in the plausibility judgment task, the size of the P600 might be correlated with participants' plausibility judgments. Therefore, I compared the grand average ERPs in expected trials that were followed by a 'plausible' judgment with unexpected trials that were followed by an 'implausible' judgment in both argument role-reversal and argument substitution sentences. This analysis, however, showed that the size of the positivity was not predicted by participants' behavioral response. In fact, the effect in the argument substitution condition was numerically slightly smaller in this analysis. Therefore, despite its sensitivity to task requirement, the size of the P600 did not bear a straightforward relation with participants' behavioral performance on the task.

Meanwhile, it is possible that a P600 effect of comparable size was elicited in both kinds of sentences, but the effect was reduced in the argument substitution sentences due to an overlapping N400 effect. This explanation, however, remains difficult to evaluate. On one hand, this explanation finds support in the observation that the late positivity in the argument substitution sentences had a later onset than that in the argument role-reversal sentences. On the other hand, previous studies have demonstrated that the presence of an N400 effect does not necessarily lead to a reduced P600 effect (e.g., Osterhout & Nicol, 1999). Given the ambiguity about the extent to which the reduced P600 effect is attributable to potential component overlap in the current study, interpretation of this aspect of the current results must therefore proceed with caution.

If there is a genuine difference between the P600's sensitivity to argument role-reversals and argument substitution, it may indicate that argument role-reversals and argument substitutions differentially modulate the cognitive processes underlying the P600. This is compatible with a recent proposal that links the P600 to error correction processes in a noisy channel model (Gibson et al., 2013). Although a linking hypothesis is required to specify what triggers error correction processes and how such processes modulate the P600 (e.g., more error corrections → more positive), the observation of a larger P600 effect to argument role-reversals than argument substitutions may indicate that comprehenders are more likely to engage in error correction in role-reversed sentences and/or that error correction is easier in role-reversed sentences.

Taken together, the results from Experiment 6 show that comprehenders reliably compute an accurate interpretation of a sentence using information about the arguments' lexical identity as well as their structural roles, but that the arguments' lexical identity has a more immediately impact on verb prediction than information about their structural roles. More generally, these results provide the first evidence that different sources of contextual information may impact linguistic prediction on different time scales.

4.5 General Discussion

In Experiment 6 I took a first step towards identifying different sources of contextual information relevant for predictive computations and compared how they impact linguistic prediction in real-time. I found that the N400 is sensitive to cloze probability differences that resulted from argument substitution but it was completely insensitive to cloze probability differences that resulted from argument role-reversals.

Taken together with the findings from Experiments 3-5 that the N400 became sensitive to argument role-reversals when the time interval for predictive computations was widened, I proposed that information about preverbal arguments' lexical identity can impact predictive computations more quickly than information about their structural roles.

The role of prediction in processing long-distance dependencies

In the Introduction of this chapter I proposed that the classic findings reported by Garnsey et al. (1989) can be re-interpreted as reflecting how the identity of a displaced element (the filler) can impact comprehenders' prediction about an upcoming verb. This proposal may be applied more broadly to the processing of other long-distance dependencies (e.g., Kazanina, Lau, Lieberman, Yoshida & Phillips, 2007; Phillips, 2006; Staub & Clifton, 2006; Traxler & Pickering, 1996; Yoshida, Dickey & Sturt, 2012). In fact, previous studies have proposed that prediction plays a role the processing various long-distance dependencies (e.g., Lau, Stroud, Plesch & Phillips, 2006; Omaki, Lau, Davidson White & Phillips, forthcoming; Staub & Clifton, 2006; Wagers, & Phillips, in press; Yoshida et al., 2012). This suggests that the processing of these dependencies is potentially very useful for the study of how predictions are computed in real-time – for example, does the absence of a 'filled-gap' effect in syntactic islands indicate that comprehenders' prediction for a gap is suppressed? What are the processing mechanisms that suppress existing predictions?

Why might arguments' lexical identity impact verb prediction more quickly than their structural roles?

When processing a simple sentence, predictions for the verb can be computed using the identity of the arguments with more or less information about their structural

roles. As soon as the individual words in the context have been processed and their meanings have been retrieved from the lexicon, predictions about an upcoming word can be generated via lexical association in long term semantic memory. For example, even without access to information about the arguments' structural roles, the lexical representation for verbs that co-occur more often with the arguments may become more activated. On the other hand, in order to compute predictions for the verb with full information about the arguments' structural roles, the syntactic relations between the arguments, in addition to the meaning of the individual arguments, must also be computed. Therefore, even if the relevant predictive computations involving argument identity and structural role information take the same amount of time, structural role information may have a delayed impact on linguistic predictions simply because it becomes available to the predictive processor later in time.

Future directions

As I discussed earlier, there are two critical assumptions in the argumentations in this dissertation. First, I assume the linking hypothesis that the size of the N400 is modulated by ease of lexical semantic access (e.g., Lau et al., 2008). Further, the proposed interpretation of the current results assumes that lexical semantic access was, at least in these studies, facilitated as a result of predictive computations (e.g., Kutas & Hillyard, 1984). Therefore, even if the first linking assumption is correct, such that a reduction in N400 amplitude in fact reflects facilitated lexical semantic access, the second assumption must also be in place in order to license inferences about the nature of predictive computations. If the second assumption were violated, then the current results would not license conclusion about the when and how linguistic predictions are being

computed. Therefore, the current proposal should be continually evaluated against independent evidence about the validity of each of these linking assumptions.

Meanwhile, although the observation that the N400 is insensitive to cloze probability differences that resulted in argument role-reversals has been reported in many studies and across languages, the effect of distance on the N400's sensitivity has only been reported in two experiments (Experiments 3 and 4 in Chapter 3). Both of these experiments were conducted in Mandarin Chinese, and the same sentence structure, presentation parameters and filler sentences were used. In order to evaluate the generalizability of these findings, future work will need to extend this paradigm to examine whether the N400 becomes sensitive to cloze probability differences that result from role-reversals when the time interval for prediction is widened in other sentence structures and other languages.

As a first step, I will extend this paradigm to the argument role-reversal sentences used in Experiment 6. To introduce a temporal delay between the arguments and the target verb, I will insert a three-word prepositional phrase (e.g., "in blue jeans") to modify the second argument (i.e., the subject). With an argument role-reversal manipulation, the same prepositional phrase will modify different head nouns across the canonical and role-reversed conditions. For example, the original item in (10) will become (11). Care will be taken to ensure that the added prepositional phrase is compatible with both nouns. The adapted materials will also be re-normed to obtain cloze probability measures for the target verbs. Keeping the same presentation rate as Experiment 6 (530ms SOA), the prepositional phrase will increase the SOA between the

second noun and the verb by 1590ms. If this provides sufficient time for information about the arguments' structural roles to impact comprehenders' predictions for the verb, then a significant N400 effect should be elicited in argument role-reversal sentences like (11).

10. A sample set of argument role-reversal sentences in Experiment 6:

    The restaurant owner forgot which customer the waitress had served …

    The restaurant owner forgot which waitress the customer had served …

11. A set of argument role-reversal sentences with an added prepositional phrase:

    The restaurant owner forgot which customer the waitress in blue jeans had served…

    The restaurant owner forgot which waitress the customer in blue jeans had served…

Further, future work is required to adjudicate between competing explanations for the observation of a significant N400 effect at the verb in argument substitution sentences. I proposed that the N400 results indicate that comprehenders can quickly use the lexical identity of the arguments to predict an upcoming verb. Crucially, this proposal implies that comprehenders use the lexical identity of the arguments and not just any nouns in the sentence context to predict the verb. However, this pattern of results is also compatible with other explanations. For instance, it is possible that the algorithms for the relevant predictive computations do not make any reference to predicate-argument structure at all. Since the expected and unexpected argument substitution sentences in Experiment 6 differed in the lexical identity of the extracted object, the N400 effect could be attributed to predictive computations that simply used all content words in the preceding context without regard to whether they constitute arguments to the event of interests.

Meanwhile, the N400 effect may also be explained by an account that makes no reference to predictive computations. In particular, it may be a consequence of how the lexicon is organized. Despite our efforts to (i) closely match the cloze probabilities between the verbs in the argument substitution and role-reversal sentences and (ii) to ensure that both extracted objects in each argument substitution item were compatible with the discourse preceding the target verb in Experiment 6, it is possible that the extracted object and the verb were on average more semantically related in the expected condition than in the unexpected condition. If this is the case, and if we assume automatic spread of activation between related items in the lexicon, then the N400 effect may the result of pre-activation of the expected verb by the extracted object through automatic spread of activation.

Future research can adjudicate between these competing accounts by examining the effect of argument substitution in sentences that are fully matched in their lexical contents. For example, to counter the confounding effects of an argument substitution manipulation (as in (12)), in a given item set the extracted object in one condition can be used as the main clause subject in the other condition and vice versa. As illustrated in (13), 'neighbor' is the extracted objected in the expected condition and the main clause subject in the unexpected condition, while for 'exterminator' the opposite is true.

12. Argument substitution manipulation in Experiment 6:

The tenant inquired which neighbor the landlord had <u>evicted</u>…

The tenant inquired which exterminator the landlord had <u>evicted</u> …

13. Lexically matched argument substitution manipulation:

The exterminator inquired which neighbor the landlord had evicted…

The neighbor inquired which exterminator the landlord had evicted …

This design will allow us to manipulate the identity of the arguments of the target verb while holding the lexical content of each sentence pair constant. If the N400 effect in Experiment 6 is fully attributable to automatic spread of activation, and assuming that linear order does not matter, then argument substitution in sentences such as (13) should not elicit an N400 effect, because the verb is preceded by the same lexical items. The same pattern would be expected if comprehenders use an algorithm that makes no reference to predicate-argument structure to compute verb-predictions. Alternatively, if comprehenders successfully identify the arguments for the embedded verb and use them to compute predictions for the verb, then the target verb should elicit a smaller N400 response in the expected condition than in the unexpected condition.

Lastly, the current proposal also predicts that converging evidence can be obtained using other measures of online linguistic prediction (e.g., anticipatory eye-movements, speeded cloze responses). For example, it predicts that comprehenders' cloze response to the sentence fragments used in Experiment 6 should be sensitive to different kinds of contextual information depending on the amount of time they are given to respond (Staub, Grant, Astheimer & Cohen, forthcoming). If comprehenders' predictions are initially only affected by the arguments' lexical identity but not by their structural roles, then we should expect their cloze responses to be (i) equally sensitive to the

arguments' lexical identity regardless of response latency, but (ii) less sensitive to the arguments' structural roles at shorter response latencies.

Conclusion

The results reported in this chapter provide further support for the proposal outlined in Chapter 3, namely, that even prominent and unambiguous contextual information such as word order may fail to impact linguistic prediction immediately. The current results provide initial evidence that different sources of contextual information may impact linguistic prediction on different time scales. With the research framework and experimental paradigm proposed in Chapters 3 and 4, future research will be able to study the mental computations that underlie linguistic prediction and how they unfold across time.

5 Aligning the eyes and the brain

5.1 Introduction

In the previous chapters I have drawn on ERP evidence to study how linguistic predictions are computed in real-time. Towards the end of Chapter 4 I outlined a future study that uses a speeded cloze task to evaluate the proposal that I put forth in this dissertation, based mostly on ERP evidence. Implicit in this proposal, as well as in the field of psycholinguistics, is the assumption that the cognitive processes underlying language comprehension can be studied using widely different methods and techniques. Naturally, different techniques provide qualitatively distinct measures (e.g., latency of button presses, ERP amplitudes, blood-oxygen level in different brain regions, etc.) and are suited for asking different kinds of research questions. Among them, ERPs and eye-movements (EMs) are excellent for capturing moment-to-moment sensitivity and have been used extensively to study real-time language comprehension. Therefore, in this chapter I turn to examine how the processing of thematic relations impacts reading eye-movements and I discuss how direct comparisons between ERP and eye-movement evidence can help to constrain theories about underlying cognitive processes.

As the techniques for studying EMs and ERPs have developed largely independent of each other, it is perhaps not surprising that earlier psycholinguistic research that used these techniques has proceeded mostly in parallel. Despite limited interactions between ERP and EM research, studies using these methods have often provided largely convergent evidence and have afforded similar conclusions about

general properties of the language comprehension system. For example, comprehenders' ERPs and EMs both showed clear sensitivity to garden path sentences (e.g., Frazier & Rayner, 1982; Osterhout, Holcomb & Swinney, 1994), suggesting that these methods tap into some common processes that underlie language comprehension.

In the past decade, there has been growing interest in directly comparing ERPs and EMs. An influential paper by Sereno and Rayner (2003) went beyond asking "Do both measures show sensitivity to this variable?" and proposed that ERPs and EMs can be used to complement each other. They argued that mis-alignments between ERP and EM results can be highly informative, as they might indicate that ERPs and EMs are sensitive to different underlying cognitive processes and can be used to constrain linking hypotheses (e.g., "What is the functional significance of the N400?"). In the paper, Sereno and Rayner suggested that ERP and EM evidence afforded different conclusions about the timing of lexical access. Noting that lexical frequency reliably impacts the first fixation on a word (which is on average 200-250ms long), they concluded that the N400 (which begins at around 250ms and peaks at 400ms) is too late to be a reflection of lexical access.

More recent research has begun to compare ERP and EM evidence more directly and explore how they align in time. On one hand, some began to do side-by-side comparisons by collecting EM and ERP data from different groups of participants using the same materials (e.g., Camblin, Gordon & Swaab, 2007; Dambacher, Gollner, Nuthmann, Jacobs, & Kliegl, 2006; Dambacher & Kliegl, 2007; Ledoux, Camblin, Swaab & Gordon, 2006). For example, Dambacher and Kliegl (2007) showed that larger

N400 amplitudes at word *N* correlated with longer fixation durations on both words *N* and *N+1*. On the other hand, some have taken up the challenge to co-register comprehenders' EMs and EEG simultaneously during natural reading (e.g., Dimigen, Sommer, Dambacher & Kliegl, 2008; Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Kretzschmar, Bornkessel-Schlesewsky & Schlesewsky, 2009). By analyzing ERPs time-locked to fixation onsets (fixation-related potentials, FRPs), Dimigen et al. (2011) found that N400 predictability effects arose around 250 ms after fixation onset and that, by the time the N400 effect reached its peak, readers had already moved on to a different word in most cases.

Since side-by-side EM/ERP comparisons and co-registration studies have only been conducted in English and German, the current study is the first to examine how EMs and ERPs align in Chinese reading comprehension. Although the generalizability of any eye-movement evidence is limited by the differences in written scripts (see below), as a first step I will assume Dambacher et al.'s (2007) results in German and consider an alignment between first fixation durations on a word and the onset of the N400 response[3]. In this chapter I present evidence from three reading eye-tracking experiments in Mandarin Chinese, which were adapted from Experiments 3 and 5 in Chapter 3. In the initial experiment (Experiment 7) I observed a striking misalignment between EMs and ERPs, which was resolved somewhat successfully in two follow-up experiments (Experiments 8 and 9). I will discuss methodological considerations as well as the

[3] Note that Dambacher et al. (2007) restricted their analyses to single fixations only (i.e., words fixated exactly once).

potential benefits and challenges of using EM and ERP evidence to constrain theories of sentence comprehension.

5.2 Basic considerations for studying EMs during reading Chinese

Naturally, since reading eye-movements are tightly linked to the physical properties of the written text, it is important to take into consideration the writing system being studied. While widely studied languages such as English and German have an alphabetic script, Chinese has a logographic writing system that is distinct in many ways. First, written Chinese is formed by strings of equally sized box-like symbols called characters, each of which represents a syllable with tonal characteristics (for a more detailed discussion about character properties and classification, see Yang, 2010). Secondly, since Chinese words are often composed of more than a single character and most characters can join with others to form multiple-character words with different meanings (Chen, Song, Lau, Wong, & Tang, 2003), the semantic and syntactic attributes of a character are often context-dependent. According to the *Lexicon of common words in contemporary Chinese* (2009), which includes 56,008 words, 6% are one-character words, 72% are two-character words, 12% are three-character words, and 10% are four-character words. Less than 0.3% of Chinese words are longer than four characters. Although there are not a lot of one-character words, many of them are closed-class words and are therefore much more frequent than multiple-character words. Further, written Chinese has no explicit marker to indicate word boundaries. Chinese readers, when asked, may at times not agree on where word boundaries are located. Therefore, compared to languages

such as English and German, where words are physically separated by spaces, the notion of a written "word" is less clear in Chinese.

Given the differences between these writing systems, results obtained from reading English and other alphabetic languages may not be fully generalizable to reading Chinese. Although previous EM studies in Chinese have demonstrated several basic similarities between EMs in reading Chinese and English (e.g., average fixation durations are about 220-250ms, fixation durations are shorter for more frequent words, more predictable words are more likely to be skipped), I draw on EM evidence from previous Chinese reading studies when possible.

5.3 Comparison between standard EM and ERP paradigms

One critical difference between common EM and ERP experiments lies in their stimulus presentation method. In reading eye-tracking experiments, sentences are presented on a computer screen and participants are free to move their eyes to read at whatever pace feels most natural. In contrast, ERP experiments commonly use a rapid serial visual presentation (RSVP) paradigm in which participants are instructed to maintain fixation at the center of the screen where the words of the sentence are presented at a fixed rate, usually around two words per second. The procedure is used mainly for two reasons. First, it minimizes eye movements, which lead to large electro-ocular artifacts in the EEG. In fact, a previous study has shown that the effects triggered by an eye movement can be order of magnitudes larger than typical psycholinguistic effects (Picton, Van Roon, Armilio, Berg, Ille & Scherg, 2000). Further, the relatively

long interval between words serves to prevent overlap of brain responses triggered by successively presented words.

The use of these different stimulus presentation methods gives rise to several important distinctions between the flow of information in ERP and EM studies, some of which will become very relevant to the interpretation of the results in Experiments 7 to 9. First of all, natural reading occurs at a variable rate (e.g., reading times vary as a function of frequency and length); RSVP presents words a fixed rate. Although some have expressed concern about the ecological validity of an RSVP paradigm for reading comprehension, ERP studies that use RSVP have been shown to yield largely the same results to those that present sentences auditorily (e.g., Federmeier & Kutas, 1999; Federmeier, McLennan, De Ochoa & Kutas, 2002). Secondly, sentences are read more quickly in natural reading (~3-5 words per second) than in a RSVP paradigm (~2 words per second). Therefore, effects on EMs at a given word are more likely to spillover to fixations on the following words. On the other hand, the slower reading rate in ERP studies gives comprehenders extra time to process and integrate different information. Thirdly, comprehenders are free to move their eyes to read different parts of a sentence as they wish (e.g., to skip words, to reread earlier parts of a sentence); they do not have such control in a RSVP paradigm, in which they also have to maintain their fixation at the center of the screen.

Last but not least, in contrast to natural reading, there is no preview of upcoming words in a RSVP paradigm. A large number of studies have demonstrated that fixation times on word *N+1* are shorter when it is visible during fixation on word *N* than when it

was masked during fixation on word *N* (e.g., McConkie & Rayner, 1975; Rayner, 1975, Rayner & Bertera, 1979; also see Rayner 1998, 2009 for reviews), which suggests that readers can begin processing a word even when it is located in the parafoveal region of the visual field (Rayner, 1998). Such *parafoveal preview benefit* has typically been assessed via the use of a gaze-contingent boundary paradigm, in which the identity of a target word is initially masked and its presentation is contingent on the reader's eye-gaze (Rayner, 1975; also see Experiment 9). Previous Chinese reading eye-tracking studies using this paradigm have indicated that Chinese readers also obtain preview benefit from characters to the right of the fixated character (e.g., Liu, Inhoff, Ye & Wu, 2002; Tsai, Lee, Tzeng, Hung & Yen, 2004; Wang, Tong, Yang & Leng, 2009; Yan, Richter, Shu & Kliegl, 2009; Yang, Wang, Xu, & Rayner, 2009; Yen, Tsai, Tzeng & Hung, 2008). Therefore, the presence of parafoveal preview in natural reading gives readers a 'head start' that is not available in an RSVP paradigm.

5.4 Experiment 7: A striking mis-alignment

The current experiment shared the same design as Experiment 3: predictability (high vs. low) × argument order (canonical vs. role-reversed). The effects of argument role-reversals were compared across sentences in which the critical verb in the canonical control condition has high vs. low cloze probability (Table 5-1). In the role-reversed condition the verbs commonly have zero cloze probability. Therefore, argument role-reversals have a greater impact on the verb's cloze probability in high-predictability sentences than in low-predictability sentences.

Comrpehenders' eye-movements were recorded while they read sentences from Experiment 3. Participants were drawn from the same population as the ERP experiments, but none of them participated in Experiment 3 (or any of the norming or ERP experiments). A total of 30 high- and 30 low-predictability item sets were selected from the top and bottom end of the cloze probability distribution of the 120 item sets from Experiment 3. This is because EMs have higher signal-to-noise ratio than ERPs and fewer trials per condition are therefore required in the current experiment compared to Experiment 3.

Based on the ERP results, which showed a P600 effect only for the role reversal, and no N400 effect, I expected to observe a clear effect of argument role-reversals at the verb and perhaps on the following word. Although the ERP results do not afford predictions about the presence of effects on specific EM measures (e.g., first fixation duration, probability of regression, regression path time), the relatively late onset of the P600 effect in ERPs suggested that argument role-reversals might be more likely to impact later measures such as regression path time than early measures such as first fixation durations.

5.4.1 Methods

Participants

Twenty-four individuals (19 females, mean age = 20.3 years) from South China Normal University participated in the current study. Data from one additional participant were excluded due to low plausibility judgment accuracy (< 75%). All participants were

native speakers of Chinese and had normal or corrected-to-normal vision. All participants gave informed consent and were paid RMB10 per hour for their participation.

Materials and Design

The experimental stimuli consisted of 60 pairs of sentences, each with a canonical and reversed argument order. All experimental sentences used the highly frequent SOV Ba(把)-construction in Mandarin Chinese (see Chapter 2). As in Experiments 3 to 5, role-reversed sentences were created by reversing the order of the pre-verbal arguments in the canonical sentences, both of which were animate. Within an individual item set the canonical and role-reversed sentences had an identical verb-argument triplet and differed only in the order of the arguments. Items were designed such that no anomaly would be evident before the critical verb. Sentences were extended beyond the target verb with words that were held constant across conditions within each item set. Following previous studies on Chinese reading eye-movement (e.g., Yang, Wang, Xu, & Rayner, 2009), a half-space was inserted between all Chinese characters (i.e., regardless of word boundary) to improve the spatial resolution of the EM measures.

Experiment 7 shared the same design as Experiment 3, in which argument order was manipulated in sentences in which the verb had high vs. low predictability (Table 5-1). Thirty high- and 30 low-predictability item sets were selected from the top and bottom end of the cloze probability distribution of the 120 item sets from Experiment 3. The verb had high predictability (average cloze = 77%, range 62 - 97%) in half of the canonical sentences and low predictability (average cloze = 3.4%, range 3.3 - 3.4%) in the other half. The verb was not predictable (0% cloze) in the role-reversed sentences.

The 60 item sets were divided into 2 lists, such that each list contained exactly one version of each item and 15 items in each condition. Thus, each participant saw each item and each condition, but never saw more than one version of the same item. The experiment also contained 60 filler items of comparable length and complexity, all of which were also used in the ERP experiment (Experiments 3-5). All filler items are grammatically well-formed, but half of them were implausible, to ensure that the overall ratio of plausible-implausible sentences in the experiment was 1:1.

Condition	Sample materials
Low-predictability Canonical	/老 刘 把 鹦 鹉/ **训 练 了**/ 好 一 段/ 时 间 。 /Mr. Liu BA parrot/ **train**/ quite/ some time. *"Mr. Liu trained the parrot for quite some time."*
Low-predictability Role-reversed	/鹦 鹉 把 老 刘/ **训 练 了**/ 好 一 段/ 时 间 。 /Parrot BA Mr. Liu /**train**/ quite/ some time. *"The parrot trained Mr. Liu for quite some time."*
High-predictability Canonical	/警 察 把 小 偷/ **抓 了**/ 回 警 局/. . . /Cop BA thief/ **arrest**/ (return to) police station/... *"The cop arrested the thief (and brought him back) to the station."*
High-predictability Role-reversed	/小 偷 把 警 察/ **抓 了**/ 回 警 局/. . . /Thief BA cop/ **arrest**/ (return to) police station/... *"The thief arrested the cop (and brought him back) to the station."*

Table 5-1. Experimental conditions and regions of interest in sample materials in Experiment 7.

Procedure

Participants were tested individually, and eye movements were recorded using an EyeLink 1000 eyetracker (SR Research, Mississauga, Ontario, Canada), interfaced with a Dell PC. The sampling rate for recordings was 1000 Hz. Stimuli were displayed on a Dell 19-in. SVGA monitor. The monitor was set to a refresh rate of 150 Hz. Participants were seated 28 inches from the computer screen. Viewing was binocular, but only the right eye was recorded. Sentences were presented in 16 pt. Song-Ti font in white on a black

background on the computer monitor. All sentences in this experiment were displayed on a single line, with a maximum length of 29 characters.

The experiment was implemented using the Eye-Track software (http://www.psych.umass.edu/eyelab/software/). A 3-point calibration procedure was performed at the beginning of each testing session, and re-calibration was carried out between trials as needed. Before the experiment began, each participant was instructed to read for comprehension in a normal manner. The participant triggered the onset of each sentence by fixating a box on the left edge of the computer screen. Each participant read five practice items before the experimental items were shown. Participants were instructed to perform a binary plausibility judgment following every experimental and filler item. The order of experimental and filler items was randomized across participants. The entire experimental session lasted approximately 30 minutes.

Analysis

The initial stages of data analysis were carried out using EyeDoctor (http://www.psych.umass.edu/eyelab/software/). Trials were omitted from the analyses if a long duration track loss occurred at any time during a trial (e.g., if there was no data for half a line of text or more). This resulted in the exclusion of 2.1% of all trials. Fixations shorter than 80 ms in duration and within one character of the previous or following fixation were incorporated into the neighboring fixation. Remaining fixations of less than 80 ms or more than 800 ms were excluded.

Three regions of interest were defined for analysis. The *argument region* consisted of the subject noun phrase, the co-verb *ba* and the object noun phrase. The

critical region consisted of the verb plus the resultative adjective and/or the aspectual marker. The *post-critical region* included the two characters immediately following the critical region. The division into regions for a sample item is shown in Table 5-1.

For each region I computed and analyzed four eye-tracking measures: *first fixation duration*, *first-pass time*, *regression path time*, and *probability of regression*. First-fixation time is the duration of the reader's first fixation in a region, provided that the reader has not previously fixated on subsequent text. Since the argument region in a given item contained the same words in different word order across conditions (e.g.., cop ba thief vs. thief ba cop), first fixations in that region likely fell on different words depending on the order of the arguments. Therefore, effects on first fixation duration in the argument region are not meaningful and will not be discussed further. First-pass time is the sum of all fixations on a critical region before the reader leaves it for the first time, either to the left or to the right. Regression path time (also known as go-past time) is the sum of all fixation durations from when the reader first fixates the region until the reader's eyes leave the region to the right, including any time spent to the left of the region after a regressive eye movement and any time spent re-reading material in the region before moving on. The probability of regression measure gives the probability that a reader makes a regressive eye movement to any preceding region after fixating the region. This measure includes only regressions made during the reader's first pass through the region; it does not include regressions made after re-fixating the region. Effects of reanalysis are often apparent in the go-past and regression measures (Staub & Rayner, 2007). Skips of a region in a particular measure were treated as missing data points.

Linear mixed effects models (Baayen, Davidson, & Bates, 2008) were used to analyze the continuous data (reading time measures) in each region, with random intercepts for subjects and items, and with congruity and predictability and their interaction as fixed effects. Mixed effects logistic regression, with the same random and fixed factors (Jaeger, 2008), was used for analyzing categorical data such as the probability of regression as well as plausibility judgments. Thus for a reading-time measure, for example, the formal specification of our model in R's lme4 package would be value ~ reversal * predictability + (1|subj) + (1|item). The significance of LMER model coefficients was determined based on highest posterior density confidence intervals computed using Markov chain Monte Carlo (MCMC) sampling (see Baayen 2008, p. 270). For mixed effects logistic regression models I report p-values based on the Wald Z statistic conventionally used in logistic regression analysis. Since predictability was manipulated across different items, main effects of predictability are not interpreted. While the results of all comparisons are shown in Table 5-3, only significant main effects of reversal and reversal × predictability interactions will be discussed in the text. The same procedures were also used to examine the effect of role-reversals in high- and low-predictability sentences separately. In this case, the formal specification of our model in R's lme4 package was value ~ reversal + (1|subj) + (1|item).

5.4.2 Results

Plausibility Judgments

Across all experimental and filler trials, participants' plausibility judgments were 90.5% accurate on average. However, due to a coding error in the experimental paradigm, item and condition information for judgment data in this experiment was lost.

Reading Eye-movements

Table 5-2 shows the mean first fixation duration, first-pass time, regression path time and probability of regression for all four conditions across the regions of interest. Table 5-3 shows the results of the mixed effects model.

In the *argument region* (subject *ba* object), a marginally significant main effect of reversal (p <.06) was found in first pass time, indicating that first pass times were on average about 30ms longer in the role-reversed condition than in the canonical condition in both high- and low- predictability sentences.

In the *critical region* (target verb), all reading time measures (first fixation duration, first pass time, and regression path time) were longer in the role-reversed condition than in the canonical condition in both high- and low- predictability sentences. Comprehenders were also more likely to regress out of the critical region during first pass in the role-reversed condition than in the canonical condition.

In the *post-critical region*, regression path times were significantly longer in the role-reversed condition than in the canonical condition in both high- and low-

predictability sentences. Comprehenders were also more likely to regress out of this region during first pass in the role-reversed condition than in the canonical condition.

Separate analyses for high- and low-predictability sentences revealed a similar pattern of results. In high-predictability sentences, role-reversals led to significantly longer first pass times ($\beta = 25$, 95% CI [3, 43], $p = .014$) and regression path times ($\beta = 73$, 95% CI [23, 122], $p = .005$), as well as marginally significantly longer first fixation times ($\beta = 11$, 95% CI [0, 22], $p = .055$) in the critical region. Role-reversals also led to significantly higher probability of regression ($\beta = 0.62$, Wald $z = 2.9$, $p = .004$) and longer regression path time ($\beta = 87$, 95% CI [3, 164], $p = .034$) in the post-critical region. Meanwhile, in low-predictability sentences, a significant effect of role-reversals was found in the probability of regression ($\beta = 0.54$, Wald $z = 2.4$, $p = .015$) in the critical region.

Measure	NP1 ba NP2	Target Verb	Verb+1
First fixation duration			
Low-predictability, Canonical	238 (6)	246 (6)	228 (7)
Low-predictability, Role-reversed	239 (7)	254 (8)	228 (8)
High-predictability, Canonical	263 (9)	238 (7)	225 (6)
High-predictability, Role-reversed	247 (7)	249 (6)	232 (7)
First-pass time			
Low-predictability, Canonical	763 (38)	316 (14)	246 (10)
Low-predictability, Role-reversed	795 (49)	324 (15)	247 (11)
High-predictability, Canonical	779 (42)	291 (14)	281 (12)
High-predictability, Role-reversed	819 (41)	317 (13)	288 (14)
Regression path time			
Low-predictability, Canonical	883 (57)	424 (27)	364 (34)
Low-predictability, Role-reversed	905 (69)	462 (24)	411 (33)
High-predictability, Canonical	921 (70)	368 (22)	454 (33)
High-predictability, Role-reversed	935 (58)	442 (28)	554 (43)
Probability of regression			
Low-predictability, Canonical	.09 (.02)	.14 (.03)	.15 (.03)
Low-predictability, Role-reversed	.09 (.02)	.20 (.03)	.20 (.03)
High-predictability, Canonical	.11 (.03)	.14 (.02)	.21 (.03)
High-predictability, Role-reversed	.08 (.02)	.16 (.03)	.32 (.04)

Table 5-2. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 7. Standard error of the mean is in parentheses.

	NP1 ba NP2			**Target Verb**			**Verb+1**		
Reading time measure	β	95% CI	*p*	β	95% CI	*p*	β	95% CI	*p*
First fixation duration									
Predictability	-8	[-13 -4]	.001	3	[-2 8]	.206	2	[-5 8]	.595
Reversal	-4	[-8 0]	.066	5	[1 9]	**.020**	2	[-2 7]	.309
Predictability × Reversal	4	[0 8]	.030	-1	[-4 3]	.744	-1	[-5 4]	.810
First-pass time									
Predictability	-10	[-56 34]	.660	9	[-2 20]	.118	-14	[-29 0]	.056
Reversal	19	[-1 37]	*.058*	8	[1 15]	**.031**	1	[-7 9]	.784
Predictability × Reversal	-2	[-20 18]	.856	-4	[-11 3]	.265	0	[-8 7]	.940
Regression path time									
Predictability	-17	[-65 29]	.467	19	[-4 42]	.123	-46	[-95 0]	.055
Reversal	9	[-11 28]	.402	28	[10 46]	**.001**	38	[10 66]	**.006**
Predictability × Reversal	2	[-18 23]	.812	-8	[-25 11]	.391	-5	[-33 22]	.721
Probability of regression	β	Wald Z	*p*	β	Wald Z	*p*	β	Wald Z	*p*
Predictability	-.04	-0.30	.763	.05	0.47	.636	.05	0.47	.636
Reversal	-.11	-1.09	.274	.19	2.31	**.021**	.19	2.31	**.021**
Predictability × Reversal	.09	0.86	.389	.09	1.06	.289	.09	1.06	.289

Table 5-3. Linear mixed-effect and logistic regression model estimates in Experiment 7.

5.4.3 Discussion

Results in Experiment 7 showed that argument role-reversals had a very clear impact on comprehenders' eye-movement records. Role-reversals not only led to longer reading times and higher probability of regression in the critical and the post-critical regions, they also led to marginally longer first pass times in the pre-critical (argument) region.

The effects of role-reversals on EMs in the current experiment appeared to be much earlier than the ERP effects observed in Experiment 3. Given previous findings that role-reversals have a pronounced effect on comprehenders' ERPs, it is perhaps not surprising that role-reversals have a clear effect on comprehenders' EMs. However, considering the fact that role-reversals consistently elicited only a P600 effect at the target verb (which begins at around 500ms post stimulus onset), it is somewhat unexpected that they affected EM measures as early as first fixation durations at the verb and even in first pass times in the pre-critical region.

In particular, the (marginally significant) effect of role-reversals on first-pass time in the pre-critical region suggested that comprehenders were sensitive to argument role-reversals even before fixating on the target verb. Since the sentences were constructed to ensure that no anomaly was evident before the target verb, it is unclear what might have contributed to the first-pass time effect in the pre-critical region. There are several non-mutually-exclusive possibilities. Firstly, this early effect might reflect comprehenders' sensitivity to the relative order of the arguments that is completely independent of the verb. Since each argument pair (e.g., *cop* and *thief*) was only used in one item set, each

ordering of the two arguments only appeared in one condition (e.g., '*cop BA thief*' only appeared in the canonical condition and '*thief BA cop*' only appeared in the role-reversed condition). Therefore, the EM effect in the pre-critical region may be fully attributable to low-level differences between conditions (e.g., the relative frequency of the order of arguments across conditions), and this effect might have spilled over and led to longer first fixations in the following (critical) region.

Meanwhile, the fist-pass time effect in the pre-critical region might also reflect genuine processing costs that are triggered by the target verb. Previous research on preview effects in reading Chinese has shown that characteristics of the character to the immediate right of fixation (characters *N+1* and perhaps also N+2) can influence the processing of the currently fixated word (character *N*; e.g., Yang et al., 2009). This is referred to as a parafoveal-on-foveal effect (see Rayner, 2009, for a review). Given that comprehenders in the current study could preview the target word only when they were fixating on the last one or two characters in the pre-critical region, the effect on first pass time might have arisen entirely from fixations near the right-edge of the region. However, since (i) the arguments were ordered differently across the canonical and role-reversed conditions, (ii) they were not matched on lexical and/or orthographic factors, and (iii) the target verb was presented immediately to the right of the second argument, the characters immediately preceding the target verb were different between conditions. In fact, the average lexical frequency of the argument immediately before the verb was higher in the role-reversed condition than in the canonical condition. Therefore, it was not possible to meaningfully compare EMs in a region immediately preceding the verb, or to expand the target region to its left, in the current experiment.

In order to examine whether there is a genuine misalignment between readers' EMs and ERPs, in Experiment 8 I inserted a temporal phrase between the second argument and the target verb to create an identical pre-critical baseline (as in Experiment 5). In Experiment 9, I tried to minimize preview benefits in reading using a gaze-contingent boundary paradigm.

5.5 Experiment 8: Alignment upon further considerations

As discussed above, a major shortcoming of Experiment 7 is that, in the experimental materials, the characters immediately to the left of the target verb were not matched between conditions. This limitation made it difficult to identify the source of the early EM effects in Experiment 7. Therefore, in the current experiment I modified the materials used in Experiment 7 to include a temporal phrase between the second argument and the verb. The temporal phrases were 4- to 5-characters long, and the same temporal phrase was used within all versions of an item set. Therefore, they could serve as a matching pre-target baseline, and they provided much flexibility for data analysis.

As in Experiment 7, I expected to see a clear effect of role-reversals on comprehenders' EMs at and following the target verb. However, different hypotheses about the cause for the early effect in Experiment 7 made different predictions regarding the EMs at the pre-critical region. First, if there is not a genuine mis-alignment with ERPs and if the early effects in Experiment 7 (the first-pass time effect in the pre-critical region as well as the first fixation effect at the verb) were due to differences in argument ordering and were completely independent of the verb, then role-reversals should not impact fixations on the temporal phrase or the first fixations on the verb in the current

experiment. Instead, the effect of role-reversals might appear in later measures such as regression path time and/or in the post-critical region. Alternatively, if the results in Experiment 7 reflect genuine early sensitivity, such that readers' EMs are sensitive to role-reversal anomalies even when the verb is only present in preview (a parafoveal-on-foveal preview effect), then role-reversals should impact first fixation durations on the verb as well as in one or more of the first pass measures in the pre-critical region.

5.5.1 Methods

Participants

Twenty-four individuals (20 females, mean age = 21 years) from South China Normal University participated in the current study. All participants were native speakers of Chinese and had normal or corrected-to-normal vision. All participants gave informed consent and were paid RMB10 per hour for their participation.

Materials and Design

Experiment 8 followed the same design as Experiment 5, in which argument order was manipulated in sentences in which the arguments and the verb were separated by an adverbial temporal expression (Table 5-4). The experimental stimuli consisted of 60 pairs of sentences, each with a canonical and reversed argument order. They were different from the items in Experiment 7 in two important ways: (1) The linear distance between the verb and its arguments was increased by inserting a temporal phrase (e.g., "*zai zuotian xiawu*", *yesterday afternoon*) between the second argument and the verb, and (2) The first arguments were always sentence-initial. The critical verbs in the canonical

condition had an average 4.1% cloze probability (range 3.3 - 6.9%) in the low predictability sentences and an average of 71.6% (range 50.0 - 96.6%) in the high predictability sentences. The critical verbs in the role-reversed condition always had 0% cloze probability. The temporal phrases had an average length of 4.5 characters (range 4 – 5). As in Experiment 7, 30 high-predictability and 30 low-predictability item sets were divided into 2 lists (15 trials/condition), each of which also contained the same 60 filler items that were used in Experiment 7.

Condition	Sample materials
Low-predictability Canonical	/老 刘 把 鹦 鹉/ 在 那 年 夏 天/ **训 练 了**/ 好 一 段/ 时 间 。 /Mr. Liu BA parrot/ ZAI that summer/ **train**/ quite/ some time. *"Mr. Liu trained the parrot for quite some time."*
Low-predictability Role-reversed	/鹦 鹉 把 老 刘/ 在 那 年 夏 天/ **训 练 了**/ 好 一 段/ 时 间 。 /Parrot BA Mr. Liu / ZAI that summer/ **train**/ quite/ some time. *"The parrot trained Mr. Liu for quite some time."*
High-predictability Canonical	/警 察 把 小 偷/ 在 那 天 傍 晚/ **抓 了**/ 回 警 局/... /Cop BA thief/ ZAI that evening/ **arrest/** (return to) police station/... *"The cop arrested the thief (and brought him back) to the station."*
High-predictability Role-reversed	/小 偷 把 警 察/ 在 那 天 傍 晚/ **抓 了**/ 回 警 局/... /Thief BA cop/ ZAI that evening/ **arrest/** (return to) police station/... *"The thief arrested the cop (and brought him back) to the station."*

Table 5-4. Experimental conditions and regions of interest in sample materials in Experiment 8.

