ABSTRACT

Title of Dissertation: THE COMPUTATION OF VERB-ARGUMENT RELATIONS IN ONLINE SENTENCE COMPREHENSION

Chia-Hsuan Liao
Doctor of Philosophy
2020

Dissertation directed by: Professor Ellen Lau
Department of Linguistics

Understanding how verbs are related to their arguments in real time is critical to building a theory of online language comprehension. This dissertation investigates the incremental processing of verb-argument relations with three interrelated approaches that use the event-related potential (ERP) methodology. First, although previous studies on verb-argument computations have mainly focused on relating nouns to simple events denoted by a simple verb, here I show by investigating compound verbs I can dissociate the timing of the subcomputations involved in argument role assignment. A set of ERP experiments in Mandarin comparing the processing of resultative compounds (*Kid bit-broke lip*: the kid bit his lip such that it broke) and coordinate compounds (*Store owner hit-scolded employee*: the store owner hit and scolded an employee) provides evidence for processing delays associated with verbs instantiating the causality relation (breaking-BY-biting) relative to the coordinate

relation (hitting-AND-scolding). Second, I develop an extension of classic ERP work on the detection of argument role-reversals (the millionaire that the servant fired) that allows me to determine the temporal stages by which argument relations are computed, from argument identification to thematic roles. Our evidence supports a three-stage model where an initial word association stage is followed by a second stage where arguments of a verb are identified, and only at a later stage does the parser start to consider argument roles. Lastly, I investigate the extent to which native language (L1) subcategorization knowledge can interfere with second language (L2) processing of verb-argument relations, by examining the ERP responses to sentences with verbs that have mismatched subcategorization constraints in L1 Mandarin and L2 English ("My sister listened the music"). The results support my hypothesis that L1 subcategorization knowledge is difficult for L2 speakers to override online, as they show some sensitivity to subcategorization violations in offline responses but not in ERPs. These data indicate that computing verb-argument relations requires accessing lexical syntax, which is vulnerable to L1 interference in L2. Together, these three ERP studies allow us to begin to put together a full model of the sub-processes by which verb-argument relations are constructed in real time in L1 and L2.

THE COMPUTATION OF VERB-ARGUMENT RELATIONS IN ONLINE SENTENCE COMPREHENSION

by

Chia-Hsuan Liao

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
2020

Advisory Committee:
Professor Ellen Lau, Chair
Professor Colin Phillips
Professor Robert Slevc
Professor Alexander Williams
Professor Yi Ting Huang, Dean's Representative

Dedication

To my family

Acknowledgements

Since I entered the realm of Neurolinguistics as an undergraduate, I have been dreaming of becoming a learned neurolinguist one day. I knew that getting on board with PhD training could pave the way, and I considered it a blessing that God led me to Maryland. As this PhD journey is coming to an end, I would like to take this opportunity to thank all the people that have shaped me into who I am today, and my friends and family that have accompanied me along the way.

First and foremost, I would like to express my deepest gratitude to Ellen Lau, my advisor and committee chair. Undoubtedly, a lot of stress was involved in pursuing a PhD degree, but working with Ellen has made the entire process as enjoyable as it could be. Ellen was able to distill the points of my half-baked experiment ideas and remind me of the big picture of my research questions. She read my work closely and gave me very constructive feedback, no matter how late I sent it to her. She even managed to keep our weekly meetings going during the pandemic, when she had to take care of her kids around the clock because of the lockdown. Even in such difficult time, she has always kept her spirits high. Her positive thinking makes me more open-minded to embrace the challenges ahead. I really appreciate Ellen's advisory style, as Ellen sees me not only as a student, but also as a person. When I was at the rock bottom, she supportively told me, "It's okay to have life events," which gave me the strength to move on. Ellen is such an incredible scientist, teacher and mentor that I could ever have asked for more in my PhD training.

I am grateful to my dissertation committee members. Colin Phillips has pushed me to articulate the interpretations of my model more clearly, and his suggestions have made the arguments in this dissertation more complete. I would like to thank Colin not only for the insights on my work, but also for his support with my scholarship application, job applications and other issues (such as my I-20 crisis). Colin's passion toward research, perspectives of language science, and dedication to students have an immense influence on me. Alexander Williams is so knowledgeable that I have benefited a lot from the discussions with him. There is an anecdote with Alexander that is worth mentioning here. When I scheduled the very first individual meeting with him, I was a bit nervous. It turned out that I had worried too much. He took me to watch the landing of a helicopter of the US army on Chapel Field. We literally discussed my resultative project with the aircraft landing right in front of us. Since then, he has been the very source to ask for advice on my work. I consider myself very lucky to have such an expert in argument structure in my committee. I am also grateful to have Yi Ting Huang and Bob Slvec as my advisory committee. I took their Cognitive Science course, and I benefited a lot from the smart questions they raised for the assigned readings and was impressed by the creative way they led the course. They taught me to think outside of the box, and I really appreciate their comments and suggestions for my dissertation.

In addition to my committee, I would like to thank the entire faculty of the Linguistics Department at Maryland. Special thank goes to Masha Polinsky, who introduced the world of heritage language study to me. We ended up having two extended projects beyond students' class work. Masha gave me the biggest flexibility to approach these research questions and to mentor undergraduate students at my own pace and offered all the needed support. I really learned a lot through this process. I would like to thank Jeff Lidz and Valentine Hacquard, with whom I learned a lot about syntactic bootstrapping, and felt more amazed by children's ability to master their native language. I am also thankful to Naomi Feldman for sharing with me her job interview experiences and Nobert Hornstein for helping me polish my job talk. Moreover, I am appreciative of Tonia Bleam and Peggy Antonisse for sharing their teaching experiences with me; the two years of working as their TA have been fruitful experiences to me. I thank Omer Preminger for the secret tip to conduct my future courses. I really appreciate Bill Idsardi's timely help with many of the administrative issues since the outbreak of the pandemic. I would also like to thank Kim Kwok for helping me with all kinds of paperwork and subject fee reimbursement, and for her friendship beyond administrative work. It is really a blessing to be enrolled in such a student-oriented program.

The idea of pursuing a PhD degree would not have emerged without the strong training from my alma mater, the Department of English at National Taiwan Normal University (NTNU). I would like to thank the teaching faculty there for nurturing me and equipping me with the knowledge to be a linguist. Special thanks are dedicated to Shiao-Hui Chan, who introduced the fantastic world of Neurolinguistics to me. Shiao-Hui detected the sparks in my eyes, and took me under her wings when I knew nothing about the brain and experiment design. I am grateful that she still let me use her lab space and facilities even after I had left NTNU. Most of the data in this dissertation were collected in her lab. In addition to equipment supply, Shiao-Hui has always been there to encourage me when something goes wrong in my life. I am very grateful to what Shiao-Hui has done for me. I would also like to thank Yeu-Ting Liu, who encouraged me to dream big and reminded me of the 'vision' I could achieve. Without his encouragement and warm support, I would not be able to finish the degree. I hope that I can be a good mentor to my students, just like what my mentors have done for me. It would be the best way to pay it forward.

Thanks to the psycholinguistics team at Maryland: Phoebe Gaston, Hanna Muller, Masato Nakamura, Rosa Lee and Aura Cruz Heredia. It has been helpful to discuss my projects with them, regardless of at which stage my projects were. Special thank goes to Aura, who kindly helped me set up the experiment at Maryland while I was collecting data in Taiwan. I would like to thank the research assistants, Anna Liddle, Katte Luckhurst, and Macie McKitrick for the help with EEG data collection.

My appreciation goes to the cohort: Suyoung Bae, Paulina Lyskawa, Phoebe Gaston, Max Papillon and Annemarie van Dooren. In additional to school work, we have had so much fun together, such as panting the walls of the first year's office, and sheltering in a pizza place at DC during a thunderstorm warning. Special thank goes to Suyoung, who was my housemate for two years. While we both once questioned ourselves if we could earn the degree, I am glad that we made it together!

Thanks to the senior cohorts and friends at Maryland: William Matchin, Nick Huang, Wing Yee Chow, Shota Momma, Eric Pelzl and Nur Başak Karataş. I have benefited a lot from the discussions with them, and I am grateful to have their inputs on my projects.

Thanks to my super officemates in the 1314F super-office: Anouk Dieuleveut, Aaron Doliana and Yu'an Yang. After a long day working in the office, nothing was more refreshing than chatting with them, along with Anouk's homemade cookies and bread.

Thanks to Zoe Schlueter and Alix Kowalski, who welcomed me at home when I just arrived in Maryland. My first year would be so banal without them and their cats. In fact, as a dog person, I did not know that I could get along well with cats until I met Bruce the hairless cat (I am not saying it was because Bruce behaved a lot like a dog, but indeed #sweaterbruceisbetterbruce).

Thanks to my friends from the "nervy" lab: Keng-yu Ken Lin, Li-Chuan Matt Ku, and Yu-Chuan Lucy Chiang. We happen to pursue our degree in different cities in the States, which makes experience sharing in the group chat a lot of fun. From sharing photos of the food I cooked, to eliciting linguistic judgment on sentences of interest, thanks for always being there. Thanks to Aymeric Collart and Shih-Chiang Johnny Hu, who helped me with EEG data collection in Taiwan, and shared all kinds of job-related information with me. Thanks to Chia-wen Lo, with whom I feel very comfortable to share ups and downs in and beyond academia; I thank her for her support in this PhD journey and I hope I am the same kind of friend to her.