Procedure

The experimental procedures were identical to those in Experiment 7. Sentences were presented in 18 pt. Song-Ti font in white on a black background on the computer monitor. All sentences in this experiment were displayed on a single line.

Analysis

A total of 1.5% of all trials were omitted from the analyses due to a long duration track loss during a trial. The procedures for data analysis were identical to those in

Experiment 7. Three regions of interest were used. Since the first argument was at a sentence-initial position in the current materials, data from the *argument region* are not reported here. The *pre-critical region* consisted of the temporal phrase. The *critical region* consisted of the verb plus the resultative adjective and/or the aspectual marker. The *post-critical region* included the word (2 to 4 characters long) immediately following the critical region. The division into regions for a sample item is shown in Table 5-4.

5.5.2 Results

Plausibility Judgments

Participants' average acceptability judgment accuracy in each condition is shown in

Table 5-5. Across both high and low predictability sentences, participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 90.9%. Participants were more accurate in rejecting role-reversed sentences than in accepting canonical sentences (canonical = 89.1%; role-reversed = 92.8%; $\beta = 0.27$, p(Wald) < .01)

	Target Response	Percent accurate (SE)
Low predictability, Canonical	Yes	88.7 (2.2)
Low predictability, Role-reversed	No	91.3 (1.8)
High predictability, Canonical	Yes	89.5 (1.8)
High predictability, Role-reversed	No	94.4 (1.2)

Table 5-5. Target response and accuracy on plausibility judgment task in Experiment 8.

Reading Eye-movements (Original region definition)

Table 5-6 shows the mean first fixation duration, first-pass time, regression path time and probability of regression for all four conditions across the regions of interest. Table 5-7 shows the results of the mixed effects model.

In the *pre-critical region* (temporal phrase), there were no significant effects involving reversal in any of the four measures. A marginally significant main effect of reversal (p =.064) was found in probability of regression, showing that comprehenders were more likely to regress out of this region during first pass in the role-reversed condition than in the canonical condition in both high- and low- predictability sentences.

In the *critical region* (target verb), regression path times were on average 40ms longer in the role-reversed condition than in the canonical condition in both high- and low-predictability sentences, but this effect was only marginally significant (p =.089).

In the *post-critical region*, there was a marginally significant predictability × reversal interaction effect in first fixation (p = .093), indicating that first fixations were longer in the role-reversed than canonical condition in the low-predictability sentences (238ms vs. 226ms) but not in the high predictability sentences (232ms vs. 235ms). Meanwhile, there was a marginally significant main effect of reversal and no significant interaction in probability of regression (p = .084), although comprehenders were also more likely to regress out of this region during first pass in the role-reversed condition than in the canonical condition in the low-predictability sentences (34% vs. 25%) but not in the high-predictability sentences (33% vs. 35%),

Separate analyses for high- and low-predictability sentences revealed a similar pattern of results. In high-predictability sentences, no significant effect of role-reversals was observed in any of the measures in any of the ROIs. Meanwhile, in low-predictability sentences, role-reversals led to higher probability of regression ($\beta = 0.43$, Wald $z = 1.9$, $p = .055$) and longer regression path time ($\beta = 102$, 95% CI [-6, 208], $p = .064$) in the post-critical region, but the effects were only marginally significant.

Measure	Pre-target	Target Verb	Verb+1
First fixation duration			
Low-predictability, Canonical	238 (7)	251 (7)	226 (6)
Low-predictability, Role-reversed	234 (8)	253 (6)	238 (7)
High-predictability, Canonical	243 (6)	252 (8)	235 (7)
High-predictability, Role-reversed	249 (8)	251 (7)	232 (7)
First-pass time			
Low-predictability, Canonical	427 (31)	307 (12)	270 (11)
Low-predictability, Role-reversed	420 (28)	313 (13)	269 (9)
High-predictability, Canonical	475 (30)	301 (15)	282 (14)
High-predictability, Role-reversed	457 (29)	303 (15)	276 (12)
Regression path time			
Low-predictability, Canonical	537 (45)	440 (32)	497 (34)
Low-predictability, Role-reversed	549 (44)	480 (43)	644 (64)
High-predictability, Canonical	584 (50)	408 (24)	656 (70)
High-predictability, Role-reversed	600 (41)	448 (32)	603 (53)
Probability of regression			
Low-predictability, Canonical	.14 (.02)	.15 (.02)	.25 (.03)
Low-predictability, Role-reversed	.17 (.03)	.19 (.04)	.34 (.04)
High-predictability, Canonical	.12 (.02)	.17 (.02)	.35 (.04)
High-predictability, Role-reversed	.17 (.02)	.20 (.03)	.33 (.05)

Table 5-6. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 8 (original region definition). Standard error of the mean is in parentheses.

	Pre-target			**Target Verb**			**Verb+1**		
Reading time measure	β	95% CI	*p*	β	95% CI	*p*	β	95% CI	*p*
First fixation duration									
Predictability	-5	[-9 -1]	.009	0	[-6 5]	.982	0	[-7 6]	.846
Reversal	1	[-3 4]	.736	0	[-5 4]	.871	1	[-4 6]	.649
Predictability × Reversal	-2	[-6 1]	.236	1	[-3 5]	.623	4	[-1 9]	.093
First-pass time									
Predictability	-22	[-39 -4]	.018	3	[-7 13]	.520	-6	[-20 10]	.443
Reversal	-8	[-19 5]	.220	2	[-5 9]	.618	-3	[-12 6]	.474
Predictability × Reversal	1	[-11 13]	.830	2	[-6 9]	.675	1	[-7 10]	.784
Regression path time									
Predictability	-26	[-56 5]	.094	12	[-12 39]	.322	-34	[-112 42]	.371
Reversal	2	[-15 22]	.796	19	[-3 41]	*.089*	26	[-12 66]	.186
Predictability × Reversal	0	[-20 18]	.960	2	[-20 23]	.853	28	[-13 66]	.174
Probability of regression	β	Wald Z	*p*	β	Wald Z	*p*	β	Wald Z	*p*
Predictability	.07	0.63	.530	-.04	-0.35	.724	-.13	-0.98	.329
Reversal	.15	1.85	*.064*	.13	1.62	.105	.13	1.73	*.084*
Predictability × Reversal	-.05	-0.66	.507	.05	0.58	.562	.10	1.36	.174

Table 5-7. Linear mixed-effect and logistic regression model estimates in Experiment 8 (original region definition).

Reading Eye-movements (Alternate region definition)

In order to examine the possibility that the effects of role-reversal were masked by the way in which the ROIs were defined, eye-tracking measures were re-computed with an alternate region definition and reanalyzed using the same procedures. Under this alternate region definition, the boundary between the pre-critical and the critical region was moved by one Chinese character and a half-space character to the left. In order words, the pre-critical region contained the temporal phrase minus one Chinese character to the right, while the critical region contained the last character of the temporal phrase and the target verb. The post-critical region was unaffected.

Table 5-8 shows the mean first fixation duration, first-pass time, regression path time and probability of regression for all four conditions across the regions of interest. Table 5-9 shows the results of the mixed effects model. Since the alternate region definition did not affect the data in the *post-critical region*, please refer to the discussion above for the results in this region.

In the *pre-critical region* (temporal phrase minus the right-most character), a marginally significant main effect of reversal (p =.051) was found in first pass times, indicating that first pass times were longer in the canonical condition than in the role-reversed condition in both high- and low- predictability sentences. This effect was not observed under the original region definition.

In the *critical region* (target verb plus one character to the left), a significant effect of reversal was found in regression path time, which was longer in the role-reversed condition than in the canonical condition in both high- and low-predictability

sentences. This effect was marginally significant (p =.089) under the original region definition. Comprehenders were also more likely to regress out of this region during first pass in the role-reversed condition than in the canonical condition. This effect likely resulted from the generally higher probability of regression in both the pre-critical and critical regions under the original region definition.

Separate analyses for high- and low-predictability sentences revealed a similar pattern of results. In high-predictability sentences, role-reversals led to significantly higher probability of regression (β = 0.51, Wald z = 2.5, p = .013) in the critical region. Meanwhile, in low-predictability sentences, role-reversals led to significantly higher probability of regression (β = 0.42, Wald z = 2, p = .047) and marginally longer regression path time (β = 61, 95% CI [-9, 124], p = .084) in the critical region, although in the pre-critical region first pass time was marginally shorter in the role-reversed condition than in the canonical condition (β = -26, 95% CI [-54, 3], p = .081).

Measure	Pre-target	Target Verb	Verb+1
First fixation duration			
Low-predictability, Canonical	237 (7)	253 (8)	226 (6)
Low-predictability, Role-reversed	233 (8)	248 (7)	238 (7)
High-predictability, Canonical	241 (6)	253 (7)	235 (7)
High-predictability, Role-reversed	249 (8)	248 (7)	232 (7)
First-pass time			
Low-predictability, Canonical	360 (24)	386 (18)	270 (11)
Low-predictability, Role-reversed	340 (19)	381 (20)	269 (9)
High-predictability, Canonical	384 (22)	383 (20)	282 (14)
High-predictability, Role-reversed	372 (26)	369 (20)	276 (12)
Regression path time			
Low-predictability, Canonical	453 (37)	521 (32)	497 (34)
Low-predictability, Role-reversed	438 (31)	576 (50)	644 (64)
High-predictability, Canonical	478 (43)	498 (29)	656 (70)
High-predictability, Role-reversed	480 (34)	540 (39)	603 (53)
Probability of regression			
Low-predictability, Canonical	.14 (.02)	.16 (.03)	.25 (.03)
Low-predictability, Role-reversed	.15 (.03)	.22 (.04)	.34 (.04)
High-predictability, Canonical	.11 (.02)	.15 (.02)	.35 (.04)
High-predictability, Role-reversed	.15 (.02)	.22 (.03)	.33 (.05)

Table 5-8. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 8 (alternate region definition). Standard error of the mean is in parentheses.

	Pre-target			**Target Verb**			**Verb+1**		
Reading time measure	β	95% CI	*p*	β	95% CI	*p*	β	95% CI	*p*
First fixation duration									
Predictability	-5	[-9 -1]	.015	0	[-5 5]	.947	0	[-7 6]	.856
Reversal	1	[-3 5]	.620	-3	[-7 1]	.213	1	[-4 6]	.643
Predictability × Reversal	-3	[-7 1]	.172	0	[-4 4]	.982	4	[-1 9]	*.086*
First-pass time									
Predictability	-16	[-30 -1]	.032	4	[-10 18]	.610	-6	[-20 9]	.425
Reversal	-10	[-21 0]	*.051*	-4	[-13 5]	.369	-3	[-12 5]	.441
Predictability × Reversal	-3	[-14 7]	.604	3	[-5 12]	.490	1	[-7 10]	.762
Regression path time									
Predictability	-19	[-45 6]	*.137*	11	[-15 38]	.399	-34	[-114 40]	.382
Reversal	-8	[-26 9]	.355	24	[2 47]	**.040**	26	[-12 67]	.191
Predictability × Reversal	-4	[-21 12]	.643	4	[-19 26]	.727	28	[-11 68]	.172
Probability of regression	β	Wald Z	*p*	β	Wald Z	*p*	β	Wald Z	*p*
Predictability	.08	0.69	.492	-.01	-0.09	.928	-.13	-0.98	.329
Reversal	.11	1.28	*.199*	.23	3.11	**.002**	.13	1.73	*.084*
Predictability × Reversal	-.06	-0.76	.448	-.04	-0.48	.631	.10	1.36	.174

Table 5-9. Linear mixed-effect and logistic regression model estimates in Experiment 8 (alternate region definition).

5.5.3 Discussion

In the current experiment, the effect of role-reversal was spread out and seemed elusive under the original region definition. The effect became clearer when the critical region was extended to include one extra character to the left of the target verb. Under this region definition, role-reversals led to higher probability of regression as well as to longer regression path times in the critical region.

The current experiment introduced one modification to the materials used in Experiment 7. The target verb was immediately preceded by its arguments in Experiment 7, such that the pre-critical region contained the same words that were ordered differently between conditions. Since a temporal phrase was placed between the second argument and the target verb, the pre-critical region in the current experiment was matched between conditions. Although all other aspects of the materials remained unchanged, these experiments yielded notably different results. Argument role-reversals affected readers' EMs in both experiments, but the time course and the prevalence of the effect were clearly different across these experiments.

First, first-pass time in the pre-critical region and first fixation durations were longer in the role-reversed condition than in the canonical condition in Experiment 7, but they did not differ between conditions in the current experiment. In fact, first pass times in the pre-critical region were numerically shorter in the role-reversed condition than in the canonical condition in the current experiment. This suggests that the early effect observed in Experiment 7 was not triggered by the target verb. Instead, it may simply

show that, even in the absence of a verb, the ease of processing of a given pair of arguments is modulated by their word order and/or structural roles.

Meanwhile, the effect of role-reversals also seemed more elusive in the current experiment than in Experiment 7. Although a significant effect of role-reversal was observed across different measures and regions in Experiment 7, it was only observed in two measures in the current experiment, namely probability of regression and regression path time, in the critical region (under *alternate region definition*). This suggests that, once baseline differences between the conditions are eliminated (e.g., by introducing a matching pre-critical region), the effect of role-reversal on EMs can be quite localized.

Qualitative alignment between EM and ERP results

The design of the current experiment was identical to that of Experiment 5, which manipulated the order of the arguments and the predictability of the verb in the canonical condition in SOV sentences in which the arguments and the verb were separated by a temporal phrase (e.g., cop_{SUBJ} BA $thief_{SUBJ}$ [ZAI yesterday evening] <u>arrest</u>). In Experiment 5, while role-reversals elicited only a P600 effect in low-predictability sentences, the same manipulation elicited a significant N400 effect as well as a P600 effect in high-predictability sentences. However, in the current experiment, no predictability × role-reversal interaction was observed in any of the measures in any ROI. In fact, when the data from high- and low- predictability sentences were analyzed separately (see above), role-reversals led to higher probability of regression in the critical region in both cases.

At first glance, this seems to be a qualitative difference between the ERP results in Experiment 5 and the EM results in the current experiment. Whereas argument role-reversals elicited qualitatively different ERP effects in high- vs. low-predictability sentences, the effects of argument role-reversals on readers' EMs were indistinguishable between these sentences. However, given the numerous differences between how these sentences are presented during ERP and EM experiments, differences in ERP and EM results must not be taken at face value.

I propose that this apparent mismatch between the EM results in the current experiment and the ERP results in Experiment 5 shows that the temporal phrase served distinct functions in these experiments. Here I will (i) review the motivations for including a temporal phrase in each of these experiments and (ii) explain why it had distinct consequences on readers' reading experience in these experiments.

In the current experiment, a temporal phrase was inserted between the arguments and the verb to create a matching pre-critical region. With readers reading 3 to 5 word per second in natural reading, effects on EMs at word *N* often spill over to fixations on word *N+1*. Therefore, the temporal phrase in the current experiment served as a buffer to neutralize any differences that might have arisen in the argument region. Meanwhile, in Experiment 5 (ERP), the same temporal phrase was included in order to increase the amount of time between the presentation of the arguments and the verb. This was motivated by the hypothesis that the N400 was insensitive to role-reversals when the verb appeared immediately following its arguments because comprehenders did not have sufficient time for computing predictions. Therefore, the temporal phrase was placed

between the arguments and the verb to widen the time interval for predictive computations.

Since Experiment 5 used an RSVP paradigm, the inclusion of a temporal phrase had a deterministic effect on when readers saw the target verb relative to the arguments. Specifically, with a fixed 600ms SOA, the time elapsed between onset of the second argument and that of the verb was increased by 1200ms in each trial, to a total of 1800ms. Meanwhile, although the inclusion of the temporal phrase also delayed readers' first fixation on the target verb, regression path times in the pre-critical region revealed that the temporal phrase only delayed readers' first fixation on the target verb by 462ms on average. Coupled with the fact that readers spent much less time reading the argument region in the current experiment than in Experiment 5, the lack of a predictability × role-reversal interaction in the current experiment may indicate that, despite the inclusion of a temporal phrase, information about the arguments' structural roles had yet to impact comprehenders' predictions about the verb when the verb was first fixated.

Temporal alignment between EM and ERP results

Since standard measures of EMs are defined in terms of text regions (e.g., duration of fixations in a given text region), they yield precise information about where experimental variables affect EMs, but they are not informative about when an effect occurs (other than during first pass vs. second pass reading). Naturally, one can determine the relative timing of EM effects based on how the measures are computed. For example, one can infer that a regression path time effect occurs later than a first fixation effect in the same region. However, the presence of an effect on standard EM measures provides

only limited information about the timing of the effect. For example, an effect on regression path time reflects the fact that it takes readers longer to go past a region in one condition than another, but it does not tell us when readers begin to show sensitivity to the difference between conditions (e.g., How many milliseconds following the onset of their first fixation in the region did they show sensitivity to the experimental manipulation?).

In the current experiment, since argument role-reversals led to increased probability of regression and regression path times in the critical region, the time course of the effect could be estimated by determining the latency of readers' regressions in the critical region. Therefore, I identified the subset of trials in which readers regressed out of the critical region during the first pass. This accounted for 18.5% of the trials in which the critical region was fixated during first pass (249 out of 1342 trials). I then computed the first pass reading time in the critical region for these trials. This analysis revealed that the average latency of regressions in the critical region was 363ms[4] (ranging from 340ms to 370ms across conditions[5]). That is, for trials in which readers regressed out of the critical region during first pass, their regressive eye-movement occurred on average 363ms following the onset of their first fixation in that region. This complements the observation that readers were more likely to regress out of the region in the role-reversed condition than in the canonical condition. By taking into consideration the amount of time needed for planning a regressive eye-movement (saccade latency is at least 150-175

[4] Since the computation of first pass time only includes fixation durations but not saccade durations, these numbers slightly underestimate the actual regression latency.

[5] Note, however, that the means suggested that the effect of argument role-reversals persisted into the post-critical region in the low-predictability sentences but not in the high-predictability sentences.

ms; Abrams & Jonides, 1988; Rayner, Slowiaczek, Clifton, & Bertera, 1983; Salthouse & Ellis, 1980; Salthouse, Ellis, Diener, & Somberg, 1981), we may estimate that argument role-reversals can impact readers' decision to initiate a regressive eye-movement at the verb by around 200ms post first-fixation onset. Compared to the onset latency of the P600 effect (~500ms post stimulus onset), the results of this follow-up analysis suggest that role-reversals may impact readers' EMs more quickly than their ERPs. However, as parafoveal preview was available in Experiment 8 (natural reading) but not in the ERP experiments (RSVP), it remains difficult to determine how much of this apparent mis-alignment is attributed to preview benefits. Therefore, in the following experiment (Experiment 9) I used a gaze-contingent boundary paradigm (Rayner, 1975) to minimize the preview benefits that are available to readers in natural reading.

5.6 Experiment 9: Contribution of parafoveal preview in reading

To further investigate whether there is a genuine misalignment between comprehenders' EMs and ERPs, in the current experiment I aimed to better match the information flow in EM and ERP experiments by minimizing the benefits of parafoveal preview in natural reading. In particular, I used a gaze-contingent boundary paradigm (Rayner, 1975) to manipulate the validity of the preview of the target verb.

In a gaze-contingent boundary paradigm, the identity of a target word is initially masked and its presentation is contingent on the reader's eye-gaze. In a given trial, an invisible, predetermined boundary is placed just to the left of a target word location that is initially occupied by a preview word. When the reader's eyes cross the boundary location, the preview word is replaced by the target word. Since this display change

occurs during a saccade, at which point vision is suppressed, readers generally do not notice the change.

Since the properties of parafoveal preview can be controlled and manipulated in this paradigm, it has been used widely to study what information readers can extract when a word is in preview. For example, Lesch, Morris, & Rayner (1992) examined whether phonological information can be obtained from the parafovea by comparing how having a homophone preview vs. an orthographic control preview may impact processing of a target word. Based on the observation that reading times on the target word (beach) is shorter with a homophone preview (beech) than with an orthographic control preview (bench), the authors inferred that readers can access phonological information of the preview word in the parafovea and thereby facilitate lexical access of the target word.

Although this paradigm allows detailed manipulation of preview information and inferences about the kind of information readers are sensitive to in preview, it is not possible to simply 'remove' parafoveal preview benefits. This is because the size of preview benefits is dependent on many factors, such as lexical frequency of the targets (e.g., English: Inhoff & Rayner, 1986; Reingold et al., 2012) and predictability (English: Balota, Pollatsek, & Rayner, 1985), the lexical status of the preview word (English: Choi & Gordon, 2012), the plausibility of the preview word in the given context (Chinese: Yang, 2010), as well as the orthographic, phonological and semantic similarity between the preview and target words (e.g., English: Pollatsek et al., 1992; Chinese: Tsai, Lee, Tzeng, Hung & Yen , 2004; Yan, Richter, Shu & Kliegl, 2009). Just as the presence of valid preview information facilitates processing, the presence of invalid preview

information may disrupt processing. For example, although we can infer that readers are sensitive to the semantic properties of a word in preview if reading times on a target word is shorter with a semantically related preview than with a semantically unrelated preview, we cannot determine the extent to which this effect is due to processing facilitation from having processed a semantically related preview word or processing disruption from having processed a semantically unrelated preview word. Therefore, this paradigm allows inferences about the kind of information that can be processed in parafovea but does not provide an estimate of an absolute parafoveal preview benefit in natural reading.

In the current study I manipulated the validity of the preview for the target verbs by using 2-character pseudowords that contained two real Chinese characters. In the valid preview condition, the target verb was never masked. In the invalid preview condition, the first two characters of the target verb were initially replaced with a two-character pseudoword. Each of the characters in the preview pseudoword was closely matched in number of strokes and character-frequency to the target character in the corresponding position (c.f. Yang, 2010).

Preview validity was fully crossed with argument role-reversal, resulting in a two (valid vs. invalid preview) by two (canonical vs. role-reversed) within-subjects design. Since the current study introduced a new manipulation (preview validity) and since the predictability manipulation in Experiments 7 and 8 never showed any interaction with argument role-reversal, only low-predictability sentences were used in the current experiment. Low-predictability sentences were chosen over high-predictability sentences to facilitate direct comparisons between EM and ERP results. This is because role-

reversals always elicited only a P600 effect in low-predictability sentences regardless of the linear distance between the arguments and the verb.

Based on previous results, I expected longer reading times on the on the target word in the invalid preview condition compared to the valid preview condition. Based on the results in Experiment 8, I expected that readers would be more likely to regress out of the target region in the role-reversed condition than in the canonical condition. In addition, as the target verb was not available in the parafovea in the invalid preview condition, the effect of argument role-reversal was expected to have a later onset in the invalid compared to the valid preview condition.

5.6.1 Methods

Thirty-six individuals (30 females, mean age = 22.1 years) from South China Normal University participated in the current study. Data from two additional participants were excluded due to significant data loss (> 50% in one or more conditions). All participants were native speakers of Chinese and had normal or corrected-to-normal vision. All participants gave informed consent and were paid RMB10 per hour for their participation.

Materials and Design

The experimental stimuli consisted of 56 item sets adapted from the low predictability items in the ERP experiment (Experiment 5). Two factors were crossed: Argument role-reversal (canonical vs. role-reversed) and Preview type (valid vs. invalid preview). As in Experiment 8, a temporal phrase (e.g., "*zai zuotian xiawu*", *yesterday*

afternoon) was placed between the second argument and the verb. The temporal phrases had an average length of 4.6 characters (range 4 – 5). A sample set of experimental items and regions of interest are shown in Table 5-10.

In all experimental stimuli the critical region had 2 target characters (plus the aspectual marker –le 了 in some cases). On average the target characters had 8.8 strokes (range 3-17) and a log character-frequency of 3.95 (range 2.3-5.7; Cai & Brysbaert, 2010). Two-character pseudowords were used as invalid preview. The two characters did not form a word, and they made no sense in the sentence context. Each of the characters in the pseudoword was closely matched in number of strokes (average 8.8; range 3-17) and log character-frequency (average 3.95; 2.3-5.7) to the target character in the corresponding position. Moreover, to avoid any orthographic and homophonic benefit from the invalid previews, the characters in invalid preview did not share any radicals (components of Chinese characters) and did not rhyme with the target character in the corresponding position. Using the gaze-contingent boundary paradigm, we presented a valid (identical) or invalid (pseudoword) preview that changed to the target characters when the reader moved his or her eyes across an invisible boundary. The boundary was placed immediately to the left of the first target character.

A Dell 17-in. SVGA monitor was used to display the stimuli. The monitor was set to a refresh rate of 150 Hz. The delay in detecting an eye movement crossing the boundary and changing the display averaged 10 ms. Since the display change occurred during a saccade, readers were not aware of the change, as confirmed by their reports. All stimuli were presented in white on a black background on the computer monitor. All

characters were printed in simple Song font (16 point). Each character was about 1 x 1 cm in size and subtended approximately 0.8° of visual angle (with the participants' eyes being 71 cm away from the monitor).

The 56 item sets were divided into 4 lists, such that each list contained exactly one version of each item and 14 items in each condition. Each list also contained 60 filler items, all of which were also used in Experiments 7 and 8.

Condition	Sample materials
Canonical	/老 刘 把 鹦 鹉/ 在 那 年 夏 天/ {巧 松 | 训 练} 了/ 好 一 段/ 时 间 。 /Mr. Liu BA parrot/ ZAI that summer/ {**#####** | **train**}/ quite/ some time. *"Mr. Liu trained the parrot for quite some time."*
Role-reversed	/鹦 鹉 把 老 刘/ 在 那 年 夏 天/ {巧 松 | 训 练} 了/ 好 一 段/ 时 间 。 /Parrot BA Mr. Liu / ZAI that summer/ {**#####** | **train**}/ quite/ some time. *"The parrot trained Mr. Liu for quite some time."*

Table 5-10. Experimental conditions and regions of interest in sample materials in Experiment 9. The pseudoword in the invalid preview condition is presented on the left of the target.

Procedure

The experimental procedures were identical to those in Experiment 8.

Analysis

Trials were examined for display change errors and irregular eye-movement patterns. Trials in which the display change (i) occurred during a fixation, (ii) was triggered by or immediately followed by a blink, (iii) was triggered by a long-distance saccade (e.g., when comprehenders quickly glanced across the screen), or (iv) was immediately followed by a fixation that fell to the left of the boundary were excluded. Data from 2 out of 38 participants were discarded because more than 50% of trials in one

or more condition were excluded. For the remaining 36 participants, a total of 19.1% of all trials were omitted from the analyses. The procedures for statistical analyses were identical to those in Experiment 8, with the exception that the factor *predictability* was replaced with the factor *preview*. Data in valid and invalid preview conditions were first analyzed together, then separately. The original region definition used in Experiment 8 was adopted here. The region definition is illustrated in Table 5-10.

5.6.2 Results

Plausibility Judgments

Participants' average acceptability judgment accuracy in each condition is shown in Table 5-11. Regardless of preview validity, participants reliably judged canonical sentences to be plausible and the role-reversed sentences to be implausible with an overall accuracy of 89.6%, with no significant difference between conditions (*p*s >.10).

	Target Response	Percent accurate (SE)
Valid Preview, Canonical	Yes	87.8 (2.1)
Valid Preview, Role-reversed	No	90.2 (1.8)
Invalid Preview, Canonical	Yes	90.4 (1.7)
Invalid Preview, Role-reversed	No	90.6 (1.7)

Table 5-11. Target response and accuracy on plausibility judgment task in Experiment 9.

Reading Eye-movements

Table 5-12 shows the mean first fixation duration, first-pass time, regression path time and probability of regression for all four conditions across the regions of interest. Table 5-13 shows the results of the mixed effects model.

In the *pre-critical region* (temporal phrase), a significant preview × reversal interaction was found in both regression path time and probability of regression. Regression path time was longer and probability of regression was higher in the role-reversed condition than in the canonical condition when the target verb was available in preview but the effect was in the opposite direction when the preview was invalid.

In the *critical region* (target verb), all reading time measures (first fixation duration, first pass time, and regression path time) were longer in the invalid preview condition than in the valid preview condition. Comprehenders were also more likely to regress out of this region during first pass in the invalid preview than in the valid preview condition. There was no significant main effect of reversal or preview × reversal interaction in any of the measures.

In the *post-critical region*, first pass times were significantly longer in the invalid preview than in the valid preview condition. Meanwhile, comprehenders were more likely to regress out of this region during first pass in the role-reversed condition than in the canonical condition. There was no significant interaction effect in any of the measures.

Separate analyses for data in the valid and invalid preview conditions revealed a similar pattern of results. In the valid preview conditions, role-reversals led to significantly higher probability of regression ($\beta = 0.45$, Wald $z = 2.2$, $p = .026$) and marginally longer first fixation durations ($\beta = -10$, 95% CI [-22, 1], $p = .073$) in the post-critical region. Meanwhile, in the invalid preview conditions, role-reversals led to significantly longer first-pass times ($\beta = 28$, 95% CI [2, 58], $p = .044$) in the critical region.

Measure	Pre-target	Target Verb	Verb+1
First fixation duration			
Valid preview, Canonical	250 (6)	255 (5)	237 (7)
Valid preview, Role-reversed	244 (6)	264 (8)	230 (8)
Invalid preview, Canonical	251 (6)	283 (9)	233 (7)
Invalid preview, Role-reversed	251 (7)	279 (9)	235 (6)
First-pass time			
Valid preview, Canonical	483 (21)	356 (13)	271 (9)
Valid preview, Role-reversed	489 (23)	362 (13)	264 (12)
Invalid preview, Canonical	483 (22)	392 (15)	277 (10)
Invalid preview, Role-reversed	491 (24)	403 (17)	285 (11)
Regression path time			
Valid preview, Canonical	555 (29)	482 (28)	479 (45)
Valid preview, Role-reversed	589 (31)	469 (23)	508 (34)
Invalid preview, Canonical	601 (29)	533 (32)	510 (44)
Invalid preview, Role-reversed	572 (28)	548 (30)	535 (40)
Probability of regression			
Valid preview, Canonical	.09 (.02)	.16 (.02)	.23 (.03)
Valid preview, Role-reversed	.11 (.02)	.15 (.02)	.28 (.03)
Invalid preview, Canonical	.13 (.02)	.21 (.02)	.21 (.03)
Invalid preview, Role-reversed	.09 (.02)	.19 (.03)	.25 (.03)

Table 5-12. Grand average reading times (in milliseconds) and probability of regression by condition for each ROI in Experiment 9. Standard error of the mean is in parentheses.

	Pre-target			**Target Verb**			**Verb+1**		
Reading time measure	β	95% CI	*p*	β	95% CI	*p*	β	95% CI	*p*
First fixation duration									
Preview	-2	[-5 2]	.371	-11	[-15 -7]	**.000**	-2	[-6 3]	.436
Reversal	-1	[-5 2]	.472	2	[-3 6]	.471	-3	[-7 1]	.154
Preview × Reversal	-2	[-6 2]	.340	3	[-1 7]	.192	-2	[-6 2]	.353
First-pass time									
Preview	-1	[-12 10]	.839	-21	[-28 -13]	**.000**	-9	[-15 -2]	**.010**
Reversal	2	[-8 13]	.673	3	[-5 10]	.496	-2	[-9 4]	.504
Preview × Reversal	0	[-11 10]	.935	-1	[-9 6]	.752	-3	[-10 3]	.309
Regression path time									
Preview	-6	[-21 10]	.469	-34	[-52 -16]	**.000**	-26	[-55 4]	*.086*
Reversal	-1	[-16 15]	.920	-1	[-19 17]	.948	18	[-12 47]	.219
Preview × Reversal	16	[1 31]	**.048**	-7	[-26 11]	.436	9	[-20 38]	.573
Probability of regression	β	Wald Z	*p*	β	Wald Z	*p*	β	Wald Z	*p*
Preview	-.01	-0.08	.939	-.17	-2.34	**.019**	.06	0.76	.450
Reversal	-.06	-0.67	.504	-.04	-0.62	.535	.18	2.36	**.018**
Preview × Reversal	.17	2.06	**.040**	-.01	-0.13	.893	.05	0.72	.470

Table 5-13. Linear mixed-effect and logistic regression model estimates in Experiment 9.

5.6.3 Discussion

The results of the current experiment can be summarized as follows: (i) preview validity had a clear effect on reading times in both critical and post-critical regions; (ii) argument role-reversals only led to higher probability of regression in the post-critical region; and somewhat surprisingly, (iii) these two factors interacted to impact probability of regression and regression path time in the pre-critical region.

First, the current results showed that readers' processing of the target verb was delayed when it was not available in preview. A significant effect of preview validity was observed across all four EM measures in the critical region as well as on first pass time in the post-critical region. This suggests that the validity of preview information had a substantial impact on readers' processing of the target verb.

Second, argument role-reversals led to increased probability of regression in the post-critical region in the current experiment. Although argument role-reversals also led to increased probability of regression in Experiment 8, a significant effect was observed only in the critical region. In order to evaluate this apparent discrepancy, the results in the valid preview condition in the current experiment were compared against the results in the low-predictability sentences in Experiment 8 under the same region definition (*original region definition*; see Section 5.5.1). Inspection of the means revealed that role-reversals led to an increased probability of regression in the post-critical region in both cases, but their effect was observed in the critical region only in Experiment 8. Given that the post-critical region was read later than the critical region, this discrepancy suggests that, even in trials in which the target word was available in preview, the effect of

argument role-reversal had a later onset in the current experiment than in Experiment 8. I will return to discuss the significance of this difference in the *General Discussion*.

Further, the current results showed a preview validity × argument role-reversal interaction on probability of regression in the pre-critical region, but not on any of the measures in either the critical or post-critical region. A closer examination of the averages revealed that the interaction effect (p = .04) was attributable to the fact that argument role-reversals had opposite effects in the invalid vs. valid preview condition, such that probability of regression increased from 9% to 11% in the valid preview condition but decreased from 13% to 9% in the invalid preview condition. Note, however, that the effect in the valid preview condition in the pre-critical region was quite small in comparison with the effect in the post-critical region and the effect in Experiment 8. Further, the decrease in regression percentages in the invalid preview condition was completely unexpected, and neither of the pairwise comparisons revealed a statistically significant difference. Therefore, here I refrain from concluding what gave rise to this interaction effect.

Meanwhile, since the processing of the target verb was delayed when it was not available in preview, readers should detect role-reversal anomalies later in time in the invalid preview condition than in the valid preview condition. Therefore, one might expect role-reversals to be associated with longer reading times and/or more regressions later in time in the invalid preview condition compared to the valid preview condition. However, if role-reversals led to the same qualitative effect (for example, increased

probability of regression) at different time points but, crucially, in the same region of text, then such timing differences would not be evident in standard EM measures.

Therefore, I followed the same reasoning outlined in Experiment 8 and analyzed the latency of readers' regressions across conditions. Since a significant effect of probability of regression was observed in the post-critical region in the current experiment, I identified a subset of trials in which (i) the critical and post-critical regions were read sequentially (i.e., readers' eyes proceeded from the critical region directly to the post-critical region without regressing out) and (ii) they regressed out of the post-critical region during first pass. This accounted for about 19.8% of all trials in which both critical and post-critical regions were fixated during first pass (233 out of 1177 trials). I then computed the total first pass time in the critical and post-critical regions for each of these trials to determine the average latency of the regression from the post-critical region[6]. This analysis revealed that these regressions occurred on average 620ms and 697ms following the onset of the first fixation in the critical region in the valid and invalid preview conditions respectively. In other words, while role-reversals lead to more regressions out of the post-critical region regardless of preview validity, the regressions in the invalid preview condition occurred on average ~80ms later than those in the valid preview condition.

[6] As noted in Experiment 8, these first pass times underestimate the actual regression latency as saccade durations were not included in the computation. Further, since fixation durations were analyzed in a larger region here compared to in Experiment 8 and there were presumably more saccades in a larger region, regression latency was underestimated to a greater extent here than in Experiment 8.

Temporal alignment between EM and ERP results

In the current experiment a boundary paradigm was used to present invalid (pseudoword) parafoveal preview for the target words in order to more directly compare the timing of the EM and ERP results. Based on the analysis reported above, argument role-reversals increased readers' likelihood of regression at around 700ms following the onset of the first fixation on the target verb when it could not be viewed in the parafovea. Taking into consideration the amount of time needed to program a regressive eye-movement (saccade latency is at least 150-175 ms; Abrams & Jonides, 1988; Rayner, Slowiaczek, Clifton, & Bertera, 1983; Salthouse & Ellis, 1980; Salthouse, Ellis, Diener, & Somberg, 1981), we may estimate that argument role-reversals can impact readers' decision to initiate a regressive eye-movement by around 500-550 ms post first-fixation onset. In Experiment 5, role-reversals elicited only a P600 effect in the low-predictability sentences, which had an onset latency of ~500ms. Therefore, the results of this follow-up analysis suggest that, when parafoveal preview is disrupted in natural reading, the effects of role-reversals on readers' EMs and ERPs were relatively closely aligned in time.

5.7 General Discussion

In this chapter I examined how argument role-reversals impact readers' eye-movements and the extent to which the effect of role-reversals on readers' EMs and ERPs aligned in time. Using the experimental materials from a previous ERP experiment (Experiment 3), the initial EM experiment (Experiment 7) suggested that readers' EMs were sensitive to argument role-reversals much more quickly than were their ERPs. By taking into considerations differences between the stimulus presentation paradigms, I

discussed why EMs and ERPs might not align even when the same experimental materials are used. In Experiment 8 I introduced a matching baseline (pre-critical region) to the experimental materials by inserting a temporal phrase between the arguments and the target verb (as in Experiment 5, ERP). By moving the target verb further away from the arguments, early EM differences attributed to different argument orders no longer spilled-over to affect EMs at the target word, and the effect of role-reversal was only observed in probability of regression and regression path times. Lastly, I used a gaze-contingent boundary paradigm (Rayner, 1975) in Experiment 9 to examine the time course of readers' sensitivity to argument role-reversals when the target verb was not present in preview. Argument role-reversal led to more regressions in the post-critical region and these regressive eye-movements occurred at around the same time as the onset of the P600 effect in the ERP experiments. These results suggested that the effects of argument role-reversal on EMs and ERPs are roughly temporally aligned.

Qualitative alignment between EMs and ERPs

With a matching pre-critical baseline, both Experiments 8 and 9 showed that the effect of argument role-reversals was rather selective, affecting only regression-related measures. Meanwhile, Experiments 1-5 (ERP) consistently showed that argument role-reversals always elicited a P600 effect. The observation that role-reversals led to more regressive eye-movements and elicited a P600 effect across experiments suggests a tentative link between the cognitive processes that trigger regressions in reading and those that underlie the P600.

Although regressions account for about 10-15% of all saccades in natural reading, very little is known about what causes them (Rayner, 1998). Many regressions are triggered by comprehension failures (Blanchard & Iran-Nejad, 1987; Ehrlich, 1983; Frazier & Rayner, 1982; Hyönä, 1995; Just & Carpenter, 1978; Shebilske & Fisher, 1983; Vauras, Hyönä, & Niemi, 1992), and regressions have traditionally been taken to reflect readers' efforts to re-read and reconfigure earlier material (Frazier & Rayner, 1982; Mitchell, Shen, Green & Hodgson, 2008). Whereas increased fixation durations and regressions may be taken to reflect the same underlying processes because many experimental manipulations that modulate one of the measures also affect the other, Staub (2010) has proposed that they reflect distinct kinds of processing difficulty. In particular, he proposed that fixation durations may be lengthened when processing is difficult but eventually succeeds (e.g., processing an infrequent or ambiguous word) whereas inter-word regressions reflect processing failures (e.g., garden path disambiguation or implausibility). In line with Staub's (2010) proposal, the implausibility that resulted from argument role-reversals led to more regressions but did not impact other reading time measures in the current study. Such selectivity is consistent with the proposal that the cognitive processes that underlie regressions may be distinguished from those that underlie increased reading times. Meanwhile, although the P600 has been extensively studied in the ERP literature, there is little consensus about what underlies it. Among existing accounts, many have linked the P600 to reanalysis processes (e.g., Friederici, 1995; Münte, Matzke & Johannes, 1997; Osterhout, Holcomb & Swinney, 1994; van de Meerendonk et al., 2010) even though they do not agree on the domain-specificity of the reanalysis processes.

Here I propose that the P600 effect and the increased regressions in the current study should be attributed to common reanalysis processes triggered by the implausibility that resulted from argument role-reversals. This predicts that variables that are hypothesized to trigger reanalysis should elicit both a P600 effect and increased regressions. Meanwhile, I hypothesize that the P600, which extends over a few hundred milliseconds, is also sensitive to cognitive processes that are not involved in reanalysis. Therefore, I predict that P600 effects that are observed in cases where little or no reanalysis is required (e.g., Kaan et al., 2000 ; Phillips et al., 2005) may be associated with different eye-movement patterns in reading.

What does temporal alignment between EMs and ERPs tell us?

One assumption implicit in the above proposal to associate regressions triggered by argument role-reversals with the P600 effect is that these effects are aligned in time. This assumption is supported by the results of the regression latency analysis in Experiment 9. Further, the observation that readers' EMs as well as ERPs were not sensitive to argument role-reversals before 500ms is consistent with the hypothesis that information about arguments' structural roles does not have an immediate impact on comprehenders' verb prediction.

Nonetheless, even mis-alignments between EMs and ERPs can be very informative (Sereno & Rayner, 2003). It may suggest that these dependent measures reflect distinct cognitive processes, and that the processes reflected by one dependent measure precede, or even lead to, the processes reflected by the other. For example, if an independent variable consistently modulated readers' ERPs earlier in time than their EMs,

then we may infer that ERPs and EMs are sensitive to distinct underlying processes and formulate hypotheses about how the processes reflected by ERPs relate to the observed eye-movement pattern. Conversely, if readers' EMs showed consistently earlier sensitivity to an independent variable than their ERPs even when the stimulus presentation conditions were matched, then we may infer that ERPs fail to detect at least some of the cognitive processes that underlie readers' EMs. Therefore, misalignments that are observed even in fully matched and carefully controlled experimental environments can be extremely informative.

Regression latency analysis

In Experiments 8 and 9 I estimated the time course of the effect of argument role-reversal on readers' EMs by analyzing the latency of their regressions. This analysis was guided by the results of standard EM analyses, such that regression latency was estimated only when argument role-reversals had a significant effect on probability of regression. However, this approach might not fully capture the timing of effects on readers' EMs for three reasons.

First, the regression latency analysis, by definition, only examined trials in which a regression was made. However, as readers can freely move their eyes in natural reading, there are presumably multiple ways in which processing difficulty can affect their eye-movements (e.g., lengthened fixation durations, increased fixations, re-reading). Further, given the prominence of argument role-reversal anomalies, it is rather unlikely that readers experienced processing difficulty only in the 20-30% of the trials in which they made regressions upon reading the target verb. Although it remains unclear why

processing difficulty results in regressions only occasionally, it is plausible that the timing of regressions in a subset of trials is not fully representative of the timing of processing difficulty in all trials. In fact, previous studies have shown that fixations that end in regressions are on average shorter in duration than those that end in progression (e.g., Altmann, Garnham & Dennis, 1992; Mitchell et al., 2008; Rayner & Sereno, 1994).

Second, since this analysis was guided by the results of standard by-region EM analyses, only regressions out of specific regions were analyzed. Incidentally, since the effect of argument role-reversals on readers' regression was observed in the critical region in Experiment 8 and in the post-critical region in Experiment 9, this analysis was conducted on regressions out of different regions across experiments. However, inspection of the means suggested that argument role-reversals also affected readers' regressions in the post-critical region in the low-predictability sentences in Experiment 8 (canonical: 25% vs. role-reversed: 34%). The latency estimates would be quite different if these regressions were included in the analysis. Therefore, we need an alternative principled way for determining which regressions to include in regression latency analysis that will not introduce unnecessary biases into the results.

Further, since regression latency analysis was conducted only on data in regions where probability of regression differed significantly across conditions, the number of data points always differed across conditions. Specifically, since there were significantly more regressions in the role-reversed condition than in the canonical condition, the latency estimate was inevitably biased towards regressions in the role-reversed condition. Although the estimated latencies did not seem to differ systematically across conditions

in the current data sets, future research will need to examine the generalizability of this observation and to explore potential solutions to this problem.

Outlook: Better ways to align EMs and ERPs?

As discussed above, one of the limitations of estimating regression latency is its limited scope – although a regression is a discrete event and its latency can be measured rather straightforwardly, fixations durations are computed over pre-defined text regions and it is not possible to meaningfully measure when fixation durations in a given text region increase. Therefore, in order to estimate the time point at which readers' EMs in general (not just their regressions) begin to show sensitivity to an independent variable, we need a continuous dependent measure that is independent of region definitions.

One potential solution is cumulative progression analysis, which was developed by Scheepers and colleague (Scheepers, Konieczny, & van Gompel, unpublished manuscript) and has since been reported in a study by Kreiner, Sturt and Garrod (2008). In contrast with standard EM measures, cumulative progression is a continuous measure that is independent of region definitions. It provides information with regard to the reader's progression rate and thereby allows one to examine processing cost over time. *Cumulative progression* is calculated by taking the first-fixation at the target region as a starting point (on both axes) and plotting the readers' progression in space (number of characters) as a function of time (sampled at a fixed interval, e.g., every 10ms). The measure is cumulative in the sense that, at any given time point, only the current rightmost character position is considered, so that the measure is updated with a new character position if and only if the reader has proceeded further to the right. The measure

remains unchanged otherwise (e.g., during a fixation or regression). Cumulative progression data in a predefined time interval (say, 1000ms) following the initial fixation in the critical region are then averaged across subjects and items for each experimental condition. In an initial study, Kreiner et al. (2008) examined how experimental variables affected readers' progression rate by fitting a non-linear regression function to individual subjects' cumulative progression data and comparing the rate parameters across conditions. Alternatively, one can examine when readers' eye progression begins to differ between conditions by finding the point of divergence between the lines (c.f., Reingold, Reichle, Glaholt & Sheridan, 2012).

Further, since cumulative progression analysis does not require a special stimulus presentation paradigm, it can be applied to EM data collected during EM/EEG co-registration. With encouraging progress being made in the past few years on overcoming the technical challenges in EM/EEG co-registration (e.g., Dimigen et al., 2011), future research can use EM/EEG co-registration to (i) examine the relations between different ERP components (e.g., the P600) and standard EM measures (e.g., probability of regressions) and (ii) compare the time course of ERP effects and the results of cumulative progression analysis results.

6 Conclusion

6.1 Overview

This dissertation set out to explore the temporal dimension of predictive computations in language comprehension. The ability to anticipate language inputs ahead of time can help to process incoming language more efficiently and is likely key to the speed and robustness of successful language comprehension. As the extent to which prediction can facilitate processing critically depends on its timing relative to the predicted event, research on the temporal properties of predictive computations (the 'when' questions) is crucial to our understanding of how the brain computes predictions during real-time comprehension (the 'how' questions). Previous research on linguistic prediction has paid little attention to the 'when' questions and has commonly assumed that all contextual information impacts comprehenders' predictions as soon as it becomes available in the input stream. In this dissertation I questioned this key assumption and asked:

1. *How do predictive computations unfold in time during language comprehension?*
2. *Do all sources of contextual information impact prediction on the same time scale?*

6.2 Synthesis of empirical findings

In Chapters 2 to 5 I presented evidence from a series of event-related potential (ERP) and reading eye-movement (EM) experiments in Mandarin Chinese and English.

Specific findings were summarized within the respective chapters. Here I synthesize the main empirical findings to answer the research questions I set out to answer:

1. *How do predictive computations unfold in time during language comprehension?*
    - *Predictive computations are not instantaneous*: Comprehenders' ERPs (in particular, their N400 response to expected vs. unexpected verbs) as well as their reading eye-movements suggest that verb prediction is initially insensitive to information about preverbal arguments' structural roles; the impact of structural role information emerges only when the time interval for predictive computations is sufficiently wide.
    - *Even prominent and unambiguous contextual information may fail to immediately impact linguistic prediction*: Even though preverbal arguments' structural roles were prominently and unambiguously marked in the verb-final *Ba*-construction in Mandarin Chinese, they fail to impact comprehenders' verb prediction when the verb appears immediately following the arguments.

2. *Do all sources of contextual information impact prediction on the same time scale?*
    - *Different sources of contextual information may impact linguistic prediction on different time scales:* The contrast between N400's sensitivity to argument substitution and its insensitivity to argument role-reversal when the verb appears shortly after its arguments suggests that information about arguments' lexical identity can impact verb prediction more quickly than information about arguments' structural roles.

6.3 Theoretical implications

This dissertation presents empirical findings that undermine the widely held assumption that all contextual information can impact linguistic prediction as soon as it becomes available in the input stream and has four main theoretical implications.

Contextual information can drive language comprehension without informing prediction

Results from the ERP experiments (Experiments 1-6) show that argument role-reversals are readily detected and consistently elicit a P600 effect, but they do not modulate the N400 unless the time interval for predictive computations is sufficiently wide and argument role-reversals greatly impact a verb's offline predictability. This contrast shows that contextual information that impacts bottom-up comprehension processes may fail to modulate top-down predictive processes.

More broadly, these findings suggest that not all contextual information may constitute inputs to predictive computations. Although certain contextual information might fail to impact linguistic prediction due to a lack of time for predictive computations (as in the case of argument role-reversals), there might also be independent constraints that determine whether a piece of contextual information becomes an input to predictive computation. For example, comprehenders may use the number feature of a subject noun phrase to predict the form (more specifically, the number feature) of the verb in a language like English (e.g., Wagers, Lau & Phillips, 2009). However, a native listener might rely on this information for prediction more or less strongly in a given context depending on whether the speaker is perceived to be a native adult speaker, a child, or a second language speaker who makes frequent agreement errors in that language. A

similar argument has been made with regard to bottom-up syntactic processing – while a significant P600 effect was elicited by gender violations made by a native speaker, errors that were by an L2 speaker with a foreign accent did not elicit a P600 effect (Hanulikova, van Goch & van Alphen, 2010).

Offline predictability measures may systematically fail to capture online predictions

The current results from both ERP and EM experiments consistently show that offline measures of a word's predictability may systematically fail to capture what and when linguistic predictions are being computed during comprehension. This is because offline predictability measures such as those obtained in language corpora and offline cloze task do not take into consideration how contextual information impacts linguistic prediction over time. Therefore, in order to draw conclusions about real-time predictive computations based on offline predictability measures, it is important to take into considerations the relevant contextual information and the time interval during which it might impact linguistic prediction before the target bottom-up input arises.

Significance of the temporal dimension of linguistic prediction

The discrepancy between offline predictability measures and online prediction further highlights the significance of the temporal dimension in the study of predictive computations. Considerations about the temporal properties of predictive computations are crucial for differentiating genuine failures to engage predictive mechanisms from cases in which certain sources of contextual information cannot impact predictions quickly enough to facilitate bottom-up processing. These considerations will likely be of particular importance for studying the role of prediction in language comprehension in

populations such as children, second language learners, elderly individuals and people with language impairments.

More broadly, prediction in any complex system (including, but not limited to language) involves mental computations that must race against rapid bottom-up input. Therefore, considerations about the time at which relevant contextual information becomes available and the time at which prediction starts to facilitate bottom-up processes are likely to have critical consequences for understanding the mental computations that underlie prediction across different domains.

A framework for studying the ingredients and steps in predictive computations

In this dissertation I put forth a framework that aims to (i) identify the relevant inputs to predictive computations and (ii) examine how each of them might impact linguistic predictions along the temporal dimension. Further, I proposed that the N400, an ERP response that is taken to reflect access to long-term semantic memory, provides snapshots of comprehenders' real-time lexical semantic predictions. Using the N400 response to a word as a diagnostic for linguistic predictions, I introduced an experimental paradigm that can be adapted to study the time course of linguistic prediction more broadly.

6.4 Recommendations for future research

Representations

In Chapter 1 I explained that a model of linguistic prediction must address questions about mental representations – in order to characterize how linguistic

predictions are computed we must also characterize the nature of representations of the inputs and outputs of predictive computations. The proposal outlined in this dissertation highlights the importance of considering subcomponents of contextual information as inputs to prediction. While I have remained largely agnostic about how different inputs might be mentally represented, this approach can be applied to study questions about the representation of different inputs to prediction more systematically. For example, future work can examine whether and how predictive computations are differentially impacted by different representations of the contextual inputs. Meanwhile, although much research on linguistic prediction has studied the outputs of prediction, little is known about the nature of output representations. For example, although the N400 has been taken to reflect lexical semantic memory access and can potentially reveal how prediction facilitates access to restored lexical semantic representations, there remain many fundamental questions about the nature of lexical semantic representations.

Further, future work will need to take into account the complex mappings between empirical measurements and mental representations. For example, individual words are convenient unit for measure predictability, but they might not be stored as individual units in long-term semantic memory. Research on sentence comprehension often uses individual words as units of prediction and has abstracted away from any potential relationships among different words (Roland, Yun, Koenig & Mauner, 2012). The predictability of a word in a sentence is commonly operationalized as the probability of that word given its preceding words. The probability is distributed over a hypothesis space that contains all possible lexical items, which is analogous to a lexicon. However, most current models have assumed a hypothesis space that is not structured – each lexical

item is a possible outcome and they bear no relationship with one another. For example, if two words (e.g., 'cookbook' and 'store') have equally low probability in a given context, then they would be assumed to be equally unpredictable, regardless of whether one is more related to the word with the highest probability (e.g., 'recipe') than the other. Crucially, such characterization of a word's predictability would fail to capture the observation that, despite matching cloze probabilities, words that are related to the most expected word are processed more quickly and elicit a smaller N400 response than words that are not (e.g., Federmeier & Kutas, 1999; Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985; see also Roland et al., 2012). Therefore, in order to better characterize the outputs of predictive computations future research should explore how stimulus predictability may interact with the structure of the hypothesis space (including, but not limited to the lexicon).

Qualitative linking hypothesis

Existing psycholinguistic research has devoted much attention to studying the linking hypotheses between empirical measures (e.g., reading times and ERP components) and underlying cognitive processes. The progress made in understanding the functional significance of the N400 since its discovery (Kutas & Hillyard, 1980) has formed the foundation for this dissertation and has greatly informed the study of linguistic prediction. Future research will be needed to better understand the linking hypotheses for other measures that may also be associated with prediction-related processes. For instance, future research will need to formulate testable hypotheses about the functional significance of the late frontal positivity and examine how it might be related to different aspects of linguistic prediction (e.g., inhibition of disconfirmed predictions, updating

current predictions). Meanwhile, the comparisons between ERP and EM results in Chapter 5 suggest a tentative link between the P600 and regressive eye-movements in reading comprehension. Future research will be needed to examine the relationships between these measures more rigorously (e.g., via EM/EEG co-registration) and explore the extent to which they reflect common underlying processes.

Quantitative linking hypotheses

Recent work has begun to establish a quantitative relationship between word predictability and reading times. By combining large behavioral data-sets with computational language modeling and non-parametric statistical techniques, Smith and Levy (2013) found that the relationship between word predictability and reading time is logarithmic across six orders of magnitude in estimated predictability. This means that the reading time difference between a word with true probability 1 and a word with true probability 0.1 is the same as the reading time difference between a word with true probability 0.01 and a word with true probability 0.001. Since comprehenders are most sensitive to differences in predictability at the low-end of the predictability scale, exactly where data are sparse, the authors suggested that '*in practice it is very difficult to assert with confidence from cloze norms that two different sets of word/context pairs are truly "equally" unpredictable in the sense that matters for real-time comprehension behavior'* (Smith & Levy, 2013, p.22).

Meanwhile, previous research has established that the amplitude of the N400 is also inversely related to a word's predictability and that the relation holds for a wide range of cloze probabilities. However, the quantitative nature of this relationship has not

been established and a linear relationship is often implicitly or explicitly assumed (e.g., deLong et al., 2005). Although the results from a recent study that varied cloze probability parametrically did not suggest a logarithmic relationship (Wlotko & Federmeier, 2012; Figure 1, p.360), the exact quantitative relationship between the N400 and predictability may have profound implications for many existing findings. For example, if the relationship between the N400 and predictability were in fact logarithmic, then differences in the N400 between conditions that were 'matched' in terms of their unpredictability (e.g., the related vs. unrelated unexpected target words in Federmeier & Kutas, 1999) would potentially be attributable to small differences in predictability that was not detected due to sparse data. This is because cloze data collected from 30-40 participants are not sensitive to small differences in predictability, e.g., a true probability of 10^{-3} and 10^{-6} would both be "0% cloze". Conversely, even 'small' difference in cloze probability in unexpected items (e.g., 3% vs. "0%" cloze) would be expected to elicit a large N400 effect. Therefore, even though existing evidence does not suggest that small cloze difference will lead to large N400 effects (e.g., Federmeier et al., 2007; Wlotko & Federmeier, 2012; Experiments 5 in Chapter 3), future research will need to establish the quantitative relationship between the N400 and word predictability more precisely.

Prediction cost upon encountering disconfirming evidence

In Chapter 1 I suggested that the ability to anticipate upcoming language inputs can help comprehenders process language more efficiently. However, given the generative nature of language – an infinite number of expressions can be generated from a limited set of elements – predictions are bound to be disconfirmed by the inputs at least some of the time. Therefore, in the search for evidence for predictive mechanisms in

language comprehension, researchers have often looked for evidence for (i) processing benefits due to correct predictions and (ii) processing costs due to incorrect predictions (for a review see Van Petten & Luka, 2012).

While the N400's robust sensitivity to word predictability suggests that it is modulated by the benefit of (partially) correct predictions (e.g., Federmeier et al., 2007, 2010; Kutas & Hillyard, 1984; Van Petten & Luka, 2012), evidence for processing costs due to prediction errors has been more elusive. One candidate is the late frontal positivity that has been observed in response to unexpected words that violate strong predictions (e.g., Federmeier et al., 2007, 2010; Van Petten & Luka, 2012). Some have proposed that this ERP component reflects revision processes that are triggered by unexpected inputs when comprehenders engage predictive mechanisms during comprehension (e.g., Federmeier et al., 2010; Wlotko, Federmeier & Kutas, 2012; Wlotko, Lee & Federmeier, 2010). Although the results reported in this dissertation provide further evidence that unexpected words that violate strong predictions elicit a late frontal positivity across studies in English and Mandarin Chinese, future research will be needed to examine what processes this component might reflect – Does it reflect processes that inhibit incorrect predictions? Does it reflect processes that revise and/or update current prediction? Does it predict comprehenders' successes in processing further inputs after initial predictions have been disconfirmed?

(More) direct evidence for predictive computations

In addition to processing costs that are associated with prediction error, the computation of predictions itself may also incur processing costs. However, direct

evidence for predictive computations (i.e., direct evidence showing when and where predictions are being computed in the brain) has remained elusive. As I discussed in Chapter 1, existing research has relied almost exclusively on evidence regarding the processing consequences of prediction to draw inferences about the nature of predictive computations. For example, many have drawn inferences about the output representations of linguistic prediction based on sensitivity to different properties of the bottom-up input (e.g., DeLong et al., 2005; Dikker et al., 2009; Dikker, Rabagliati, Farmer & Pylkkänen, 2010; Fedemeier & Kutas, 1999; Van Berkum et al., 2005; Wicha et al., 2004). Meanwhile, the research reported in this dissertation draws inferences about the time course of predictive computations by examining when facilitative effects of prediction (N400 effects) emerge. This approach has been extremely useful and has led to many important discoveries about the nature of linguistic prediction. However, since this approach is bounded by measurable outcomes of predictive computations, it remains extremely difficult, even with techniques that provide high temporal resolution, to examine the intermediate computations involved (e.g., What are the intermediate steps for mapping arguments' lexical identity to verb prediction?) and how they unfold on a smaller time scale (e.g., When is structural role information computed and how much time does it take to compute verb prediction using such information?). Therefore, while this approach will continue to allow new inferences about the properties of linguistic prediction based on its processing consequences, future research will be needed to develop other approaches that can provide more direct evidence for the computation of linguistic prediction itself.

6.5 General conclusion

In this dissertation I highlight the importance of the temporal dimension of linguistic prediction and have presented the first evidence regarding how linguistic prediction may develop in real time. Contrary to the widely held assumption that contextual information can impact linguistic prediction as soon as it arises in the input, the current results show that even prominent and unambiguous contextual information may fail to impact linguistic prediction immediately. Further, I have provided initial evidence that different sources of contextual information may impact predictive computations on a different time scale. More generally, I present a research framework for studying the ingredients and steps involved in predictive computations.

Appendix (Experimental materials)

Experiment 1

a - Animacy-congruous, Combinable (Control)
b - Animacy-violated, Combinable (Role-reversed)
c - Animacy-congruous, Non-combinable
d - Animacy-violated, Non-combinable

The animacy-congruous conditions shared the same set of verbs, all of which freely allow inanimate direct objects. The animacy-violated conditions shared a different set of verbs. The verbs in these two conditions are accurately described as 'animacy incongruous', but post-test searches suggested that the incongruity may be graded for some verbs. 80% of verbs straightforwardly disallowed inanimate objects. Of the remainder, 10% had a strong bias against inanimate objects, and the other 10% had a weaker bias against inanimate objects.

1a. 在 这个作家 把 这本小说 读完了 之后， 太阳 就 开始 下山了。
1b. 在 这个作家 把 这本小说 感动了 之后， 太阳 就 开始 下山了。
1c. 在 这个作家 把 这本小说 打破了 之后， 太阳 就 开始 下山了。
1d. 在 这个作家 把 这本小说 剪伤了 之后， 太阳 就 开始 下山了。
2a. 在 这只小白兔 把 那块红萝卜 吃光了 之后， 有 一阵风 吹过。
2b. 在 这只小白兔 把 那块红萝卜 喂饱了 之后， 有 一阵风 吹过。
2c. 在 这只小白兔 把 那块红萝卜 修好了 之后， 有 一阵风 吹过。
2d. 在 这只小白兔 把 那块红萝卜 迷住了 之后， 有 一阵风 吹过。
3a. 在 这个画家 把 这幅自画像 挂起了 之后， 客人 就 到访了。
3b. 在 这个画家 把 这幅自画像 迷倒了 之后， 客人 就 到访了。
3c. 在 这个画家 把 这幅自画像 喝光了 之后， 客人 就 到访了。
3d. 在 这个画家 把 这幅自画像 治愈了 之后， 客人 就 到访了。
4a. 在 这只猫儿 把 这个鱼缸 打翻了 之后， 主人 就 回来了。
4b. 在 这只猫儿 把 这个鱼缸 迷住了 之后， 主人 就 回来了。
4c. 在 这只猫儿 把 这个鱼缸 服错了 之后， 主人 就 回来了。
4d. 在 这只猫儿 把 这个鱼缸 气死了 之后， 主人 就 回来了。
5a. 在 那位病人 把 这种新药物 服错了 之后， 医生 就很 担心了。
5b. 在 那位病人 把 这种新药物 治愈了 之后， 医生 就很 担心了。
5c. 在 那位病人 把 这种新药物 解答了 之后， 医生 就很 担心了。
5d. 在 那位病人 把 这种新药物 捧红了 之后， 医生 就很 担心了。
6a. 在 那位模特儿 把 那一头白发 染黑了 之后， 陈老板 就 很高兴了。
6b. 在 那位模特儿 把 那一头白发 气死了 之后， 陈老板 就 很高兴了。
6c. 在 那位模特儿 把 那一头白发 推翻了 之后， 陈老板 就 很高兴了。
6d. 在 那位模特儿 把 那一头白发 绊倒了 之后， 陈老板 就 很高兴了。
7a. 在 那个小孩 把 这个玩具 丢失了 之后， 妈妈 就 生气了。
7b. 在 那个小孩 把 这个玩具 逗乐了 之后， 妈妈 就 生气了。
7c. 在 那个小孩 把 这个玩具 化解了 之后， 妈妈 就 生气了。
7d. 在 那个小孩 把 这个玩具 难倒了 之后， 妈妈 就 生气了。
8a. 在 这位科学家 把 新的理论 推翻了 之后， 大家 都 争相 访问他。
8b. 在 这位科学家 把 新的理论 捧红了 之后， 大家 都 争相 访问他。

8c. 在 这位科学家 把 新的理论 染黑了 之后， 大家 都 争相 访问他。
8d. 在 这位科学家 把 新的理论 弄伤了 之后， 大家 都 争相 访问他。
9a. 在 这位专家 把 这个危机 化解了 之后， 大家 都 很佩服他。
9b. 在 这位专家 把 这个危机 难到了 之后， 大家 都 很佩服他。
9c. 在 这位专家 把 这个危机 擦乾了 之后， 大家 都 很佩服他。
9d. 在 这位专家 把 这个危机 满足了 之后， 大家 都 很佩服他。
10a. 在 高材生 把 数学题 解答了 之后， 老师 感到 很高兴。
10b. 在 高材生 把 数学题 难倒了 之后， 老师 感到 很高兴。
10c. 在 高材生 把 数学题 挂起了 之后， 老师 感到 很高兴。
10d. 在 高材生 把 数学题 困住了 之后， 老师 感到 很高兴。
11a. 在 那个技工 把 这台机器 修好了 之后， 老板 十分 满意。
11b. 在 那个技工 把 这台机器 弄伤了 之后， 老板 十分 满意。
11c. 在 那个技工 把 这台机器 扑灭了 之后， 老板 十分 满意。
11d. 在 那个技工 把 这台机器 害苦了 之后， 老板 十分 满意。
12a. 在 收藏家 把 这帧照片 挂起了 之后， 地震 就 发生了。
12b. 在 收藏家 把 这帧照片 满足了 之后， 地震 就 发生了。
12c. 在 收藏家 把 这帧照片 完成了 之后， 地震 就 发生了。
12d. 在 收藏家 把 这帧照片 压伤了 之后， 地震 就 发生了。
13a. 在 那位市民 把 家居火警 扑灭了 之后， 就 开始 下大雨了。
13b. 在 那位市民 把 家居火警 困住了 之后， 就 开始 下大雨了。
13c. 在 那位市民 把 家居火警 夺去了 之后， 就 开始 下大雨了。
13d. 在 那位市民 把 家居火警 蒙蔽了 之后， 就 开始 下大雨了。
14a. 在 特务 把 这项任务 完成了 之后， 战争 就 结束了。
14b. 在 特务 把 这项任务 害苦了 之后， 战争 就 结束了。
14c. 在 特务 把 这项任务 举起了 之后， 战争 就 结束了。
14d. 在 特务 把 这项任务 关上了 之后， 战争 就 结束了。
15a. 在 那个邮差 把 包裹 送出了 之后， 车祸 就 发生了。
15b. 在 那个邮差 把 包裹 压伤了 之后， 车祸 就 发生了。
15c. 在 那个邮差 把 包裹 做完了 之后， 车祸 就 发生了。
15d. 在 那个邮差 把 包裹 气死了 之后， 车祸 就 发生了。
16a. 在 海盗 把 宝藏 夺去了 之后， 那艘船 就 翻了。
16b. 在 海盗 把 宝藏 蒙蔽了 之后， 那艘船 就 翻了。
16c. 在 海盗 把 宝藏 否决了 之后， 那艘船 就 翻了。
16d. 在 海盗 把 宝藏 害惨了 之后， 那艘船 就 翻了。
17a. 在 那个囚犯 把 那个手扣 破坏了 之后， 陈警长 就 发怒了。
17b. 在 那个囚犯 把 那个手扣 关上了 之后， 陈警长 就 发怒了。
17c. 在 那个囚犯 把 那个手扣 弄混了 之后， 陈警长 就 发怒了。
17d. 在 那个囚犯 把 那个手扣 压伤了 之后， 陈警长 就 发怒了。
18a. 在 那个议员 把 这份提案 否决了 之后， 总统 就 来到了。
18b. 在 那个议员 把 这份提案 气死了 之后， 总统 就 来到了。
18c. 在 那个议员 把 这份提案 运走了 之后， 总统 就 来到了。
18d. 在 那个议员 把 这份提案 毒死了 之后， 总统 就 来到了。
19a. 在 这只大象 把 池塘里的水 喝光了 之后， 其他动物 就 离开了。
19b. 在 这只大象 把 池塘里的水 滋润了 之后， 其他动物 就 离开了。
19c. 在 这只大象 把 池塘里的水 丢失了 之后， 其他动物 就 离开了。
19d. 在 这只大象 把 池塘里的水 逗乐了 之后， 其他动物 就 离开了。
20a. 在 商人 把 商品 售出了 之后， 田老板 就 高兴了。
20b. 在 商人 把 商品 害惨了 之后， 田老板 就 高兴了。
20c. 在 商人 把 商品 传开了 之后， 田老板 就 高兴了。

20d. 在 商人 把 商品 迷倒了 之后， 田老板 就 高兴了。
21a. 在 那搬运工人 把 这些货物 运走了 之后， 天色 就 变坏了。
21b. 在 那搬运工人 把 这些货物 压伤了 之后， 天色 就 变坏了。
21c. 在 那搬运工人 把 这些货物 拍下了 之后， 天色 就 变坏了。
21d. 在 那搬运工人 把 这些货物 平定了 之后， 天色 就 变坏了。
22a. 在 这个记者 把 这突发消息 传开了 之后， 市民 就 很担心了。
22b. 在 这个记者 把 这突发消息 吓坏了 之后， 市民 就 很担心了。
22c. 在 这个记者 把 这突发消息 烧着了 之后， 市民 就 很担心了。
22d. 在 这个记者 把 这突发消息 刑伤了 之后， 市民 就 很担心了。
23a. 在 这个摄影师 把 这里的风景 拍下了 之后， 天色 就 开始 变坏了。
23b. 在 这个摄影师 把 这里的风景 迷倒了 之后， 天色 就 开始 变坏了。
23c. 在 这个摄影师 把 这里的风景 购下了 之后， 天色 就 开始 变坏了。
23d. 在 这个摄影师 把 这里的风景 蒙蔽了 之后， 天色 就 开始 变坏了。
24a. 在 那群暴徒 把 这辆坦克车 烧着了 之后， 军队 就 来到了。
24b. 在 那群暴徒 把 这辆坦克车 平定了 之后， 军队 就 来到了。
24c. 在 那群暴徒 把 这辆坦克车 输光了 之后， 军队 就 来到了。
24d. 在 那群暴徒 把 这辆坦克车 缠住了 之后， 军队 就 来到了。
25a. 在 这个收藏家 把 这件古董 购下了 之后， 恶霸 就 现身了。
25b. 在 这个收藏家 把 这件古董 逗乐了 之后， 恶霸 就 现身了。
25c. 在 这个收藏家 把 这件古董 背熟了 之后， 恶霸 就 现身了。
25d. 在 这个收藏家 把 这件古董 累坏了 之后， 恶霸 就 现身了。
26a. 在 这群赌徒 把 那笔钱 输光了 之后， 高利贷集团 就 乘虚而入了。
26b. 在 这群赌徒 把 那笔钱 蒙蔽了 之后， 高利贷集团 就 乘虚而入了。
26c. 在 这群赌徒 把 那笔钱 破坏了 之后， 高利贷集团 就 乘虚而入了。
26d. 在 这群赌徒 把 那笔钱 送走了 之后， 高利贷集团 就 乘虚而入了。
27a. 在 这些参赛者 把 游戏规则 背熟了 之后， 比赛 就 要开始了。
27b. 在 这些参赛者 把 游戏规则 吓怕了 之后， 比赛 就 要开始了。
27c. 在 这些参赛者 把 游戏规则 售出了 之后， 比赛 就 要开始了。
27d. 在 这些参赛者 把 游戏规则 绊倒了 之后， 比赛 就 要开始了。
28a. 在 那家庭主妇 把 这些家务 做完了 之后， 弟弟 就 打翻了 颜料。
28b. 在 那家庭主妇 把 这些家务 累坏了 之后， 弟弟 就 打翻了 颜料。
28c. 在 那家庭主妇 把 这些家务 藏起了 之后， 弟弟 就 打翻了 颜料。
28d. 在 那家庭主妇 把 这些家务 撞伤了 之后， 弟弟 就 打翻了 颜料。
29a. 在 那位太空人 把 这支火箭 检查了 之后， 工程师 就 放心了。
29b. 在 那位太空人 把 这支火箭 送走了 之后， 工程师 就 放心了。
29c. 在 那位太空人 把 这支火箭 跨过了 之后， 工程师 就 放心了。
29d. 在 那位太空人 把 这支火箭 打动了 之后， 工程师 就 放心了。
30a. 在 那个青年 把 这批毒品 藏起了 之后， 警察 就 来到了。
30b. 在 那个青年 把 这批毒品 害惨了 之后， 警察 就 来到了。
30c. 在 那个青年 把 这批毒品 听熟了 之后， 警察 就 来到了。
30d. 在 那个青年 把 这批毒品 滑倒了 之后， 警察 就 来到了。
31a. 在 那个推销员 把 这辆房車 售出了 之后， 天色 就 开始 昏暗了。
31b. 在 那个推销员 把 这辆房車 撞伤了 之后， 天色 就 开始 昏暗了。
31c. 在 那个推销员 把 这辆房車 挂起了 之后， 天色 就 开始 昏暗了。
31d. 在 那个推销员 把 这辆房車 害惨了 之后， 天色 就 开始 昏暗了。
32a. 在 这个运动员 把 这啦啦队歌 听熟了 之后， 比赛 就 结束了。
32b. 在 这个运动员 把 这啦啦队歌 打动了 之后， 比赛 就 结束了。
32c. 在 这个运动员 把 这啦啦队歌 售出了 之后， 比赛 就 结束了。
32d. 在 这个运动员 把 这啦啦队歌 推倒了 之后， 比赛 就 结束了。

33a. 在 这清洁女工 把 湿滑的地面 擦乾了 之后， 天上 就 开始 下雨了。
33b. 在 这清洁女工 把 湿滑的地面 滑倒了 之后， 天上 就 开始 下雨了。
33c. 在 这清洁女工 把 湿滑的地面 送出了 之后， 天上 就 开始 下雨了。
33d. 在 这清洁女工 把 湿滑的地面 逗乐了 之后， 天上 就 开始 下雨了。
34a. 在 这位投资者 把 那些股票 售出了 之后， 股市 就 开始 上扬了。
34b. 在 这位投资者 把 那些股票 害惨了 之后， 股市 就 开始 上扬了。
34c. 在 这位投资者 把 那些股票 踢开了 之后， 股市 就 开始 上扬了。
34d. 在 这位投资者 把 那些股票 卷走了 之后， 股市 就 开始 上扬了。
35a. 在 那个官员 把 选举结果 篡改了 之后， 总统 就 被暗杀了。
35b. 在 那个官员 把 选举结果 推倒了 之后， 总统 就 被暗杀了。
35c. 在 那个官员 把 选举结果 抱紧了 之后， 总统 就 被暗杀了。
35d. 在 那个官员 把 选举结果 压倒了 之后， 总统 就 被暗杀了。
36a. 在 那个小孩 把 那小石块 踢开了 之后， 小狗 就 跑过来了。
36b. 在 那个小孩 把 那小石块 绊倒了 之后， 小狗 就 跑过来了。
36c. 在 那个小孩 把 那小石块 忘却了 之后， 小狗 就 跑过来了。
36d. 在 那个小孩 把 那小石块 闷坏了 之后， 小狗 就 跑过来了。
37a. 在 滑浪者 把 大浪 克服了 之后， 众人 就 大叫起来。
37b. 在 滑浪者 把 大浪 卷走了 之后， 众人 就 大叫起来。
37c. 在 滑浪者 把 大浪 搬走了 之后， 众人 就 大叫起来。
37d. 在 滑浪者 把 大浪 感动了 之后， 众人 就 大叫起来。
38a. 在 举重选手 把 哑铃 举起了 之后， 大家 都 很震惊。
38b. 在 举重选手 把 哑铃 压倒了 之后， 大家 都 很震惊。
38c. 在 举重选手 把 哑铃 讲解了 之后， 大家 都 很震惊。
38d. 在 举重选手 把 哑铃 保护了 之后， 大家 都 很震惊。
39a. 在 影评人 把 那文艺片 批评了 之后， 大家 都 很失望。
39b. 在 影评人 把 那文艺片 闷坏了 之后， 大家 都 很失望。
39c. 在 影评人 把 那文艺片 穿好了 之后， 大家 都 很失望。
39d. 在 影评人 把 那文艺片 指证了 之后， 大家 都 很失望。
40a. 在 那个牧师 把 这圣经故事 讲解了 之后， 音乐 就 响起来了。
40b. 在 那个牧师 把 这圣经故事 感动了 之后， 音乐 就 响起来了。
40c. 在 那个牧师 把 这圣经故事 销毁了 之后， 音乐 就 响起来了。
40d. 在 那个牧师 把 这圣经故事 炸死了 之后， 音乐 就 响起来了。
41a. 在 那冲锋队员 把 避弹衣 穿好了 之后， 敌人 就 来袭了。
41b. 在 那冲锋队员 把 避弹衣 保护了 之后， 敌人 就 来袭了。
41c. 在 那冲锋队员 把 避弹衣 引爆了 之后， 敌人 就 来袭了。
41d. 在 那冲锋队员 把 避弹衣 刺伤了 之后， 敌人 就 来袭了。
42a. 在 这个罪犯 把 那些证据 销毁了 之后， 证人 就 消失了。
42b. 在 这个罪犯 把 那些证据 指证了 之后， 证人 就 消失了。
42c. 在 这个罪犯 把 那些证据 打断了 之后， 证人 就 消失了。
42d. 在 这个罪犯 把 那些证据 闷坏了 之后， 证人 就 消失了。
43a. 在 这位运动员 把 世界纪录 打破了 之后， 欢众 在 不断地 喝彩了。
43b. 在 这位运动员 把 世界纪录 吓坏了 之后， 欢众 在 不断地 喝彩了。
43c. 在 这位运动员 把 世界纪录 检查了 之后， 欢众 在 不断地 喝彩了。
43d. 在 这位运动员 把 世界纪录 滋润了 之后， 欢众 在 不断地 喝彩了。
44a. 在 恐怖分子 把 炸弹 引爆了 之后， 警方 就 展开 调查。
44b. 在 恐怖分子 把 炸弹 炸死了 之后， 警方 就 展开 调查。
44c. 在 恐怖分子 把 炸弹 剪辑了 之后， 警方 就 展开 调查。
44d. 在 恐怖分子 把 炸弹 浮起了 之后， 警方 就 展开 调查。
45a. 在 那个武侠 把 那把宝剑 打断了 之后， 众人 都 吃了一惊。