It is a blessing to be surrounded by a group of supportive and considerate friends. Heartfelt thanks go to Cindy Tsai, Rebecca Chu and Ky To. We literally spent every weekend in the McKeldin library, if not in "Cindy's building." In addition to helping with one another's homework, we did many crazy things together, such as appreciating cherry blossoms at midnight in DC. With Keqin He and Yuling Guo joining us later, we had more fun at Parkside. I really miss those good old days. Thank you for making my five years in Maryland unforgettable.

I would like to express my gratitude to the big family of the Maryland Chinese Bible Study Group: Rebecca Chu, Cindy Tsai, Yuling Guo, Rachel Chang, Hope Shi, Yanlin Qiao, Peipei Hu, Yuyun Peng, Chuan-Fu Lin, Grace Ho, Tracy Zheng, Debbie Chu, Jing Geng, Hongen Zhou, Shanyun Zheng, Sam Kan, Keith Keenan, Grace Kung, Jiali Liang, Joshua Lo, Esther Lo, Tim Koo, Titus Kan, Jianli Liu, Jason Chen and Yong Yang. In particular, Rachel has been a sweet and considerate friend, and I could not imagine how my life would have been without her. Hope invited me to her house, cooked for me and offered to study with me when I was in the lowest spirits. Yanlin has kept me in her prayers and I really appreciate that. Peipei has a great sense of humor and is always enthusiastic; I can always regain strength after chatting with her. Thanks to Yuyun for her prayers at my weakest moment. I am grateful to my godparents Der-Chen Chang and Shimu; I would like to thank them for their prayers and encouragement, which support me to get over difficulties. They all have made Maryland my home away from home.

There were more than 600 participants participating in the work in this dissertation. One third of them let me put an EEG cap on their head. Their contribution is acknowledged here. Lastly, I would like to acknowledge that some of the projects in this dissertation were funded by the William Orr Dingwall Dissertation Fellowship.

While Maryland is tens of thousands of miles away from home, encouragements from my family are the source of my energy. I would like to thank my uncles and aunts for their prayers, and for sending me care packages from Taiwan. My deepest gratitude goes to my parents, who support me to pursue my

dreams without a second word and to my sister, who is always there for me to cheer me up. They are my foundation and support, and this dissertation is dedicated to all of them.

While there have been ups and downs in this journey, God has been my strength and hope. He has empowered me with inner strength through his Spirit (Ephesians 3:16). Reflecting on the paths that I have gone through, I know I am blessed beyond measure. I give thanks to the Lord for He is good all the time. May His name be glorified.

Table of Contents

Dedication ii

Acknowledgements iii

Table of Contents x

List of Tables xiii

List of Figures xiv

Chapter 1: Introduction 1
- 1.1 Overview 1
- 1.2 Incremental processing of verb-argument relations: State of the art 3
- 1.3 Research questions in the dissertation 8
- 1.4 Presuppositions of this dissertation 11
  - 1.4.1 Lexicalized vs. non-lexicalized analyses of argument structure 11
  - 1.4.2 Predictive mechanisms in sentence processing 13
- 1.5 The EEG (electroencephalography) technique 14
  - 1.5.1 N400 14
  - 1.5.2 P600 16
- 1.6 The methodological approach: Taking the timing of prediction as a chronometer for linguistic computations 17
- 1.7 Outline of this dissertation 18

Chapter 2: Enough time to get results? An ERP investigation of prediction with complex events 19
- 2.1 Introduction 19
  - 2.1.1 "Slow prediction" and argument structure 20
  - 2.1.2 The current study 22
- 2.2 Experiment 1 27
  - 2.2.1 Participants 29
  - 2.2.2 Materials 30
  - 2.2.3 Procedure 32
  - 2.2.4 Data acquisition and analysis 33
  - 2.2.5 Results 35
    - 2.2.5.1 *Behavioral data* 35
    - 2.2.5.2 *ERP data* 36
  - 2.2.6 Discussion 38
- 2.3 Experiment 2 40
  - 2.3.1 Participants 42
  - 2.3.2 Materials 42
  - 2.3.3 Procedure 46
  - 2.3.4 Data acquisition and analysis 46
  - 2.3.5 Results 47
    - 2.3.5.1 *Behavioral data* 47
    - 2.3.5.2 *ERP data* 47

2.3.6 Discussion 49
2.4 Experiment 3 51
2.4.1 Participants 51
2.4.2 Materials 52
2.4.3 Procedure 53
2.4.4 Data acquisition and analysis 54
2.4.5 Results 54
2.4.5.1 *Behavioral data* 54
2.4.5.2 *ERP data* 55
2.4.6 Discussion 56
2.5 General Discussion 57
2.5.1 Slow prediction due to the computation of a complex resultative predicate 58
2.5.2 Slow prediction due to memory search for an optimal candidate 60
2.5.3 ERP responses to verbs 63
2.5.4 Implications for L2 acquisition and processing 66
2.6 Conclusion 67

Chapter 3: The time course of argument-verb computation in online sentence comprehension: Evidence from the N400 69
3.1 Introduction 69
3.1.1 Slow vs. fast prediction in sentence comprehension 70
3.1.2 Bag of Words vs. Bag of Arguments hypotheses in argument-verb computation 73
3.1.3 The current study 76
3.2 Experiment 4 77
3.2.1 Participants 78
3.2.2 Materials 79
3.2.3 Procedure 81
3.2.4 Data acquisition and analysis 82
3.2.5 Results 84
3.2.5.1 *Behavioral data* 84
3.2.5.2 *ERP data* 84
3.2.6 Discussion 86
3.3 Experiment 5 87
3.3.1 Participants 88
3.3.2 Materials 88
3.3.3 Procedure 89
3.3.4 Data acquisition and analysis 89
3.3.5 Results 90
3.3.5.1 *Behavioral data* 90
3.3.5.2 *ERP data* 90
3.3.6 Discussion 91
3.4 Experiment 6 93
3.4.1 Participants 93
3.4.2 Materials 94
3.4.3 Procedure 94
3.4.4 Data acquisition and analysis 95
3.4.5 Results 95
3.4.5.1 *Behavioral data* 95
3.4.5.2 *ERP data* 95
3.4.6 Discussion 97

3.5 General Discussion 98
3.5.1 Toward a processing model of argument-verb relation computations............. 101
3.5.2 Reconciling these results with prior role reversal findings............................ 104
3.6 Conclusion 107

Chapter 4: ERP sensitivity to subcategorization violations in L2..................... 108
4.1 Introduction.. 108
4.1.1 Prior behavioral studies investigating the processing of subcategorization in L2 109
4.1.2 ERP studies of subcategorization violation in L1 and L2 speakers................ 112
4.1.3 The current study 116
4.2 Experiment 7... 118
4.2.1 Participants.. 118
4.2.2 Materials.. 118
4.2.3 Procedure 121
4.2.4 Data acquisition and analysis.. 122
4.2.5 Results... 124
4.2.5.1 *Behavioral data*... 124
4.2.5.2 *ERP data* 125
4.2.5.2.1 Verb subcategorization violations 125
4.2.5.2.2 Phrase structure violations.. 127
4.3 General Discussion 128
4.3.1 Responses to subcategorization violations in L1 and L2 129
4.3.2 Knowledge and processing accounts 131
4.3.3 An ELAN response to subcategorization violations in L1 (?) 140
4.3.4 Responses to phrase structure violations in L1 and L2................................... 142
4.4. Conclusion 145

Chapter 5: Conclusion.. 146
5.1 Summary... 146
5.2 Outstanding questions... 149
5.2.1 Connecting the dots, across chapters 149
5.2.2 Broader connections to prior work on verb-argument relation processing...... 151
5.3 Future work 154
5.3.1 Temporal dynamics of computing complex events 155
5.3.2 Types of argument information that feed into parsing along the temporal scale 156
5.3.3 Computations of verb-argument relations in different populations 159

Appendix (Experiment materials).. 161

Bibliography 194

List of Tables

Table 1: Example stimulus in each condition in Experiment 1 (averaged cloze probability in parenthesis).. 25

Table 2: Example stimulus in each condition in Experiment 2 (averaged cloze probability in parenthesis) 45

Table 3: Example stimulus in each condition in Experiment 4 80

Table 4: Example stimulus in each condition in Experiment 5 89

Table 5: Example stimuli in each condition in Experiment 7................................... 120

Table 6: Accuracy rate of each condition for the L2 group...................................... 123

List of Figures

Figure 1: Presentation of stimuli in Experiment 1 and Experiment 2 33

Figure 2: Top: Grand average ERPs from the verb to the noun at Cz and Pz in Experiment 1. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 1 (Left: Causative minus Resultative; Right: Simple minus Resultative) 37

Figure 3: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 1. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 1 (Low cloze minus High cloze) 38

Figure 4: Top: Grand average ERPs from the verb to the noun of the Resultative set (Left) and Coordinate set (Right) at the Cz electrode in Experiment 2. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 2 (Left: R-Simple minus Resultative; Right: C-Simple minus Coordinate). 48

Figure 5: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 2. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 2 (Low cloze minus High cloze) 49

Figure 6: Presentation of stimuli in Experiment 3 54

Figure 7: Top: Grand average ERPs from the verb to the noun of the Resultative set (Left) and Coordinate set (Right) at the Cz electrode in Experiment 3. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 3 (Left: R-Simple minus Resultative; Right: C-Simple minus Coordinate). 55

Figure 8: Left: Grand average ERPs to cloze sanity check at Cz in Experiment 3. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 3 (Low cloze minus High cloze) 56

Figure 9: Left: Topographic distribution of Resultative effect across different time windows in Experiments 1-3. Right: Grand average ERPs from the verb to the noun of Resultative and R-Simple conditions at the Pz electrode in Experiments 1-3 64

Figure 10: Left: Topographic distribution of Coordinate effect across different time windows in Experiments 1-3. Right: Grand average ERPs from the verb to the noun of Coordinate and C-Simple conditions at the Pz electrode in Experiments 1-3 65

Figure 11: Presentation of stimuli in Experiment 4 82

Figure 12: Grand average ERPs to predictability effect of Baseline and Complement at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 4 (Complement minus Baseline) 85

Figure 13: Left: Grand average ERPs to cloze control items at Cz in Experiment 1. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 4 (Low minus High cloze) 86

Figure 14: Grand average ERPs to cloze manipulations of Baseline and Reversal at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 5 (Reversal minus Baseline) 91

Figure 15: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 5. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 5 (Low minus High cloze) 91

Figure 16: Presentation of stimuli in Experiment 6 96

Figure 17: Grand average ERPs to predictability effect of Baseline and Complement at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 6 (Complement minus Baseline)......................... 96

Figure 18: Left: Grand average ERPs to cloze control items at Cz in Experiment 6. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 6 (Low minus High cloze) 94

Figure 19: The three-stage processing model of argument-verb computations..... ...102

Figure 20: Left: Grand average ERPs of the Subcategorization conditions from the determiner to the nouns at electrode Pz. Right: The topographic distributions in the 300-500 and 600-900 ms intervals at the determiner in L1 and L2 speakers 126

Figure 21: Left: Grand average ERPs of the Phrase structure conditions from the preposition to the following word at electrode Pz. Right: The topographic distributions in the 300-500 and 600-900 ms intervals at the preposition in L1 and L2 speakers 127

Figure 22: A schematic diagram of online processing upon reading a verb. The figure is adapted from the Revised Hierarchical Model (Kroll & Stewart, 1994) 135

Figure 23: Left: Grand average ERPs of the Phrase structure conditions from the preposition to the following word at electrode F3. Right: The topographic distributions in the 100-200 ms interval at the preposition in L1 and L2 speakers.. 141

Chapter 1 Introduction

1.1 Overview

Verbs play a key role in human language. Understanding how verbs are related to "argument" phrases in a sentence, like its subject or object, is critical to building a theory of online language comprehension. How many such arguments we find, and what grammatical form they take, depends importantly on the syntactic category of the verb. For example, the sentence "The farmer fled from the wolves" is acceptable, while "*The wolf chased from the farmers" is not, due in part to grammatical differences between "flee" and "chase." In addition, it is the verb that tells us what sort of event the clause describes: a fleeing, a chasing, a running, a hunting, and so on. This in turn informs our understanding of the semantic relations associated with the subject or object. The subject of an active clause with "flee" names the thing that flees, while the subject of an active clause with "chase" names the thing that chases. Finally, which type of semantic relation is associated with a given grammatical relation also depends on the verb. When the verb is "flee," the subject of an active clause links to a relation we might classify as an agent relation: a thing that flees is the agent of an action. But this is not so with "fear." A thing that fears is instead the experiencer of a mental state. So the relation linked to the subject might be classified differently, as an experiencer. In these ways verbs are highly informative about both the syntax and the semantics of the dependent phrases in their grammatical context.

How do comprehenders compute verb-argument relations in real time, such that they are able to interpret who is doing what to whom in a sentence? Verb-argument relations, as an interface of syntax and semantics, provide a window to investigate how syntax and semantics interact to construct the representations of a sentence over time. The dissertation investigates the incremental processes of verb-argument computations. Previous studies on verb-argument computations mainly focus on relating nouns to simple events denoted by a simple verb. In this dissertation, I will show that investigating the processing of compound verbs denoting complex events—up to now an understudied area in psycholinguistics—gives us the opportunity to dissociate the timing of some of the subcomputations involved in argument role assignment. In particular, with a compound verb, the parser has to combine the verbs into a single complex predicate denoting a complex event, and then derive the corresponding set of argument roles. Another goal of the dissertation is to map out the temporal stages by which argument relations are computed, from argument identification to thematic roles. In addition to constructing the temporal scale of verb-argument computations in native speakers, I extend the model by evaluating how bilinguals resolve the conflict online when argument structures between their two languages differ from each other.

The remainder of the introduction is organized as follows. Below in Section 1.2 I will provide an overview of the computation questions that have been addressed in earlier work, and the state-of-the-art psycholinguistics findings that speak to these questions. Then in Section 1.3 I will introduce research questions in the dissertation. In framing these research questions, I find it useful to adopt several assumptions about the grammatical representation of verb-argument relations; these are discussed in Section 1.4. Since the research questions are centered on online computation and its timing, in Section 1.5 I will briefly introduce some basic background on the EEG (electroencephalography) technique, whose temporal resolution allows us to track neural computations that support language comprehension at the millisecond level. Section 1.6 introduces the methodological approach in this dissertation. Section 1.7 gives an outline of this dissertation.

1.2 Incremental processing of verb-argument relations: State of the art

How incremental is the online computation of verb-argument relations? Existing psycholinguistic models have proposed the following, even though the exact details vary among different models (Bornkessel & Schlesewsky, 2006; MacDonald, Pearlmutter, & Seidenberg, 1994): When an argument precedes a verb, its argument role will not be confirmed until the verb is encountered, since argument roles are partially determined by the verb. But what the parser can do incrementally upon encountering the argument is to consider its structural position, as well as what kinds of things it denotes, and make the best estimate of what argument role will be assigned to the argument. After the verb is presented, the pre-activated argument role is then checked against the actual list of semantic relations permitted by the verb. If there is a mismatch between the pre-activated argument role and the list of semantic relations, then the parser will have to reanalyze the thematic relations between the argument and the verb to repair it. By contrast, the mapping of an argument role to a post-verbal argument may be more straightforward. When the verb is encountered, its argument structure information can be accessed to constrain the role of an upcoming argument immediately.

While existing psycholinguistic models agree with the processes stated above, each of them has its own focus in studying the computation of verb-argument relations. For example, for Bornkessel and Schlesewsky (2006), their goal is to construct a model that can account for cross-linguistic variations in verb-argument computations. MacDonald, et al. (1994) believe that ambiguity resolution is involved as comprehenders are trying to compute verb-argument relations online, because in many cases, a verb can assign more than one thematic relation to its argument. Other researchers examine situations when initial parsing goes wrong in order to determine how the process of reanalyzing argument relations is implemented (e.g. Christianson, Hollingworth, Halliwell, & Ferreira, 2001). Below I review some of this work in more detail.

To begin with, Bornkessel and Schlesewsky (2006) propose the extended argument dependency model (the eADM), with the goal to account for cross-linguistic variation. Their idea is that computing verb-argument relations is an elementary component in sentence processing, which should be universal among languages in the world. However, languages vary regarding their restrictions on arguments of different structural positions, and therefore the processing system needs to take this variation into account in computing argument role assignment estimates. For example, Mayan languages like Mam and Jakeltek have animacy hierarchy strictly encoded in syntactic structure, such that an argument of higher animacy rank will take on a structurally higher position (Craig, 1977; Minkoff, 2000). In these languages, "the dog bit the man" is not grammatical, because humans outrank animals in animacy hierarchy. Rather, the intended meaning should be expressed as “the man was bit[ten by] the dog.” Interestingly, if the arguments are inanimate, sentences like “the rock hit the window” and “the window was hit by the rock” are both acceptable, because the two arguments are of the same animacy rank. Different from Mayan languages, other languages such as English do not have such strict restrictions. In English, even though it is still more likely for a subject argument to be animate (Snider & Zaenen, 2006), the animacy hierarchy can be violated (e.g. “the news shocked John”). Prominence assignment in eADM is designed to address such cross-linguistic variation. Although comprehenders of all languages go through identical stages to compute the relations between arguments and a verb, each language has its own prominence hierarchy regarding the set of argument information (such as animacy, definiteness, person). By adjusting the prominence hierarchy, the eADM model is then able to cover cross-linguistic variation in verb-argument computations, such that the pre-verbal argument role estimates take this variation into account.

MacDonald, et al. (1994) emphasize the fact that verb-argument computations usually require ambiguity resolution, because a noun can be linked to a variety of argument roles and a verb usually permits more than one theta grid. MacDonald et al. investigate how this ambiguity resolution is accomplished. Even to process a simple sentence like “John cooked,” comprehenders have to tackle ambiguity resolution, as the parser would have to evaluate semantic and morpho-syntactic features of the argument and the verb respectively in order to accurately compute their relations. In particular, “John,” as an animate argument, is very likely to be an agent. In addition to semantic features of an argument, MacDonald et al. believe that frequency of usage and, potentially, the discourse context can determine which theta grid will be selected for an initial analysis. Here, the verb "cooked" is more frequently used in a transitive context with an active voice, thus the <agent, patient> theta grid is more activated than other ones. In other words, in this model, argument structure and sentence structure are confirmed when a verb is encountered. To compute the verb-argument relations, the argument "John" is linked to the subject position as an agent role of a cooking event. The parser has built up the sentence structure and expected the object argument to receive the patient role. As the sentence turns out to be intransitive, the parser will have to reanalyze the sentence by selecting the <agent> theta grid and its corresponding syntactic structure.