45b. 在 那个武侠 把 那把宝剑 刺伤了 之后， 众人 都 吃了一惊。
45c. 在 那个武侠 把 那把宝剑 克服了 之后， 众人 都 吃了一惊。
45d. 在 那个武侠 把 那把宝剑 迷倒了 之后， 众人 都 吃了一惊。
46a. 在 陈导演 把 那出电影 剪辑了 之后， 天 就 快亮了。
46b. 在 陈导演 把 那出电影 闷坏了 之后， 天 就 快亮了。
46c. 在 陈导演 把 那出电影 收好了 之后， 天 就 快亮了。
46d. 在 陈导演 把 那出电影 保护了 之后， 天 就 快亮了。
47a. 在 那个习泳者 把 浮板 抱紧了 之后， 海浪 就 开始 大了。
47b. 在 那个习泳者 把 浮板 浮起了 之后， 海浪 就 开始 大了。
47c. 在 那个习泳者 把 浮板 繫上了 之后， 海浪 就 开始 大了。
47d. 在 那个习泳者 把 浮板 揭发了 之后， 海浪 就 开始 大了。
48a. 在 裁缝 把 剪刀 收好了 之后， 客人 就 离开了。
48b. 在 裁缝 把 剪刀 剪伤了 之后， 客人 就 离开了。
48c. 在 裁缝 把 剪刀 批评了 之后， 客人 就 离开了。
48d. 在 裁缝 把 剪刀 难倒了 之后， 客人 就 离开了。
49a. 在 赛车手 把 安全带 繫上了 之后， 车祸 就 发生了。
49b. 在 赛车手 把 安全带 保护了 之后， 车祸 就 发生了。
49c. 在 赛车手 把 安全带 吃掉了 之后， 车祸 就 发生了。
49d. 在 赛车手 把 安全带 砸伤了 之后， 车祸 就 发生了。
50a. 在 这个奸商 把 那份文件 销毁了 之后， 形势 就 扭转了。
50b. 在 这个奸商 把 那份文件 揭发了 之后， 形势 就 扭转了。
50c. 在 这个奸商 把 那份文件 吃光了 之后， 形势 就 扭转了。
50d. 在 这个奸商 把 那份文件 害惨了 之后， 形势 就 扭转了。
51a. 在 滑冰好手 把 那双滑冰鞋 穿上了 之后， 比赛 就 暂停了。
51b. 在 滑冰好手 把 那双滑冰鞋 刓伤了 之后， 比赛 就 暂停了。
51c. 在 滑冰好手 把 那双滑冰鞋 打翻了 之后， 比赛 就 暂停了。
51d. 在 滑冰好手 把 那双滑冰鞋 吓怕了 之后， 比赛 就 暂停了。
52a. 在 那些工人 把 这块大石 搬走了 之后， 天上 就 开始 下雨了。
52b. 在 那些工人 把 这块大石 砸伤了 之后， 天上 就 开始 下雨了。
52c. 在 那些工人 把 这块大石 打开了 之后， 天上 就 开始 下雨了。
52d. 在 那些工人 把 这块大石 改变了 之后， 天上 就 开始 下雨了。
53a. 在 障碍赛选手 把 这个栏 跨过了 之后， 评判 就 站起来了。
53b. 在 障碍赛选手 把 这个栏 绊倒了 之后， 评判 就 站起来了。
53c. 在 障碍赛选手 把 这个栏 篡改了 之后， 评判 就 站起来了。
53d. 在 障碍赛选手 把 这个栏 烧伤了 之后， 评判 就 站起来了。
54a. 在 飞行表演员 把 降落伞 打开了 之后， 观众 就 欢呼起来了。
54b. 在 飞行表演员 把 降落伞 缠住了 之后， 观众 就 欢呼起来了。
54c. 在 飞行表演员 把 降落伞 扑灭了 之后， 观众 就 欢呼起来了。
54d. 在 飞行表演员 把 降落伞 弄脏了 之后， 观众 就 欢呼起来了。
55a. 在 那退伍军人 把 战争的经历 忘却了 之后， 老太太 就 开始 生病了。
55b. 在 那退伍军人 把 战争的经历 改变了 之后， 老太太 就 开始 生病了。
55c. 在 那退伍军人 把 战争的经历 售出了 之后， 老太太 就 开始 生病了。
55d. 在 那退伍军人 把 战争的经历 弄湿了 之后， 老太太 就 开始 生病了。
56a. 在 那个消防员 把 这场火 扑灭了 之后， 大家 都 在欢呼。
56b. 在 那个消防员 把 这场火 烧伤了 之后， 大家 都 在欢呼。
56c. 在 那个消防员 把 这场火 擦乾了 之后， 大家 都 在欢呼。
56d. 在 那个消防员 把 这场火 出卖了 之后， 大家 都 在欢呼。
57a. 在 那个小孩 把 这些颜料 弄混了 之后， 老师 就 生气了。
57b. 在 那个小孩 把 这些颜料 弄脏了 之后， 老师 就 生气了。

57c. 在 那个小孩 把 这些颜料 穿上了 之后， 老师 就 生气了。
57d. 在 那个小孩 把 这些颜料 吓坏了 之后， 老师 就 生气了。
58a. 在 王小姐 把 雨水 擦乾了 之后， 太阳 就 出来了。
58b. 在 王小姐 把 雨水 弄湿了 之后， 太阳 就 出来了。
58c. 在 王小姐 把 雨水 销毁了 之后， 太阳 就 出来了。
58d. 在 王小姐 把 雨水 感动了 之后， 太阳 就 出来了。
59a. 在 这个卧底 把 那个窃听器 装好了 之后， 行动 就 取消了。
59b. 在 这个卧底 把 那个窃听器 出卖了 之后， 行动 就 取消了。
59c. 在 这个卧底 把 那个窃听器 读完了 之后， 行动 就 取消了。
59d. 在 这个卧底 把 那个窃听器 喂饱了 之后， 行动 就 取消了。
60a. 在 这头小狗 把 那块肉 吃掉了 之后， 主人 就 回家了。
60b. 在 这头小狗 把 那块肉 毒死了 之后， 主人 就 回家了。
60c. 在 这头小狗 把 那块肉 装好了 之后， 主人 就 回家了。
60d. 在 这头小狗 把 那块肉 吓坏了 之后， 主人 就 回家了。

Experiment 2

a - Canonical control
b - Role-reversed (Animacy-congruous)

1a. 在 大象 把 蚂蚁 踩死了 之后， 小明 伤心得 哭了。
1b. 在 蚂蚁 把 大象 踩死了 之后， 小明 伤心得 哭了。
2a. 在 出租司机 把 路人 撞伤了 之后， 小强 就 立即 报警。
2b. 在 路人 把 出租司机 撞伤了 之后， 小强 就 立即 报警。
3a. 在 那个屠夫 把 这只鸡 宰了 之后， 肉贩 就 来到了。
3b. 在 这只鸡 把 那个屠夫 宰了 之后， 肉贩 就 来到了。
4a. 在 邓老师 把 这个学生 惩罚了 之后， 校长 就 显得 有点担心了。
4b. 在 这个学生 把 邓老师 惩罚了 之后， 校长 就 显得 有点担心了。
5a. 在 陈探长 把 这个疑犯 拘捕了 之后， 天色 就 变坏了。
5b. 在 这个疑犯 把 陈探长 拘捕了 之后， 天色 就 变坏了。
6a. 在 这头鲨鱼 把 那个泳客 咬伤了 之后， 警报 就 响起来了。
6b. 在 那个泳客 把 这头鲨鱼 咬伤了 之后， 警报 就 响起来了。
7a. 在 老伯伯 把 金鱼 喂饱了 之后， 天色 就 变坏了。
7b. 在 金鱼 把 老伯伯 喂饱了 之后， 天色 就 变坏了。
8a. 在 兽医 把 那头小狗 治好了 之后， 小玲 高兴得 睡不著觉。
8b. 在 那头小狗 把 兽医 治好了 之后， 小玲 高兴得 睡不著觉。
9a. 在 驯兽师 把 那头狮子 驯服了 之后， 大家 都 很佩服他。
9b. 在 那头狮子 把 驯兽师 驯服了 之后， 大家 都 很佩服他。
10a. 在 这个猎人 把 那只飞鸟 击毙了 之后， 天上 就 开始 下雨了。
10b. 在 那只飞鸟 把 这个猎人 击毙了 之后， 天上 就 开始 下雨了。
11a. 在 这只花猫 把 那只麻雀 吃掉了 之后， 孩子们 都 很害怕。
11b. 在 那只麻雀 把 这只花猫 吃掉了 之后， 孩子们 都 很害怕。
12a. 在 吴管家 把 那只流浪狗 赶走了 之后， 王老板 很满意。
12b. 在 那只流浪狗 把 吴管家 赶走了 之后， 王老板 很满意。
13a. 在 何婆婆 把 这个孤儿 养大了 之后， 大家 都 很佩服她。
13b. 在 这个孤儿 把 何婆婆 养大了 之后， 大家 都 很佩服她。
14a. 在 那个保姆 把 这个婴儿 抱住了 之后， 天 就 快亮了。
14b. 在 这个婴儿 把 那个保姆 抱住了 之后， 天 就 快亮了。

15a. 在 熊猫妈妈 把 小熊猫 压伤了 之后， 饲养员 就 有点担心了。
15b. 在 小熊猫 把 熊猫妈妈 压伤了 之后， 饲养员 就 有点担心了。
16a. 在 那头秃鹰 把 那只小鸡 叼走了 之后， 太阳 就 出来了。
16b. 在 那只小鸡 把 那头秃鹰 叼走了 之后， 太阳 就 出来了。
17a. 在 那头大狗 把 那婴孩 吓坏了 之后， 众人 都 很担心。
17b. 在 那婴孩 把 那头大狗 吓坏了 之后， 众人 都 很担心。
18a. 在 那头缉毒犬 把 那个毒贩 认出了 之后， 警报 就 响起来了。
18b. 在 那个毒贩 把 那头缉毒犬 认出了 之后， 警报 就 响起来了。
19a. 在 这只青蛙 把 那些苍蝇 吃掉了 之后， 天上 就 开始 下雨了。
19b. 在 那些苍蝇 把 这只青蛙 吃掉了 之后， 天上 就 开始 下雨了。
20a. 在 那个奸商 把 这群黑工 剥削了 之后， 众人 都 很气愤。
20b. 在 这群黑工 把 那个奸商 剥削了 之后， 众人 都 很气愤。
21a. 在 锺老板 把 这群员工 犒劳了 之后， 大家 都 很高兴。
21b. 在 这群员工 把 锺老板 犒劳了 之后， 大家 都 很高兴。
22a. 在 目击证人 把 凶徒 举报了 之后， 车祸 就 发生了。
22b. 在 凶徒 把 目击证人 举报了 之后， 车祸 就 发生了。
23a. 在 针炙师 把 那位客人 刺痛了 之后， 大家 都 有点害怕。
23b. 在 那位客人 把 针炙师 刺痛了 之后， 大家 都 有点害怕。
24a. 在 丁老板 把 这些员工 辞去了 之后， 众人 都 吃了一惊。
24b. 在 这些员工 把 丁老板 辞去了 之后， 众人 都 吃了一惊。
25a. 在 董大夫 把 那个病人 治愈了 之后， 众人 都 很高兴。
25b. 在 那个病人 把 董大夫 治愈了 之后， 众人 都 很高兴。
26a. 在 那位贵妇 把 那头小狗 装扮好了 之后， 天上 就 开始 下雨了。
26b. 在 那头小狗 把 那位贵妇 装扮好了 之后， 天上 就 开始 下雨了。
27a. 在 评委 把 那位得奖者 夸奖了 之后， 大家 都 很佩服他。
27b. 在 那位得奖者 把 评委 夸奖了 之后， 大家 都 很佩服他。
28a. 在 那个农夫 把 乌鸦 赶走了 之后， 地震 就 发生了。
28b. 在 乌鸦 把 那个农夫 赶走了 之后， 地震 就 发生了。
29a. 在 灭虫专家 把 蟑螂 灭绝了 之后， 其他昆虫 就 出现了。
29b. 在 蟑螂 把 灭虫专家 灭绝了 之后， 其他昆虫 就 出现了。
30a. 在 那位商人 把 这些高官 都收买了 之后， 战事 就 展开了。
30b. 在 这些高官 把 那位商人 都收买了 之后， 战事 就 展开了。
31a. 在 这个考官 把 这些考生 难倒了 之后， 其他考生 都 吃了一惊。
31b. 在 这些考生 把 这个考官 难倒了 之后， 其他考生 都 吃了一惊。
32a. 在 魔术师 把 小猫 变走了 之后， 众人 都 很吃惊。
32b. 在 小猫 把 魔术师 变走了 之后， 众人 都 很吃惊。
33a. 在 饲养员 把 这些动物 喂饱了 之后， 太阳 就 落山了。
33b. 在 这些动物 把 饲养员 喂饱了 之后， 太阳 就 落山了。
34a. 在 皇帝 把 贪官 罢免了 之后， 众人 都 很高兴。
34b. 在 贪官 把 皇帝 罢免了 之后， 众人 都 很高兴。
35a. 在 猎人 把 雪橇狗 排成一队 之后， 天色 就 开始 变坏了。
35b. 在 雪橇狗 把 猎人 排成一队 之后， 天色 就 开始 变坏了。
36a. 在 摄影师 把 那只企鹅 拍下了 之后， 众人 都 想看看。
36b. 在 那只企鹅 把 摄影师 拍下了 之后， 众人 都 想看看。
37a. 在 那个画家 把 那群骏马 画了下来 之后， 众人 都 想看看。
37b. 在 那群骏马 把 那个画家 画了下来 之后， 众人 都 想看看。
38a. 在 骑师 把 这头野马 驯服了 之后， 众人 都 很高兴。
38b. 在 这头野马 把 骑师 驯服了 之后， 众人 都 很高兴。
39a. 在 小弟弟 把 萤火虫 装起来了 之后， 天上 就 开始 下雨了。

39b. 在 萤火虫 把 小弟弟 装起来了 之后， 天上 就 开始 下雨了。
40a. 在 那个渔民 把 那一网鱼 捞起来了 之后， 太阳 就 落山了。
40b. 在 那一网鱼 把 那个渔民 捞起来了 之后， 太阳 就 落山了。
41a. 在 这帮童党 把 那个乖学生 教坏了 之后， 大家 都 很愤怒。
41b. 在 那个乖学生 把 这帮童党 教坏了 之后， 大家 都 很愤怒。
42a. 在 这个法官 把 那个犯人 判刑了 之后， 众人 都 很满意。
42b. 在 那个犯人 把 这个法官 判刑了 之后， 众人 都 很满意。
43a. 在 民众 把 总统 推倒了 之后， 警报 就 响起来了。
43b. 在 总统 把 民众 推倒了 之后， 警报 就 响起来了。
44a. 在 那个恶霸 把 小弟弟 欺负了 之后， 小弟弟 就 生病了。
44b. 在 小弟弟 把 那个恶霸 欺负了 之后， 小弟弟 就 生病了。
45a. 在 那个慈善家 把 那个孤儿 收养了 之后， 众人 都 很高兴。
45b. 在 那个孤儿 把 那个慈善家 收养了 之后， 众人 都 很高兴。
46a. 在 刽子手 把 那个犯人 处决了 之后， 地震 就 发生了。
46b. 在 那个犯人 把 刽子手 处决了 之后， 地震 就 发生了。
47a. 在 那个有钱人 把 佣人 累坏了 之后， 大家 都 批评他。
47b. 在 佣人 把 那个有钱人 累坏了 之后， 大家 都 批评他。
48a. 在 那个小丑 把 观众 逗乐了 之后， 太阳 就 落山了。
48b. 在 观众 把 那个小丑 逗乐了 之后， 太阳 就 落山了。
49a. 在 那些凶徒 把 人质 绑住了 之后， 地震 就 发生了。
49b. 在 人质 把 那些凶徒 绑住了 之后， 地震 就 发生了。
50a. 在 那位师傅 把 那个学徒 培养成材了 之后， 众人 都 很高兴。
50b. 在 那个学徒 把 那位师傅 培养成材了 之后， 众人 都 很高兴。
51a. 在 心理医生 把 那个病人 催眠了 之后， 天色 就 开始 变坏了。
51b. 在 那个病人 把 心理医生 催眠了 之后， 天色 就 开始 变坏了。
52a. 在 那个化妆师 把 女演员 装扮了 之后， 众人 都 赞美她。
52b. 在 女演员 把 那个化妆师 装扮了 之后， 众人 都 赞美她。
53a. 在 那个主人 把 那只小狗 遗失了 之后， 太阳 就 落山了。
53b. 在 那只小狗 把 那个主人 遗失了 之后， 太阳 就 落山了。
54a. 在 那个守卫 把 那可疑人物 扣查了 之后， 众人 都 松了一口气。
54b. 在 那可疑人物 把 那个守卫 扣查了 之后， 众人 都 松了一口气。
55a. 在 灭虫专家 把 那些老鼠 毒死了 之后， 天上 就 开始 下雨了。
55b. 在 那些老鼠 把 灭虫专家 毒死了 之后， 天上 就 开始 下雨了。
56a. 在 消防员 把 被困的小孩 救出来了 之后， 众人 都 很高兴。
56b. 在 被困的小孩 把 消防员 救出来了 之后， 众人 都 很高兴。
57a. 在 按摩师 把 那个客人 弄痛了 之后， 其他客人 都 有点担心。
57b. 在 那个客人 把 按摩师 弄痛了 之后， 其他客人 都 有点担心。
58a. 在 那个奸商 把 民市 欺骗了 之后， 众人 都 很担心。
58b. 在 民市 把 那个奸商 欺骗了 之后， 众人 都 很担心。
59a. 在 那个村民 把 那条鱼 烤熟了 之后， 所有人 都 走过来了。
59b. 在 那条鱼 把 那个村民 烤熟了 之后， 所有人 都 走过来了。
60a. 在 那个球星 把 这班球迷 迷倒了 之后， 球队的收入 增加了。
60b. 在 这班球迷 把 那个球星 迷倒了 之后， 球队的收入 增加了。

Experiment 3

a - Canonical control
b - Role-reversed (Animacy-congruous)

High predictability

1a. 故事中， 屠夫 把 那只猪 宰了 以后 卖到 很好的 价钱。
1b. 故事中， 那只猪 把 屠夫 宰了 以后 卖到 很好的 价钱。
2a. 到了最后， 消防员 把 被困的小孩 解救了 出来， 并 送到了 安全地点。
2b. 到了最后， 被困的小孩 把 消防员 解救了 出来， 并 送到了 安全地点。
3a. 晚会上， 小丑 把 观众 逗乐了 之后 便 表现得 十分神气。
3b. 晚会上， 观众 把 小丑 逗乐了 之后 便 表现得 十分神气。
4a. 记录片中， 青蛙 把 蚊子 吃到 肚子里， 并没有 任何 动静。
4b. 记录片中， 蚊子 把 青蛙 吃到 肚子里， 并没有 任何 动静。
5a. 故事中， 救生员 把 那个溺水者 救起 到岸上， 成为了 大英雄。
5b. 故事中， 那个溺水者 把 救生员 救起 到岸上， 成为了 大英雄。
6a. 传说中， 武松 把 猛虎 打死了 以后 成了 大英雄。
6b. 传说中， 猛虎 把 武松 打死了 以后 成了 大英雄。
7a. 上个月， 那个画家 把 那群骏马 画下来 以后 便 立即 展示 在 画廊中。
7b. 上个月， 那群骏马 把 那个画家 画下来 以后 便 立即 展示 在 画廊中。
8a. 前几天， 警犬 把 失踪少女 找回来 并 送到了 警察局。
8b. 前几天， 失踪少女 把 警犬 找回来 并 送到了 警察局。
9a. 去年， 董大夫 把 那个病人 治愈 以后 便 决定 退休了。
9b. 去年， 那个病人 把 董大夫 治愈 以后 便 决定 退休了。
10a. 晚会上， 相声演员 把 来宾 逗得 哈哈大笑， 全场 掌声雷动。
10b. 晚会上， 来宾 把 相声演员 逗得 哈哈大笑， 全场 掌声雷动。
11a. 昨天晚上， 警察 把 人质 解救了 出来， 成了 救人英雄。
11b. 昨天晚上， 人质 把 警察 解救了 出来， 成了 救人英雄。
12a. 去年初， 行骗者 把 村民 骗了 以后 便 逃之夭夭了。
12b. 去年初， 村民 把 行骗者 骗了 以后 便 逃之夭夭了。
13a. 昨天晚上， 警察 把 小偷 抓了 回警局， 并 教训了 他 一顿。
13b. 昨天晚上， 小偷 把 警察 抓了 回警局， 并 教训了 他 一顿。
14a. 上一次， 造型师 把 女演员 打扮得 很美， 获得 不少的 赞赏。
14b. 上一次， 女演员 把 造型师 打扮得 很美， 获得 不少的 赞赏。
15a. 事件当中， 谈判专家 把 劫匪 说服 并 解救了 人质。
15b. 事件当中， 劫匪 把 谈判专家 说服 并 解救了 人质。
16a. 后来， 何婆婆 把 这个孤儿 收养后 并 带着 他 离开了 小镇。
16b. 后来， 这个孤儿 把 何婆婆 收养后 并 带着 他 离开了 小镇。
17a. 故事中， 巫师 把 王子 变成了 青蛙， 世上 只有 小公主 才能 解救 他。
17b. 故事中， 王子 把 巫师 变成了 青蛙， 世上 只有 小公主 才能 解救 他。
18a. 上个月， 摄影师 把 那群企鹅 拍下来 以后 刊登 在 杂志的 封面上。
18b. 上个月， 那群企鹅 把 摄影师 拍下来 以后 刊登 在 杂志的 封面上。
19a. 昨天， 那个画家 把 模特儿 描画得 非常 完美， 同学们 都 惊叹不已。
19b. 昨天， 模特儿 把 那个画家 描画得 非常 完美， 同学们 都 惊叹不已。
20a. 故事中， 慈善家 把 孤儿 领养 回家， 并且 供 他 上学。
20b. 故事中， 孤儿 把 慈善家 领养 回家， 并且 供 他 上学。
21a. 昨天， 货车司机 把 路人 撞伤了 以后 立即 送 他 到 医院 包扎。
21b. 昨天， 路人 把 货车司机 撞伤了 以后 立即 送 他 到 医院 包扎。
22a. 刚才， 这只花猫 把 那只麻雀 吃得 连 根 羽毛 都 没留下。
22b. 刚才， 那只麻雀 把 这只花猫 吃得 连 根 羽毛 都 没留下。
23a. 电影中， 专家 把 传染病人 隔离了 起来， 防止 病菌 传播。
23b. 电影中， 传染病人 把 专家 隔离了 起来， 防止 病菌 传播。

24a. 去年初， 奸商 把 投资者 骗得 血本无归， 转眼间 逃去无踪。
24b. 去年初， 投资者 把 奸商 骗得 血本无归， 转眼间 逃去无踪。
25a. 电影中， 牧羊人 把 羊群 赶到 草原上 吃草， 不慌不忙的。
25b. 电影中， 羊群 把 牧羊人 赶到 草原上 吃草， 不慌不忙的。
26a. 前几天， 大狗 把 孩子 咬伤了 以后 一天 都 不肯 进食。
26b. 前几天， 孩子 把 大狗 咬伤了 以后 一天 都 不肯 进食。
27a. 去年， 奸商 把 市民 欺骗了 以后 便 臭名远播了。
27b. 去年， 市民 把 奸商 欺骗了 以后 便 臭名远播了。
28a. 上星期， 灭虫专家 把 那些老鼠 消灭 干净， 并且 放了 灭鼠药 以 预防。
28b. 上星期， 那些老鼠 把 灭虫专家 消灭 干净， 并且 放了 灭鼠药 以 预防。
29a. 去年初， 针炙师 把 女病人 治好 以后 心里 十分 满足。
29b. 去年初， 女病人 把 针炙师 治好 以后 心里 十分 满足。
30a. 上星期， 凶犯 把 目击证人 杀死 在 沙发上， 便 逃去无踪了。
30b. 上星期， 目击证人 把 凶犯 杀死 在 沙发上， 便 逃去无踪了。
31a. 刚才， 蚂蚁 把 大象 咬了 一口， 可是 它 完全 没 感觉到。
31b. 刚才， 大象 把 蚂蚁 咬了 一口， 可是 它 完全 没 感觉到。
32a. 昨天， 那个农夫 把 豺狼 打死了 以后 便 向 邻居 炫耀一番。
32b. 昨天， 豺狼 把 那个农夫 打死了 以后 便 向 邻居 炫耀一番。
33a. 昨天， 猎人 把 飞鸟 打下来 以后 便 立即 卖到 很好的 价钱。
33b. 昨天， 飞鸟 把 猎人 打下来 以后 便 立即 卖到 很好的 价钱。
34a. 这些年来， 驯兽师 把 狮子 驯服得 像 一只小猫 一样 乖巧。
34b. 这些年来， 狮子 把 驯兽师 驯服得 像 一只小猫 一样 乖巧。
35a. 上星期， 法官 把 那个犯人 判死刑 以后 听到 很多 反对的 声音。
35b. 上星期， 那个犯人 把 法官 判死刑 以后 听到 很多 反对的 声音。
36a. 晚上， 小明 把 萤火虫 放飞 到 大自然， 心里 很高兴。
36b. 晚上， 萤火虫 把 小明 放飞 到 大自然， 心里 很高兴。
37a. 昨天晚上， 刑警 把 嫌犯 抓住了 并且 立即 带回 警局。
37b. 昨天晚上， 嫌犯 把 刑警 抓住了 并且 立即 带回 警局。
38a. 电影中， 战友 把 烈士 埋葬 在 山坡上， 并 在 坟前 放上 鲜花。
38b. 电影中， 烈士 把 战友 埋葬 在 山坡上， 并 在 坟前 放上 鲜花。
39a. 故事中， 心理医生 把 精神病人 治好 以后 仍然 很关心 他的 状况。
39b. 故事中， 精神病人 把 心理医生 治好 以后 仍然 很关心 他的 状况。
40a. 上星期， 兽医 把 那条小狗 治好后 将 它 送回 主人 怀里。
40b. 上星期， 那条小狗 把 兽医 治好后 将 它 送回 主人 怀里。
41a. 上星期， 消防员 把 那个伤者 救了 出 火海， 自己 却 受了 重伤。
41b. 上星期， 那个伤者 把 消防员 救了 出 火海， 自己 却 受了 重伤。
42a. 昨天下午， 老陈 把 河豚 放回 到 河里， 并 没有 告诉 任何人。
42b. 昨天下午， 河豚 把 老陈 放回 到 河里， 并 没有 告诉 任何人。
43a. 昨天， 魔术师 把 白鸽 变走 以后 全场观众 都 很兴奋。
43b. 昨天， 白鸽 把 魔术师 变走 以后 全场观众 都 很兴奋。
44a. 昨天， 老张 把 害虫 消灭 干净了， 觉得 很有 成就感。
44b. 昨天， 害虫 把 老张 消灭 干净了， 觉得 很有 成就感。
45a. 昨天晚上， 警方 把 疑犯 抓住了 以后 立即 带往 派出所。
45b. 昨天晚上， 疑犯 把 警方 抓住了 以后 立即 带往 派出所。
46a. 昨天， 那个恶霸 把 小弟弟 打了 一顿， 其他 孩子 都 怕得 不敢作声。
46b. 昨天， 小弟弟 把 那个恶霸 打了 一顿， 其他 孩子 都 怕得 不敢作声。
47a. 上星期， 这匹野马 把 骑师 摔了 下来， 而且 踩了 一脚。
47b. 上星期， 骑师 把 这匹野马 摔了 下来， 而且 踩了 一脚。
48a. 昨天下午， 老鹰 把 野鸡 吃掉了 以后 便 飞走了。

48b. 昨天下午， 野鸡 把 老鹰 吃掉了 以后 便 飞走了。
49a. 故事里， 刽子手 把 死囚 杀死了 以后 便 严肃地 离开了。
49b. 故事里， 死囚 把 刽子手 杀死了 以后 便 严肃地 离开了。
50a. 今天， 老伯伯 把 金鱼 放进 鱼缸里， 并 换上 新鲜的 水。
50b. 今天， 金鱼 把 老伯伯 放进 鱼缸里， 并 换上 新鲜的 水。
51a. 上星期， 警察 把 灾民 安置 到了 安全的 地方， 十分 尽责。
51b. 上星期， 灾民 把 警察 安置 到了 安全的 地方， 十分 尽责。
52a. 事件中， 那些歹徒 把 人质 杀害了 以后 便 与 警方 展开 枪战。
52b. 事件中， 人质 把 那些歹徒 杀害了 以后 便 与 警方 展开 枪战。
53a. 去年， 坏学生 把 小弟弟 带坏了 并且 经常 指使 他 去 偷东西。
53b. 去年， 小弟弟 把 坏学生 带坏了 并且 经常 指使 他 去 偷东西。
54a. 昨天晚上， 蟑螂 把 黄老板 吓得 连 饭 都 不敢 吃了。
54b. 昨天晚上， 黄老板 把 蟑螂 吓得 连 饭 都 不敢 吃了。
55a. 刚才， 熊猫宝宝 把 饲养人员 逗得 笑弯了腰， 心情 豁然 开朗。
55b. 刚才， 饲养人员 把 熊猫宝宝 逗得 笑弯了腰， 心情 豁然 开朗。
56a. 去年， 那个毒贩 把 那条缉毒犬 杀死 在 路旁， 警方 查了 很久 才 破案。
56b. 去年， 那条缉毒犬 把 那个毒贩 杀死 在 路旁， 警方 查了 很久 才 破案。
57a. 传说中， 皇帝 把 佟妃 打入冷宫 以后 继续 花天酒地。
57b. 传说中， 佟妃 把 皇帝 打入冷宫 以后 继续 花天酒地。
58a. 刚才， 仓鼠 把 男孩 吓倒 在 地上， 不知道 溜 到 哪里 去了。
58b. 刚才， 男孩 把 仓鼠 吓倒 在 地上， 不知道 溜 到 哪里 去了。
59a. 上个月， 吴管家 把 那只流浪狗 收留 下来 并且 很用心 照顾 它。
59b. 上个月， 那只流浪狗 把 吴管家 收留 下来 并且 很用心 照顾 它。
60a. 电影中， 斗牛勇士 把 蛮牛 制服了 之后 便 拔刀 杀了 它。
60b. 电影中， 蛮牛 把 斗牛勇士 制服了 之后 便 拔刀 杀了 它。

Low predictability

61a. 事件当中， 那群旅客 把 恐怖分子 制服了 以后 大家 才 稍微 松一口气。
61b. 事件当中， 恐怖分子 把 那群旅客 制服了 以后 大家 才 稍微 松一口气。
62a. 故事中， 铁木真 把 部落民族 聚集 起来， 宣布 重要的 消息。
62b. 故事中， 部落民族 把 铁木真 聚集 起来， 宣布 重要的 消息。
63a. 昨天， 那位顾客 把 售货员 投诉了 一番， 场面 十分 尴尬。
63b. 昨天， 售货员 把 那位顾客 投诉了 一番， 场面 十分 尴尬。
64a. 那年， 人们 把 鲨鱼 捕获了 之后 就 开始 庆祝。
64b. 那年， 鲨鱼 把 人们 捕获了 之后 就 开始 庆祝。
65a. 报道中， 巡逻队 把 藏羚羊 送回 保护区， 并且 好好地 照顾。
65b. 报道中， 藏羚羊 把 巡逻队 送回 保护区， 并且 好好地 照顾。
66a. 那年， 贪官 把 包青天 陷害得 苦不堪言， 激起了 百姓的 愤怒。
66b. 那年， 包青天 把 贪官 陷害得 苦不堪言， 激起了 百姓的 愤怒。
67a. 当天， 狱卒 把 那个囚犯 放走了 之后 还是 决定 自首了。
67b. 当天， 那个囚犯 把 狱卒 放走了 之后 还是 决定 自首了。
68a. 那天， 得奖者 把 评委 感动了 还 流下了 热泪。
68b. 那天， 评委 把 得奖者 感动了 还 流下了 热泪。
69a. 这段时间， 导演 把 新演员 培训了 一番， 十分 用心。
69b. 这段时间， 新演员 把 导演 培训了 一番， 十分 用心。
70a. 昨天， 猎人 把 雪橇狗 喂饱了 以后 便 出外 打猎了。
70b. 昨天， 雪橇狗 把 猎人 喂饱了 以后 便 出外 打猎了。
71a. 去年， 秘书 把 董事 迷惑得 神魂颠倒了， 得到 不少 好处。

71b. 去年， 董事 把 秘书 迷惑得 神魂颠倒了， 得到 不少 好处。
72a. 上个月， 经理人 把 新乐队 包装了 一番， 并 开始 着手 准备 出专辑。
72b. 上个月， 新乐队 把 经理人 包装了 一番， 并 开始 着手 准备 出专辑。
73a. 故事中， 佣人 把 富翁 服侍了 多年， 最后 却 被 赶出去了。
73b. 故事中， 富翁 把 佣人 服侍了 多年， 最后 却 被 赶出去了。
74a. 昨天， 皇后 把 宫女 处罚了 一顿， 并且 公开了 她的 恶行。
74b. 昨天， 宫女 把 皇后 处罚了 一顿， 并且 公开了 她的 恶行。
75a. 前几天， 贵妇 把 波斯猫 买了 回家 为 自己 解闷。
75b. 前几天， 波斯猫 把 贵妇 买了 回家 为 自己 解闷。
76a. 上星期， 何太太 把 搬运工人 请来 帮忙， 一点 也 不吝啬。
76b. 上星期， 搬运工人 把 何太太 请来 帮忙， 一点 也 不吝啬。
77a. 今天， 工人 把 那个商人 告上 法庭， 誓要 讨回 公道。
77b. 今天， 那个商人 把 工人 告上 法庭， 誓要 讨回 公道。
78a. 故事中， 皇后 把 白雪公主 骗了 过来， 并且 毒害了。
78b. 故事中， 白雪公主 把 皇后 骗了 过来， 并且 毒害了。
79a. 當時， 群众 把 那个叛徒 揪出来 毒打， 毫不 留手。
79b. 當時， 那个叛徒 把 群众 揪出来 毒打， 毫不 留手。
80a. 上星期， 老太太 把 猫咪 丢了 在 公园， 使 家人 十分 担心。
80b. 上星期， 猫咪 把 老太太 丢了 在 公园， 使 家人 十分 担心。
81a. 昨天， 护士 把 伤兵 包扎 妥当 并且 扶到 病床上。
81b. 昨天， 伤兵 把 护士 包扎 妥当 并且 扶到 病床上。
82a. 昨天， 小孩 把 保姆 累坏了 之后， 就 跑去 找 爷爷 玩了。
82b. 昨天， 保姆 把 小孩 累坏了 之后， 就 跑去 找 爷爷 玩了。
83a. 这年初， 地主 把 童工 卖给了 人贩， 从中 谋取 暴利。
83b. 这年初， 童工 把 地主 卖给了 人贩， 从中 谋取 暴利。
84a. 电影里， 妈妈 把 新生儿 遗弃了 在 教堂 门口， 心痛 欲绝。
84b. 电影里， 新生儿 把 妈妈 遗弃了 在 教堂 门口， 心痛 欲绝。
85a. 昨天， 熊猫妈妈 把 小熊猫 生了 下来， 大家 都 为此 感到 很高兴。
85b. 昨天， 小熊猫 把 熊猫妈妈 生了 下来， 大家 都 为此 感到 很高兴。
86a. 那天， 乞丐 把 吴管家 缠住了 不愿走， 希望 能 讨得 几块钱。
86b. 那天， 吴管家 把 乞丐 缠住了 不愿走， 希望 能 讨得 几块钱。
87a. 第二天， 那个奸商 把 这群黑工 藏在 小黑屋里， 让 他们 干累活。
87b. 第二天， 这群黑工 把 那个奸商 藏在 小黑屋里， 让 他们 干累活。
88a. 那天， 乌鸦 把 那个农夫 惹火了 以后 还 不走， 真的 不知死活。
88b. 那天， 那个农夫 把 乌鸦 惹火了 以后 还 不走， 真的 不知死活。
89a. 那年， 地方官 把 首相 贿赂了 以后 变得 有恃无恐。
89b. 那年， 首相 把 地方官 贿赂了 以后 变得 有恃无恐。
90a. 电影里， 学徒 把 那位师傅 超越了 以后 就 得意忘形了。
90b. 电影里， 那位师傅 把 学徒 超越了 以后 就 得意忘形了。
91a. 昨天， 小弟弟 把 狮子 吵醒了 以后 害怕得 只懂 躲在 爸爸 身后。
91b. 昨天， 狮子 把 小弟弟 吵醒了 以后 害怕得 只懂 躲在 爸爸 身后。
92a. 上星期， 那个毕业生 把 周老板 恭维了 一番 就 得到了 录用。
92b. 上星期， 周老板 把 那个毕业生 恭维了 一番 就 得到了 录用。
93a. 刚才， 按摩师 把 老顾客 得罪了 以后 连忙 道歉。
93b. 刚才， 老顾客 把 按摩师 得罪了 以后 连忙 道歉。
94a. 去年， 高官 把 那位商人 制裁了 以后 大家 都 很高兴。
94b. 去年， 那位商人 把 高官 制裁了 以后 大家 都 很高兴。
95a. 传闻说， 将军 把 士兵 提拔 成为 第九小队的 组长。
95b. 传闻说， 士兵 把 将军 提拔 成为 第九小队的 组长。

96a. 几天前， 小潘 把 那个女孩 娶了 回家 当 媳妇， 还 请 大家 吃了 一顿饭。
96b. 几天前， 那个女孩 把 小潘 娶了 回家 当 媳妇， 还 请 大家 吃了 一顿饭。
97a. 上星期， 私家侦探 把 那个政客 揭发了 并且 刊登了 照片 作 证据。
97b. 上星期， 那个政客 把 私家侦探 揭发了 并且 刊登了 照片 作 证据。
98a. 那天， 可疑人物 把 门卫 引开 之后 就 盗走了 宝物。
98b. 那天， 门卫 把 可疑人物 引开 之后 就 盗走了 宝物。
99a. 事实上， 暴君 把 民众 统治了 已有 几十年。
99b. 事实上， 民众 把 暴君 统治了 已有 几十年。
100a. 访问里， 演员 把 国家元首 讽刺了 一番， 十分 到位。
100b. 访问里， 国家元首 把 演员 讽刺了 一番， 十分 到位。
101a. 这一年， 考生们 把 考官 收买 起来了， 真是 难以置信。
101b. 这一年， 考官 把 考生们 收买 起来了， 真是 难以置信。
102a. 去年初， 老刘 把 鹦鹉 训练了 好一段 时间， 但是 还未 成功。
102b. 去年初， 鹦鹉 把 老刘 训练了 好一段 时间， 但是 还未 成功。
103a. 上星期， 钟老板 把 这群员工 奖励了 一番， 这 让 大家 都 很是意外。
103b. 上星期， 这群员工 把 钟老板 奖励了 一番， 这 让 大家 都 很是意外。
104a. 刚才， 那个主人 把 那只小狗 打扮了 一番， 然后 才 开心地 出门。
104b. 刚才， 那只小狗 把 那个主人 打扮了 一番， 然后 才 开心地 出门。
105a. 上星期， 那个村民 把 那条鱼 送给了 村长， 希望 討得 他的 歡心。
105b. 上星期， 那条鱼 把 那个村民 送给了 村长， 希望 討得 他的 歡心。
106a. 书上说， 老作家 把 新人 捧红了 以后 便 退隐 文坛了。
106b. 书上说， 新人 把 老作家 捧红了 以后 便 退隐 文坛了。
107a. 昨天， 教练 把 选手 鼓励了 一番 之后 比赛 就 开始了。
107b. 昨天， 选手 把 教练 鼓励了 一番 之后 比赛 就 开始了。
108a. 昨天， 海洋学家 把 鲸鱼 解剖了 以后 写了 一份报告。
108b. 昨天， 鲸鱼 把 海洋学家 解剖了 以后 写了 一份报告。
109a. 这几年来， 爷爷 把 孙子 宠坏了 但是 他 觉得 是 理所当然的。
109b. 这几年来， 孙子 把 爷爷 宠坏了 但是 他 觉得 是 理所当然的。
110a. 晚上， 父母 把 孩子 哄睡了 以后 才 能 好好 休息 一下。
110b. 晚上， 孩子 把 父母 哄睡了 以后 才 能 好好 休息 一下。
111a. 上星期， 女歌手 把 狗仔队 耍了 之后 就 消失了。
111b. 上星期， 狗仔队 把 女歌手 耍了 之后 就 消失了。
112a. 电影里， 孩子们 把 妈妈 歌颂了 一番， 场面 非常 感人。
112b. 电影里， 妈妈 把 孩子们 歌颂了 一番， 场面 非常 感人。
113a. 昨天下午， 奶奶 把 孙女 喂饱了 以后 抱 到 床上 小睡。
113b. 昨天下午， 孙女 把 奶奶 喂饱了 以后 抱 到 床上 小睡。
114a. 那天， 丁老板 把 张主任 开除了 以后 觉得 很惭愧。
114b. 那天， 张主任 把 丁老板 开除了 以后 觉得 很惭愧。
115a. 后来， 那位候选人 把 黑帮头目 绳之 以法， 赢得了 大家的 支持。
115b. 后来， 黑帮头目 把 那位候选人 绳之 以法， 赢得了 大家的 支持。
116a. 那天， 那位教授 把 学生们 表扬了 一番， 也 讲了 一些 励志的 话。
116b. 那天， 学生们 把 那位教授 表扬了 一番， 也 讲了 一些 励志的 话。
117a. 每天傍晚， 那对夫妇 把 小狗 喂饱 以后 便 一起 去 公园 散步了。
117b. 每天傍晚， 小狗 把 那对夫妇 喂饱 以后 便 一起 去 公园 散步了。
118a. 这几天， 见习生 把 老板 惹怒了 却 一直 没有 发现。
118b. 这几天， 老板 把 见习生 惹怒了 却 一直 没有 发现。
119a. 去年， 总编辑 把 那模特儿 捧红 成为 一线明星， 并 赚了 一笔钱。
119b. 去年， 那模特儿 把 总编辑 捧红 成为 一线明星， 并 赚了 一笔钱。
120a. 那天， 那位贵妇 把 那条小狗 送给了 邻居 并 托 她 好好 照顾。

120b. 那天， 那条小狗 把 那位贵妇 送给了 邻居 并 托 她 好好 照顾。

Experiment 4

a – Short distance, canonical control
b – Short distance, Role-reversed (Animacy-congruous)
c – Long distance, canonical control
d – Long distance, Role-reversed (Animacy-congruous)