In other cases, like garden path sentences, the parser will have to go further to reanalyze the relations between a verb and its arguments. Consider the famous sentence "the horse raced past the barn fell" as an example. An initial reading would treat the verb "raced" as a simple verb with past tense, which assigns an agent role to the argument "horse." Then when reaching the disambiguating verb "fell," readers realize that "raced" is in fact a passive participle and "horse" now becomes a patient and is the direct object of the subordinate clause. Much work has examined exactly how this process of reanalyzing argument relations is implemented online, notably asking whether suppressing/overwriting interpretative representations that involve reanalyzing argument relations might be particularly difficult (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Sturt, 2007).

Another line of research has asked whether argument role assignment is always faithful to syntax, investigating how and when syntactic and semantic information interacts to construct representations of a sentence over time. One way to approach this question is to reverse the thematic roles of arguments in a sentence, and study whether comprehenders might temporarily mis-parse it and/or generate the wrong interpretation. For example, in “the hearty meal was devouring” (Kim & Osterhout, 2005), “the hearty meal” is supposed to be a patient of the devouring event, rather than an agent. Much of this line of research has been carried out with event-related potentials (ERP) measurements focusing on the amplitude of the N400 response, which is often taken to serve as a neural indicator of lexical/conceptual pre-activation (Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008). Surprisingly, although “devouring” should not be an expected event when “the hearty meal” serves as an agent, the N400 modulation is not sensitive to it. Such a finding has been replicated in different languages with different verb final structures (Chow & Phillips, 2013; Chow, Smith, Lau, & Phillips, 2016; Hoeks, Stowe, & Doedens, 2004; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Momma, Sakai, & Phillips, 2015). Some authors take these results to support theories in which argument role assignment can diverge from syntax, for example an independent syntax-free system that derives thoughts from word associations (Kim & Osterhout, 2005), or a system in which the parser spells out an implausible analysis, but plausibility heuristics come up with a plausible interpretation (van Herten, Chwilla, & Kolk 2006). However, a number of other authors have argued that these conclusions are premature, noting that the absence of appropriate lexical expectations does not entail that an incorrect sentence meaning has been generated (Brouwer, Fitz, & Hoeks, 2012). Chow, Lau, Wang, & Phillips (2018) posit that the problem should be attributed to limited amount of time to compute thematic relations (i.e. “meal-as-an-agent”), and/or using the thematic information to update predictions of the upcoming verb online. I am inspired by Chow et al.’s (2018) timing account, and will further map out the time course of when different pieces of argument information contribute to prediction of a verb in this dissertation (see more details below and Chapter 3).

Taken together, previous work has investigated incremental processing of verb-argument relations from different perspectives. However, not many of them have provided the temporal information by which such relations are computed. In addition, previous studies on verb-argument computations mainly focus on relating nouns to simple events denoted by a simple verb. In this dissertation, I will investigate the processing of compound verbs denoting complex events, which gives us the opportunity to dissociate the timing of some of the subcomputations involved in argument role assignment.

1.3 Research questions in the dissertation

Building on prior work, I am going to investigate three different but equally important processing questions on verb-argument computations in this dissertation. The three research questions are listed below, and I will introduce each of them in the following paragraphs:

1. How quickly can complex verb-argument relations be computed to impact the prediction of a subsequent argument?
2. What are the temporal stages by which argument relations are computed, from

argument identification to thematic roles?

3. How much does L1 argument structure interfere with L2 processing?

To address the first question, in Chapter 2, I will take advantage of the substantial argument structure differences provided by Mandarin in this domain to investigate the timing of argument structure computation. In a Mandarin resultative construction like *Mom washed-ruined the clothes*, the second noun phrase *clothes* is understood to name both what was washed and what was thereby ruined. In contrast, the Mandarin coordinate construction like *Storeowner hit-scolded the employee, employee* is understood to name both what/whom was hit and what/whom was scolded. Then how do comprehenders rapidly compute these kinds of complex relationships to reach the intended message? My work has taken advantage of recent advances in cognitive science indicating that human processing is extremely predictive (more details below in Section 1.4.2). If I set up situations in which accurate prediction depends on having computed certain verb-argument relations, then evidence of successful prediction can tell me how quickly those relations are computed. Here this approach allows me to temporally disassociate the subcomputations required for resultative and coordinate complex verbs. With the same approach, I shall be able to investigate subcomputations of other types of compound verbs in the future, such that I can map out the time course of argument role computations for complex events.

The next chapter (Chapter 3) aims to identify the time course for different kinds of argument information to be computed for argument-verb relation computation. As briefly discussed in Section 1.2, no N400 difference is found between role reversal sentences ("The waitress that the customer served") and their canonical baseline ("The customer that the waitress served"). This observation has often been taken to provide insight into the speed of argument structure computation—in other words, that argument role computation is particularly slow to be incorporated into predictions for the verb (Chow, Momma, et al., 2016). Chow, Smith et al. (2016) propose initial verb prediction is driven by the identity of the arguments in the same clause as the verb, but not argument roles (the Bag of Arguments hypothesis). Here I aim to both provide stronger evidence for this hypothesis, and to map out a more detailed time course by which noun phrases are identified as arguments of the verb. I achieve this by a more tightly controlled manipulation of the presence of a clause boundary, as in "[The waitress thought [the customer served ____]]", and evaluating sensitivity to this boundary on the N400 to the verb in two experiments that varied in stimulus-onset asynchrony (SOA). On the basis of a series of EEG experiments, I will propose a processing model of the stages comprehenders go through to compute argument-verb relations.

In the last experimental chapter of the dissertation (Chapter 4), I begin to extend my investigation of the time course of online argument role computation to the case of late second language learners. My focus here is on the impact of conflicting L1 argument structure knowledge on the accuracy of early L2 argument structure computation. Event concepts of common verbs are likely to be the same for speakers of different languages (e.g. eat, sleep), but the argument structure of a verb is linguistic knowledge, and could vary from language to language. For example, in

English, “bark” takes only one argument whereas it can take two arguments in Mandarin. Therefore, English native speakers reject sentences like “The dog is barking the girl” but those sentences might be accepted by English learners whose L1 is Mandarin, as the direct translation is grammatical in Mandarin. Previous studies suggest that L2 speakers, despite being more vulnerable in constructing hierarchical details of sentence structures, are able to rapidly and accurately compute verb-argument relations (Clahsen & Felser, 2006). However, even though verb-argument computation is a local structure, it involves accessing lexical syntax, which might be prone to L1 transfer when processing in real time. The results may have implications for how L2 argument structure is represented in the lexicon.

1.4 Presuppositions of this dissertation

In the dissertation, I will assume that (1) argument structure information is encoded in verbs and (2) the mechanism of sentence processing is predictive. In the sections below, I will first present different linguistic analyses of argument structure (lexicalized vs. non-lexicalized views) based on Williams (2015). Then, I will briefly touch on the debate about whether sentence comprehension is predictive. Empirical studies will be reviewed to show that the engagement of a predictive mechanism in sentence processing is currently the dominant view.

1.4.1 Lexicalized vs. non-lexicalized analyses of argument structure

The lexicalized view suggests that argument structure is encoded in a verb, and that the verb is the head of the structures it projects (e.g. Chomsky, 1981; 1995).

For example, under the traditional Government and Binding framework, the verb "chase" has the following information stored in the lexicon:

(1) Chase: Verb [Subcategorization frame <Noun_1, Noun_2>, Theta grid <Θ_1, Θ_2>]

To derive the expression "chased the dog," the verb "chase" would check if the category of "the dog" matches the values in the argument list (i.e. information in the brackets in (1)). "Chase" requires a noun as its dependent, and "the dog" belongs to the noun category, so "chase" can have a syntactic argument "the dog." Since the verb "chase" is the head, the result of such combination still belongs to the category of verbs.

By contrast, the non-lexicalized view suggests that argument information is not coupled with a verb. The specific dependencies can be stated in constructions (e.g. Goldberg, 1995). Constructions can be seen as structured templates that have functions and meanings, but they are not phonologically interpreted. For example, in a transitive construction, it is the construction (i.e. subject-verb-object) that endows a sentence transitive interpretation. Since argument structure is not specified in the lexicon, the representation of a verb can be presented below.

(2) Chase: Verb [*t*]

In this example, *t* is a simple contextual feature that corresponds to the transitive construction. Note that such information is very vague; all the information about what transitive structures look like is stored as part of the transitive construction itself, not individual verbs. Sentences are derived via filling words into constructions.

As the goal of the dissertation is to investigate how and when verb-argument relations are computed in sentence processing, rather than teasing apart predictions from the lexicalized and the non-lexicalized views, for convenience of exposition in the dissertation, I will assume the lexicalized view of argument structure encoding. However, it is probable that our findings could be restated in the non-lexicalized view where what is encoded with the verb is not argument structure per se but information about which constructions the verb can or is likely to occur in.

1.4.2 Predictive mechanisms in sentence processing

Among psycholinguists, the extent to which sentence comprehension is predictive has long been debated. Some classic arguments for prediction are the fact that listeners can respond to another interlocutor immediately in conversations, and sometimes can even fill in a particular word that the other person fails to produce (Schegloff, 2000). By contrast, arguments against prediction are based on the "low payoff" intuition, that the benefits gained from a successful prediction might not equal the costs of frequently revising wrong predictions (Forster, 1981; Jackendoff, 2002). However, in recent decades, more and more empirical findings show that comprehension involves some degree of prediction. Behavioral and ERP work has shown that predictive sentence contexts have a robust facilitatory influence on the processing of the subsequent word (see Kutas, Delong, & Smith, 2011; Van Petten & Luka, 2012, for review, and more discussions on ERP work in Section 1.5.2 below).

Eye tracking work has also been able to demonstrate such effects prior to the critical word. For example, given a scene of a man, a girl, a motorcycle, and a carousel, and presented with the sentence frame “the man likes to ride ____,” participants tend to look at the picture of a motorcycle, whereas given the context “the girl likes to ride ____,” participants tend to look more at the picture of the carousel (Kamide, Altmann, & Haywood, 2003). These examples and others have been taken to suggest that comprehenders quickly integrate information from the context to predict what is coming next.

1.5 The EEG (electroencephalography) technique

Since all the questions addressed in this dissertation are about the time course of the incremental processes of verb-argument computation, the EEG technique is an appropriate tool to probe these questions. A major advantage of EEG is that it provides a direct measure of real time brain activity at the millisecond level. In addition, participants can read/listen to the experiment materials for comprehension without making an explicit response. Below, I will briefly review the key properties of two ERP components which will play a central role in the experiments conducted here.

1.5.1 N400

The amplitude of the N400 response in ERP has frequently been used to track the prediction of lexical and conceptual material (Kutas & Federmeier, 2000; Lau, Phillips, & Poeppel, 2008). The N400 component peaks between 300-600 ms after the onset of the stimulus presentation, and is negatively correlated with the predictability of a target word (Kutas & Hillyard, 1984). Predictability is usually operationalized in these experiments by the construct of cloze probability, which is the percentage of responses that a given word occurred in a separate offline completion task (Taylor, 1953). For example, given a sentence context like "He was afraid that doing drugs would damage his ____," a majority of participants in the offline norming might complete the sentence with "brain" and a minority with "reputation," and the predicted high-cloze continuation "brain" would then elicit a significantly smaller N400 than the less predicted low-cloze "reputation" (Thornhill & Van Petten, 2012). These kinds of effects have been frequently replicated (Federmeier & Kutas, 1999; Federmeier, 2007; Thornhill & Van Petten, 2012). Many authors have taken these results to indicate that linguistic input is predicted in context (although it is worth noting that a non-predictive explanation is possible in which these effects are due to variations in integration difficulty after the bottom-up input is encountered).

Whether N400 reductions reflect conceptual pre-activation, lexical pre-activation, or both, is still an open question. Consistent with a conceptual component, N400 responses are observed for meaningful pictures and environmental sounds as well as spoken and written words, and N400 modulations are observed when sentences are completed with pictures (Kutas & Federmeier, 2011). Federmeier and Kutas (1999) observe N400 reduction for unexpected completions that were semantically related to the expected completion ("They wanted to make the hotel look more like a tropical resort, so along the driveway they planted rows of palms/pines/tulips"), consistent with the idea that the conceptual features themselves are pre-activated by the context, although other accounts are also possible (see Thornhill & Van Petten, 2012, for similar results). Consistent with a lexical pre-activation component, work by Laszlo and Federmeier (2009) shows N400 sensitivity to unexpected words that are orthographically related to the expected ending. Brothers, Swaab, & Traxler (2015) show that on a trial-by-trial basis, N400s are reduced earlier and to a greater extent for words that participants have specifically predicted than words that are simply contextually supported. Together, I take these different lines of work to suggest that N400 effects reflect a combination of pre-activating conceptual features and pre-activating specific lexical items. N400 amplitude will thus be my key dependent measure of context-based prediction in Chapters 2 and 3.

1.5.2 P600

The P600 response lacks a clear peak, but is usually more prominent in the 500-900 ms interval after the onset of a problematic word over the parietal sites (Osterhout & Holcomb, 1992; Hagroot, Brown, & Groothusen, 1993). The P600 effect has often been associated with grammaticality violations, such as phrase structure violations, argument structure violations and agreement errors (Friederici & Frisch, 2000; Hagroot, et al., 1993;). However, more and more studies reveal that P600 effect can be observed by grammatical but structurally complicated sentences, such as garden path sentences, or some kinds of semantic incongruity as in argument role reversal sentences (Hagoort & Brown 2000; Kim & Osterhout, 2005; Osterhout, Holcomb, & Swinney, 1994). More interestingly, the P600 effect can also be elicited by errors in musical harmony, such as when a chord is played out of key with the rest of a musical phrase (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). The above evidence suggests that the P600 effect is sensitive to rule-governed sequences, and it is a domain-general response. Taken together, I take the P600 as an index of difficulty in structure analysis (Osterhout & Holcomb, 1992; Kaan & Swaab, 2003). P600 amplitude will be my key dependent measure indexing the detection of subcategorization frame violations in Chapter 4.

1.6 The methodological approach: Taking the timing of prediction as a chronometer for linguistic computations

In this dissertation, I am going to take advantage of predictive mechanisms in sentence comprehension as a tool for studying the time course of argument role computation. Since successful prediction relies on completing linguistic analysis in the context, the timing of prediction can be used to probe how long it takes to compute particular linguistic analyses. This approach has been successfully implemented in recent work (Federmeier, Kutas, & Schul, 2010; Chow, et al., 2018; Momma, et al., 2015). For example, in Chow et al. (2018), the authors use the N400 as an index of prediction to estimate how long it takes to update predictions of an upcoming verb on the basis of argument roles. They manipulate the linear distance between an argument and a verb, by varying the position of an adverbial temporal phrase in a sentence. In the short-distance conditions, an adverbial temporal phrase is placed at the beginning of the sentence, and the argument is immediately followed by the verb (*Last week, thief ba* <u>*cop*</u> <u>*arrest*</u> meaning: "The thief arrested the cop last week"); in the long-distance condition, the adverbial phrase is placed between the argument and the verb (*Thief ba* <u>*cop*</u> *last week* <u>*arrest*</u>, with the meaning being identical as the short-distance condition), which creates a buffer (around 1800 ms) to formulate the prediction of an upcoming verb. The results show that relative to their canonical baseline ((*Last week*) *the cop ba the thief* (*last week*) *arrested*), there is no N400 effect in short-distance conditions, but critically the N400 response is recovered in long-distance conditions. This kind of data provides an initial framework for developing a detailed time course model of top-down interpretation processes. We will use the same approach to investigate the time course of verb-argument computations in Chapters 2 and 3.

1.7 Outline of this dissertation

The structure of the dissertation is as follows. Chapter 2 examines the time course of verb-argument computations with complex events. Chapter 3 maps out the temporal stages by which argument relations are computed, from argument identification to thematic roles. Chapter 4 investigates how the L2 speakers react to the conflict when argument structure between L1 and L2 does not match. Chapter 5 concludes the dissertation by summarizing the results of each chapter and discussing outstanding questions for future explorations.

Chapter 2: Enough time to get results? An ERP investigation of prediction with complex events

Part of this chapter has been published as Liao, C-H & Lau E. (2020). Enough time to get results? An ERP investigation of prediction with complex events. *Language, Cognition, and Neuroscience.* 1-21.

2.1 Introduction

This chapter uses the timing of prediction to investigate the mechanisms by which complex verb relations are computed online. Mandarin Chinese has a highly productive system of compound verbs—such as coordinate verbs (X hit-scolded Y, meaning X hit and scolded Y) and resultative verbs (X bit-broke Y, meaning X bit Y and in doing so caused Y to break)—which require mechanisms to combine the verbs into a single complex predicate denoting a complex event, and to derive the corresponding set of argument roles. If each individual verb of the compound predicts a very different object than the compound as a whole, how long does that update take, and does it depend on the nature of the meaning relation? While many studies will be needed to develop a full model of these highly complex processes, we hope to show that this new method provides a way to successfully dissociate some of them experimentally. There exists another body of work that has asked about how comprehenders deal with simple verb-object structures that require more complex semantics. For example, with coercion structures, additional analysis is needed, such that "begin the book" is understood as "begin 'doing something' with the book"

(Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010); with light-verb constructions such as "John gave Mary a kiss," the verb "gave" denotes a general sense of transfer, and the event nominal "kiss" conveys the action itself. Therefore, "give" and "kiss" would share the arguments in this context: "John" is the agent of "give" and "kiss," and "Mary" is the recipient of "give" and the patient of "kiss" (Wittenberg, Paczynski, Wiese, Jackendoff, & Kuperberg, 2014). However, to our knowledge, relatively little is known thus far about the processing algorithms by which complex verb structures are interpreted, even though they are pervasive in many languages. Here we take a preliminary step towards disentangling the fine-grained linguistic and conceptual subcomputations that are likely to be required, by comparing the speed of prediction update associated with coordinate compounds and resultative compounds in Mandarin. Our results show that prediction for the object noun is not immediately updated by information from Resultative verbs. Resultative verb structures appear to require additional or longer-lasting computations.

2.1.1 "Slow prediction" and argument structure

Although prior work has shown that comprehenders use contextual information to predict specific lexical forms, recent studies have argued that predictions are not always fast and accurate (Chow, Lau, Wang, & Phillips, 2018; Chow, Smith, Lau, & Phillips, 2016; Momma, Sakai, & Phillips, 2015). These studies were investigating a longstanding puzzle in the literature about why reversing the thematic roles of noun phrases in a sentence usually does not modulate N400 amplitude. For example, Chow et al. (2018) tested Mandarin sentences such as *Cop* *ba thief arrest* "the cop arrested the thief" and the role-reversed *Thief ba cop arrest* "the thief arrested the cop." Although the cloze probability in the canonical sentences was much higher than that of the role-reversal sentences, there was no N400 difference between the two conditions. This insensitivity of the N400 to differences in predictability caused by argument role reversals has been observed many times across many languages, and numerous hypotheses have been proposed to account for it (Hoeks, Stowe, & Doedens, 2004; Kim & Osterhout, 2005; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Brouwer, Fitz, & Hoeks, 2012).