1a. 那天下午， 这只花猫 把 那只麻雀 吃得 连 根 羽毛 都 没留下。
1b. 那天下午， 那只麻雀 把 这只花猫 吃得 连 根 羽毛 都 没留下。
1c. 这只花猫 把 那只麻雀 在 那天下午 吃得 连 根 羽毛 都 没留下。
1d. 那只麻雀 把 这只花猫 在 那天下午 吃得 连 根 羽毛 都 没留下。
2a. 昨天晚上， 货车司机 把 路人 撞伤了 以后 立即 送 他 到 医院 包扎。
2b. 昨天晚上， 路人 把 货车司机 撞伤了 以后 立即 送 他 到 医院 包扎。
2c. 货车司机 把 路人 在 昨天晚上 撞伤了 以后 立即 送 他 到 医院 包扎。
2d. 路人 把 货车司机 在 昨天晚上 撞伤了 以后 立即 送 他 到 医院 包扎。
3a. 去年春天， 慈善家 把 孤儿 领养 回家， 并且 供 他 上学。
3b. 去年春天， 孤儿 把 慈善家 领养 回家， 并且 供 他 上学。
3c. 慈善家 把 孤儿 在 去年春天 领养 回家， 并且 供 他 上学。
3d. 孤儿 把 慈善家 在 去年春天 领养 回家， 并且 供 他 上学。
4a. 昨天早上， 牧羊人 把 羊群 赶到 草原上 吃草， 不慌不忙的。
4b. 昨天早上， 羊群 把 牧羊人 赶到 草原上 吃草， 不慌不忙的。
4c. 牧羊人 把 羊群 在 昨天早上 赶到 草原上 吃草， 不慌不忙的。
4d. 羊群 把 牧羊人 在 昨天早上 赶到 草原上 吃草， 不慌不忙的。
5a. 去年五月， 奸商 把 投资者 骗得 血本无归， 转眼间 逃去无踪。
5b. 去年五月， 投资者 把 奸商 骗得 血本无归， 转眼间 逃去无踪。
5c. 奸商 把 投资者 在 去年五月 骗得 血本无归， 转眼间 逃去无踪。
5d. 投资者 把 奸商 在 去年五月 骗得 血本无归， 转眼间 逃去无踪。
6a. 上星期， 专家 把 传染病人 隔离了 起来， 防止 病菌 传播。
6b. 上星期， 传染病人 把 专家 隔离了 起来， 防止 病菌 传播。
6c. 专家 把 传染病人 在 上星期 隔离了 起来， 防止 病菌 传播。
6d. 传染病人 把 专家 在 上星期 隔离了 起来， 防止 病菌 传播。
7a. 上星期， 大狗 把 孩子 咬伤了 以后 一天 都 不肯 进食。
7b. 上星期， 孩子 把 大狗 咬伤了 以后 一天 都 不肯 进食。
7c. 大狗 把 孩子 在 上星期 咬伤了 以后 一天 都 不肯 进食。
7d. 孩子 把 大狗 在 上星期 咬伤了 以后 一天 都 不肯 进食。
8a. 昨天晚上， 灭虫专家 把 那些老鼠 消灭 干净， 并且 放了 灭鼠药 以 预防。
8b. 昨天晚上， 那些老鼠 把 灭虫专家 消灭 干净， 并且 放了 灭鼠药 以 预防。
8c. 灭虫专家 把 那些老鼠 在 昨天晚上 消灭 干净， 并且 放了 灭鼠药 以 预防。
8d. 那些老鼠 把 灭虫专家 在 昨天晚上 消灭 干净， 并且 放了 灭鼠药 以 预防。
9a. 去年底， 奸商 把 市民 欺骗了 以后 便 臭名远播了。
9b. 去年底， 市民 把 奸商 欺骗了 以后 便 臭名远播了。
9c. 奸商 把 市民 在 去年底 欺骗了 以后 便 臭名远播了。
9d. 市民 把 奸商 在 去年底 欺骗了 以后 便 臭名远播了。
10a. 那天早上， 蚂蚁 把 大象 咬了 一口， 可是 它 完全 没 感觉到。
10b. 那天早上， 大象 把 蚂蚁 咬了 一口， 可是 它 完全 没 感觉到。
10c. 蚂蚁 把 大象 在 那天早上 咬了 一口， 可是 它 完全 没 感觉到。

10d. 大象 把 蚂蚁 在 那天早上 咬了 一口， 可是 它 完全 没 感觉到。
11a. 昨天晚上， 那头小狗 把 那位贵妇 咬了 一口， 并 继续 追着 她 跑。
11b. 昨天晚上， 那位贵妇 把 那头小狗 咬了 一口， 并 继续 追着 她 跑。
11c. 那头小狗 把 那位贵妇 在 昨天晚上 咬了 一口， 并 继续 追着 她 跑。
11d. 那位贵妇 把 那头小狗 在 昨天晚上 咬了 一口， 并 继续 追着 她 跑。
12a. 去年初， 凶犯 把 目击证人 杀死 在沙发上， 便 逃去无踪了。
12b. 去年初， 目击证人 把 凶犯 杀死 在沙发上， 便 逃去无踪了。
12c. 凶犯 把 目击证人 在 去年初 杀死 在沙发上， 便 逃去无踪了。
12d. 目击证人 把 凶犯 在 去年初 杀死 在沙发上， 便 逃去无踪了。
13a. 上个月， 针灸师 把 女病人 治好 以后 心里 十分 满足。
13b. 上个月， 女病人 把 针灸师 治好 以后 心里 十分 满足。
13c. 针灸师 把 女病人 在 上个月 治好 以后 心里 十分 满足。
13d. 女病人 把 针灸师 在 上个月 治好 以后 心里 十分 满足。
14a. 上星期， 猎人 把 飞鸟 打下来 以后 便 立即 卖到 很好的 价钱。
14b. 上星期， 飞鸟 把 猎人 打下来 以后 便 立即 卖到 很好的 价钱。
14c. 猎人 把 飞鸟 在 上星期 打下来 以后 便 立即 卖到 很好的 价钱。
14d. 飞鸟 把 猎人 在 上星期 打下来 以后 便 立即 卖到 很好的 价钱。
15a. 上星期， 法官 把 那个犯人 判死刑 以后 听到 很多 反对的 声音。
15b. 上星期， 那个犯人 把 法官 判死刑 以后 听到 很多 反对的 声音。
15c. 法官 把 那个犯人 在 上星期 判死刑 以后 听到 很多 反对的 声音。
15d. 那个犯人 把 法官 在 上星期 判死刑 以后 听到 很多 反对的 声音。
16a. 去年初， 驯兽师 把 狮子 驯服得 像 一只小猫 一样 乖巧。
16b. 去年初， 狮子 把 驯兽师 驯服得 像 一只小猫 一样 乖巧。
16c. 驯兽师 把 狮子 在 去年初 驯服得 像 一只小猫 一样 乖巧。
16d. 狮子 把 驯兽师 在 去年初 驯服得 像 一只小猫 一样 乖巧。
17a. 昨天早上， 那个农夫 把 豺狼 打死了 以后 便 向 邻居 炫耀一番。
17b. 昨天早上， 豺狼 把 那个农夫 打死了 以后 便 向 邻居 炫耀一番。
17c. 那个农夫 把 豺狼 在 昨天早上 打死了 以后 便 向 邻居 炫耀一番。
17d. 豺狼 把 那个农夫 在 昨天早上 打死了 以后 便 向 邻居 炫耀一番。
18a. 昨天晚上， 小明 把 萤火虫 放飞 到 大自然， 心里 很高兴。
18b. 昨天晚上， 萤火虫 把 小明 放飞 到 大自然， 心里 很高兴。
18c. 小明 把 萤火虫 在 昨天晚上 放飞 到 大自然， 心里 很高兴。
18d. 萤火虫 把 小明 在 昨天晚上 放飞 到 大自然， 心里 很高兴。
19a. 上星期， 黑帮头目 把 那位候选人 杀掉 并 故意 造成是 自杀的 景象。
19b. 上星期， 那位候选人 把 黑帮头目 杀掉 并 故意 造成是 自杀的 景象。
19c. 黑帮头目 把 那位候选人 在 上星期 杀掉 并 故意 造成是 自杀的 景象。
19d. 那位候选人 把 黑帮头目 在 上星期 杀掉 并 故意 造成是 自杀的 景象。
20a. 昨天晚上， 刑警 把 嫌犯 抓住了 并且 立即 带回 警局。
20b. 昨天晚上， 嫌犯 把 刑警 抓住了 并且 立即 带回 警局。
20c. 刑警 把 嫌犯 在 昨天晚上 抓住了 并且 立即 带回 警局。
20d. 嫌犯 把 刑警 在 昨天晚上 抓住了 并且 立即 带回 警局。
21a. 去年初， 战友 把 烈士 埋葬 在山坡上， 并 在 坟前 放上 鲜花。
21b. 去年初， 烈士 把 战友 埋葬 在山坡上， 并 在 坟前 放上 鲜花。
21c. 战友 把 烈士 在 去年初 埋葬 在山坡上， 并 在 坟前 放上 鲜花。
21d. 烈士 把 战友 在 去年初 埋葬 在山坡上， 并 在 坟前 放上 鲜花。
22a. 上星期， 兽医 把 那条小狗 治好后 将 它 送回 主人 怀里。
22b. 上星期， 那条小狗 把 兽医 治好后 将 它 送回 主人 怀里。
22c. 兽医 把 那条小狗 在 上星期 治好后 将 它 送回 主人 怀里。
22d. 那条小狗 把 兽医 在 上星期 治好后 将 它 送回 主人 怀里。

23a. 去年初， 心理医生 把 精神病人 治好 以后 仍然 很关心 他的 状况。
23b. 去年初， 精神病人 把 心理医生 治好 以后 仍然 很关心 他的 状况。
23c. 心理医生 把 精神病人 在 去年初 治好 以后 仍然 很关心 他的 状况。
23d. 精神病人 把 心理医生 在 去年初 治好 以后 仍然 很关心 他的 状况。
24a. 昨天下午， 吴管家 把 乞丐 赶出去 然后 赶紧 把 门 关上。
24b. 昨天下午， 乞丐 把 吴管家 赶出去 然后 赶紧 把 门 关上。
24c. 吴管家 把 乞丐 在 昨天下午 赶出去 然后 赶紧 把 门 关上。
24d. 乞丐 把 吴管家 在 昨天下午 赶出去 然后 赶紧 把 门 关上。
25a. 今天早上， 老陈 把 河豚 放回 到河里， 并 没有 告诉 任何人。
25b. 今天早上， 河豚 把 老陈 放回 到河里， 并 没有 告诉 任何人。
25c. 老陈 把 河豚 在 今天早上 放回 到河里， 并 没有 告诉 任何人。
25d. 河豚 把 老陈 在 今天早上 放回 到河里， 并 没有 告诉 任何人。
26a. 上星期， 消防员 把 那个伤者 救了 出火海， 自己 却 受了 重伤。
26b. 上星期， 那个伤者 把 消防员 救了 出火海， 自己 却 受了 重伤。
26c. 消防员 把 那个伤者 在 上星期 救了 出火海， 自己 却 受了 重伤。
26d. 那个伤者 把 消防员 在 上星期 救了 出火海， 自己 却 受了 重伤。
27a. 昨天下午， 狮子 把 小弟弟 吓得 嚎啕大哭， 姐姐 在一旁 安抚 他。
27b. 昨天下午， 小弟弟 把 狮子 吓得 嚎啕大哭， 姐姐 在一旁 安抚 他。
27c. 狮子 把 小弟弟 在 昨天下午 吓得 嚎啕大哭， 姐姐 在一旁 安抚 他。
27d. 小弟弟 把 狮子 在 昨天下午 吓得 嚎啕大哭， 姐姐 在一旁 安抚 他。
28a. 上星期， 老张 把 害虫 消灭 干净了， 觉得 很有 成就感。
28b. 上星期， 害虫 把 老张 消灭 干净了， 觉得 很有 成就感。
28c. 老张 把 害虫 在 上星期 消灭 干净了， 觉得 很有 成就感。
28d. 害虫 把 老张 在 上星期 消灭 干净了， 觉得 很有 成就感。
29a. 昨天晚上， 魔术师 把 白鸽 变走 以后 全场观众 都 很兴奋。
29b. 昨天晚上， 白鸽 把 魔术师 变走 以后 全场观众 都 很兴奋。
29c. 魔术师 把 白鸽 在 昨天晚上 变走 以后 全场观众 都 很兴奋。
29d. 白鸽 把 魔术师 在 昨天晚上 变走 以后 全场观众 都 很兴奋。
30a. 上个月， 刽子手 把 死囚 杀死了 以后 便 严肃地 离开了。
30b. 上个月， 死囚 把 刽子手 杀死了 以后 便 严肃地 离开了。
30c. 刽子手 把 死囚 在 上个月 杀死了 以后 便 严肃地 离开了。
30d. 死囚 把 刽子手 在 上个月 杀死了 以后 便 严肃地 离开了。
31a. 昨天下午， 老鹰 把 野鸡 吃掉了 以后 便 飞走了。
31b. 昨天下午， 野鸡 把 老鹰 吃掉了 以后 便 飞走了。
31c. 老鹰 把 野鸡 在 昨天下午 吃掉了 以后 便 飞走了。
31d. 野鸡 把 老鹰 在 昨天下午 吃掉了 以后 便 飞走了。
32a. 上星期， 这匹野马 把 骑师 摔了 下来， 而且 踩了 一脚。
32b. 上星期， 骑师 把 这匹野马 摔了 下来， 而且 踩了 一脚。
32c. 这匹野马 把 骑师 在 上星期 摔了 下来， 而且 踩了 一脚。
32d. 骑师 把 这匹野马 在 上星期 摔了 下来， 而且 踩了 一脚。
33a. 昨天下午， 那个恶霸 把 小弟弟 打了 一顿， 其他 孩子 都 怕得 不敢作声。
33b. 昨天下午， 小弟弟 把 那个恶霸 打了 一顿， 其他 孩子 都 怕得 不敢作声。
33c. 那个恶霸 把 小弟弟 在 昨天下午 打了 一顿， 其他 孩子 都 怕得 不敢作声。
33d. 小弟弟 把 那个恶霸 在 昨天下午 打了 一顿， 其他 孩子 都 怕得 不敢作声。
34a. 上星期， 警方 把 疑犯 抓住了 以后 立即 带往 派出所。
34b. 上星期， 疑犯 把 警方 抓住了 以后 立即 带往 派出所。
34c. 警方 把 疑犯 在 上星期 抓住了 以后 立即 带往 派出所。
34d. 疑犯 把 警方 在 上星期 抓住了 以后 立即 带往 派出所。
35a. 今天早上， 老伯伯 把 金鱼 放进 鱼缸里， 并 换上 新鲜的 水。

35b. 今天早上， 金鱼 把 老伯伯 放进 鱼缸里， 并 换上 新鲜的 水。
35c. 老伯伯 把 金鱼 在 今天早上 放进 鱼缸里， 并 换上 新鲜的 水。
35d. 金鱼 把 老伯伯 在 今天早上 放进 鱼缸里， 并 换上 新鲜的 水。
36a. 去年初， 坏学生 把 小弟弟 带坏了 并且 经常 指使 他 去 偷东西。
36b. 去年初， 小弟弟 把 坏学生 带坏了 并且 经常 指使 他 去 偷东西。
36c. 坏学生 把 小弟弟 在 去年初 带坏了 并且 经常 指使 他 去 偷东西。
36d. 小弟弟 把 坏学生 在 去年初 带坏了 并且 经常 指使 他 去 偷东西。
37a. 昨天上午， 警察 把 灾民 安置 到了 安全的 地方， 十分 尽责。
37b. 昨天上午， 灾民 把 警察 安置 到了 安全的 地方， 十分 尽责。
37c. 警察 把 灾民 在 昨天上午 安置 到了 安全的 地方， 十分 尽责。
37d. 灾民 把 警察 在 昨天上午 安置 到了 安全的 地方， 十分 尽责。
38a. 今天早上， 那些歹徒 把 人质 杀害了 以后 便 与 警方 展开 枪战。
38b. 今天早上， 人质 把 那些歹徒 杀害了 以后 便 与 警方 展开 枪战。
38c. 那些歹徒 把 人质 在 今天早上 杀害了 以后 便 与 警方 展开 枪战。
38d. 人质 把 那些歹徒 在 今天早上 杀害了 以后 便 与 警方 展开 枪战。
39a. 昨天晚上， 蟑螂 把 黄老板 吓得 连 饭 都 不敢 吃了。
39b. 昨天晚上， 黄老板 把 蟑螂 吓得 连 饭 都 不敢 吃了。
39c. 蟑螂 把 黄老板 在 昨天晚上 吓得 连 饭 都 不敢 吃了。
39d. 黄老板 把 蟑螂 在 昨天晚上 吓得 连 饭 都 不敢 吃了。
40a. 昨天早上， 熊猫宝宝 把 饲养人员 逗得 笑弯了腰， 心情 豁然 开朗。
40b. 昨天早上， 饲养人员 把 熊猫宝宝 逗得 笑弯了腰， 心情 豁然 开朗。
40c. 熊猫宝宝 把 饲养人员 在 昨天早上 逗得 笑弯了腰， 心情 豁然 开朗。
40d. 饲养人员 把 熊猫宝宝 在 昨天早上 逗得 笑弯了腰， 心情 豁然 开朗。
41a. 上个月， 吴管家 把 那只流浪狗 收留 下来 并且 很用心 照顾 它。
41b. 上个月， 那只流浪狗 把 吴管家 收留 下来 并且 很用心 照顾 它。
41c. 吴管家 把 那只流浪狗 在 上个月 收留 下来 并且 很用心 照顾 它。
41d. 那只流浪狗 把 吴管家 在 上个月 收留 下来 并且 很用心 照顾 它。
42a. 上星期， 仓鼠 把 男孩 吓倒 在地上， 不知道 溜 到 哪里 去了。
42b. 上星期， 男孩 把 仓鼠 吓倒 在地上， 不知道 溜 到 哪里 去了。
42c. 仓鼠 把 男孩 在 上星期 吓倒 在地上， 不知道 溜 到 哪里 去了。
42d. 男孩 把 仓鼠 在 上星期 吓倒 在地上， 不知道 溜 到 哪里 去了。
43a. 那年初， 皇帝 把 佟妃 打入冷宫 以后 继续 花天酒地。
43b. 那年初， 佟妃 把 皇帝 打入冷宫 以后 继续 花天酒地。
43c. 皇帝 把 佟妃 在 那年初 打入冷宫 以后 继续 花天酒地。
43d. 佟妃 把 皇帝 在 那年初 打入冷宫 以后 继续 花天酒地。
44a. 上个月， 巡逻队 把 藏羚羊 救了 回来， 令人 十分 佩服。
44b. 上个月， 藏羚羊 把 巡逻队 救了 回来， 令人 十分 佩服。
44c. 巡逻队 把 藏羚羊 在 上个月 救了 回来， 令人 十分 佩服。
44d. 藏羚羊 把 巡逻队 在 上个月 救了 回来， 令人 十分 佩服。
45a. 那天晚上， 狱卒 把 那个囚犯 打了 一顿， 使 他 重伤。
45b. 那天晚上， 那个囚犯 把 狱卒 打了 一顿， 使 他 重伤。
45c. 狱卒 把 那个囚犯 在 那天晚上 打了 一顿， 使 他 重伤。
45d. 那个囚犯 把 狱卒 在 那天晚上 打了 一顿， 使 他 重伤。
46a. 这几天， 猫咪 把 老太太 逗得 很开心， 使 她 不再 愁眉深锁。
46b. 这几天， 老太太 把 猫咪 逗得 很开心， 使 她 不再 愁眉深锁。
46c. 猫咪 把 老太太 在 这几天 逗得 很开心， 使 她 不再 愁眉深锁。
46d. 老太太 把 猫咪 在 这几天 逗得 很开心， 使 她 不再 愁眉深锁。
47a. 那天下午， 斗牛勇士 把 蛮牛 制服了 之后 便 拔刀 杀了 它。
47b. 那天下午， 蛮牛 把 斗牛勇士 制服了 之后 便 拔刀 杀了 它。

47c. 斗牛勇士 把 蛮牛 在 那天下午 制服了 之后 便 拔刀 杀了 它。
47d. 蛮牛 把 斗牛勇士 在 那天下午 制服了 之后 便 拔刀 杀了 它。
48a. 这些年来， 父母 把 孩子 抚养 长大 是 很辛苦的。
48b. 这些年来， 孩子 把 父母 抚养 长大 是 很辛苦的。
48c. 父母 把 孩子 在 这些年来 抚养 长大 是 很辛苦的。
48d. 孩子 把 父母 在 这些年来 抚养 长大 是 很辛苦的。
49a. 今天早上， 妈妈 把 新生儿 抱来 放在 怀里， 脸上 露出了 欣慰的 笑容。
49b. 今天早上， 新生儿 把 妈妈 抱来 放在 怀里， 脸上 露出了 欣慰的 笑容。
49c. 妈妈 把 新生儿 在 今天早上 抱来 放在 怀里， 脸上 露出了 欣慰的 笑容。
49d. 新生儿 把 妈妈 在 今天早上 抱来 放在 怀里， 脸上 露出了 欣慰的 笑容。
50a. 那年初， 穷书生 把 仙女 娶回家 之后， 他的 人生 便 开始 转运 了。
50b. 那年初， 仙女 把 穷书生 娶回家 之后， 他的 人生 便 开始 转运 了。
50c. 穷书生 把 仙女 在 那年初 娶回家 之后， 他的 人生 便 开始 转运 了。
50d. 仙女 把 穷书生 在 那年初 娶回家 之后， 他的 人生 便 开始 转运 了。
51a. 那天下午， 猫咪 把 老鼠 吃掉了 之后， 挺着 胀胀的 肚子 回到 院子里。
51b. 那天下午， 老鼠 把 猫咪 吃掉了 之后， 挺着 胀胀的 肚子 回到 院子里。
51c. 猫咪 把 老鼠 在 那天下午 吃掉了 之后， 挺着 胀胀的 肚子 回到 院子里。
51d. 老鼠 把 猫咪 在 那天下午 吃掉了 之后， 挺着 胀胀的 肚子 回到 院子里。
52a. 那年初， 那个叛徒 把 群众 出卖了 一次， 大家 就 不再 愿意 相信 他。
52b. 那年初， 群众 把 那个叛徒 出卖了 一次， 大家 就 不再 愿意 相信 他。
52c. 那个叛徒 把 群众 在 那年初 出卖了 一次， 大家 就 不再 愿意 相信 他。
52d. 群众 把 那个叛徒 在 那年初 出卖了 一次， 大家 就 不再 愿意 相信 他。
53a. 上星期， 女歌手 把 狗仔队 甩了 很远， 成功地 避开了 他们。
53b. 上星期， 狗仔队 把 女歌手 甩了 很远， 成功地 避开了 他们。
53c. 女歌手 把 狗仔队 在 上星期 甩了 很远， 成功地 避开了 他们。
53d. 狗仔队 把 女歌手 在 上星期 甩了 很远， 成功地 避开了 他们。
54a. 那天下午， 爷爷 把 孙子 抱在 怀里 不肯 放手， 生怕 别人 抢了 似的。
54b. 那天下午， 孙子 把 爷爷 抱在 怀里 不肯 放手， 生怕 别人 抢了 似的。
54c. 爷爷 把 孙子 在 那天下午 抱在 怀里 不肯 放手， 生怕 别人 抢了 似的。
54d. 孙子 把 爷爷 在 那天下午 抱在 怀里 不肯 放手， 生怕 别人 抢了 似的。
55a. 那天早上， 那条缉毒犬 把 那个毒贩 找出来 之后 显得 十分 神气。
55b. 那天早上， 那个毒贩 把 那条缉毒犬 找出来 之后 显得 十分 神气。
55c. 那条缉毒犬 把 那个毒贩 在 那天早上 找出来 之后 显得 十分 神气。
55d. 那个毒贩 把 那条缉毒犬 在 那天早上 找出来 之后 显得 十分 神气。
56a. 去年初， 民众 把 暴君 推下台 之后， 通过 讨论 推选 新的 君主。
56b. 去年初， 暴君 把 民众 推下台 之后， 通过 讨论 推选 新的 君主。
56c. 民众 把 暴君 在 去年初 推下台 之后， 通过 讨论 推选 新的 君主。
56d. 暴君 把 民众 在 去年初 推下台 之后， 通过 讨论 推选 新的 君主。
57a. 昨天晚上， 熊猫妈妈 把 小熊猫 抱在 怀里， 不让 饲养人员 靠近。
57b. 昨天晚上， 小熊猫 把 熊猫妈妈 抱在 怀里， 不让 饲养人员 靠近。
57c. 熊猫妈妈 把 小熊猫 在 昨天晚上 抱在 怀里， 不让 饲养人员 靠近。
57d. 小熊猫 把 熊猫妈妈 在 昨天晚上 抱在 怀里， 不让 饲养人员 靠近。
58a. 今天早上， 鹦鹉 把 老刘 逗了 一番， 使 他 哈哈大笑。
58b. 今天早上， 老刘 把 鹦鹉 逗了 一番， 使 他 哈哈大笑。
58c. 鹦鹉 把 老刘 在 今天早上 逗了 一番， 使 他 哈哈大笑。
58d. 老刘 把 鹦鹉 在 今天早上 逗了 一番， 使 他 哈哈大笑。
59a. 那天下午， 东北虎 把 盗猎者 咬了 一口， 然后 顺利 逃避开 了。
59b. 那天下午， 盗猎者 把 东北虎 咬了 一口， 然后 顺利 逃避开 了。
59c. 东北虎 把 盗猎者 在 那天下午 咬了 一口， 然后 顺利 逃避开 了。

59d. 盗猎者 把 东北虎 在 那天下午 咬了 一口， 然后 顺利 逃避开 了。
60a. 上星期， 人们 把 鲨鱼 放回 海中， 看着 它 慢慢地 游走。
60b. 上星期， 鲨鱼 把 人们 放回 海中， 看着 它 慢慢地 游走。
60c. 人们 把 鲨鱼 在 上星期 放回 海中， 看着 它 慢慢地 游走。
60d. 鲨鱼 把 人们 在 上星期 放回 海中， 看着 它 慢慢地 游走。
61a. 昨天下午， 这个学生 把 邓老师 气坏了 却 不肯 道歉。
61b. 昨天下午， 邓老师 把 这个学生 气坏了 却 不肯 道歉。
61c. 这个学生 把 邓老师 在 昨天下午 气坏了 却 不肯 道歉。
61d. 邓老师 把 这个学生 在 昨天下午 气坏了 却 不肯 道歉。
62a. 昨天早上， 贵妇 把 波斯猫 抱在 手里 仔细 观察， 看看 它 有没有 生病。
62b. 昨天早上， 波斯猫 把 贵妇 抱在 手里 仔细 观察， 看看 它 有没有 生病。
62c. 贵妇 把 波斯猫 在 昨天早上 抱在 手里 仔细 观察， 看看 它 有没有 生病。
62d. 波斯猫 把 贵妇 在 昨天早上 抱在 手里 仔细 观察， 看看 它 有没有 生病。
63a. 去年五月， 钟老板 把 这群员工 开除了 出去， 并且 没有给 任何的 额外 补贴。
63b. 去年五月， 这群员工 把 钟老板 开除了 出去， 并且 没有给 任何的 额外 补贴。
63c. 钟老板 把 这群员工 在 去年五月 开除了 出去， 并且 没有给 任何的 额外 补贴。
63d. 这群员工 把 钟老板 在 去年五月 开除了 出去， 并且 没有给 任何的 额外 补贴。
64a. 昨天晚上， 可疑人物 把 门卫 打倒 在地上， 然后 快速地 跑了 进去。
64b. 昨天晚上， 门卫 把 可疑人物 打倒 在地上， 然后 快速地 跑了 进去。
64c. 可疑人物 把 门卫 在 昨天晚上 打倒 在地上， 然后 快速地 跑了 进去。
64d. 门卫 把 可疑人物 在 昨天晚上 打倒 在地上， 然后 快速地 跑了 进去。
65a. 昨天晚上， 小潘 把 那个 女孩 送回家 然后 自己 再 打车 回家。
65b. 昨天晚上， 那个 女孩 把 小潘 送回家 然后 自己 再 打车 回家。
65c. 小潘 把 那个 女孩 在 昨天晚上 送回家 然后 自己 再 打车 回家。
65d. 那个 女孩 把 小潘 在 昨天晚上 送回家 然后 自己 再 打车 回家。
66a. 那天早上， 皇后 把 白雪公主 毒害了 以后 便 以为 自己 是 世上 最美的人。
66b. 那天早上， 白雪公主 把 皇后 毒害了 以后 便 以为 自己 是 世上 最美的人。
66c. 皇后 把 白雪公主 在 那天早上 毒害了 以后 便 以为 自己 是 世上 最美的人。
66d. 白雪公主 把 皇后 在 那天早上 毒害了 以后 便 以为 自己 是 世上 最美的人。
67a. 昨天下午， 恐怖分子 把 那群旅客 劫持了 一个 小时， 并 要求 跟 警方 谈条件。
67b. 昨天下午， 那群旅客 把 恐怖分子 劫持了 一个 小时， 并 要求 跟 警方 谈条件。
67c. 恐怖分子 把 那群旅客 在 昨天下午 劫持了 一个 小时， 并 要求 跟 警方 谈条件。
67d. 那群旅客 把 恐怖分子 在 昨天下午 劫持了 一个 小时， 并 要求 跟 警方 谈条件。
68a. 昨天晚上， 护士 把 伤兵 治好 以后 叮嘱 他 要 多休息 。
68b. 昨天晚上， 伤兵 把 护士 治好 以后 叮嘱 他 要 多休息 。
68c. 护士 把 伤兵 在 昨天晚上 治好 以后 叮嘱 他 要 多休息 。
68d. 伤兵 把 护士 在 昨天晚上 治好 以后 叮嘱 他 要 多休息 。
69a. 昨天下午， 那位顾客 把 售货员 骂了 一顿， 并且 要求 他 鞠躬 道歉。
69b. 昨天下午， 售货员 把 那位顾客 骂了 一顿， 并且 要求 他 鞠躬 道歉。
69c. 那位顾客 把 售货员 在 昨天下午 骂了 一顿， 并且 要求 他 鞠躬 道歉。
69d. 售货员 把 那位顾客 在 昨天下午 骂了 一顿， 并且 要求 他 鞠躬 道歉。
70a. 去年六月， 那位商人 把 高官 贿赂了 以后 得到 不少的 好处。
70b. 去年六月， 高官 把 那位商人 贿赂了 以后 得到 不少的 好处。
70c. 那位商人 把 高官 在 去年六月 贿赂了 以后 得到 不少的 好处。
70d. 高官 把 那位商人 在 去年六月 贿赂了 以后 得到 不少的 好处。
71a. 今天早上， 那只小狗 把 那个主人 逗乐了 并且 使 他的 烦恼 一扫而空。
71b. 今天早上， 那个主人 把 那只小狗 逗乐了 并且 使 他的 烦恼 一扫而空。
71c. 那只小狗 把 那个主人 在 今天早上 逗乐了 并且 使 他的 烦恼 一扫而空。
71d. 那个主人 把 那只小狗 在 今天早上 逗乐了 并且 使 他的 烦恼 一扫而空。

72a. 今天早上， 那一网鱼 把 渔民 乐得 开怀， 实在 值得 庆祝 一下。
72b. 今天早上， 渔民 把 那一网鱼 乐得 开怀， 实在 值得 庆祝 一下。
72c. 那一网鱼 把 渔民 在 今天早上 乐得 开怀， 实在 值得 庆祝 一下。
72d. 渔民 把 那一网鱼 在 今天早上 乐得 开怀， 实在 值得 庆祝 一下。
73a. 上星期， 那个村民 把 那条鱼 吃掉了 以后 便 生病了。
73b. 上星期， 那条鱼 把 那个村民 吃掉了 以后 便 生病了。
73c. 那个村民 把 那条鱼 在 上星期 吃掉了 以后 便 生病了。
73d. 那条鱼 把 那个村民 在 上星期 吃掉了 以后 便 生病了。
74a. 昨天晚上， 雪橇狗 把 猎人 带回家 以后 便 得到 奖赏了。
74b. 昨天晚上， 猎人 把 雪橇狗 带回家 以后 便 得到 奖赏了。
74c. 雪橇狗 把 猎人 在 昨天晚上 带回家 以后 便 得到 奖赏了。
74d. 猎人 把 雪橇狗 在 昨天晚上 带回家 以后 便 得到 奖赏了。
75a. 去年春天， 小狗 把 那对夫妇 咬伤了 不止 一次， 这让 主人 感到 非常 头疼。
75b. 去年春天， 那对夫妇 把 小狗 咬伤了 不止 一次， 这让 主人 感到 非常 头疼。
75c. 小狗 把 那对夫妇 在 去年春天 咬伤了 不止 一次， 这让 主人 感到 非常 头疼。
75d. 那对夫妇 把 小狗 在 去年春天 咬伤了 不止 一次， 这让 主人 感到 非常 头疼。
76a. 去年春天， 总司令 把 部队 调遣 到边境， 以 防守 敌人的 进攻。
76b. 去年春天， 部队 把 总司令 调遣 到边境， 以 防守 敌人的 进攻。
76c. 总司令 把 部队 在 去年春天 调遣 到边境， 以 防守 敌人的 进攻。
76d. 部队 把 总司令 在 去年春天 调遣 到边境， 以 防守 敌人的 进攻。
77a. 那年初， 皇帝 把 贪官 杀了 之后， 任命了 今年的 状元 为 新的 官员。
77b. 那年初， 贪官 把 皇帝 杀了 之后， 任命了 今年的 状元 为 新的 官员。
77c. 皇帝 把 贪官 在 那年初 杀了 之后， 任命了 今年的 状元 为 新的 官员。
77d. 贪官 把 皇帝 在 那年初 杀了 之后， 任命了 今年的 状元 为 新的 官员。
78a. 昨天下午， 那个农夫 把 乌鸦 赶了 出去， 然后 在农场上 立了 一个 稻草人。
78b. 昨天下午， 乌鸦 把 那个农夫 赶了 出去， 然后 在农场上 立了 一个 稻草人。
78c. 那个农夫 把 乌鸦 在 昨天下午 赶了 出去， 然后 在农场上 立了 一个 稻草人。
78d. 乌鸦 把 那个农夫 在 昨天下午 赶了 出去， 然后 在农场上 立了 一个 稻草人。
79a. 昨天下午， 新演员 把 导演 气坏了 却 完全 没发现。
79b. 昨天下午， 导演 把 新演员 气坏了 却 完全 没发现。
79c. 新演员 把 导演 在 昨天下午 气坏了 却 完全 没发现。
79d. 导演 把 新演员 在 昨天下午 气坏了 却 完全 没发现。
80a. 昨天晚上， 演员 把 国家元首 逗笑了 而且 获得了 不少的 赞美。
80b. 昨天晚上， 国家元首 把 演员 逗笑了 而且 获得了 不少的 赞美。
80c. 演员 把 国家元首 在 昨天晚上 逗笑了 而且 获得了 不少的 赞美。
80d. 国家元首 把 演员 在 昨天晚上 逗笑了 而且 获得了 不少的 赞美。
81a. 昨天早上， 裁判 把 那个参赛者 罚下场 以后 便 宣布 比赛 暂停。
81b. 昨天早上， 那个参赛者 把 裁判 罚下场 以后 便 宣布 比赛 暂停。
81c. 裁判 把 那个参赛者 在 昨天早上 罚下场 以后 便 宣布 比赛 暂停。
81d. 那个参赛者 把 裁判 在 昨天早上 罚下场 以后 便 宣布 比赛 暂停。
82a. 昨天下午， 大象 把 小鹿 吓得 发抖 之后 瘫倒在 地上 不敢动。
82b. 昨天下午， 小鹿 把 大象 吓得 发抖 之后 瘫倒在 地上 不敢动。
82c. 大象 把 小鹿 在 昨天下午 吓得 发抖 之后 瘫倒在 地上 不敢动。
82d. 小鹿 把 大象 在 昨天下午 吓得 发抖 之后 瘫倒在 地上 不敢动。
83a. 昨天下午， 那只猪 把 屠夫 撞倒 在地上， 然后 拼命地 往外面 跑。
83b. 昨天下午， 屠夫 把 那只猪 撞倒 在地上， 然后 拼命地 往外面 跑。
83c. 那只猪 把 屠夫 在 昨天下午 撞倒 在地上， 然后 拼命地 往外面 跑。
83d. 屠夫 把 那只猪 在 昨天下午 撞倒 在地上， 然后 拼命地 往外面 跑。
84a. 上星期， 那位师傅 把 学徒 骂了 一顿， 并且 惩罚 他 立刻 重新 做一遍。

84b. 上星期， 学徒 把 那位师傅 骂了 一顿， 并且 惩罚 他 立刻 重新 做一遍。
84c. 那位师傅 把 学徒 在 上星期 骂了 一顿， 并且 惩罚 他 立刻 重新 做一遍。
84d. 学徒 把 那位师傅 在 上星期 骂了 一顿， 并且 惩罚 他 立刻 重新 做一遍。
85a. 那年春天， 孙悟空 把 唐僧 送到 西天 取经， 一路上 他们 经历了 重重困难。
85b. 那年春天， 唐僧 把 孙悟空 送到 西天 取经， 一路上 他们 经历了 重重困难。
85c. 孙悟空 把 唐僧 在 那年春天 送到 西天 取经， 一路上 他们 经历了 重重困难。
85d. 唐僧 把 孙悟空 在 那年春天 送到 西天 取经， 一路上 他们 经历了 重重困难。
86a. 今天早上， 那群骏马 把 那个画家 吸引 住了， 他 真想 立刻 坐下来 作画。
86b. 今天早上， 那个画家 把 那群骏马 吸引 住了， 他 真想 立刻 坐下来 作画。
86c. 那群骏马 把 那个画家 在 今天早上 吸引 住了， 他 真想 立刻 坐下来 作画。
86d. 那个画家 把 那群骏马 在 今天早上 吸引 住了， 他 真想 立刻 坐下来 作画。
87a. 今天早上， 那条大狗 把 那婴孩 吓哭了 以后 便 没趣地 走开了。
87b. 今天早上， 那婴孩 把 那条大狗 吓哭了 以后 便 没趣地 走开了。
87c. 那条大狗 把 那婴孩 在 今天早上 吓哭了 以后 便 没趣地 走开了。
87d. 那婴孩 把 那条大狗 在 今天早上 吓哭了 以后 便 没趣地 走开了。
88a. 去年春天， 经理人 把 新乐队 解散了 以后 让 他们 单独 发展。
88b. 去年春天， 新乐队 把 经理人 解散了 以后 让 他们 单独 发展。
88c. 经理人 把 新乐队 在 去年春天 解散了 以后 让 他们 单独 发展。
88d. 新乐队 把 经理人 在 去年春天 解散了 以后 让 他们 单独 发展。
89a. 去年底， 那模特儿 把 总编辑 迷倒了 以后 便 立即 成为 封面女郎。
89b. 去年底， 总编辑 把 那模特儿 迷倒了 以后 便 立即 成为 封面女郎。
89c. 那模特儿 把 总编辑 在 去年底 迷倒了 以后 便 立即 成为 封面女郎。
89d. 总编辑 把 那模特儿 在 去年底 迷倒了 以后 便 立即 成为 封面女郎。
90a. 昨天早上， 球迷 把 这个球星 围了 起来， 想要 跟他 合照 和 要签名。
90b. 昨天早上， 这个球星 把 球迷 围了 起来， 想要 跟他 合照 和 要签名。
90c. 球迷 把 这个球星 在 昨天早上 围了 起来， 想要 跟他 合照 和 要签名。
90d. 这个球星 把 球迷 在 昨天早上 围了 起来， 想要 跟他 合照 和 要签名。
91a. 昨天晚上， 来宾 把 相声演员 称赞了 一番， 并且 请求 演员们 再 表演 一小段。
91b. 昨天晚上， 相声演员 把 来宾 称赞了 一番， 并且 请求 演员们 再 表演 一小段。
91c. 来宾 把 相声演员 在 昨天晚上 称赞了 一番， 并且 请求 演员们 再 表演 一小段。
91d. 相声演员 把 来宾 在 昨天晚上 称赞了 一番， 并且 请求 演员们 再 表演 一小段。
92a. 今天早上， 妈妈 把 孩子们 带到 院子里， 让他们 在那里 玩耍。
92b. 今天早上， 孩子们 把 妈妈 带到 院子里， 让他们 在那里 玩耍。
92c. 妈妈 把 孩子们 在 今天早上 带到 院子里， 让他们 在那里 玩耍。
92d. 孩子们 把 妈妈 在 今天早上 带到 院子里， 让他们 在那里 玩耍。
93a. 上个月， 老师傅 把 学徒 教会了 之后， 打算 让他 留在 作坊里 继续 帮忙。
93b. 上个月， 学徒 把 老师傅 教会了 之后， 打算 让他 留在 作坊里 继续 帮忙。
93c. 老师傅 把 学徒 在 上个月 教会了 之后， 打算 让他 留在 作坊里 继续 帮忙。
93d. 学徒 把 老师傅 在 上个月 教会了 之后， 打算 让他 留在 作坊里 继续 帮忙。
94a. 昨天下午， 老顾客 把 按摩师 夸奖了 一番， 按摩师 高兴得 笑不拢嘴。
94b. 昨天下午， 按摩师 把 老顾客 夸奖了 一番， 按摩师 高兴得 笑不拢嘴。
94c. 老顾客 把 按摩师 在 昨天下午 夸奖了 一番， 按摩师 高兴得 笑不拢嘴。
94d. 按摩师 把 老顾客 在 昨天下午 夸奖了 一番， 按摩师 高兴得 笑不拢嘴。
95a. 去年春天， 富翁 把 佣人 解雇了 之后 便 立即 请来了 一个 新人 。
95b. 去年春天， 佣人 把 富翁 解雇了 之后 便 立即 请来了 一个 新人 。
95c. 富翁 把 佣人 在 去年春天 解雇了 之后 便 立即 请来了 一个 新人 。
95d. 佣人 把 富翁 在 去年春天 解雇了 之后 便 立即 请来了 一个 新人 。
96a. 那天下午， 那个溺水者 把 救生员 拉下 水里， 由于 水流 太急 他们 难以上岸。
96b. 那天下午， 救生员 把 那个溺水者 拉下 水里， 由于 水流 太急 他们 难以上岸。