The work by Chow and colleagues argues that the explanation for the role-reversal results at least partly depends on how *quickly* predictions can be generated on the basis of the context. Chow et al. (2018) manipulated the linear distance between an argument and a verb, via varying the position of an adverbial temporal phrase in a sentence. In the short-distance conditions, the adverbial temporal phrase was placed at the beginning of the sentence, and the argument was immediately followed by the verb (*Last week, cop ba thief arrest*); in the long-distance condition, the adverbial phrase was placed between the argument and the verb (*Cop ba thief last week arrest)*, which created a buffer (around 1800 ms) to formulate the prediction of an upcoming verb. The results showed that there was no N400 effect in short-distance conditions, but critically the N400 response was recovered in long-distance conditions. Momma et al. (2015) report similar findings in Japanese. Together, these data argue that argument roles can be used to predict an upcoming verb if sufficient time is provided; the corollary implication is that not all information in the context impacts prediction immediately. Chow, Momma, et al. (2016) discuss several reasons that argument roles might impact predictive computations slower than other kinds of information: (1) in the absence of the verb, argument roles like agent and patient are not directly observable from the syntactic structure but must be inferred, (2) the semantic memory database of event schemas that support correct verb predictions may not be organized in such a way that it can be rapidly probed with cues like *cop-as-agent* or *thief-as-patient*. In a separate paper, they are able to use similar logic to demonstrate that it is the argument *roles* in particular that are slow to impact prediction, as comprehenders appear quick to identify which noun phrases in the sentence are arguments of the verb at all and to preferentially weight these arguments in computing predictions for the verb (Chow, Smith, et al., 2016).

For the current research, the key takeaway from the prior work by Chow et al. (2018) is that we can estimate the temporal dynamics of argument structure computations by using N400 designs that vary the timing between the word-to-be-predicted and the critical elements of the context that could contribute to that prediction, and thus gain insight into the processes that relate the linguistic input with conceptual representations in long-term memory. In the current study, our goal is to use the same kind of approach to investigate the online computation of more complex argument structures and their corresponding event structures, by taking advantage of some convenient properties of compound verbs in Mandarin.

2.1.2 The current study

Compounding is a very productive word formation process in Mandarin. In fact, according to Huang (1998), stems of all lexical categories, except for prepositions, can be combined to form a compound. In the current study, we investigate the argument and event structure computations required to process compound verbs composed of two verbal morphemes (V1-V2); in particular, compound verbs whose two verbal morphemes are involved in a causal/resultative relation (i.e., V1 resulting in V2). According to Williams (2014), V1 of a resultative compound verb is a means predicate, whereas V2 a result predicate. In the most common type of resultative compound verb (Shen & Mochizuki, 2010), V1 is a transitive verb and V2 predicates the object of V1, indicating how the object of V1 was affected by the event described by V1. For example, in sentence (1), the subject of the complex predicate *washed-ruined* names the washer, while the object names both what is washed and what is ruined:

(1) 媽媽 洗壞了 衣服. (Transitive)

Mom washed-ruined le the clothes

"Mom washed the clothes so that the clothes were ruined.

While the literature on the processing of verb-argument relations often characterizes the problem as a relatively straightforward one of mapping arguments to a verb and participants to an event, resultatives are one of many cases that remind us that languages regularly make use of structures that go beyond this simple characterization. In the interpretation of resultatives and other compound verbs, participants are related to events, but events are also related to other events. In resultatives, this relation has a specifically causal dimension: the result described by

V2 is in some way caused by the event described by V1. The goal of our study was to begin to map the time course of the syntactic and semantic processes that are engaged by these more complex relations, in order to bring new insights to our understanding of the components of argument structure computation in general. As a starting point, we hypothesized that the extra complexity of the argument and event structure in resultatives would require extra processing time, delaying updates to predictions about upcoming arguments.

In the three experiments reported here, the basic logic was the following. We created subject-verb-object item sets where the amplitude of the N400 response to the object noun was the dependent measure of interest. All versions of a given item had the same object noun, which was carefully selected to have a relatively low cloze probability in control conditions (~10%), but a relatively high cloze probability in the resultative condition (e.g. Table 1). The key question of interest across the three experiments is how much processing time is required for comprehenders to be able to take advantage of the predictability of the resultative context to reduce N400 responses on the object.

Condition	Verb	Sentence context	Target	Cloze
Simple	V1-ASP	小孩 咬過了 *The kid bit-ASP le*	嘴唇 *lip*	Low (10%)
		“The kid had bitten his lip”		
Resultative	V1-V2	小孩 咬破了 *The kid bit-broke le*	嘴唇 *lip*	High (39%)
		“The kid bit his lip such that his lip was broken.”		
Causative	Made-V2	小孩 弄破了 *The kid made-broken le*	嘴唇 *lip*	Low (9%)
		“The kid did something to his lip such that his lip was broken.”		

Table 1: Example stimulus in each condition in Experiment 1 (averaged cloze probability in parenthesis)

Given the prior literature discussed above, we assume that in a simple context like *A kid had bitten _____,* upon recognizing the word *bite* as a simple verb and retrieving its meaning from the lexicon, comprehenders can rapidly generate a prediction for the object based on the verb alone, searching for frequent *biting* events in semantic and episodic memory and identifying the patient of the event as the likely upcoming object noun. If Chow, Momma, et al. (2016) are correct, comprehenders can also rapidly identify the pre-verbal noun as an argument and use it to search memory specifically for *biting* events that have *kid* as a participant. By contrast, if the verb is a resultative compound verb, then comprehenders would have to additionally analyze the correct structure of the compound, evaluate the event relation of the two verbal morphemes, disambiguate the thematic structure, and generate a representation of the complex event, such that the parser could probe memory for schemas or episodes involving the proper agents and patients for the complex event. For example, in *The kid bit-broke* _____, comprehenders would have to recognize the verb-verb sequence as a resultative compound verb, evaluate the relations of *biting* and *breaking* events, disambiguate thematic structure involved with *biting* <agent, patient> and *breaking* <agent, theme>, and generate a representation of a broken-by-biting event, where the subject should be an animate agent to perform the biting event, and the object should be a patient that could be broken by biting. Then they need to be able to successfully probe long-term memory for broken-by-biting events, make inferences with items in that database as premises, potentially constrained to those involving *a kid*. Our goal was thus to begin to home in on how much time it might take to use this extra information coded by a resultative compound verb to generate predictions about the object.

Before proceeding to the experiments, some basic background on the resultative construction in Mandarin is in order. Our study focuses on the *washed-ruined* or “transitive” type of resultative. This type of resultative has the following features: (1) the clause has both a subject and an object, (2) the object names the thing that enters the V2 result, (3) V1 is a transitive verb, and (4) the subject and object name the agent and patient of the V1 event. Still, it is worth noting the existence of other resultative types with different argument relations. In “unergative” resultatives, V1 is still a transitive verb, but V2 predicates the subject of V1, which thus bears an agent-experiencer role, as in *Mom washed-tired the clothes (*媽媽 洗累了 衣服). Argument roles assigned by V1 are also not restricted to agent and patient roles; in

*Boy upset-cried Mom (*男孩 氣哭了 媽媽), V1 assigns an experiencer role to the object. V1 is also not restricted to being a transitive verb; in *Mom coughed-hoarse voice* (媽媽 咳啞了 嗓子), both V1 *coughed* and V2 *hoarse* are intransitive verbs, but combining them together forms a transitive complex predicate. In our materials, the intended parse of the "transitive" resultative was encouraged through the higher frequency of this type of resultative, verb subcategorization preferences, and plausibility.

A considerable number of existing studies have investigated the role of factors such as lexical frequency, semantic transparency, morphological headedness in Mandarin compound verb word recognition (Kuo, 2006; Zhang & Peng, 1992; see Myers, 2006 for a review), but few have examined the processing of these verbs in a sentence context. To our knowledge, Lin and Jaeger's conference paper (2014) is the only study that has examined the factors of structural probability and thematic role order of resultatives in sentence context. Their eye-tracking results showed that transitive resultatives had the shortest first-pass and total fixation time at the post-verb critical region, indicating that the transitive is the easiest one to process compared with other types of resultative verbs.

2.2 Experiment 1

In Experiment 1, we asked whether comprehenders could use predictions afforded by a resultative compound verb to facilitate processing on the object noun when reading with a stimulus-onset asynchrony of 800 ms from verb to object (e.g. *A kid had bit-broke his lip*). In the Resultative condition, the compound verbs were always composed of a transitive V1 and an intransitive V2, in which V2 predicated the object of V1 and indicated the result of V1. Objects were selected to be strongly predictable by the resultative context, as determined by offline completion norming. In this first experiment we included two baseline conditions in which the context did not strongly predict the object. The Simple condition contained a simple verb (V1-asp, e.g. *A kid had bit his lip*). The Causative condition was included to rule out the possibility that any facilitation in the Resultative condition was due to unintended associative priming from V2 alone. Since V2 itself was intransitive, we added a transitive light verb *to make*, to form a transitive complex predicate (e.g. *A kid had made-broken his lip*). To match the number of characters of the verbs in the Resultative and the Causative conditions, we added an experiential aspect marker *guo* after V1 (meaning 'have V1-ed') in the Simple condition. All of the verbs thus had the same word length.

Experiment 1 used an 800 ms stimulus-onset asynchrony (SOA) between each word (600 ms on, 200 ms off), where the compound verb was presented on a single screen, as is natural in Mandarin. In other words, from the onset of the verb, comprehenders had 800 ms to process the verb and to predict an upcoming object noun. We note that although in English studies the typical SOA used for RSVP is shorter than 800 ms, such a slow presentation rate is relatively common in Mandarin (e.g., Zhou et al., 2010; Su et al., 2018). With no clear prior evidence about what time range might be required for complex argument/event structure processing, we chose to begin with an 800 ms SOA as it is a slow enough presentation rate to not be consciously taxing, but has been successfully used to identify certain slower aspects of argument role computation/prediction (Momma et al., 2015). If 800 ms SOA is enough time for participants to compute the resultative structure and use it to generate predictions about the object, then ERP responses to the critical object noun should track the offline cloze probability, with reduced N400 amplitude in the Resultative condition relative to the Causative and Simple conditions. However, if prediction on the basis of the Resultative takes longer than is afforded by an 800 ms SOA, then we would see no N400 differences among the three types of verbs. As this second case predicts a null effect, we also included a sanity check comparison in a separate set of items to show that N400 effects are indeed elicited for predictable and unpredictable object nouns following simple verbs.

2.2.1 Participants

Forty-nine naïve young adults (28 females, 20-35 years old, mean: 24) participated in the study at National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders. Of the 49 participants, 20 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves[1]. The reported results were obtained from the remaining 29 participants (15 females, 20-34 years old, mean age: 24). All of them consented to participate in the experiment. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland, College Park.

[1] The rejection rate was unusually high because (1) the epoch was fairly long (-100 to 1600 ms), and (2) the air conditioner in the lab was broken during data collection section. 10 out of the 20 excluded participants were removed because of sweat artifact.

2.2.2 Materials

Our stimuli were sentences of SVO structure, with the verbs varying among the following three conditions: Resultative (*bit-broken*), Causative (*made-broken*) and Simple (*bit-asp*), and the rest of the sentence being the same. Note that even though the subject and objects were kept identical, we intended to make the object in the Resultative context more predictable than that in the Causative and Simple contexts (see Table 1).

We started by finding resultative verbs from *A dictionary of Chinese verb-resultative complement phrases* (Wang, Jiao, & Pang, 1987). We selected an initial list of high frequency resultative compound verbs (n = 186) as the critical verbs for our Resultative condition. Based on the verbs (V1-V2) in the Resultative condition, we created our Causative and Simple conditions. The verbs in the Causative condition were resultative complex predicates whose V1 was a causative light verb *make* and V2 was taken from the Resultative condition. As for the Simple condition, its verbs were literally simple predicates. We took V1 from the Resultative condition and added an experiential aspect marker *guo* after V1 to match the number of characters in Resultative and Causative conditions. Note that resultative compound verbs in Mandarin are usually accomplishment or achievement verbs; they are telic predicates that describe an event as having a natural endpoint (Tai, 1984), and they frequently occur with the perfective aspect marker *le*. We thus added the perfective aspect marker *le* at the end of each verb in all experiment conditions to make them sound more natural in a sentence context.

In total we created 186 triplets of verbs, with one Resultative verb, one Causative verb and one Simple verb in each triplet. We added a subject noun phrase in each set, such that the subject noun phrase was the same among different conditions. The 186 triplets of subject-verb frames, 558 sentence frames in total, were divided into nine lists. Each list had 62 frames that were critical to the current study, and none of the frames were repeated among the lists. We had another 360 filler sentence frames, which were stimuli for another experiment, were divided into nine lists (so 40 filler frames per list) to pair up with the current study. Therefore, each list contained 102 sentences. 225 participants were recruited for the cloze norming (25 participants per list); none of the participants took part in the ERP experiment. Cloze norming data were collected online via Ibex Farm (http://spellout.net/ibexfarm/). We presented the context of a sentence frame all at once and the sentence frame would remain on the screen. Participants were instructed to provide the best continuations for the sentence frames. When computing the cloze probability of the target objects, we counted near synonyms (e.g., 馬路 *road* and 道路 *roadway*), nouns that were further specified by a modifier, (e.g., 秀髮 *beautiful hair* and 頭髮 *hair*), and words that contained a functional morpheme (e.g., 刀 and 刀子 *knife*) as the same lexical item.

Through cloze norming, our goal was to select sentence frames in which a given object noun phrase was more predicted by the Resultative condition than by the Causative or Simple conditions. Sentence frames that did not meet this criterion were excluded. The finalized stimuli comprised 90 triplets, with the average cloze probability for the target noun being 39% (range: 16%-80%) in the Resultative condition, 9% (range: 0%-36%) in the Causative condition, and 11% (range: 0%-36%) in the Simple condition (See Table 1). After finalizing the target words, we added more contexts following the target object nouns to make the sentences slightly longer and sound more natural. Each sentence consisted of six to nine words, with each word being one to four characters long. As a sanity check, we also included a set of 60 sentences from Liao and Chan (2016) with a similar cloze probability contrast (high cloze: 40%; low cloze: 0%), but where the predictability was driven by multiple features of the context and not just the verb. However, note that the cloze target of these sanity check sentences was in the sentence final position.

Due to the fact that we had three lists for the experiment manipulations and two lists for the sanity check sentences, six experimental lists were constructed such that no sentence context or target was repeated within the same list. The presentation order of the sentence stimuli was randomized within each list. Participants were randomly assigned to one of the six lists.

2.2.3 Procedure

Participants sat in front of a computer screen with their hands on a keyboard. Sentences were segmented into words; the complex verb and aspect marker were always presented as a single word on the same screen (see example in (2)), which were presented one word at a time in a white font (traditional Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 600 ms. Each word appeared on the screen for 600 ms, with a 200 ms inter-stimulus interval, for a stimulus-onset asynchrony (SOA) of

800 ms (See Figure 1 for details). At the end of 20% of the trials, a comprehension question would show up on the screen, and the participant had to answer via button pressing in order to proceed to the next trial. Prior to the experimental session, participants were presented with six practice trials with feedback to familiarize themselves with the task. The experimental session was divided into three blocks of 50 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 90 minutes.

(2) Sentence segmentation for stimulus presentation:

小孩/咬破了/嘴唇/

The kid/ bit-broke le/ (his) lip/

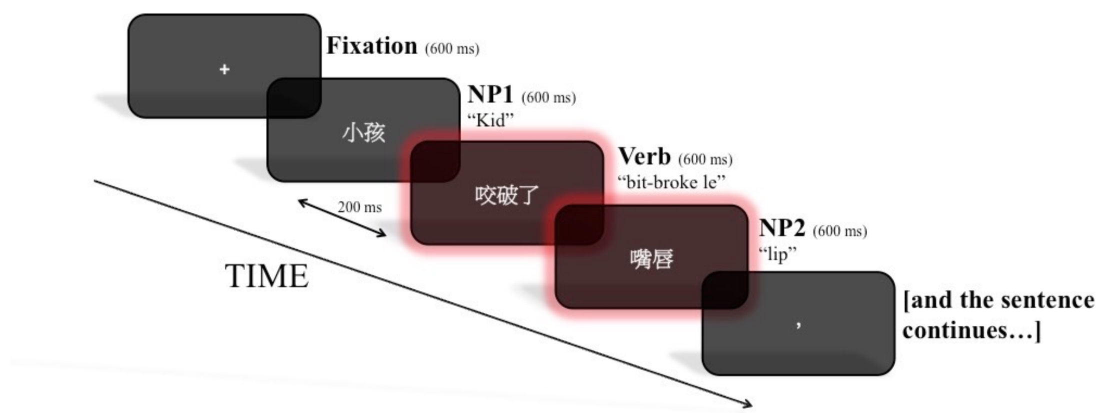

Figure 1: Presentation of stimuli in Experiment 1 and Experiment 2.

2.2.4 Data acquisition and analysis

E-prime 2.0 (Psychology Software Tools Incorporated) was used to present the experimental stimuli, record participants' behavioral data, and send the event codes to the digitization computer. EEG was recorded from 30 electrodes placed according to the 10/20 system (FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T3, C3, CZ, C4, T4, TP7, CP3, CPZ, CP4, TP8, T5, P3, PZ, P4, T6, O1, OZ, O2). Each channel was referenced to an average of the left and right mastoids for both online and off-line analyses. Four additional electrodes (two on the outer canthus of each eye and two on the upper and lower ridge of the left eye) were placed to monitor blinks and horizontal eye movements. The impedance of all the electrodes was kept below 5 kΩ. EEG signals were continuously digitized at 1000 Hz, filtered between DC to 100 Hz (NuAmps, NeuroScan Incorporated).

ERP analyses were time-locked from the onset of the verb. The EEG data were processed with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) in Matlab (MathWorks, Inc.). A linear derivation file was first imported to convert the four monopolar eye-movement monitoring channels to two bipolar channels (VEOG and HEOG). We applied a notch filter at 60 Hz and an Infinite Impulse Response (IIR) filter with the band-pass value set between 0.1 Hz to 30 Hz, 12 dB/oct. Then the continuous EEG file was epoched (1) from -100 to 1600 ms, from the onset of the verb until the end of the object noun phrase, for all the experimental conditions and (2) from -100 to 800 ms for the sanity check items. Baseline correction was applied with the pre-stimulus -100 to 0 ms interval. After baseline correction, artifact rejection was carried out by reviewing the epochs both automatically and manually: At each channel, a 200-ms window was moved across the data (100 ms before and 1600 ms after the stimulus) in 100-ms increments and any epoch where the peak-to-peak voltage exceeded 70 μV was rejected. We then reviewed the data, and if needed, adjusted the voltage threshold for individual subjects. Epochs contaminated by excessive blinking, body movements, skin potentials, and amplifier saturation were rejected. The overall rejection rates (including sanity check items) across participants was 20.3 ± 11.3% (mean ± SD); participants with greater than 40% trials rejected were excluded from further analysis. The rejection rates of each critical condition were: Resultative: 22.6 ± 11.9% (mean ± SD), Causative: 22.1 ± 11.0%, Simple: 22 ± 10.3%.