96c. 那个溺水者 把 救生员 在 那天下午 拉下 水里， 由于 水流 太急 他们 难以上岸。
96d. 救生员 把 那个溺水者 在 那天下午 拉下 水里， 由于 水流 太急 他们 难以上岸。
97a. 今天早上， 评委 把 得奖者 表扬了 一番， 然后 颁发 奖状 和 奖金。
97b. 今天早上， 得奖者 把 评委 表扬了 一番， 然后 颁发 奖状 和 奖金。
97c. 评委 把 得奖者 在 今天早上 表扬了 一番， 然后 颁发 奖状 和 奖金。
97d. 得奖者 把 评委 在 今天早上 表扬了 一番， 然后 颁发 奖状 和 奖金。
98a. 去年春天， 保姆 把 小孩 照顾 得很好， 黄太太 特别 信任 她。
98b. 去年春天， 小孩 把 保姆 照顾 得很好， 黄太太 特别 信任 她。
98c. 保姆 把 小孩 在 去年春天 照顾 得很好， 黄太太 特别 信任 她。
98d. 小孩 把 保姆 在 去年春天 照顾 得很好， 黄太太 特别 信任 她。
99a. 今天早上， 那个病人 把 董大夫 感谢了 一番， 場面 十分 感人。
99b. 今天早上， 董大夫 把 那个病人 感谢了 一番， 場面 十分 感人。
99c. 那个病人 把 董大夫 在 今天早上 感谢了 一番， 場面 十分 感人。
99d. 董大夫 把 那个病人 在 今天早上 感谢了 一番， 場面 十分 感人。
100a. 去年五月， 军人 把 囚犯 关起来 並且 进行 严密的 监察。
100b. 去年五月， 囚犯 把 军人 关起来 並且 进行 严密的 监察。
100c. 军人 把 囚犯 在 去年五月 关起来 並且 进行 严密的 监察。
100d. 囚犯 把 军人 在 去年五月 关起来 並且 进行 严密的 监察。
101a. 昨天早上， 村民 把 行骗者 打了 一顿， 要 为 自己 出口气。
101b. 昨天早上， 行骗者 把 村民 打了 一顿， 要 为 自己 出口气。
101c. 村民 把 行骗者 在 昨天早上 打了 一顿， 要 为 自己 出口气。
101d. 行骗者 把 村民 在 昨天早上 打了 一顿， 要 为 自己 出口气。
102a. 昨天下午， 那个奸商 把 这群黑工 骗了 过来， 极为 自私。
102b. 昨天下午， 这群黑工 把 那个奸商 骗了 过来， 极为 自私。
102c. 那个奸商 把 这群黑工 在 昨天下午 骗了 过来， 极为 自私。
102d. 这群黑工 把 那个奸商 在 昨天下午 骗了 过来， 极为 自私。
103a. 去年五月， 那个商人 把 工人 辞退了 以后 便 破产了。
103b. 去年五月， 工人 把 那个商人 辞退了 以后 便 破产了。
103c. 那个商人 把 工人 在 去年五月 辞退了 以后 便 破产了。
103d. 工人 把 那个商人 在 去年五月 辞退了 以后 便 破产了。
104a. 去年春天， 海洋学家 把 鲸鱼 研究了 一年多， 却 没有 任何 重大的 发现。
104b. 去年春天， 鲸鱼 把 海洋学家 研究了 一年多， 却 没有 任何 重大的 发现。
104c. 海洋学家 把 鲸鱼 在 去年春天 研究了 一年多， 却 没有 任何 重大的 发现。
104d. 鲸鱼 把 海洋学家 在 去年春天 研究了 一年多， 却 没有 任何 重大的 发现。
105a. 去年春天， 模特儿 把 那个画家 迷住了 并且 欺骗了 他的 感情。
105b. 去年春天， 那个画家 把 模特儿 迷住了 并且 欺骗了 他的 感情。
105c. 模特儿 把 那个画家 在 去年春天 迷住了 并且 欺骗了 他的 感情。
105d. 那个画家 把 模特儿 在 去年春天 迷住了 并且 欺骗了 他的 感情。
106a. 昨天下午， 女演员 把 造型师 骂了 一顿， 完全 不顾 仪态。
106b. 昨天下午， 造型师 把 女演员 骂了 一顿， 完全 不顾 仪态。
106c. 女演员 把 造型师 在 昨天下午 骂了 一顿， 完全 不顾 仪态。
106d. 造型师 把 女演员 在 昨天下午 骂了 一顿， 完全 不顾 仪态。
107a. 昨天早上， 总经理 把 助手 辞退了 以后 便 立即 找到 一个 新的。
107b. 昨天早上， 助手 把 总经理 辞退了 以后 便 立即 找到 一个 新的。
107c. 总经理 把 助手 在 昨天早上 辞退了 以后 便 立即 找到 一个 新的。
107d. 助手 把 总经理 在 昨天早上 辞退了 以后 便 立即 找到 一个 新的。
108a. 今天早上， 选手 把 教练 气得 差点 晕倒了， 气氛 很紧张。
108b. 今天早上， 教练 把 选手 气得 差点 晕倒了， 气氛 很紧张。
108c. 选手 把 教练 在 今天早上 气得 差点 晕倒了， 气氛 很紧张。

108d. 教练 把 选手 在 今天早上 气得 差点 晕倒了， 气氛 很紧张。
109a. 去年初， 那群企鹅 把 摄影师 逗乐了 并且 出现在 新一期 杂志的 封面上。
109b. 去年初， 摄影师 把 那群企鹅 逗乐了 并且 出现在 新一期 杂志的 封面上。
109c. 那群企鹅 把 摄影师 在 去年初 逗乐了 并且 出现在 新一期 杂志的 封面上。
109d. 摄影师 把 那群企鹅 在 去年初 逗乐了 并且 出现在 新一期 杂志的 封面上。
110a. 昨天早上， 董事 把 秘书 辞退了 以后 感到 有一点 后悔。
110b. 昨天早上， 秘书 把 董事 辞退了 以后 感到 有一点 后悔。
110c. 董事 把 秘书 在 昨天早上 辞退了 以后 感到 有一点 后悔。
110d. 秘书 把 董事 在 昨天早上 辞退了 以后 感到 有一点 后悔。
111a. 昨天晚上， 猛虎 把 武松 咬伤了 但却 难逃 一劫。
111b. 昨天晚上， 武松 把 猛虎 咬伤了 但却 难逃 一劫。
111c. 猛虎 把 武松 在 昨天晚上 咬伤了 但却 难逃 一劫。
111d. 武松 把 猛虎 在 昨天晚上 咬伤了 但却 难逃 一劫。
112a. 今天早上， 考生们 把 考官 惹怒了 以后 十分 担心， 全都 不敢 作声。
112b. 今天早上， 考官 把 考生们 惹怒了 以后 十分 担心， 全都 不敢 作声。
112c. 考生们 把 考官 在 今天早上 惹怒了 以后 十分 担心， 全都 不敢 作声。
112d. 考官 把 考生们 在 今天早上 惹怒了 以后 十分 担心， 全都 不敢 作声。
113a. 去年底， 私家侦探 把 那个政客 调查了 一个月， 终于 查到 他的 婚外情。
113b. 去年底， 那个政客 把 私家侦探 调查了 一个月， 终于 查到 他的 婚外情。
113c. 私家侦探 把 那个政客 在 去年底 调查了 一个月， 终于 查到 他的 婚外情。
113d. 那个政客 把 私家侦探 在 去年底 调查了 一个月， 终于 查到 他的 婚外情。
114a. 那天下午， 包青天 把 贪官 杀头 以后 便 严厉地 警告 其他 官员。
114b. 那天下午， 贪官 把 包青天 杀头 以后 便 严厉地 警告 其他 官员。
114c. 包青天 把 贪官 在 那天下午 杀头 以后 便 严厉地 警告 其他 官员。
114d. 贪官 把 包青天 在 那天下午 杀头 以后 便 严厉地 警告 其他 官员。
115a. 昨天晚上， 童工 把 地主 戏弄了 一番， 连累 其他人 也 受苦了。
115b. 昨天晚上， 地主 把 童工 戏弄了 一番， 连累 其他人 也 受苦了。
115c. 童工 把 地主 在 昨天晚上 戏弄了 一番， 连累 其他人 也 受苦了。
115d. 地主 把 童工 在 昨天晚上 戏弄了 一番， 连累 其他人 也 受苦了。
116a. 上星期， 学生们 把 那位教授 气坏了 却 认为 自己 没错。
116b. 上星期， 那位教授 把 学生们 气坏了 却 认为 自己 没错。
116c. 学生们 把 那位教授 在 上星期 气坏了 却 认为 自己 没错。
116d. 那位教授 把 学生们 在 上星期 气坏了 却 认为 自己 没错。
117a. 那些年来， 部落民族 把 铁木真 奉为 英雄， 所有人 都 十分 敬佩 他。
117b. 那些年来， 铁木真 把 部落民族 奉为 英雄， 所有人 都 十分 敬佩 他。
117c. 部落民族 把 铁木真 在 那些年来 奉为 英雄， 所有人 都 十分 敬佩 他。
117d. 铁木真 把 部落民族 在 那些年来 奉为 英雄， 所有人 都 十分 敬佩 他。
118a. 今天早上， 劫匪 把 谈判专家 劫持了 几个小时， 完全 不肯 放人。
118b. 今天早上， 谈判专家 把 劫匪 劫持了 几个小时， 完全 不肯 放人。
118c. 劫匪 把 谈判专家 在 今天早上 劫持了 几个小时， 完全 不肯 放人。
118d. 谈判专家 把 劫匪 在 今天早上 劫持了 几个小时， 完全 不肯 放人。
119a. 那天晚上， 奶奶 把 孙女 抱在 怀里， 耐心地 哄 她 睡觉。
119b. 那天晚上， 孙女 把 奶奶 抱在 怀里， 耐心地 哄 她 睡觉。
119c. 奶奶 把 孙女 在 那天晚上 抱在 怀里， 耐心地 哄 她 睡觉。
119d. 孙女 把 奶奶 在 那天晚上 抱在 怀里， 耐心地 哄 她 睡觉。
120a. 去年八月， 周老板 把 那个毕业生 录用了 以后， 店里 生意 便 越来越好了。
120b. 去年八月， 那个毕业生 把 周老板 录用了 以后， 店里 生意 便 越来越好了。
120c. 周老板 把 那个毕业生 在 去年八月 录用了 以后， 店里 生意 便 越来越好了。
120d. 那个毕业生 把 周老板 在 去年八月 录用了 以后， 店里 生意 便 越来越好了。

Experiment 5

a - Canonical control
b - Role-reversed (Animacy-congruous)

<u>High predictability sentences</u>

1a. 屠夫 把 那只猪 在 上星期 宰了 以后 卖到 很好的 价钱。
1b. 那只猪 把 屠夫 在 上星期 宰了 以后 卖到 很好的 价钱。
2a. 消防员 把 被困的小孩 在 昨天黄昏 解救了 出来， 并 送到了 安全地点。
2b. 被困的小孩 把 消防员 在 昨天黄昏 解救了 出来， 并 送到了 安全地点。
3a. 小丑 把 观众 在 昨天晚上 逗乐了 之后 便 表现得 十分神气。
3b. 观众 把 小丑 在 昨天晚上 逗乐了 之后 便 表现得 十分神气。
4a. 青蛙 把 蚊子 在 转眼间 吃到 肚子里， 并没有 任何 动静。
4b. 蚊子 把 青蛙 在 转眼间 吃到 肚子里， 并没有 任何 动静。
5a. 救生员 把 那个溺水者 在 日落前 救起 到岸上， 成为了 大英雄。
5b. 那个溺水者 把 救生员 在 日落前 救起 到岸上， 成为了 大英雄。
6a. 武松 把 猛虎 在 那年秋天 打死了 以后 成了 大英雄。
6b. 猛虎 把 武松 在 那年秋天 打死了 以后 成了 大英雄。
7a. 那个画家 把 那群骏马 在 去年底 画下来 以后 便 立即 展示 在 画廊中。
7b. 那群骏马 把 那个画家 在 去年底 画下来 以后 便 立即 展示 在 画廊中。
8a. 警犬 把 失踪少女 在 昨天晚上 找回来 并 送到了 警察局。
8b. 失踪少女 把 警犬 在 昨天晚上 找回来 并 送到了 警察局。
9a. 董大夫 把 那个病人 在 去年底 治愈 以后 便 决定 退休了。
9b. 那个病人 把 董大夫 在 去年底 治愈 以后 便 决定 退休了。
10a. 相声演员 把 来宾 在 昨天晚上 逗得 哈哈大笑， 全场 掌声雷动。
10b. 来宾 把 相声演员 在 昨天晚上 逗得 哈哈大笑， 全场 掌声雷动。
11a. 警察 把 人质 在 昨天晚上 解救了 出来， 成了 救人英雄。
11b. 人质 把 警察 在 昨天晚上 解救了 出来， 成了 救人英雄。
12a. 行骗者 把 村民 在 去年初 骗了 以后 便 逃之夭夭了。
12b. 村民 把 行骗者 在 去年初 骗了 以后 便 逃之夭夭了。
13a. 警察 把 小偷 在 那天傍晚 抓了 回警局， 并 教训了 他 一顿。
13b. 小偷 把 警察 在 那天傍晚 抓了 回警局， 并 教训了 他 一顿。
14a. 造型师 把 女演员 在 上星期 打扮得 很美， 获得 不少的 赞赏。
14b. 女演员 把 造型师 在 上星期 打扮得 很美， 获得 不少的 赞赏。
15a. 谈判专家 把 劫匪 在 昨天黄昏 说服 并 解救了 人质。
15b. 劫匪 把 谈判专家 在 昨天黄昏 说服 并 解救了 人质。
16a. 何婆婆 把 这个孤儿 在 去年初 收养后 并 带着 他 离开了 小镇。
16b. 这个孤儿 把 何婆婆 在 去年初 收养后 并 带着 他 离开了 小镇。
17a. 巫师 把 王子 在 转眼间 变成了 青蛙， 世上 只有 小公主 才能 解救 他。
17b. 王子 把 巫师 在 转眼间 变成了 青蛙， 世上 只有 小公主 才能 解救 他。
18a. 摄影师 把 那群企鹅 在 去年底 拍下来 以后 刊登 在 杂志的 封面上。
18b. 那群企鹅 把 摄影师 在 去年底 拍下来 以后 刊登 在 杂志的 封面上。
19a. 那个画家 把 模特儿 在 上星期 描画得 非常 完美， 同学们 都 惊叹不已。
19b. 模特儿 把 那个画家 在 上星期 描画得 非常 完美， 同学们 都 惊叹不已。
20a. 慈善家 把 孤儿 在 那年冬天 领养 回家， 并且 供 他 上学。
20b. 孤儿 把 慈善家 在 那年冬天 领养 回家， 并且 供 他 上学。
21a. 货车司机 把 路人 在 那天下午 撞伤了 以后 立即 送 他 到 医院 包扎。
21b. 路人 把 货车司机 在 那天下午 撞伤了 以后 立即 送 他 到 医院 包扎。

22a. 这只花猫 把 那只麻雀 在 昨天黄昏 吃得 连 根 羽毛 都 没留下。
22b. 那只麻雀 把 这只花猫 在 昨天黄昏 吃得 连 根 羽毛 都 没留下。
23a. 专家 把 传染病人 在 上周末 隔离了 起来， 防止 病菌 传播。
23b. 传染病人 把 专家 在 上周末 隔离了 起来， 防止 病菌 传播。
24a. 奸商 把 投资者 在 那年冬天 骗得 血本无归， 转眼间 逃去无踪。
24b. 投资者 把 奸商 在 那年冬天 骗得 血本无归， 转眼间 逃去无踪。
25a. 牧羊人 把 羊群 在 今天早上 赶到 草原上 吃草， 不慌不忙的。
25b. 羊群 把 牧羊人 在 今天早上 赶到 草原上 吃草， 不慌不忙的。
26a. 大狗 把 孩子 在 上星期 咬伤了 以后 一天 都 不肯 进食。
26b. 孩子 把 大狗 在 上星期 咬伤了 以后 一天 都 不肯 进食。
27a. 奸商 把 市民 在 那年初 欺骗了 以后 便 臭名远播了。
27b. 市民 把 奸商 在 那年初 欺骗了 以后 便 臭名远播了。
28a. 灭虫专家 把 那些老鼠 在 上周末 消灭 干净， 并且 放了 灭鼠药 以 预防。
28b. 那些老鼠 把 灭虫专家 在 上周末 消灭 干净， 并且 放了 灭鼠药 以 预防。
29a. 针炙师 把 女病人 在 去年初 治好 以后 心里 十分 满足。
29b. 女病人 把 针炙师 在 去年初 治好 以后 心里 十分 满足。
30a. 凶犯 把 目击证人 在 去年底 杀死 在 沙发上， 便 逃去无踪了。
30b. 目击证人 把 凶犯 在 去年底 杀死 在 沙发上， 便 逃去无踪了。
31a. 蚂蚁 把 大象 在 那天中午 咬了 一口， 可是 它 完全 没 感觉到。
31b. 大象 把 蚂蚁 在 那天中午 咬了 一口， 可是 它 完全 没 感觉到。
32a. 那个农夫 把 豺狼 在 那天下午 打死了 以后 便 向 邻居 炫耀一番。
32b. 豺狼 把 那个农夫 在 那天下午 打死了 以后 便 向 邻居 炫耀一番。
33a. 猎人 把 飞鸟 在 昨天早上 打下来 以后 便 立即 卖到 很好的 价钱。
33b. 飞鸟 把 猎人 在 昨天早上 打下来 以后 便 立即 卖到 很好的 价钱。
34a. 驯兽师 把 狮子 在 去年初 驯服得 像 一只小猫 一样 乖巧。
34b. 狮子 把 驯兽师 在 去年初 驯服得 像 一只小猫 一样 乖巧。
35a. 法官 把 那个犯人 在 去年秋天 判死刑 以后 听到 很多 反对的 声音。
35b. 那个犯人 把 法官 在 去年秋天 判死刑 以后 听到 很多 反对的 声音。
36a. 小明 把 萤火虫 在 那天早上 放飞 到 大自然， 心里 很高兴。
36b. 萤火虫 把 小明 在 那天早上 放飞 到 大自然， 心里 很高兴。
37a. 刑警 把 嫌犯 在 昨天晚上 抓住了 并且 立即 带回 警局。
37b. 嫌犯 把 刑警 在 昨天晚上 抓住了 并且 立即 带回 警局。
38a. 战友 把 烈士 在 那年秋天 埋葬 在 山坡上， 并 在 坟前 放上 鲜花。
38b. 烈士 把 战友 在 那年秋天 埋葬 在 山坡上， 并 在 坟前 放上 鲜花。
39a. 心理医生 把 精神病人 在 去年夏天 治好 以后 仍然 很关心 他的 状况。
39b. 精神病人 把 心理医生 在 去年夏天 治好 以后 仍然 很关心 他的 状况。
40a. 兽医 把 那条小狗 在 上星期 治好后 将 它 送回 主人 怀里。
40b. 那条小狗 把 兽医 在 上星期 治好后 将 它 送回 主人 怀里。
41a. 消防员 把 那个伤者 在 那天傍晚 救了 出 火海， 自己 却 受了 重伤。
41b. 那个伤者 把 消防员 在 那天傍晚 救了 出 火海， 自己 却 受了 重伤。
42a. 老陈 把 河豚 在 上个月 放回 到 河里， 并 没有 告诉 任何人。
42b. 河豚 把 老陈 在 上个月 放回 到 河里， 并 没有 告诉 任何人。
43a. 魔术师 把 白鸽 在 昨天晚上 变走 以后 全场观众 都 很兴奋。
43b. 白鸽 把 魔术师 在 昨天晚上 变走 以后 全场观众 都 很兴奋。
44a. 老张 把 害虫 在 上周末 消灭 干净了， 觉得 很有 成就感。
44b. 害虫 把 老张 在 上周末 消灭 干净了， 觉得 很有 成就感。
45a. 警方 把 疑犯 在 昨天傍晚 抓住了 以后 立即 带往 派出所。
45b. 疑犯 把 警方 在 昨天傍晚 抓住了 以后 立即 带往 派出所。
46a. 那个恶霸 把 小弟弟 在 那年夏天 打了 一顿， 其他 孩子 都 怕得 不敢作声。

46b. 小弟弟 把 那个恶霸 在 那年夏天 打了 一顿， 其他 孩子 都 怕得 不敢作声。
47a. 这匹野马 把 骑师 在 上星期 摔了 下来， 而且 踩了 一脚。
47b. 骑师 把 这匹野马 在 上星期 摔了 下来， 而且 踩了 一脚。
48a. 老鹰 把 野鸡 在 昨天下午 吃掉了 以后 便 飞走了。
48b. 野鸡 把 老鹰 在 昨天下午 吃掉了 以后 便 飞走了。
49a. 刽子手 把 死囚 在 那天黄昏 杀死了 以后 便 严肃地 离开了。
49b. 死囚 把 刽子手 在 那天黄昏 杀死了 以后 便 严肃地 离开了。
50a. 老伯伯 把 金鱼 在 今天早上 放进 鱼缸里， 并 换上 新鲜的 水。
50b. 金鱼 把 老伯伯 在 今天早上 放进 鱼缸里， 并 换上 新鲜的 水。
51a. 警察 把 灾民 在 上星期 安置 到了 安全的 地方， 十分 尽责。
51b. 灾民 把 警察 在 上星期 安置 到了 安全的 地方， 十分 尽责。
52a. 那些歹徒 把 人质 在 去年五月 杀害了 以后 便 与 警方 展开 枪战。
52b. 人质 把 那些歹徒 在 去年五月 杀害了 以后 便 与 警方 展开 枪战。
53a. 坏学生 把 小弟弟 在 那年底 带坏了 并且 经常 指使 他 去 偷东西。
53b. 小弟弟 把 坏学生 在 那年底 带坏了 并且 经常 指使 他 去 偷东西。
54a. 蟑螂 把 黄老板 在 昨天晚上 吓得 连 饭 都 不敢 吃了。
54b. 黄老板 把 蟑螂 在 昨天晚上 吓得 连 饭 都 不敢 吃了。
55a. 熊猫宝宝 把 饲养人员 在 昨天下午 逗得 笑弯了腰， 心情 豁然 开朗。
55b. 饲养人员 把 熊猫宝宝 在 昨天下午 逗得 笑弯了腰， 心情 豁然 开朗。
56a. 那个毒贩 把 那条缉毒犬 在 去年初 杀死 在 路旁， 警方 查了 很久 才 破案。
56b. 那条缉毒犬 把 那个毒贩 在 去年初 杀死 在 路旁， 警方 查了 很久 才 破案。
57a. 皇帝 把 佟妃 在 那年底 打入冷宫 以后 继续 花天酒地。
57b. 佟妃 把 皇帝 在 那年底 打入冷宫 以后 继续 花天酒地。
58a. 仓鼠 把 男孩 在 昨天早上 吓倒 在 地上， 不知道 溜 到 哪里 去了。
58b. 男孩 把 仓鼠 在 昨天早上 吓倒 在 地上， 不知道 溜 到 哪里 去了。
59a. 吴管家 把 那只流浪狗 在 去年四月 收留 下来 并且 很用心 照顾 它。
59b. 那只流浪狗 把 吴管家 在 去年四月 收留 下来 并且 很用心 照顾 它。
60a. 斗牛勇士 把 蛮牛 在 昨天下午 制服了 之后 便 拔刀 杀了 它。
60b. 蛮牛 把 斗牛勇士 在 昨天下午 制服了 之后 便 拔刀 杀了 它。

Low predictability sentences

61a. 那群旅客 把 恐怖分子 在 昨天黄昏 制服了 以后 大家 才 稍微 松一口气。
61b. 恐怖分子 把 那群旅客 在 昨天黄昏 制服了 以后 大家 才 稍微 松一口气。
62a. 铁木真 把 部落民族 在 那年秋天 聚集 起来， 宣布 重要的 消息。
62b. 部落民族 把 铁木真 在 那年秋天 聚集 起来， 宣布 重要的 消息。
63a. 那位顾客 把 售货员 在 昨天下午 投诉了 一番， 场面 十分 尴尬。
63b. 售货员 把 那位顾客 在 昨天下午 投诉了 一番， 场面 十分 尴尬。
64a. 人们 把 鲨鱼 在 去年六月 捕获了 之后 就 开始 庆祝。
64b. 鲨鱼 把 人们 在 去年六月 捕获了 之后 就 开始 庆祝。
65a. 巡逻队 把 藏羚羊 在 去年八月 送回 保护区， 并且 好好地 照顾。
65b. 藏羚羊 把 巡逻队 在 去年八月 送回 保护区， 并且 好好地 照顾。
66a. 贪官 把 包青天 在 那年春天 陷害得 苦不堪言， 激起了 百姓的 愤怒。
66b. 包青天 把 贪官 在 那年春天 陷害得 苦不堪言， 激起了 百姓的 愤怒。
67a. 狱卒 把 那个囚犯 在 去年底 放走了 之后 还是 决定 自首了。
67b. 那个囚犯 把 狱卒 在 去年底 放走了 之后 还是 决定 自首了。
68a. 得奖者 把 评委 在 昨天晚上 感动了 还 流下了 热泪。
68b. 评委 把 得奖者 在 昨天晚上 感动了 还 流下了 热泪。
69a. 导演 把 新演员 在 去年夏天 培训了 一番， 十分 用心。

69b. 新演员 把 导演 在 去年夏天 培训了 一番， 十分 用心。
70a. 猎人 把 雪橇狗 在 昨天早上 喂饱了 以后 便 出外 打猎了。
70b. 雪橇狗 把 猎人 在 昨天早上 喂饱了 以后 便 出外 打猎了。
71a. 秘书 把 董事 在 去年初 迷惑得 神魂颠倒了， 得到 不少 好处。
71b. 董事 把 秘书 在 去年初 迷惑得 神魂颠倒了， 得到 不少 好处。
72a. 经理人 把 新乐队 在 上个月 包装了 一番， 并 开始 着手 准备 出专辑。
72b. 新乐队 把 经理人 在 上个月 包装了 一番， 并 开始 着手 准备 出专辑。
73a. 佣人 把 富翁 在 这几年来 服侍得 很不错， 最后 却 被 赶出去了。
73b. 富翁 把 佣人 在 这几年来 服侍得 很不错， 最后 却 被 赶出去了。
74a. 皇后 把 宫女 在 那年初 处罚了 一顿， 并且 公开了 她的 恶行。
74b. 宫女 把 皇后 在 那年初 处罚了 一顿， 并且 公开了 她的 恶行。
75a. 贵妇 把 波斯猫 在 去年三月 买了 回家 为 自己 解闷。
75b. 波斯猫 把 贵妇 在 去年三月 买了 回家 为 自己 解闷。
76a. 何太太 把 搬运工人 在 上星期 请来 帮忙， 一点 也 不吝啬。
76b. 搬运工人 把 何太太 在 上星期 请来 帮忙， 一点 也 不吝啬。
77a. 工人 把 那个商人 在 去年底 告上 法庭， 誓要 讨回 公道。
77b. 那个商人 把 工人 在 去年底 告上 法庭， 誓要 讨回 公道。
78a. 皇后 把 白雪公主 在 那年夏天 骗了 过来， 并且 毒害了。
78b. 白雪公主 把 皇后 在 那年夏天 骗了 过来， 并且 毒害了。
79a. 群众 把 那个叛徒 在 那天傍晚 揪出来 毒打， 毫不 留手。
79b. 那个叛徒 把 群众 在 那天傍晚 揪出来 毒打， 毫不 留手。
80a. 老太太 把 猫咪 在 上周末 丢了 在 公园， 使 家人 十分 担心。
80b. 猫咪 把 老太太 在 上周末 丢了 在 公园， 使 家人 十分 担心。
81a. 护士 把 伤兵 在 昨天下午 包扎 妥当 并且 扶到 病床上。
81b. 伤兵 把 护士 在 昨天下午 包扎 妥当 并且 扶到 病床上。
82a. 小孩 把 保姆 在 那天下午 累坏了 之后， 就 跑去 找 爷爷 玩了。
82b. 保姆 把 小孩 在 那天下午 累坏了 之后， 就 跑去 找 爷爷 玩了。
83a. 地主 把 童工 在 那年秋天 卖给了 人贩， 从中 谋取 暴利。
83b. 童工 把 地主 在 那年秋天 卖给了 人贩， 从中 谋取 暴利。
84a. 妈妈 把 新生儿 在 那年冬天 遗弃了 在 教堂 门口， 心痛 欲绝。
84b. 新生儿 把 妈妈 在 那年冬天 遗弃了 在 教堂 门口， 心痛 欲绝。
85a. 熊猫妈妈 把 小熊猫 在 昨天晚上 生了 下来， 大家 都 为此 感到 很高兴。
85b. 小熊猫 把 熊猫妈妈 在 昨天晚上 生了 下来， 大家 都 为此 感到 很高兴。
86a. 乞丐 把 吴管家 在 那天早上 缠住了 不愿走， 希望 能 讨得 几块钱。
86b. 吴管家 把 乞丐 在 那天早上 缠住了 不愿走， 希望 能 讨得 几块钱。
87a. 那个奸商 把 这群黑工 在 那年冬天 藏在 小黑屋里， 让 他们 干累活。
87b. 这群黑工 把 那个奸商 在 那年冬天 藏在 小黑屋里， 让 他们 干累活。
88a. 乌鸦 把 那个农夫 在 今天早上 惹火了 以后 还 不走， 真的 不知死活。
88b. 那个农夫 把 乌鸦 在 今天早上 惹火了 以后 还 不走， 真的 不知死活。
89a. 地方官 把 首相 在 去年初 贿赂了 以后 变得 有恃无恐。
89b. 首相 把 地方官 在 去年初 贿赂了 以后 变得 有恃无恐。
90a. 学徒 把 那位师傅 在 这几年来 超越了 以后 就 得意忘形了。
90b. 那位师傅 把 学徒 在 这几年来 超越了 以后 就 得意忘形了。
91a. 小弟弟 把 狮子 在 昨天下午 吵醒了 以后 害怕得 只懂 躲在 爸爸 身后。
91b. 狮子 把 小弟弟 在 昨天下午 吵醒了 以后 害怕得 只懂 躲在 爸爸 身后。
92a. 那个毕业生 把 周老板 在 上个月 恭维了 一番 就 得到了 录用。
92b. 周老板 把 那个毕业生 在 上个月 恭维了 一番 就 得到了 录用。
93a. 按摩师 把 老顾客 在 上星期 得罪了 以后 连忙 道歉。
93b. 老顾客 把 按摩师 在 上星期 得罪了 以后 连忙 道歉。

94a. 高官 把 那位商人 在 去年底 制裁了 以后 大家 都 很高兴。
94b. 那位商人 把 高官 在 去年底 制裁了 以后 大家 都 很高兴。
95a. 将军 把 士兵 在 去年十月 提拔 成为 第九小队的 组长。
95b. 士兵 把 将军 在 去年十月 提拔 成为 第九小队的 组长。
96a. 小潘 把 那个女孩 在 去年八月 娶了 回家 当 媳妇， 还 请 大家 吃了 一顿饭。
96b. 那个女孩 把 小潘 在 去年八月 娶了 回家 当 媳妇， 还 请 大家 吃了 一顿饭。
97a. 私家侦探 把 那个政客 在 上个月 揭发了 并且 刊登了 照片 作 证据。
97b. 那个政客 把 私家侦探 在 上个月 揭发了 并且 刊登了 照片 作 证据。
98a. 可疑人物 把 门卫 在 那天晚上 引开 之后 就 盗走了 宝物。
98b. 门卫 把 可疑人物 在 那天晚上 引开 之后 就 盗走了 宝物。
99a. 暴君 把 民众 在 这几年来 统治得 一塌糊涂， 民不聊生。
99b. 民众 把 暴君 在 这几年来 统治得 一塌糊涂， 民不聊生。
100a. 演员 把 国家元首 在 上星期 讽刺了 一番， 十分 到位。
100b. 国家元首 把 演员 在 上星期 讽刺了 一番， 十分 到位。
101a. 考生们 把 考官 在 去年底 收买 起来了， 真是 难以置信。
101b. 考官 把 考生们 在 去年底 收买 起来了， 真是 难以置信。
102a. 老刘 把 鹦鹉 在 那年夏天 训练了 好一段 时间， 但是 还未 成功。
102b. 鹦鹉 把 老刘 在 那年夏天 训练了 好一段 时间， 但是 还未 成功。
103a. 钟老板 把 这群员工 在 今天早上 奖励了 一番， 这 让 大家 都 很是意外。
103b. 这群员工 把 钟老板 在 今天早上 奖励了 一番， 这 让 大家 都 很是意外。
104a. 那个主人 把 那只小狗 在 上周末 打扮了 一番， 然后 才 开心地 出门。
104b. 那只小狗 把 那个主人 在 上周末 打扮了 一番， 然后 才 开心地 出门。
105a. 那个村民 把 那条鱼 在 上星期 送给了 村长， 希望 討得 他的 歡心。
105b. 那条鱼 把 那个村民 在 上星期 送给了 村长， 希望 討得 他的 歡心。
106a. 老作家 把 新人 在 去年春天 捧红了 以后 便 退隐 文坛了。
106b. 新人 把 老作家 在 去年春天 捧红了 以后 便 退隐 文坛了。
107a. 教练 把 选手 在 今天早上 鼓励了 一番 之后 比赛 就 开始了。
107b. 选手 把 教练 在 今天早上 鼓励了 一番 之后 比赛 就 开始了。
108a. 海洋学家 把 鲸鱼 在 去年底 解剖了 以后 写了 一份报告。
108b. 鲸鱼 把 海洋学家 在 去年底 解剖了 以后 写了 一份报告。
109a. 爷爷 把 孙子 在 这几年来 宠坏了 但是 他 觉得 是 理所当然的。
109b. 孙子 把 爷爷 在 这几年来 宠坏了 但是 他 觉得 是 理所当然的。
110a. 父母 把 孩子 在 昨天晚上 哄睡了 以后 才 能 好好 休息 一下。
110b. 孩子 把 父母 在 昨天晚上 哄睡了 以后 才 能 好好 休息 一下。
111a. 女歌手 把 狗仔队 在 上周末 耍了 之后 就 消失了。
111b. 狗仔队 把 女歌手 在 上周末 耍了 之后 就 消失了。
112a. 孩子们 把 妈妈 在 上周末 歌颂了 一番， 场面 非常 感人。
112b. 妈妈 把 孩子们 在 上周末 歌颂了 一番， 场面 非常 感人。
113a. 奶奶 把 孙女 在 那天下午 喂饱了 以后 抱 到 床上 小睡。
113b. 孙女 把 奶奶 在 那天下午 喂饱了 以后 抱 到 床上 小睡。
114a. 丁老板 把 张主任 在 去年初 开除了 以后 觉得 很惭愧。
114b. 张主任 把 丁老板 在 去年初 开除了 以后 觉得 很惭愧。
115a. 那位候选人 把 黑帮头目 在 去年底 绳之 以法， 赢得了 大家的 支持。
115b. 黑帮头目 把 那位候选人 在 去年底 绳之 以法， 赢得了 大家的 支持。
116a. 那位教授 把 学生们 在 上星期 表扬了 一番， 也 讲了 一些 励志的 话。
116b. 学生们 把 那位教授 在 上星期 表扬了 一番， 也 讲了 一些 励志的 话。
117a. 那对夫妇 把 小狗 在 昨天黄昏 喂饱 以后 便 一起 去 公园 散步了。
117b. 小狗 把 那对夫妇 在 昨天黄昏 喂饱 以后 便 一起 去 公园 散步了。
118a. 见习生 把 老板 在 上星期 惹怒了 却 一直 没有 发现。

118b. 老板 把 见习生 在 上星期 惹怒了 却 一直 没有 发现。
119a. 总编辑 把 那模特儿 在 那年初 捧红 成为 一线明星， 并 赚了 一笔钱。
119b. 那模特儿 把 总编辑 在 那年初 捧红 成为 一线明星， 并 赚了 一笔钱。
120a. 那位贵妇 把 那条小狗 在 上个月 送给了 邻居 并 托 她 好好 照顾。
120b. 那条小狗 把 那位贵妇 在 上个月 送给了 邻居 并 托 她 好好 照顾。

Experiment 6

a – Expected
b - Unxpected

Argument Substitution

1a. The aquarium visitor wondered which fish the penguins had eaten during the performance.
1b. The aquarium visitor wondered which trainer the penguins had eaten during the performance.
2a. The storyteller explained which dragon the prince had slain with his sword.
2b. The storyteller explained which princess the prince had slain with his sword.
3a. The superintendent overheard which tenant the landlord had evicted at the end of May.
3b. The superintendent overheard which realtor the landlord had evicted at the end of May.
4a. The butler speculated which guests the master had invited to the party.
4b. The butler speculated which servants the master had invited to the party.
5a. The priest mentioned which nonbeliever the nun had converted on her mission trip.
5b. The priest mentioned which archbishop the nun had converted on her mission trip.
6a. The sheriff boasted which horse the cowboy had ridden across the town.
6b. The sheriff boasted which bandit the cowboy had ridden across the town.
7a. Mr. Stevens revealed which clerk the manager had fired for being lazy.
7b. Mr. Stevens revealed which customer the manager had fired for being lazy.
8a. The secret serviceman concealed which ex-president the agent had protected from the terrorists.
8b. The secret serviceman concealed which assassin the agent had protected from the terrorists.
9a. The Roman emperor asked which god the gladiator had worshipped for strength in battle.
9b. The Roman emperor asked which adversary the gladiator had worshipped for strength in battle.
10a. The author revealed which maiden the hero had saved from the tower.
10b. The author revealed which ogre the hero had saved from the tower.
11a. The columnist publicized which billionaire the supermodel had married on public television.
11b. The columnist publicized which stylist the supermodel had married on public television.
12a. The Iranian scholar researched which camels the Persians had ridden across the desert.
12b. The Iranian scholar researched which soldiers the Persians had ridden across the desert.
13a. The beggar saw which leper the saint had cured in the street.
13b. The beggar saw which prophet the saint had cured in the street.
14a. The historian recorded which patriot the king had knighted at the ceremony.
14b. The historian recorded which traitor the king had knighted at the ceremony.
15a. The monk recalled which sinner the priest had forgiven in church yesterday.
15b. The monk recalled which cardinal the priest had forgiven in church yesterday.
16a. The scientist noted which antelope the lion had eaten for his dinner.
16b. The scientist noted which cub the lion had eaten for his dinner.
17a. The secretary confirmed which illustrator the author had hired for the new book.
17b. The secretary confirmed which readers the author had hired for the new book.
18a. The stenographer recorded which defendant the judge had sentenced to fifteen years.
18b. The stenographer recorded which lawyer the judge had sentenced to fifteen years.
19a. Spencer guessed which agent the actress had hired to promote her career.
19b. Spencer guessed which admirer the actress had hired to promote her career.
20a. The editor listed which orphans the philanthropist had adopted from the faraway place.

20b. The editor listed which volunteers the philanthropist had adopted from the faraway place.
21a. The farmer remembered which pig the butcher had slaughtered in his kitchen.
21b. The farmer remembered which vegetarian the butcher had slaughtered in his kitchen.
22a. The homeowner asked which insects the exterminator had killed when he came over.
22b. The homeowner asked which plumber the exterminator had killed when he came over.
23a. The barkeeper knew which barmaids the regulars had tipped for their drinks.
23b. The barkeeper knew which troublemakers the regulars had tipped for their drinks.
24a. The ethnographer asked which deer the Indians had hunted with bows and arrows.
24b. The ethnographer asked which babies the Indians had hunted with bows and arrows.
25a. The historian knew which princess the king had married at the royal palace.
25b. The historian knew which prince the king had married at the royal palace.
26a. The journalist reported which opponent the politician had defeated by a wide margin.
26b. The journalist reported which voters the politician had defeated by a wide margin.
27a. The parent noticed which insects the child had collected in a jar.
27b. The parent noticed which teacher the child had collected in a jar.
28a. The scribe recorded which loyalists the monarch had rewarded for their actions.
28b. The scribe recorded which rebels the monarch had rewarded for their actions.
29a. The teacher remembered which delinquent the principal had suspended for a week.
29b. The teacher remembered which overacheiver the principal had suspended for a week.
30a. Sally heard which patient the doctor had treated first thing this morning.
30b. Sally heard which nurse the doctor had treated first thing this morning.
31a. The beach bum saw which fish the seagulls had caught along the beach.
31b. The beach bum saw which picnickers the seagulls had caught along the beach.
32a. The circus-goer explained which juggler the clown had tripped in the center ring.
32b. The circus-goer explained which child the clown had tripped in the center ring.
33a. The football fan heard which cheerleader the quarterback had dated after the season.
33b. The football fan heard which linebacker the quarterback had dated after the season.
34a. The seasonal laborer marked which cows the farmer had milked early this morning.
34b. The seasonal laborer marked which chickens the farmer had milked early this morning.
35a. The secretary knew which manager the CEO had promoted from the old office.
35b. The secretary knew which investor the CEO had promoted from the old office.
36a. General Jones heard which terrorist the corporal had captured during the battle.
36b. General Jones heard which comrade the corporal had captured during the battle.
37a. Jason saw which nerd the teacher had praised for his good behavior.
37b. Jason saw which rascal the teacher had praised for his good behavior.
38a. Jim forgot which addict the therapist had treated in his last session.
38b. Jim forgot which doctor the therapist had treated in his last session.
39a. The announcer stated which teammate the player had replaced for the upcoming game.
39b. The announcer stated which rival the player had replaced for the upcoming game.
40a. The crafts fair manager asked which cows the leatherworker had skinned to make his bags.
40b. The crafts fair manager asked which customers the leatherworker had skinned to make his bags.
41a. The police woman recognized which officer the lieutenant had promoted to a higher position.
41b. The police woman recognized which thief the lieutenant had promoted to a higher position.
42a. The reviewer forgot which researchers the authors had cited in their paper.
42b. The reviewer forgot which mice the authors had cited in their paper.
43a. The swim team captain recalled which champion the swimmer had defeated at last week's race.
43b. The swim team captain recalled which lifeguard the swimmer had defeated at last week's race.
44a. The zoologist marked which pup the seal had birthed over the weekend.
44b. The zoologist marked which walrus the seal had birthed over the weekend.
45a. The captain realized which stowaway the sailor had hidden in the cargo hold.
45b. The captain realized which pirate the sailor had hidden in the cargo hold.
46a. The father recalled which girl the boy had dated in junior high.
46b. The father recalled which dog the boy had dated in junior high.
47a. The Japanese scientists showed which salespeople the robots had replaced in the last decade.
47b. The Japanese scientists showed which patients the robots had replaced in the last decade.
48a. The principal forgot which student the teacher had disciplined after school yesterday.

48b. The principal forgot which parent the teacher had disciplined after school yesterday.
49a. The shepherd saw which lamb the sheep had birthed by the barn.
49b. The shepherd saw which predator the sheep had birthed by the barn.
50a. The snoop determined which bouncer the alcoholic had punched last Friday night.
50b. The snoop determined which psychologist the alcoholic had punched last Friday night.
51a. Carla inquired which exterminator the landlord had used in the apartment.
51b. Carla inquired which neighbor the landlord had used in the apartment.
52a. Karen learned which teenager the mother had grounded for disobeying her.
52b. Karen learned which baby the mother had grounded for disobeying her.
53a. Sandra observed which biologists the zookeeper had consulted about the food.
53b. Sandra observed which gorillas the zookeeper had consulted about the food.
54a. The anthropologist discovered which animal the caveman had hunted with flint arrows.
54b. The anthropologist discovered which goddess the caveman had hunted with flint arrows.
55a. The mother wondered which trick-or-treaters the neighbor had scared with his costume.
55b. The mother wondered which milkman the neighbor had scared with his costume.
56a. The movie buff speculated which actor the stuntman had portrayed in the opening scene.
56b. The movie buff speculated which director the stuntman had portrayed in the opening scene.
57a. The researcher recorded which chimpanzee the biologist had tagged to be studied.
57b. The researcher recorded which assistant the biologist had tagged to be studied.
58a. The sentry knew which guard the captain had fired from the position.
58b. The sentry knew which townsperson the captain had fired from the position.
59a. The team manager knew which athlete the officials had penalized for being unsportsmanlike.
59b. The team manager knew which umpire the officials had penalized for being unsportsmanlike.
60a. The theater owner described which spectators the magician had amazed with his tricks.
60b. The theater owner described which rabbit the magician had amazed with his tricks.

Argument Role-reversal

61a. The librarian documented which celebrities the journalist had interviewed for the magazine.
61b. The librarian documented which journalist the celebrities had interviewed for the magazine.
62a. The old widower remembered which villager the ghost had haunted for many years.
62b. The old widower remembered which ghost the villager had haunted for many years.
63a. The firefighter reported which victim the paramedic had saved after the fire.
63b. The firefighter reported which paramedic the victim had saved after the fire.
64a. The historian documented which prince the assassin had killed in the 10th century.
64b. The historian documented which assassin the prince had killed in the 10th century.
65a. The park ranger documented which eagle the hunter had shot with a rifle.
65b. The park ranger documented which hunter the eagle had shot with a rifle.
66a. The judo master recognized which defender the assailant had attacked late last night.
66b. The judo master recognized which assailant the defender had attacked late last night.
67a. The naturalist observed which predators the deer had avoided by sleeping in the daytime.
67b. The naturalist observed which deer the predators had avoided by sleeping in the daytime.
68a. The restaurant owner forgot which customer the waitress had served during dinner yesterday.
68b. The restaurant owner forgot which waitress the customer had served during dinner yesterday.
69a. The head nun explained which friar the temptress had seduced after Sunday mass.
69b. The head nun explained which temptress the friar had seduced after Sunday mass.
70a. The housekeeper showed which mouse the cat had killed under the table.
70b. The housekeeper showed which cat the mouse had killed under the table.
71a. The nanny knew which housekeeper the billionaire had hired from their conversation.
71b. The nanny knew which billionaire the housekeeper had hired from their conversation.
72a. The researcher learned which whale the biologist had studied in the lab.
72b. The researcher learned which biologist the whale had studied in the lab.
73a. Tracy announced which applicant the administrator had chosen after the interview.
73b. Tracy announced which administrator the applicant had chosen after the interview.
74a. The birthday boy saw which friend the clown had scared at the party.

74b. The birthday boy saw which clown the friend had scared at the party.
75a. The columnist publicized which photographer the celebrity had punched in the nose.
75b. The columnist publicized which celebrity the photographer had punched in the nose.
76a. The documentarian showed which piglets the agriculturalist had raised at the farm.
76b. The documentarian showed which agriculturalist the piglets had raised at the farm.
77a. The police officer noted which jeweler the burglar had robbed on his spree.
77b. The police officer noted which burglar the jeweler had robbed on his spree.
78a. The royal cook asked which lord the chef had served for twenty years.
78b. The royal cook asked which chef the lord had served for twenty years.
79a. The trapeze artist indicated which lion the trainer had tamed for the circus.
79b. The trapeze artist indicated which trainer the lion had tamed for the circus.
80a. The wildlife curator mentioned which ape the zookeeper had fed some fresh fruit.
80b. The wildlife curator mentioned which zookeeper the ape had fed some fresh fruit.
81a. The aid worker realized which refugees the philanthropist had helped to escape starvation.
81b. The aid worker realized which philanthropist the refugees had helped to escape starvation.
82a. The camper reported which girl the bear had mauled in the forest.
82b. The camper reported which bear the girl had mauled in the forest.
83a. The dance instructor noticed which onlookers the performers had impressed in the audience.
83b. The dance instructor noticed which performers the onlookers had impressed in the audience.
84a. The high schooler heard which geek the jock had bullied in the hall.
84b. The high schooler heard which jock the geek had bullied in the hall.
85a. The parent saw which child the lifeguard had rescued from the pool.
85b. The parent saw which lifeguard the child had rescued from the pool.
86a. The ringmaster recognized which cheetah the tamer had trained three years ago.
86b. The ringmaster recognized which tamer the cheetah had trained three years ago.
87a. The travel agent recorded which innkeeper the guests had liked in his notebook.
87b. The travel agent recorded which guests the innkeeper had liked in his notebook.
88a. The bird watcher saw which photographer the hawk had attacked in the woods.
88b. The bird watcher saw which hawk the photographer had attacked in the woods.
89a. The costume designer confirmed which actor the barber had shaved for the part.
89b. The costume designer confirmed which barber the actor had shaved for the part.
90a. The Indian king identified which prince the elephant had trampled during the long journey.
90b. The Indian king identified which elephant the prince had trampled during the long journey.
91a. The investigator discovered which couple the abductor had taken during their vacation.
91b. The investigator discovered which abductor the couple had taken during their vacation.
92a. The jail keeper forgot which guard the prisoner had attacked in his cell.
92b. The jail keeper forgot which prisoner the guard had attacked in his cell.
93a. The sailor saw which whale the man had harpooned on the starboard side.
93b. The sailor saw which man the whale had harpooned on the starboard side.
94a. The sheriff recalled which locals the gangsters had robbed in a dark alleyway.
94b. The sheriff recalled which gangsters the locals had robbed in a dark alleyway.
95a. The undead king recognized which woman the zombie had bitten during the fight.
95b. The undead king recognized which zombie the woman had bitten during the fight.
96a. Tonia explained which housewife the conman had swindled over the phone.
96b. Tonia explained which conman the housewife had swindled over the phone.
97a. The broadcaster explained which contestant the judge had disqualified from the show.
97b. The broadcaster explained which judge the contestant had disqualified from the show.
98a. The campaign volunteer speculated which mayor the voters had elected by a landslide.
98b. The campaign volunteer speculated which voters the mayor had elected by a landslide.
99a. The family counselor understood which daughter the stepparent had abused on a daily basis.
99b. The family counselor understood which stepparent the daughter had abused on a daily basis.
100a. The newscaster confirmed which fugitive the policeman had arrested following the robbery.
100b. The newscaster confirmed which policeman the fugitive had arrested following the robbery.
101a. The opera composer revealed which pirate the singer had portrayed with the greatest finesse.
101b. The opera composer revealed which singer the pirate had portrayed with the greatest finesse.
102a. The prison warden guessed which jailer the inmate had stabbed with a knife.

102b. The prison warden guessed which inmate the jailer had stabbed with a knife.
103a. The producer observed which viewers the acrobats had impressed with their tricks.
103b. The producer observed which acrobats the viewers had impressed with their tricks.
104a. The queen recognized which nobleman the comedian had entertained with his jokes.
104b. The queen recognized which comedian the nobleman had entertained with his jokes.
105a. The reporter inquired which animals the groomer had cleaned with special shampoo.
105b. The reporter inquired which groomer the animals had cleaned with special shampoo.
106a. The ape researcher identified which chimpanzee the poacher had shot in the jungle.
106b. The ape researcher identified which poacher the chimpanzee had shot in the jungle.
107a. The boatman saw which shark the diver had speared just off the shore.
107b. The boatman saw which diver the shark had speared just off the shore.
108a. The committee member guessed which advisors the president had appointed to the Cabinet.
108b. The committee member guessed which president the advisors had appointed to the Cabinet.
109a. The farmer indicated which donkey the townsperson had bought in the neighboring village.
109b. The farmer indicated which townsperson the donkey had bought in the neighboring village.
110a. The military man knew which soldiers the general had promoted after their tour.
110b. The military man knew which general the soldiers had promoted after their tour.
111a. The parole officer confirmed which delinquent the psychologist had treated several times already.
111b. The parole officer confirmed which psychologist the delinquent had treated several times already.
112a. The policeman knew which suspect the detective had arrested for the recent crime.
112b. The policeman knew which detective the suspect had arrested for the recent crime.
113a. The priest knew which demon the exorcist had expelled with a cross.
113b. The priest knew which exorcist the demon had expelled with a cross.
114a. Jack observed which passenger the conductor had removed from the train.
114b. Jack observed which conductor the passenger had removed from the train.
115a. Teresa knew which neighbor the parrot had bitten at the block party.
115b. Teresa knew which parrot the neighbor had bitten at the block party.
116a. The art critic identified which model the artist had drawn in the studio.
116b. The art critic identified which artist the model had drawn in the studio.
117a. The cotton farmer wondered which master the slave had escaped by running through the woods.
117b. The cotton farmer wondered which slave the master had escaped by running through the woods.
118a. The farmhand recorded which goats the breeder had sold at the market.
118b. The farmhand recorded which breeder the goats had sold at the market.
119a. The four-year-old knew which hero the dragon had eaten for dinner last night.
119b. The four-year-old knew which dragon the hero had eaten for dinner last night.
120a. The security guard investigated which patrolman the loiterer had evaded for three hours.
120b. The security guard investigated which loiterer the patrolman had evaded for three hours.

Experiment 7

a - Canonical control
b - Role-reversed (Animacy-congruous)

<u>High predictability</u>

1a. 故事中，/屠夫 把 那只猪/宰了/以后/卖到很好的价钱。
1b. 故事中，/那只猪 把 屠夫/宰了/以后/卖到很好的价钱。
2a. 到了最后，/消防员 把 被困的小孩/解救了/出来，/并送到了安全地点。
2b. 到了最后，/被困的小孩 把 消防员/解救了/出来，/并送到了安全地点。
3a. 晚会上，/小丑 把 观众/逗乐了/之后/便表现得十分神气。
3b. 晚会上，/观众 把 小丑/逗乐了/之后/便表现得十分神气。
4a. 记录片中，/青蛙 把 蚊子/吃到/肚子里，/并没有任何动静。

4b. 记 录 片 中 ， / 蚊 子 把 青 蛙 / 吃 到 /肚 子 里 ， /并 没 有 任 何 动 静 。
5a. 故 事 中 ， / 救 生 员 把 那 个 溺 水 者 / 救 起 /到 岸 上 ， /成 为 了 大 英 雄 。
5b. 故 事 中 ， / 那 个 溺 水 者 把 救 生 员 / 救 起 /到 岸 上 ， /成 为 了 大 英 雄 。
6a. 传 说 中 ， / 武 松 把 猛 虎 / 打 死 了 /以 后 /成 了 大 英 雄 。
6b. 传 说 中 ， / 猛 虎 把 武 松 / 打 死 了 /以 后 /成 了 大 英 雄 。
7a. 上 个 月 ， / 那 个 画 家 把 那 群 骏 马 / 画 下 来 /以 后 /便 立 即 展 示 在 画 廊 中 。
7b. 上 个 月 ， / 那 群 骏 马 把 那 个 画 家 / 画 下 来 /以 后 /便 立 即 展 示 在 画 廊 中 。
8a. 前 几 天 ， / 警 犬 把 失 踪 少 女 / 找 回 来 /并 送 到 了 /警 察 局 。
8b. 前 几 天 ， / 失 踪 少 女 把 警 犬 / 找 回 来 /并 送 到 了 /警 察 局 。
9a. 去 年 ， / 董 大 夫 把 那 个 病 人 / 治 愈 /以 后 /便 决 定 退 休 了 。
9b. 去 年 ， / 那 个 病 人 把 董 大 夫 / 治 愈 /以 后 /便 决 定 退 休 了 。
10a. 晚 会 上 ， / 相 声 演 员 把 来 宾 / 逗 得 /哈 哈 大 笑 ， /全 场 掌 声 雷 动 。
10b. 晚 会 上 ， / 来 宾 把 相 声 演 员 / 逗 得 /哈 哈 大 笑 ， /全 场 掌 声 雷 动 。
11a. 昨 天 晚 上 ， / 警 察 把 人 质 / 解 救 了 /出 来 ， /成 了 救 人 英 雄 。
11b. 昨 天 晚 上 ， / 人 质 把 警 察 / 解 救 了 /出 来 ， /成 了 救 人 英 雄 。
12a. 去 年 初 ， / 行 骗 者 把 村 民 / 骗 了 /以 后 /便 逃 之 夭 夭 了 。
12b. 去 年 初 ， / 村 民 把 行 骗 者 / 骗 了 /以 后 /便 逃 之 夭 夭 了 。
13a. 昨 天 晚 上 ， / 警 察 把 小 偷 / 抓 了 /回 警 局 ， /并 教 训 了 他 一 顿 。
13b. 昨 天 晚 上 ， / 小 偷 把 警 察 / 抓 了 /回 警 局 ， /并 教 训 了 他 一 顿 。
14a. 上 一 次 ， / 造 型 师 把 女 演 员 / 打 扮 得 /很 美 ， /获 得 不 少 的 赞 赏 。
14b. 上 一 次 ， / 女 演 员 把 造 型 师 / 打 扮 得 /很 美 ， /获 得 不 少 的 赞 赏 。
15a. 事 件 当 中 ， / 谈 判 专 家 把 劫 匪 / 说 服 /并 解 救 了 /人 质 。
15b. 事 件 当 中 ， / 劫 匪 把 谈 判 专 家 / 说 服 /并 解 救 了 /人 质 。
16a. 后 来 ， / 何 婆 婆 把 这 个 孤 儿 / 收 养 后 /并 带 着 /他 离 开 了 小 镇 。
16b. 后 来 ， / 这 个 孤 儿 把 何 婆 婆 / 收 养 后 /并 带 着 /他 离 开 了 小 镇 。
17a. 故 事 中 ， / 巫 师 把 王 子 / 变 成 了 /青 蛙 ， /世 上 只 有 小 公 主 才 能 解 救 他 。
17b. 故 事 中 ， / 王 子 把 巫 师 / 变 成 了 /青 蛙 ， /世 上 只 有 小 公 主 才 能 解 救 他 。
18a. 上 个 月 ， / 摄 影 师 把 那 群 企 鹅 / 拍 下 来 /以 后 /刊 登 在 杂 志 的 封 面 上 。
18b. 上 个 月 ， / 那 群 企 鹅 把 摄 影 师 / 拍 下 来 /以 后 /刊 登 在 杂 志 的 封 面 上 。
19a. 昨 天 ， / 那 个 画 家 把 模 特 儿 / 描 画 得 /非 常 完 美 ， /同 学 们 都 惊 叹 不 已 。
19b. 昨 天 ， / 模 特 儿 把 那 个 画 家 / 描 画 得 /非 常 完 美 ， /同 学 们 都 惊 叹 不 已 。
20a. 故 事 中 ， / 慈 善 家 把 孤 儿 / 领 养 /回 家 ， /并 且 供 他 上 学 。
20b. 故 事 中 ， / 孤 儿 把 慈 善 家 / 领 养 /回 家 ， /并 且 供 他 上 学 。
21a. 昨 天 ， / 货 车 司 机 把 路 人 / 撞 伤 了 /以 后 /立 即 送 他 到 医 院 包 扎 。
21b. 昨 天 ， / 路 人 把 货 车 司 机 / 撞 伤 了 /以 后 /立 即 送 他 到 医 院 包 扎 。
22a. 刚 才 ， / 这 只 花 猫 把 那 只 麻 雀 / 吃 得 /连 根 羽 毛 /都 没 留 下 。
22b. 刚 才 ， / 那 只 麻 雀 把 这 只 花 猫 / 吃 得 /连 根 羽 毛 /都 没 留 下 。
23a. 电 影 中 ， / 专 家 把 传 染 病 人 / 隔 离 了 /起 来 ， /防 止 病 菌 传 播 。
23b. 电 影 中 ， / 传 染 病 人 把 专 家 / 隔 离 了 /起 来 ， /防 止 病 菌 传 播 。
24a. 去 年 初 ， / 奸 商 把 投 资 者 / 骗 得 /血 本 无 归 ， /转 眼 间 逃 去 无 踪 。
24b. 去 年 初 ， / 投 资 者 把 奸 商 / 骗 得 /血 本 无 归 ， /转 眼 间 逃 去 无 踪 。
25a. 电 影 中 ， / 牧 羊 人 把 羊 群 / 赶 到 /草 原 上 /吃 草 ， 不 慌 也 不 忙 。
25b. 电 影 中 ， / 羊 群 把 牧 羊 人 / 赶 到 /草 原 上 /吃 草 ， 不 慌 也 不 忙 。
26a. 前 几 天 ， / 大 狗 把 孩 子 / 咬 伤 了 /以 后 /一 天 都 不 肯 进 食 。
26b. 前 几 天 ， / 孩 子 把 大 狗 / 咬 伤 了 /以 后 /一 天 都 不 肯 进 食 。
27a. 去 年 ， / 奸 商 把 市 民 / 欺 骗 了 /以 后 /便 臭 名 远 播 了 。
27b. 去 年 ， / 市 民 把 奸 商 / 欺 骗 了 /以 后 /便 臭 名 远 播 了 。
28a. 上 星 期 ， / 灭 虫 专 家 把 那 些 老 鼠 / 消 灭 干 净 ， /并 且 /放 了 灭 鼠 药 以 预 防 。
28b. 上 星 期 ， / 那 些 老 鼠 把 灭 虫 专 家 / 消 灭 干 净 ， /并 且 /放 了 灭 鼠 药 以 预 防 。

29a. 去 年 初 ， / 针 炙 师 把 女 病 人 / 治 好 /以 后 /心 里 十 分 满 足 。
29b. 去 年 初 ， / 女 病 人 把 针 炙 师 / 治 好 /以 后 /心 里 十 分 满 足 。
30a. 上 星 期 ， / 凶 犯 把 目 击 证 人 / 杀 死 /在 沙 发 上 ， /便 逃 去 无 踪 了 。
30b. 上 星 期 ， / 目 击 证 人 把 凶 犯 / 杀 死 /在 沙 发 上 ， /便 逃 去 无 踪 了 。

Low predictability

31a. 昨 天 ， / 小 弟 弟 把 狮 子 / 吵 醒 了 /以 后 /害 怕 得 只 懂 躲 在 爸 爸 身 后 。
31b. 昨 天 ， / 狮 子 把 小 弟 弟 / 吵 醒 了 /以 后 /害 怕 得 只 懂 躲 在 爸 爸 身 后 。
32a. 上 星 期 ， / 那 个 毕 业 生 把 周 老 板 / 恭 维 了 /一 番 /就 得 到 了 录 用 。
32b. 上 星 期 ， / 周 老 板 把 那 个 毕 业 生 / 恭 维 了 /一 番 /就 得 到 了 录 用 。
33a. 刚 才 ， / 按 摩 师 把 老 顾 客 / 得 罪 了 /以 后 /连 忙 道 歉 。
33b. 刚 才 ， / 老 顾 客 把 按 摩 师 / 得 罪 了 /以 后 /连 忙 道 歉 。
34a. 去 年 ， / 高 官 把 那 位 商 人 / 制 裁 了 /以 后 /大 家 都 很 高 兴 。
34b. 去 年 ， / 那 位 商 人 把 高 官 / 制 裁 了 /以 后 /大 家 都 很 高 兴 。
35a. 传 闻 说 ， / 将 军 把 士 兵 / 提 拔 /成 为 /第 九 小 队 的 组 长 。
35b. 传 闻 说 ， / 士 兵 把 将 军 / 提 拔 /成 为 /第 九 小 队 的 组 长 。
36a. 几 天 前 ， / 小 潘 把 那 个 女 孩 / 娶 了 /回 家 /当 媳 妇 ， 还 请 大 家 吃 了 一 顿 饭 。
36b. 几 天 前 ， / 那 个 女 孩 把 小 潘 / 娶 了 /回 家 /当 媳 妇 ， 还 请 大 家 吃 了 一 顿 饭 。
37a. 上 星 期 ， / 私 家 侦 探 把 那 个 政 客 / 揭 发 了 /并 且 /刊 登 了 照 片 作 证 据 。
37b. 上 星 期 ， / 那 个 政 客 把 私 家 侦 探 / 揭 发 了 /并 且 /刊 登 了 照 片 作 证 据 。
38a. 那 天 ， / 可 疑 人 物 把 门 卫 / 引 开 /之 后 /就 盗 走 了 宝 物 。
38b. 那 天 ， / 门 卫 把 可 疑 人 物 / 引 开 /之 后 /就 盗 走 了 宝 物 。
39a. 事 实 上 ， / 暴 君 把 民 众 / 统 治 了 /已 有 /几 十 年 。
39b. 事 实 上 ， / 民 众 把 暴 君 / 统 治 了 /已 有 /几 十 年 。
40a. 访 问 里 ， / 演 员 把 国 家 元 首 / 讽 刺 了 /一 番 ， /十 分 到 位 。
40b. 访 问 里 ， / 国 家 元 首 把 演 员 / 讽 刺 了 /一 番 ， /十 分 到 位 。
41a. 这 一 年 ， / 考 生 们 把 考 官 / 收 买 /起 来 了 ， /真 是 难 以 置 信 。
41b. 这 一 年 ， / 考 官 把 考 生 们 / 收 买 /起 来 了 ， /真 是 难 以 置 信 。
42a. 去 年 初 ， / 老 刘 把 鹦 鹉 / 训 练 了 /好 一 段 /时 间 ， 但 是 还 未 成 功 。
42b. 去 年 初 ， / 鹦 鹉 把 老 刘 / 训 练 了 /好 一 段 /时 间 ， 但 是 还 未 成 功 。
43a. 上 星 期 ， / 钟 老 板 把 这 群 员 工 / 奖 励 了 /一 番 ， /这 让 大 家 都 很 是 意 外 。
43b. 上 星 期 ， / 这 群 员 工 把 钟 老 板 / 奖 励 了 /一 番 ， /这 让 大 家 都 很 是 意 外 。
44a. 刚 才 ， / 那 个 主 人 把 那 只 小 狗 / 打 扮 了 /一 番 ， /然 后 才 开 心 地 出 门 。
44b. 刚 才 ， / 那 只 小 狗 把 那 个 主 人 / 打 扮 了 /一 番 ， /然 后 才 开 心 地 出 门 。
45a. 上 星 期 ， / 那 个 村 民 把 那 条 鱼 / 送 给 了 /村 长 ， /希 望 討 得 他 的 歡 心 。
45b. 上 星 期 ， / 那 条 鱼 把 那 个 村 民 / 送 给 了 /村 长 ， /希 望 討 得 他 的 歡 心 。
46a. 书 上 说 ， / 老 作 家 把 新 人 / 捧 红 了 /以 后 /便 退 隐 文 坛 了 。
46b. 书 上 说 ， / 新 人 把 老 作 家 / 捧 红 了 /以 后 /便 退 隐 文 坛 了 。
47a. 昨 天 ， / 教 练 把 选 手 / 鼓 励 了 /一 番 /之 后 比 赛 就 开 始 了 。
47b. 昨 天 ， / 选 手 把 教 练 / 鼓 励 了 /一 番 /之 后 比 赛 就 开 始 了 。
48a. 昨 天 ， / 海 洋 学 家 把 鲸 鱼 / 解 剖 了 /以 后 /写 了 一 份 报 告 。
48b. 昨 天 ， / 鲸 鱼 把 海 洋 学 家 / 解 剖 了 /以 后 /写 了 一 份 报 告 。
49a. 这 几 年 来 ， / 爷 爷 把 孙 子 / 宠 坏 了 /但 是 /他 觉 得 是 理 所 必 然 的 。
49b. 这 几 年 来 ， / 孙 子 把 爷 爷 / 宠 坏 了 /但 是 /他 觉 得 是 理 所 必 然 的 。
50a. 晚 上 ， / 父 母 把 孩 子 / 哄 睡 了 /以 后 /才 能 好 好 休 息 一 下 。
50b. 晚 上 ， / 孩 子 把 父 母 / 哄 睡 了 /以 后 /才 能 好 好 休 息 一 下 。
51a. 上 星 期 ， / 女 歌 手 把 狗 仔 队 / 耍 了 /之 后 /就 消 失 了 。
51b. 上 星 期 ， / 狗 仔 队 把 女 歌 手 / 耍 了 /之 后 /就 消 失 了 。

52a. 电影里，/孩子们把妈妈/歌颂了/一番，/场面非常感人。
52b. 电影里，/妈妈把孩子们/歌颂了/一番，/场面非常感人。
53a. 昨天下午，/奶奶把孙女/喂饱了/以后/抱到床上小睡。
53b. 昨天下午，/孙女把奶奶/喂饱了/以后/抱到床上小睡。
54a. 那天，/丁老板把张主任/开除了/以后/觉得很惭愧。
54b. 那天，/张主任把丁老板/开除了/以后/觉得很惭愧。
55a. 后来，/那位候选人把黑帮头目/绳之/以法，/赢得了大家的支持。
55b. 后来，/黑帮头目把那位候选人/绳之/以法，/赢得了大家的支持。
56a. 那天，/那位教授把学生们/表扬了/一番，/也讲了一些励志的话。
56b. 那天，/学生们把那位教授/表扬了/一番，/也讲了一些励志的话。
57a. 每天傍晚，/那对夫妇把小狗/喂饱/以后/便一起去公园散步了。
57b. 每天傍晚，/小狗把那对夫妇/喂饱/以后/便一起去公园散步了。
58a. 这几天，/见习生把老板/惹怒了/却一直/没有发现。
58b. 这几天，/老板把见习生/惹怒了/却一直/没有发现。
59a. 去年，/总编辑把那模特儿/捧红/成为/一线明星，并赚了一笔钱。
59b. 去年，/那模特儿把总编辑/捧红/成为/一线明星，并赚了一笔钱。
60a. 那天，/那位贵妇把那条小狗/送给了/邻居/并托她好好照顾。
60b. 那天，/那条小狗把那位贵妇/送给了/邻居/并托她好好照顾。

Experiment 8

a - Canonical control
b - Role-reversed (Animacy-congruous)

<u>High predictability</u>

1a. /救生员把那个溺水者/在日落前/救起/到岸上，/十分英勇。/
1b. /这只花猫把那只麻雀/在昨天黄昏/吃得/连根/羽毛都没留下。/
2a. /那个画家把那群骏马/在去年底/画下来/并且/展示在画廊中。/
2b. /奸商把投资者/在那年冬天/骗得/血本/无归。/
3a. /那个溺水者把救生员/在日落前/救起/到岸上，/十分英勇。/
3b. /那只麻雀把这只花猫/在昨天黄昏/吃得/连根/羽毛都没留下。/
4a. /模特儿把那个画家/在上星期/描画得/非常/完美。/
4b. /烈士把战友/在那年秋天/埋葬/在山坡上，/并放上鲜花。/
5a. /劫匪把谈判专家/在昨天黄昏/说服/并解救了/人质。/
5b. /那个犯人把法官/在去年秋天/判死刑/以后/惹来很多不满。/
6a. /那个画家把模特儿/在上星期/描画得/非常/完美。/
6b. /战友把烈士/在那年秋天/埋葬/在山坡上，/并放上鲜花。/
7a. /小偷把警察/在那天傍晚/抓了/回警局，/并教训了他一顿。/
7b. /飞鸟把猎人/在昨天早上/打下来/以后/便立即卖出去了。/
8a. /来宾把相声演员/在昨天晚上/逗得/哈哈/大笑。/
8b. /那些老鼠把灭虫专家/在上周末/消灭/干净，/非常厉害。/
9a. /猛虎把武松/在那年秋天/打死了/以后/成了大英雄。/
9b. /传染病人把专家/在上周末/隔离了/起来，/防止病菌传播。/
10a. /董大夫把那个病人/在去年底/治愈/以后/便退休了。/
10b. /大狗把孩子/在上星期/咬伤了/以后/一天都不肯进食。/
11a. /那群骏马把那个画家/在去年底/画下来/并且/展示在画廊中。/
11b. /投资者把奸商/在那年冬天/骗得/血本/无归。/

12a. /摄 影 师 把 那 群 企 鹅 /在 去 年 底 /拍 下 来 /以 后 /便 成 名 了 。 /
12b. /小 明 把 萤 火 虫 /在 那 天 早 上 /放 飞 /到 大 自 然 ， /心 里 很 高 兴 。 /
13a. /造 型 师 把 女 演 员 /在 上 星 期 /打 扮 得 /很 美 ， /获 得 不 少 赞 赏 。 /
13b. /驯 兽 师 把 狮 子 /在 去 年 初 /驯 服 得 /像 一 只 /小 猫 一 样 乖 巧 。 /
14a. /谈 判 专 家 把 劫 匪 /在 昨 天 黄 昏 /说 服 /并 解 救 了 /人 质 。 /
14b. /法 官 把 那 个 犯 人 /在 去 年 秋 天 /判 死 刑 /以 后 /惹 来 很 多 不 满 。 /
15a. /失 踪 少 女 把 警 犬 /在 昨 天 晚 上 /找 回 来 /并 送 到 了 /警 察 局 。 /
15b. /羊 群 把 牧 羊 人 /在 今 天 早 上 /赶 到 /草 原 上 /吃 草 。 /
16a. /人 质 把 警 察 /在 昨 天 晚 上 /解 救 了 /出 来 ， /成 了 救 人 英 雄 。 /
16b. /目 击 证 人 把 凶 犯 /在 去 年 底 /杀 死 /在 沙 发 上 /便 逃 去 无 踪 了 。 /
17a. /那 群 企 鹅 把 摄 影 师 /在 去 年 底 /拍 下 来 /以 后 /便 成 名 了 。 /
17b. /萤 火 虫 把 小 明 /在 那 天 早 上 /放 飞 /到 大 自 然 ， /心 里 很 高 兴 。 /
18a. /那 个 病 人 把 董 大 夫 /在 去 年 底 /治 愈 /以 后 /便 退 休 了 。 /
18b. /孩 子 把 大 狗 /在 上 星 期 /咬 伤 了 /以 后 /一 天 都 不 肯 进 食 。 /
19a. /相 声 演 员 把 来 宾 /在 昨 天 晚 上 /逗 得 /哈 哈 /大 笑 。 /
19b. /灭 虫 专 家 把 那 些 老 鼠 /在 上 周 末 /消 灭 /干 净 ， /非 常 厉 害 。 /
20a. /警 察 把 小 偷 /在 那 天 傍 晚 /抓 了 /回 警 局 ， /并 教 训 了 他 一 顿 。 /
20b. /猎 人 把 飞 鸟 /在 昨 天 早 上 /打 下 来 /以 后 /便 立 即 卖 出 去 了 。 /
21a. /那 只 猪 把 屠 夫 /在 上 星 期 /宰 了 /以 后 /卖 到 很 好 的 价 钱 。 /
21b. /路 人 把 货 车 司 机 /在 那 天 下 午 /撞 伤 了 /以 后 /立 即 送 他 到 医 院 。 /
22a. /慈 善 家 把 孤 儿 /在 那 年 冬 天 /领 养 /回 家 ， /并 供 他 上 学 。 /
22b. /魔 术 师 把 白 鸽 /在 昨 天 晚 上 /变 走 /以 后 /全 场 观 众 都 很 兴 奋 。 /
23a. /孤 儿 把 慈 善 家 /在 那 年 冬 天 /领 养 /回 家 ， /并 供 他 上 学 。 /
23b. /白 鸽 把 魔 术 师 /在 昨 天 晚 上 /变 走 /以 后 /全 场 观 众 都 很 兴 奋 。 /
24a. /警 犬 把 失 踪 少 女 /在 昨 天 晚 上 /找 回 来 /并 送 到 了 /警 察 局 。 /
24b. /牧 羊 人 把 羊 群 /在 今 天 早 上 /赶 到 /草 原 上 /吃 草 。 /
25a. /警 察 把 人 质 /在 昨 天 晚 上 /解 救 了 /出 来 ， /成 了 救 人 英 雄 。 /
25b. /凶 犯 把 目 击 证 人 /在 去 年 底 /杀 死 /在 沙 发 上 /便 逃 去 无 踪 了 。 /
26a. /行 骗 者 把 村 民 /在 去 年 初 /骗 了 /以 后 /便 逃 之 夭 夭 了 。 /
26b. /那 个 农 夫 把 豺 狼 /在 那 天 下 午 /打 死 了 /以 后 /便 炫 耀 了 一 番 。 /
27a. /村 民 把 行 骗 者 /在 去 年 初 /骗 了 /以 后 /便 逃 之 夭 夭 了 。 /
27b. /豺 狼 把 那 个 农 夫 /在 那 天 下 午 /打 死 了 /以 后 /便 炫 耀 了 一 番 。 /
28a. /女 演 员 把 造 型 师 /在 上 星 期 /打 扮 得 /很 美 ， /获 得 不 少 赞 赏 。 /
28b. /狮 子 把 驯 兽 师 /在 去 年 初 /驯 服 得 /像 一 只 /小 猫 一 样 乖 巧 。 /
29a. /武 松 把 猛 虎 /在 那 年 秋 天 /打 死 了 /以 后 /成 了 大 英 雄 。 /
29b. /专 家 把 传 染 病 人 /在 上 周 末 /隔 离 了 /起 来 ， /防 止 病 菌 传 播 。 /
30a. /屠 夫 把 那 只 猪 /在 上 星 期 /宰 了 /以 后 /卖 到 很 好 的 价 钱 。 /
30b. /货 车 司 机 把 路 人 /在 那 天 下 午 /撞 伤 了 /以 后 /立 即 送 他 到 医 院 。 /

<u>Low predictability</u>

31a. /那 位 商 人 把 高 官 /在 去 年 底 /制 裁 了 /以 后 /大 家 都 很 高 兴 。 /
31b. /学 生 们 把 那 位 教 授 /在 上 星 期 /表 扬 了 /一 番 ， /非 常 高 兴 。 /
32a. /妈 妈 把 新 生 儿 /在 那 年 冬 天 /遗 弃 /在 教 堂 /门 口 ， 心 痛 欲 绝 。 /
32b. /钟 老 板 把 这 群 员 工 /在 今 天 早 上 /奖 励 了 /一 番 ， /提 升 士 气 。 /
33a. /乞 丐 把 吴 管 家 /在 那 天 早 上 /缠 住 了 /不 愿 走 ， /想 要 讨 钱 。 /
33b. /教 练 把 选 手 /在 今 天 早 上 /鼓 励 了 /一 番 /后 比 赛 就 开 始 了 。 /
34a. /首 相 把 地 方 官 /在 去 年 初 /贿 赂 了 /以 后 /变 得 有 恃 无 恐 。 /
34b. /孩 子 把 父 母 /在 昨 天 晚 上 /哄 睡 了 /以 后 /才 能 休 息 一 下 。 /

35a. /那 个 女 孩 把 小 潘 /在 去 年 八 月 /娶 了 /回 家 /当 媳 妇 。 /
35b. /老 板 把 见 习 生 /在 上 星 期 /惹 怒 了 /却 一 直 /没 发 现 。 /
36a. /学 徒 把 那 位 师 傅 /在 这 几 年 来 /超 越 了 /以 后 /就 得 意 忘 形 了 。 /
36b. /女 歌 手 把 狗 仔 队 /在 上 周 末 /耍 了 /之 后 /就 消 失 了 。 /
37a. /老 顾 客 把 按 摩 师 /在 上 星 期 /得 罪 了 /以 后 /连 忙 道 歉 。 /
37b. /黑 帮 头 目 把 那 位 候 选 人 /在 去 年 底 /绳 之 /以 法 ， /赢 尽 民 心 。 /
38a. /按 摩 师 把 老 顾 客 /在 上 星 期 /得 罪 了 /以 后 /连 忙 道 歉 。 /
38b. /那 位 候 选 人 把 黑 帮 头 目 /在 去 年 底 /绳 之 /以 法 ， /赢 尽 民 心 。 /
39a. /高 官 把 那 位 商 人 /在 去 年 底 /制 裁 了 /以 后 /大 家 都 很 高 兴 。 /
39b. /那 位 教 授 把 学 生 们 /在 上 星 期 /表 扬 了 /一 番 ， /非 常 高 兴 。 /
40a. /地 主 把 童 工 /在 那 年 秋 天 /卖 给 了 /人 贩 ， /从 中 谋 取 暴 利 。 /
40b. /老 刘 把 鹦 鹉 /在 那 年 夏 天 /训 练 了 /好 一 段 /时 间 。 /
41a. /将 军 把 士 兵 /在 去 年 十 月 /提 拔 /成 为 /小 队 组 长 。 /
41b. /那 对 夫 妇 把 小 狗 /在 昨 天 黄 昏 /喂 饱 /以 后 /便 去 散 步 。 /
42a. /新 生 儿 把 妈 妈 /在 那 年 冬 天 /遗 弃 /在 教 堂 /门 口 ， 心 痛 欲 绝 。 /
42b. /这 群 员 工 把 钟 老 板 /在 今 天 早 上 /奖 励 了 /一 番 ， /提 升 士 气 。 /
43a. /熊 猫 妈 妈 把 小 熊 猫 /在 昨 天 晚 上 /生 了 /下 来 ， /大 家 都 很 高 兴 。 /
43b. /那 个 村 民 把 那 条 鱼 /在 上 星 期 /送 给 了 /村 长 ， /想 要 討 好 他 。 /
44a. /小 熊 猫 把 熊 猫 妈 妈 /在 昨 天 晚 上 /生 了 /下 来 ， /大 家 都 很 高 兴 。 /
44b. /那 条 鱼 把 那 个 村 民 /在 上 星 期 /送 给 了 /村 长 ， /想 要 討 好 他 。 /
45a. /小 潘 把 那 个 女 孩 /在 去 年 八 月 /娶 了 /回 家 /当 媳 妇 。 /
45b. /见 习 生 把 老 板 /在 上 星 期 /惹 怒 了 /却 一 直 /没 发 现 。 /
46a. /私 家 侦 探 把 那 个 政 客 /在 上 个 月 /揭 发 了 /并 且 /刊 登 了 证 据 。 /
46b. /总 编 辑 把 那 模 特 儿 /在 那 年 初 /捧 红 /成 为 /一 线 明 星 。 /
47a. /那 个 奸 商 把 这 群 黑 工 /在 那 年 冬 天 /藏 在 /小 黑 屋 /里 面 。 /
47b. /海 洋 学 家 把 鲸 鱼 /在 去 年 底 /解 剖 了 /以 后 /写 了 一 份 报 告 。 /
48a. /那 位 师 傅 把 学 徒 /在 这 几 年 来 /超 越 了 /以 后 /就 得 意 忘 形 了 。 /
48b. /狗 仔 队 把 女 歌 手 /在 上 周 末 /耍 了 /之 后 /就 消 失 了 。 /
49a. /周 老 板 把 那 个 毕 业 生 /在 上 个 月 /恭 维 了 /一 番 /就 得 到 录 用 。 /
49b. /张 主 任 把 丁 老 板 /在 去 年 初 /开 除 了 /以 后 /觉 得 很 惭 愧 。 /
50a. /小 孩 把 保 姆 /在 那 天 下 午 /累 坏 了 /之 后 /便 跑 去 荡 秋 千 了 。 /
50b. /考 生 们 把 考 官 /在 去 年 底 /收 买 /起 来 了 ， /险 些 被 发 现 。 /
51a. /保 姆 把 小 孩 /在 那 天 下 午 /累 坏 了 /之 后 /便 跑 去 荡 秋 千 了 。 /
51b. /考 官 把 考 生 们 /在 去 年 底 /收 买 /起 来 了 ， /险 些 被 发 现 。 /
52a. /童 工 把 地 主 /在 那 年 秋 天 /卖 给 了 /人 贩 ， /从 中 谋 取 暴 利 。 /
52b. /鹦 鹉 把 老 刘 /在 那 年 夏 天 /训 练 了 /好 一 段 /时 间 。 /
53a. /士 兵 把 将 军 /在 去 年 十 月 /提 拔 /成 为 /小 队 组 长 。 /
53b. /小 狗 把 那 对 夫 妇 /在 昨 天 黄 昏 /喂 饱 /以 后 /便 去 散 步 。 /
54a. /吴 管 家 把 乞 丐 /在 那 天 早 上 /缠 住 了 /不 愿 走 ， /想 要 讨 钱 。 /
54b. /选 手 把 教 练 /在 今 天 早 上 /鼓 励 了 /一 番 /后 比 赛 就 开 始 了 。 /
55a. /那 个 政 客 把 私 家 侦 探 /在 上 个 月 /揭 发 了 /并 且 /刊 登 了 证 据 。 /
55b. /那 模 特 儿 把 总 编 辑 /在 那 年 初 /捧 红 /成 为 /一 线 明 星 。 /
56a. /可 疑 人 物 把 门 卫 /在 那 天 晚 上 /引 开 /之 后 /就 盗 走 了 宝 物 。 /
56b. /那 位 贵 妇 把 那 条 小 狗 /在 上 个 月 /送 给 了 /邻 家 的 /陈 太 太 。 /
57a. /这 群 黑 工 把 那 个 奸 商 /在 那 年 冬 天 /藏 在 /小 黑 屋 /里 面 。 /
57b. /鲸 鱼 把 海 洋 学 家 /在 去 年 底 /解 剖 了 /以 后 /写 了 一 份 报 告 。 /
58a. /门 卫 把 可 疑 人 物 /在 那 天 晚 上 /引 开 /之 后 /就 盗 走 了 宝 物 。 /
58b. /那 条 小 狗 把 那 位 贵 妇 /在 上 个 月 /送 给 了 /邻 家 的 /陈 太 太 。 /
59a. /地 方 官 把 首 相 /在 去 年 初 /贿 赂 了 /以 后 /变 得 有 恃 无 恐 。 /

59b. /父 母 把 孩 子 /在 昨 天 晚 上 /哄 睡 了 /以 后 /才 能 休 息 一 下 。 /
60a. /那 个 毕 业 生 把 周 老 板 /在 上 个 月 /恭 维 了 /一 番 /就 得 到 录 用 。 /
60b. /丁 老 板 把 张 主 任 /在 去 年 初 /开 除 了 /以 后 /觉 得 很 惭 愧 。 /

Experiment 9

a - Canonical control
b - Role-reversed (Animacy-congruous)

The pseudoword in the invalid preview condition is presented in parentheses.

1a. /那 群 旅 客 把 恐 怖 分 子 /在 昨 天 黄 昏 /制 服 了 /以 后 /警 方 才 到 场 。 / (姐房)
1b. /恐 怖 分 子 把 那 群 旅 客 /在 昨 天 黄 昏 /制 服 了 /以 后 /警 方 才 到 场 。 / (姐房)
2a. /铁 木 真 把 部 落 民 族 /在 那 年 秋 天 /聚 集 /起 来 /宣 布 重 要 的 消 息 。 / (境曾)
2b. /部 落 民 族 把 铁 木 真 /在 那 年 秋 天 /聚 集 /起 来 /宣 布 重 要 的 消 息 。 / (境曾)
3a. /那 位 顾 客 把 售 货 员 /在 昨 天 下 午 /投 诉 了 /一 番 ， /场 面 尴 尬 。 / (极作)
3b. /售 货 员 把 那 位 顾 客 /在 昨 天 下 午 /投 诉 了 /一 番 ， /场 面 尴 尬 。 / (极作)
4a. /人 们 把 鲨 鱼 /在 去 年 六 月 /捕 获 了 /之 后 /就 开 始 庆 祝 。 / (课速)
4b. /鲨 鱼 把 人 们 /在 去 年 六 月 /捕 获 了 /之 后 /就 开 始 庆 祝 。 / (课速)
5a. /巡 逻 队 把 藏 羚 羊 /在 去 年 八 月 /送 回 /保 护 区 /好 好 照 顾 。 / (音自)
5b. /藏 羚 羊 把 巡 逻 队 /在 去 年 八 月 /送 回 /保 护 区 /好 好 照 顾 。 / (音自)
6a. /贪 官 把 包 青 天 /在 那 年 春 天 /陷 害 得 /苦 不 堪 言 ， /激 怒 了 百 姓 。 / (胸通)
6b. /包 青 天 把 贪 官 /在 那 年 春 天 /陷 害 得 /苦 不 堪 言 ， /激 怒 了 百 姓 。 / (胸通)
7a. /狱 卒 把 那 个 囚 犯 /在 去 年 底 /放 走 了 /之 后 /还 是 决 定 自 首 。 / (爸快)
7b. /那 个 囚 犯 把 狱 卒 /在 去 年 底 /放 走 了 /之 后 /还 是 决 定 自 首 。 / (爸快)
8a. /得 奖 者 把 评 委 /在 昨 天 晚 上 /感 动 了 /还 流 下 /了 热 泪 。 / (错年)
8b. /评 委 把 得 奖 者 /在 昨 天 晚 上 /感 动 了 /还 流 下 /了 热 泪 。 / (错年)
9a. /导 演 把 新 演 员 /在 去 年 夏 天 /培 训 了 /一 番 ， /十 分 用 心 。 / (渐巧)
9b. /新 演 员 把 导 演 /在 去 年 夏 天 /培 训 了 /一 番 ， /十 分 用 心 。 / (渐巧)
10a. /猎 人 把 雪 橇 狗 /在 昨 天 早 上 /喂 饱 了 /以 后 /便 去 打 猎 了 。 / (超绅)
10b. /雪 橇 狗 把 猎 人 /在 昨 天 早 上 /喂 饱 了 /以 后 /便 去 打 猎 了 。 / (超绅)
11a. /秘 书 把 董 事 /在 去 年 初 /迷 惑 得 /神 魂 颠 倒 ， /得 到 不 少 好 处 。 / (政辜)
11b. /董 事 把 秘 书 /在 去 年 初 /迷 惑 得 /神 魂 颠 倒 ， /得 到 不 少 好 处 。 / (政辜)
12a. /经 理 人 把 新 乐 队 /在 上 个 月 /包 装 了 /一 番 ， /准 备 出 专 辑 。 / (兄期)
12b. /新 乐 队 把 经 理 人 /在 上 个 月 /包 装 了 /一 番 ， /准 备 出 专 辑 。 / (兄期)
13a. /佣 人 把 富 翁 /在 这 几 年 来 /服 侍 得 /很 好 ， /最 后 却 被 辞 去 了 。 / (官枝)
13b. /富 翁 把 佣 人 /在 这 几 年 来 /服 侍 得 /很 好 ， /最 后 却 被 辞 去 了 。 / (官枝)
14a. /皇 后 把 宫 女 /在 那 年 初 /处 罚 了 /一 顿 ， /并 且 公 开 其 恶 行 。 / (必祖)
14b. /宫 女 把 皇 后 /在 那 年 初 /处 罚 了 /一 顿 ， /并 且 公 开 其 恶 行 。 / (必祖)
15a. /何 太 太 把 搬 运 工 人 /在 上 星 期 /请 来 /帮 忙 ， /毫 不 吝 啬 。 / (高没)
15b. /搬 运 工 人 把 何 太 太 /在 上 星 期 /请 来 /帮 忙 ， /毫 不 吝 啬 。 / (高没)
16a. /工 人 把 那 个 商 人 /在 去 年 底 /告 上 /法 庭 ， /誓 要 讨 回 公 道 。 / (听子)
16b. /那 个 商 人 把 工 人 /在 去 年 底 /告 上 /法 庭 ， /誓 要 讨 回 公 道 。 / (听子)
17a. /皇 后 把 白 雪 公 主 /在 那 年 夏 天 /欺 骗 了 /过 来 ， /并 且 毒 害 了 。 / (椅蒂)
17b. /白 雪 公 主 把 皇 后 /在 那 年 夏 天 /欺 骗 了 /过 来 ， /并 且 毒 害 了 。 / (椅蒂)
18a. /群 众 把 那 个 叛 徒 /在 那 天 傍 晚 /揪 出 来 /毒 打 ， /毫 不 留 手 。 / (颊让)
18b. /那 个 叛 徒 把 群 众 /在 那 天 傍 晚 /揪 出 来 /毒 打 ， /毫 不 留 手 。 / (颊让)
19a. /老 太 太 把 猫 咪 /在 上 周 末 /丢 失 了 /在 公 园 ， /真 叫 人 担 心 。 / (乔必)

19b. /猫 咪 把 老 太 太 /在 上 周 末 /丢 失 了 /在 公 园 ， /真 叫 人 担 心 。 / (乔必)
20a. /护 士 把 伤 兵 /在 昨 天 下 午 /包 扎 /妥 当 /并 扶 到 病 床 上 。 / (务尤)
20b. /伤 兵 把 护 士 /在 昨 天 下 午 /包 扎 /妥 当 /并 扶 到 病 床 上 。 / (务尤)
21a. /小 孩 把 保 姆 /在 那 天 下 午 /累 坏 了 /之 后 /便 跑 去 荡 秋 千 了 。 / (猪近)
21b. /保 姆 把 小 孩 /在 那 天 下 午 /累 坏 了 /之 后 /便 跑 去 荡 秋 千 了 。 / (猪近)
22a. /地 主 把 童 工 /在 那 年 秋 天 /卖 给 了 /人 贩 ， /从 中 谋 取 暴 利 。 / (店点)
22b. /童 工 把 地 主 /在 那 年 秋 天 /卖 给 了 /人 贩 ， /从 中 谋 取 暴 利 。 / (店点)
23a. /妈 妈 把 新 生 儿 /在 那 年 冬 天 /遗 弃 /在 教 堂 /门 口 ， 心 痛 欲 绝 。 / (牌呆)
23b. /新 生 儿 把 妈 妈 /在 那 年 冬 天 /遗 弃 /在 教 堂 /门 口 ， 心 痛 欲 绝 。 / (牌呆)
24a. /乞 丐 把 吴 管 家 /在 那 天 早 上 /缠 住 了 /不 愿 走 ， /想 要 讨 钱 。 / (遥何)
24b. /吴 管 家 把 乞 丐 /在 那 天 早 上 /缠 住 了 /不 愿 走 ， /想 要 讨 钱 。 / (遥何)
25a. /那 个 奸 商 把 这 群 黑 工 /在 那 年 冬 天 /藏 在 /小 黑 屋 /里 面 。 / (糟有)
25b. /这 群 黑 工 把 那 个 奸 商 /在 那 年 冬 天 /藏 在 /小 黑 屋 /里 面 。 / (糟有)
26a. /乌 鸦 把 那 个 农 夫 /在 今 天 早 上 /惹 火 了 /以 后 /飞 得 远 远 的 。 / (辈内)
26b. /那 个 农 夫 把 乌 鸦 /在 今 天 早 上 /惹 火 了 /以 后 /飞 得 远 远 的 。 / (辈内)
27a. /地 方 官 把 首 相 /在 去 年 初 /贿 赂 了 /以 后 /变 得 有 恃 无 恐 。 / (柴皱)
27b. /首 相 把 地 方 官 /在 去 年 初 /贿 赂 了 /以 后 /变 得 有 恃 无 恐 。 / (柴皱)
28a. /学 徒 把 那 位 师 傅 /在 这 几 年 来 /超 越 了 /以 后 /就 得 意 忘 形 了 。 / (程游)
28b. /那 位 师 傅 把 学 徒 /在 这 几 年 来 /超 越 了 /以 后 /就 得 意 忘 形 了 。 / (程游)
29a. /小 弟 弟 把 狮 子 /在 昨 天 下 午 /吵 醒 了 /以 后 /害 怕 得 哭 了 。 / (层嘴)
29b. /狮 子 把 小 弟 弟 /在 昨 天 下 午 /吵 醒 了 /以 后 /害 怕 得 哭 了 。 / (层嘴)
30a. /那 个 毕 业 生 把 周 老 板 /在 上 个 月 /恭 维 了 /一 番 /就 得 到 录 用 。 / (素假)
30b. /周 老 板 把 那 个 毕 业 生 /在 上 个 月 /恭 维 了 /一 番 /就 得 到 录 用 。 / (素假)
31a. /按 摩 师 把 老 顾 客 /在 上 星 期 /得 罪 了 /以 后 /连 忙 道 歉 。 / (着简)
31b. /老 顾 客 把 按 摩 师 /在 上 星 期 /得 罪 了 /以 后 /连 忙 道 歉 。 / (着简)
32a. /高 官 把 那 位 商 人 /在 去 年 底 /制 裁 了 /以 后 /大 家 都 很 高 兴 。 / (姐琴)
32b. /那 位 商 人 把 高 官 /在 去 年 底 /制 裁 了 /以 后 /大 家 都 很 高 兴 。 / (姐琴)
33a. /将 军 把 士 兵 /在 去 年 十 月 /提 拔 /成 为 /小 队 组 长 。 / (棒径)
33b. /士 兵 把 将 军 /在 去 年 十 月 /提 拔 /成 为 /小 队 组 长 。 / (棒径)
34a. /私 家 侦 探 把 那 个 政 客 /在 上 个 月 /揭 发 了 /并 且 /刊 登 了 证 据 。 / (裙只)
34b. /那 个 政 客 把 私 家 侦 探 /在 上 个 月 /揭 发 了 /并 且 /刊 登 了 证 据 。 / (裙只)
35a. /可 疑 人 物 把 门 卫 /在 那 天 晚 上 /引 开 /之 后 /就 盗 走 了 宝 物 。 / (介天)
35b. /门 卫 把 可 疑 人 物 /在 那 天 晚 上 /引 开 /之 后 /就 盗 走 了 宝 物 。 / (介天)
36a. /暴 君 把 民 众 /在 这 几 年 来 /统 治 得 /一 塌 糊 涂 ， /民 不 聊 生 。 / (院往)
36b. /民 众 把 暴 君 /在 这 几 年 来 /统 治 得 /一 塌 糊 涂 ， /民 不 聊 生 。 / (院往)
37a. /演 员 把 国 家 元 首 /在 上 星 期 /讽 刺 了 /一 番 ， /十 分 到 位 。 / (朴坦)
37b. /国 家 元 首 把 演 员 /在 上 星 期 /讽 刺 了 /一 番 ， /十 分 到 位 。 / (朴坦)
38a. /考 生 们 把 考 官 /在 去 年 底 /收 买 /起 来 了 ， /险 些 被 发 现 。 / (至光)
38b. /考 官 把 考 生 们 /在 去 年 底 /收 买 /起 来 了 ， /险 些 被 发 现 。 / (至光)
39a. /老 刘 把 鹦 鹉 /在 那 年 夏 天 /训 练 了 /好 一 段 /时 间 。 / (巧松)
39b. /鹦 鹉 把 老 刘 /在 那 年 夏 天 /训 练 了 /好 一 段 /时 间 。 / (巧松)
40a. /钟 老 板 把 这 群 员 工 /在 今 天 早 上 /奖 励 了 /一 番 ， /提 升 士 气 。 / (香肚)
40b. /这 群 员 工 把 钟 老 板 /在 今 天 早 上 /奖 励 了 /一 番 ， /提 升 士 气 。 / (香肚)
41a. /那 个 主 人 把 那 只 小 狗 /在 上 周 末 /打 扮 了 /一 番 /才 出 门 。 / (正犹)
41b. /那 只 小 狗 把 那 个 主 人 /在 上 周 末 /打 扮 了 /一 番 /才 出 门 。 / (正犹)
42a. /那 个 村 民 把 那 条 鱼 /在 上 星 期 /送 给 了 /村 长 ， /想 要 討 好 他 。 / (音点)
42b. /那 条 鱼 把 那 个 村 民 /在 上 星 期 /送 给 了 /村 长 ， /想 要 討 好 他 。 / (音点)
43a. /老 作 家 把 新 人 /在 去 年 春 天 /捧 红 了 /以 后 /便 退 隐 文 坛 了 。 / (崎冲)
43b. /新 人 把 老 作 家 /在 去 年 春 天 /捧 红 了 /以 后 /便 退 隐 文 坛 了 。 / (崎冲)

44a. /教 练 把 选 手 /在 今 天 早 上 /鼓 励 了 /一 番 /后 比 赛 就 开 始 了 。 / (键肚)
44b. /选 手 把 教 练 /在 今 天 早 上 /鼓 励 了 /一 番 /后 比 赛 就 开 始 了 。 / (键肚)
45a. /海 洋 学 家 把 鲸 鱼 /在 去 年 底 /解 剖 了 /以 后 /写 了 一 份 报 告 。 / (新哨)
45b. /鲸 鱼 把 海 洋 学 家 /在 去 年 底 /解 剖 了 /以 后 /写 了 一 份 报 告 。 / (新哨)
46a. /爷 爷 把 孙 子 /在 这 几 年 来 /宠 坏 了 /但 是 /他 觉 得 没 有 问 题 。 / (弥况)
46b. /孙 子 把 爷 爷 /在 这 几 年 来 /宠 坏 了 /但 是 /他 觉 得 没 有 问 题 。 / (弥况)
47a. /父 母 把 孩 子 /在 昨 天 晚 上 /哄 睡 了 /以 后 /才 能 休 息 一 下 。 / (峡简)
47b. /孩 子 把 父 母 /在 昨 天 晚 上 /哄 睡 了 /以 后 /才 能 休 息 一 下 。 / (峡简)
48a. /孩 子 们 把 妈 妈 /在 上 周 末 /歌 颂 了 /一 番 ， /场 面 很 感 人 。 / (精栓)
48b. /妈 妈 把 孩 子 们 /在 上 周 末 /歌 颂 了 /一 番 ， /场 面 很 感 人 。 / (精栓)
49a. /奶 奶 把 孙 女 /在 那 天 下 午 /喂 饱 了 /以 后 /抱 到 床 上 小 睡 。 / (曾浅)
49b. /孙 女 把 奶 奶 /在 那 天 下 午 /喂 饱 了 /以 后 /抱 到 床 上 小 睡 。 / (曾浅)
50a. /丁 老 板 把 张 主 任 /在 去 年 初 /开 除 了 /以 后 /觉 得 很 惭 愧 。 / (天哈)
50b. /张 主 任 把 丁 老 板 /在 去 年 初 /开 除 了 /以 后 /觉 得 很 惭 愧 。 / (天哈)
51a. /那 位 候 选 人 把 黑 帮 头 目 /在 去 年 底 /绳 之 /以 法 ， /赢 尽 民 心 。 / (蛇已)
51b. /黑 帮 头 目 把 那 位 候 选 人 /在 去 年 底 /绳 之 /以 法 ， /赢 尽 民 心 。 / (蛇已)
52a. /那 位 教 授 把 学 生 们 /在 上 星 期 /表 扬 了 /一 番 ， /非 常 高 兴 。 / (或污)
52b. /学 生 们 把 那 位 教 授 /在 上 星 期 /表 扬 了 /一 番 ， /非 常 高 兴 。 / (或污)
53a. /那 对 夫 妇 把 小 狗 /在 昨 天 黄 昏 /喂 饱 /以 后 /便 去 散 步 。 / (曾浅)
53b. /小 狗 把 那 对 夫 妇 /在 昨 天 黄 昏 /喂 饱 /以 后 /便 去 散 步 。 / (曾浅)
54a. /见 习 生 把 老 板 /在 上 星 期 /惹 怒 了 /却 一 直 /没 发 现 。 / (辈冠)
54b. /老 板 把 见 习 生 /在 上 星 期 /惹 怒 了 /却 一 直 /没 发 现 。 / (辈冠)
55a. /总 编 辑 把 那 模 特 儿 /在 那 年 初 /捧 红 /成 为 /一 线 明 星 。 / (崎冲)
55b. /那 模 特 儿 把 总 编 辑 /在 那 年 初 /捧 红 /成 为 /一 线 明 星 。 / (崎冲)
56a. /那 位 贵 妇 把 那 条 小 狗 /在 上 个 月 /送 给 了 /邻 家 的 /陈 太 太 。 / (音点)
56b. /那 条 小 狗 把 那 位 贵 妇 /在 上 个 月 /送 给 了 /邻 家 的 /陈 太 太 。 / (音点)

References

Abrams, R. A., & Jonides, J. (1988). Programming saccadic eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 428-443.

Altmann, G. T. M., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language, 31*, 685 – 712.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17, 364-390.

Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280-289.

Bar, M. (2011). *Predictions in the Brain: Using Our Past to Generate a Future.* Oxford University Press, USA.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience, 29*, 8447-8451.

Bever, T. (1970). The cognitive basis for linguistic structures. In: R. Hayes (Ed.) *Cognition and language development*. Wiley, New York, pp. 279-362.

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, *63*, 489-505.

Blanchard, H. E., & Iran-Nejad, A. (1987). Comprehension processes and eye movement patterns in the reading of surprise-ending stories. *Discourse Processes, 10,* 127-138.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59, 55-73.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and Language*, 117, 133-152.

Boston, M. F., Vasishth, S., Patil, U. & Kliegl, R. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2, 1-12.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127-143.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5, 34-44.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *Plos ONE*, 5(6), e10729.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *Plos ONE, 5(6), e10729.*

Camblin, C. C., Gordon, C. G., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103-128.

Camblin, C.C., Gordon, P.C., Swaab, T.Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory & Language, 56,* 103-128**.**

Caramazza, A. & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language, 3*, 572-582.

Carpenter, P.A., & Just, M.A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, 82, 45-73.

Choi, W., & Gordon, P. C. (2012). Coordination of Word Recognition and Oculomotor Control During Reading: The Role of Implicit Lexical Decisions. Journal of Experimental Psychology: Human Perception and Performance. Advance online publication. doi: 10.1037/a0030432

Coulson, S., King, J. and Kutas, M. (1998). Expect the Unexpected: Event-Related Brain Response to Morphosyntactic Violations. *Language and Cognitive Processes*, 13(1), 21-58.

Curran, T., Tucker, D. M., Kutas, M. & Posner, M.T. (1993). Topography of the N400: brain electrical activity reflecting semantic expectancy. *Electroencephalography and clinical Neurophysiology, 88*, 188-209.

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. Brain Research, 1155, 147–162.

Davidson, P.R. & Wolpert, D.M. (2005). Widespread access to predictive models in motor system: a short review. *Journal of Neural Engineering*, 2, S313-S319.

Deacon, D., Hewitt, S., Yang, C. M., & Nagata, M. (2000). Event-related potential indices of semantic priming using masked and unmasked words: Evidence that the N400 does not reflect a post-lexical process. *Cognitive Brain Research*, 9, 137-146.

Delong, K.A., Urbach, T.P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117-1121.

Dikker, S. & Pylkkänen, L. (2011). Before the N400: effects of lexical-semantic violations in visual cortex. *Brain and Language*, *118*, 23-28.

Dikker, S., Rabagliati, H., & Pylkkänen, L. (2009). Sensitivity to Syntax in Visual Cortex. *Cognition*, 110, 293-321.

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science*, *21*, 629-34.

Dimigen, O., Sommer, W., Dambacher, M., & Kliegl, R. (2008). Simultaneous recording of eye movements and ERPs indicates early access to word meaning in natural, left-to-right reading. International Journal of Psychology, 43(3–4), 47.

Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Co-registration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140, 552-572.

Ditman, T., Holcomb, P.J., & Kuperberg, G.R. (2007). The contributions of lexico-semantic and discourse information to the resolution of ambiguous categorical anaphors. *Language and Cognitive Processes*, 22, 793-827.

Ehrlich, K. (1983). Eye movements in pronoun assignment: A study of sentence integration. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 253-268). New York: Academic Press.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*, 149-161.

Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002).The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39,* 133–146.

Federmeier, K.D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44, 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.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.

Ferreira, F., Bailey, K.G.D. & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, 11, 11-15.

Ferreria, F., & M Henderson, J. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language*, 30(6), 725-745.

Fischler, I., Bloom, P., Childers, D., Roucos, S., & Perry, N. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, 20, 400-409.

Fodor. J.D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry*, 9, 427- 473.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A Competence Based Theory of Syntactic Closure. In J. Bresnan (Ed.) *The Mental Representation of Grammatical Relations*. Cambridge, MA. The MIT Press.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (ed.), *Attention and Performance XII: The Psychology of Reading*, (pp. 559-586). Hillsdale, NJ: Lawrence Erlbaum Associates.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178-210.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain and Language, 50*, 259-281.

Friederici, A. D., Hahne, A., & von Cramon, D. Y. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: Electrophysiological evidence for a double dissociation. *Brain and Language*, 62(3), 311-341.

Friederici, A.D., Steinhauer, K., Frisch, S., (1999). Lexical integration: sequential effects of syntactic and semantic information. *Memory and Cognition*, 27, 438-453.

Gibson, E., Stearns, L., Bergen, L., Eddy, M., & Fedorenko, E. (2013). The P600 indexes rational error correction within a noisy-channel model of human communication. Talk presented at the 26th annual CUNY Human Sentence Processing Conference, Columbia, SC.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883-899.

Hagoort, P., Baggio, G., & Willems, R. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences*, 4th ed. (pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The Syntactic Positive Shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439-484.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438-441.

Hahne, A., & Friederici, A.D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, 13, 339-356.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of NAACL-2001* (pp. 159–166). Stroudsburg, PA: Association for Computational Linguistics.

Hanulikova, A., van Goch, M., & van Alphen, P. (2010). When grammatical errors do not matter: An ERP study on the effect of foreign-accent on syntactic processing. Talk presented at the 23rd annual CUNY Human Sentence Processing Conference, New York, NY.

Hawkins, J. (2004). *On Intelligence*. New York: Henry Holt.

Hoeks, J.C.J., Stowe, L.A., Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59-73.

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, *14*, 1125-1138.

Hyönä, J. (1995). An eye movement analysis of topic-shift effect during repeated reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1365-1373.

Jackendoff, R. (2002). *Foundations of Language*. Oxford: Oxford University Press.

Just, M.A., & Carpenter, P.A. (1978). Inference processing during reading: Reflections from eye fixations, In J.W. Senders, D. E Fisher, & R. A. Monty (Eds.), *Eye movements and the higher psychological functions (pp.* 157-174). Hillsdale, N J: Eflbaum.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 5*, 159–201.

Kazanina, N., Lau, E., Lieberman, M., Yoshida, M., & Phillips, C. (2007). The Effect of Syntactic Constraints on the Processing of Backwards Anaphora. *Journal of Memory and Language*, *56*, 384-409

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, 45(2), 153–168.

Kim, A. & Sikos, L. (2011). Conflict and surrender during sentence processing: An ERP study of syntax-semantics interaction. *Brain and Language*, 118, 15-22.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52, 205-225.

Kolk, H.H.J., Chwilla, D.J., van Herten, M., & Oor, P. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language*, 85, 1-36.

Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58,* 239-261.

Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. Neuroreport, 20(18), 1613-1618.

Kuperberg GR. Language in schizophrenia Part 1: An Introduction. *Language and Linguistic Compass*. 2010; 4(8): 576-589.

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, 21(5), 489-530.

Kuperberg, G., Kreher, D.A., Sitnikova, T., Caplan, D., & Holcomb, P. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language*, 100, 223-238.

Kuperberg, G.K. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23-49.

Kuperberg, G.R., Sitnikova, T., Caplan, D., Holcomb, P.J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, 217, 117-129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4, 463-470.

Kutas, M., & Federmeier, K. D. (2011) Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.

Kutas, M., & Hillyard, S.A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.

Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.

Kutas, M., van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994-2005. In M. Traxler & M.A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659-724). New York: Elsevier.

Laszlo, S., & Federmeier, K. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, 61, 326-338.

Lau, E., Stroud, C., Plesch, S., & Phillips, C. (2006). The Role of Structural Prediction in Rapid Syntactic Analysis. *Brain & Language, 98,* 74-88 (2006).

Lau, E.F., Holcomb, P.J., & Kuperberg, G.R. (2013). Dissociating N400 effects of prediction from association in single word contexts. *Journal of Cognitive Neuroscience*, *25*, 484-502.

Lau, E.F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9, 920-933.

Ledoux, K., Camblin, C. C., Swaab, T. Y., & Gordon, P. C. (2006). Reading words in discourse: The modulation of lexical priming effects by message-level context. *Behavioral and Cognitive Neuroscience Reviews, 5,* 107-127.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Liu, W., Inhoff A. W.,Ye, Y., & Wu, C. (2002). Use of parafoveally visible characters during the reading of Chinese sentences. *Journal of Experimental Psychology:Human Perception and Performance,* 28, 1213-1227.

Llinás, R.R. (2002). *I of the vortex: from neurons to self*. Cambridge, MA: MIT Press.

MacDonald, M.C., Pearlmutter, N.J., & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676-703.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71-102.

McElree, B. & Griffith, T. (1995). Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 134-157.

Mitchell, D., Shen, X., Green, M. J., Hodgson, T. L. (2008). Accounting for regressive eye-movements in models of sentence processing: A reappraisal of the Selective Reanalysis hypothesis, *Journal of Memory and Language, 59*, 266-293.

Moreno, E. and Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Brain Research: Cognitive Brain Research*, *22*, 205-220.

Nieuwland, M.S., & Kuperberg, G.R. (2008). When the truth isn't too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19, 1213-1218.

Oishi, H., & Sakamoto, T., (2010). Immediate interaction between syntactic and semantic outputs: evidence from event-related potentials in Japanese sentence processing. Poster presented at the 22nd annual CUNY Human Sentence Processing Conference, Davis, CA.

Omaki, A., Lau, E. F., Davidson White, I., & Phillips, C. (forthcoming). Hyper-active gap filling: Pre-verbal object gap creation in English filler-gap dependency processing.

Osterhout, L., & Holcomb, P.J. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Osterhout, L., & Nicol, J. (1999). On the distinctiveness, independence, and time course of the brain responses to syntactic and semantic anomalies. *Language and Cognitive Processes*, 14, 283-317.

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.

Paczynski, M., & Kuperberg, G.R. (2011). Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. *Language and Cognitive Processes*, 26 (9), 1402-1456.

Paczynski, M., & Kuperberg, G.R. (2012). Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of Memory and Language*, 67 (4), 426-448.

Phillips, C. (2006). The real-time status of island phenomena. *Language*, *82*, 795-823.

Phillips, C., Kazanina, N., & Abada, S. (2005). ERP Effects of the Processing of Syntactic Long-distance Dependencies. *Cognitive Brain Research, 22*, 407-428.

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, 447-475.

Picton, T. W., Van Roon, P., Armilio, M. L., Berg, P., Ille, N., & Scherg, M. (2000). The correction of ocular artifacts: A topographic perspective. Clinical Neurophysiology, 111, 53–65.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7,* 65-81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., Slowiaczek, M.L., Clifton, C., & Bertera, J.H. (1983). Latency of sequential eye movements: Implications for reading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 912-922.

Reingold, E. M., Reichle, E. D., Glaholt, M. G., & Sheridan, H. (2012). Direct lexical control of eye movements in reading: Evidence from a survival analysis of fixation durations. *Cognitive Psychology, 65*, 177-206.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In Proceedings of the 2009 conference on empirical methods in natural language processing (pp. 324–333).

Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition, 122*, 267-279.

Rugg, M. D. (1985) The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology*, 22, 642–647.

Salthouse, T. A., Ellis, C. L., Diener, D. C., & Somberg, B. L. (1981). Stimulus processing during eye fixations. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 611-623.

Salthouse, T.A., & Ellis, C.L. (1980). Determinants of eye-fixation duration. *American Journal of Psychology*, 93, 207-234.

Schwanenflugel, P.J., & LaCount, K.L. (1988). Semantic relatedness and the scope of facilitation for upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 344 ‑ 354.

Schwanenflugel, P.J., & Shoben, E.J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language*, *24*, 232 ‑ 252.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493.

Shebilske, W. L., & Fisher, D. E (1983). Eye movements and context effects during reading of extended discourse. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes (pp.* 153-179). New York: Academic Press

Slobin, D. (1966). Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior, 5*, 219-227.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.

Staub, A., & Clifton, C., Jr. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 425-436.

Staub, A., Grant, M., Astheimer, L., & Cohen, A. (forthcoming). The Cloze Race Model: A computational account of lexical prediction in the cloze task.

Stroud, C. (2008). Structural and semantic selectivity in the electrophysiology of sentence comprehension. Ph.D. Dissertation, University of Maryland College Park.

Stroud, C., & Phillips, C. (2012). Examining the evidence for an independent semantic analyzer: An ERP study in Spanish. *Brain and Language*, 120 (2), 108-126.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*, 1004–1006.

Townsend, D.J., & Bever, T.G. (2001). *Sentence Comprehension: The Integration of Habits and Rules.* Cambridge, MA: MIT Press.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, 33, 285-318.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 19(3), 528-553.

Trueswell, J.C., Sekerina, I., Hill, N.M. & Logrip, M.L. (1999). The kindergartenpath effect: studying on-line sentence processing in young children. *Cognition*, *73*, 89-134.

Tsai, J-L., Lee, C-Y., Tzeng, O.J.l., Hung, D.L., & Yen, N-S. (2004). Use of phonological codes from Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain and Language, 91,* 235-244.

Urbach, T. F., & Kutas, M. (2010). Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. *Journal of Memory and Language*, 63, 158-179.

Van Berkum, J. J. A. (2009). The neuropragmatics of 'simple' utterance comprehension: An ERP review. In U. Sauerland, & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276-316). Basingstoke: Palgrave Macmillan.

Van Berkum, J.J.A., 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, 443-467.

Van Berkum, J.J.A., Van den Brink, D., Tesink, C.M.J.Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580-591.

Van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. Th.W.M. (2009). Monitoring in language perception. *Language and Linguistics Compass*, 1211-1224.

Van de Meerendonk, N., Kolk, H. H. J., Vissers, C. Th.W.M., & Chwilla, D. J. (2010). Monitoring in Language Perception: Mild and Strong Conflicts Elicit Different ERP Patterns. *Journal of Cognitive Neuroscience*, 22, 67-82.

Van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13(7), 967-985.

Van Herten, M., Chwilla, D.J., & Kolk, H.H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience*, 18, 1181-1197.

Van Herten, M., Kolk, H.H.J., & Chwilla, D. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, 22, 241-255.

Van Petten, C., & Luka, B. J. (2012) Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176–190.

Vauras, M., Hy6n~i, J., & Niemi, E (1992). Comprehending coherent and incoherent texts: Evidence from eye movement patterns and recall performance. *Journal of Research in Reading, 15,* 39-54.

Wagers, M. W., & Phillips, C. (in press). Going the distance: Memory and decision making in active dependency construction. *Quarterly Journal of Experimental Psychology*.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Attraction in comprehension: representations and processes. *Journal of Memory and Language, 61*, 206-237.

Wang. S., Tong, X., Yang, J., & Leng, Y., (2009). Semantic codes are obtained before word fixation in Chinese sentences reading: Evidence from Eye-movement. *Acta Psychologica Sinica* (China), 41, 220-232.

Wicha, N.Y.Y., Moreno, E.M., Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy,

and gender agreement in Spanish sentence reading. Journal of Cognitive Neuroscience, 16, 1272-1288.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *Neuroimage*, *62*, 356-366.

Wlotko, E. W., Federmeier, K. D., & Kutas, M. (2012). To predict or not to predict: Age-related differences in the use of sentential context. *Psychology and Aging*, *27*, 975-988.

Wlotko, E., Lee, C., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass*, *4*, *623-638*.

Wolpert, D.M. (1997). Computational approaches to motor control. Trends in Cognitive Sciences 1: 209-216.

Yan, M., Richter, E., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review,* 16, 561-566.

Yang, J. (2010). Word recognition in the parafovea: An eye movement investigation of Chinese reading. Ph.D. Dissertation, University of Massachusetts Amherst.

Yang, J., Wang, S., Xu, Y., & Rayner, K. (2009). Do Chinese readers obtain preview benefit from character n+2? Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 1192-1204.

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, 1203, 103-115.

Yen, M-H., Tsai, J-L., Tzeng, O. J.I., & Hung, D. L. (2008). Eye movements and parafoveal word processing in reading Chinese. *Memory & Cognition*, 36, 1033-1045.

Yoshida, M., Dickey, M. W., & Sturt, P. (2012). Predictive Processing of Syntactic Structure: Sluicing and Ellipsis in Real-Time Sentence Processing. *Language and Cognitive Processes*, *28*, 272-302.