Our hypotheses centered around the N400 response at the object noun phrase. We selected nine electrodes over the central-parietal area (C3, CZ, C4, CP3, CPZ, CP4, P3, PZ, P4), known to show the most prominent N400 effect, and averaged them as our single clustered region of interest (ROI). We carried out a repeated-measure Type III ANOVA on the mean amplitudes in the measurement time windows of 1100-1300 ms, which was 300-500 ms after the onset of the noun, evaluating effects of Verb type (Resultative, Causative, Simple). When Mauchly's test of Sphericity was violated, Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to adjust the p-values.

In the sanity check items that were designed to replicate standard N400 effects of cloze probability, we carried out a paired t-test over the same set of electrodes evaluating the effect of predictability (High-cloze, Low-cloze).

2.2.5 Results

2.2.5.1 Behavioral data

The overall accuracy rate to the comprehension questions was 93 % (80%-100%), showing that participants were paying attention during the experiment.

2.2.5.2 ERP data

In order to ensure a clean baseline, we time-locked ERPs to the onset of the verb, where the three conditions first differed from one another, even though our interest of analysis would focus on the N400 responses at the noun. We ran statistic analyses on a pre-defined clustered ROI. However, when we visually inspected the data, we observed somewhat inconsistent patterns across electrodes: although the N400 responses to Causative condition were numerically more negative than Resultative among electrodes in our ROI, the N400 responses to Simple were more negative than Resultative over some electrodes (e.g. Cz) but not others (e.g. Pz). Figure 2 shows the grand average ERPs to Resultative, Causative and Simple conditions across several electrodes (Cz, Pz) that usually show robust N400 effects.

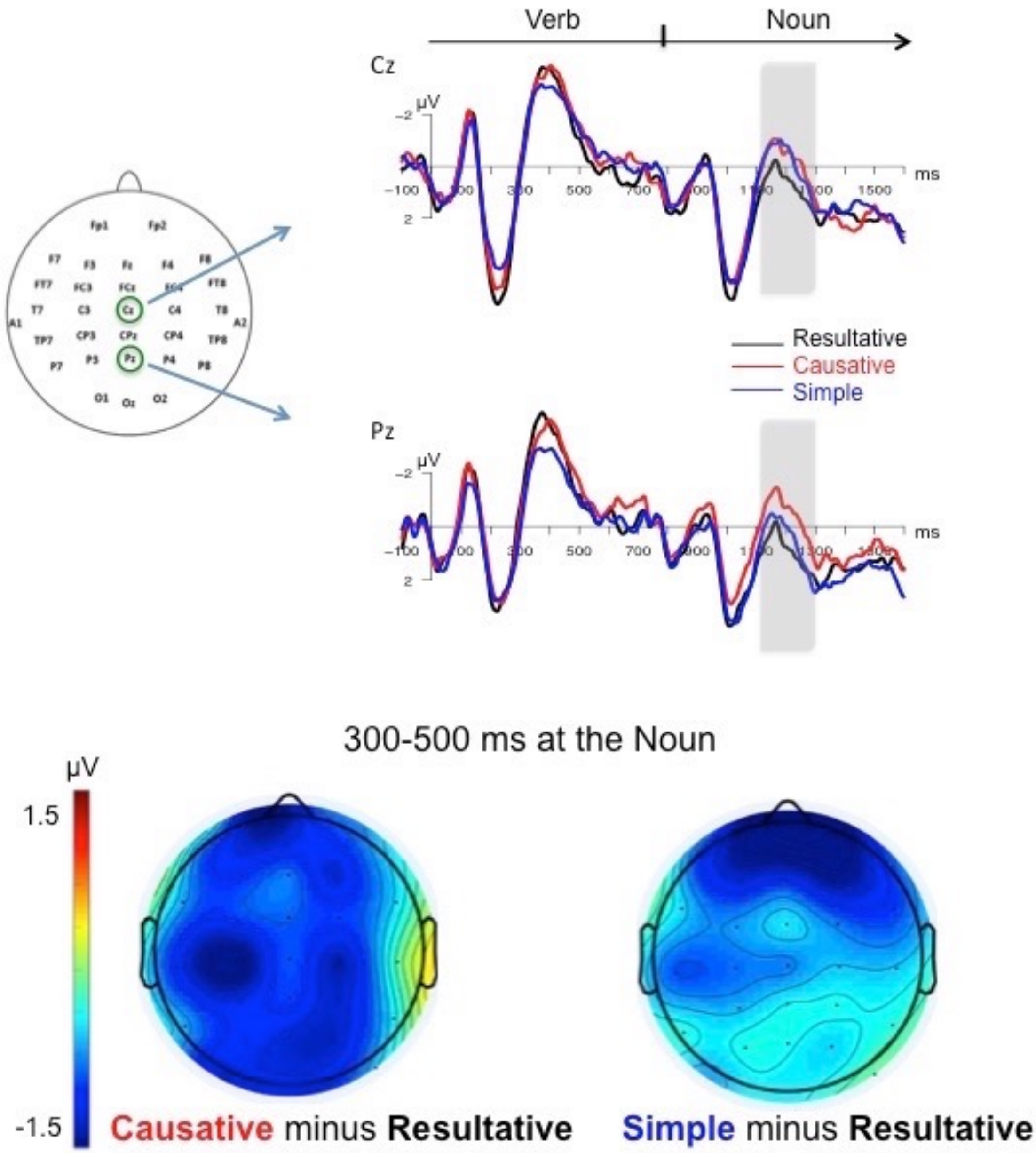

Figure 2: Top: Grand average ERPs from the verb to the noun at Cz and Pz in Experiment 1. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 1 (Left: Causative minus Resultative; Right: Simple minus Resultative).

Statistical analyses during the N400 time-window showed a main effect of Verb type ($F(2, 56) = 3.70$, $p < 0.05$). Follow-up paired-t-tests reveal that the N400 response to the object in the Resultative condition was significantly smaller than the

Causative ($t(28) = -2.55$, $p < 0.05$), but when compared the N400 response to the object in the Resultative condition to the Simple condition, there was no significant difference ($t(28) = -1.09$, $p = 0.28$).

Plotted in Figure 3 are the grand average ERPs to the Predictability effect in High- and Low-cloze sanity check sentences. During the N400 time window, there was a significant main effect of cloze ($t(28) = 26.10$, $p < 0.001$), which confirms the clear impression from visual inspection that the high-cloze continuations elicited reduced N400 amplitude in comparison to the low-cloze continuations.

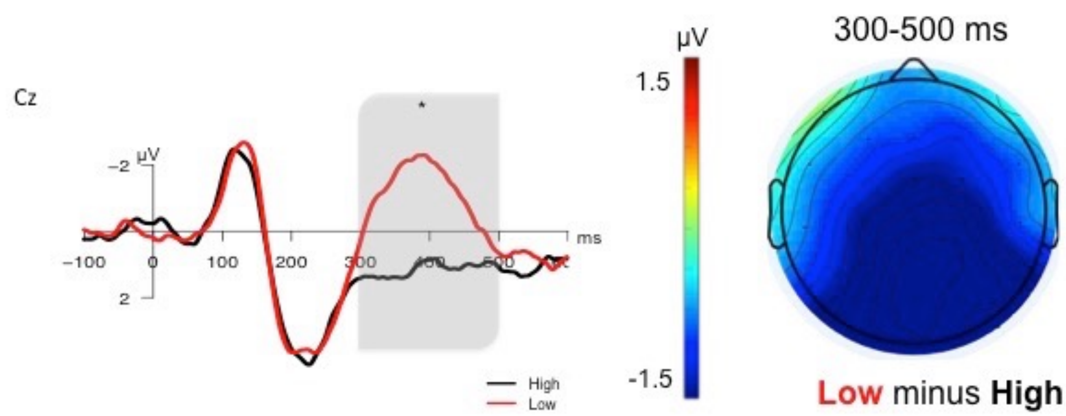

Figure 3: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 1. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 1 (Low cloze minus High cloze).

2.2.6 Discussion

Experiment 1 was designed to investigate how quickly the computation of a resultative compound verb can impact predictions of an upcoming noun. We used an 800 ms stimulus-onset asynchrony rate and asked whether the cues encoded in the resultative compound verbs could be used to update predictions in time to facilitate processing of the subsequent noun. We used materials in which offline cloze probability was high for the Resultative condition and low for the Causative and

Simple conditions, so that rapid use of resultative cues for prediction should result in a reduced N400 for the noun in the Resultative condition relative to the other two. In contrast, if an 800 ms SOA is not enough time to use resultative cues to update predictions, we expected that all three conditions should elicit relatively similar N400 amplitudes.

However, it is difficult to draw strong conclusions about either possibility from these results, as they fit neither of these predicted patterns. As shown in Figure 2, centro-parietal electrodes did show a reduced N400 response to the object in the Resultative condition than in the Causative condition, and this difference was significant in a follow-up pairwise comparison. However, the N400 contrast between the Resultative and the Simple conditions were not significant, even though the cloze probability to the object of the Causative and the Simple conditions were quite similar (Causative: 9%; Simple: 11%). In fact, if we took a closer look, we found that some anterior electrodes seemed to fit the "fast" prediction pattern, with smaller N400 in Resultative relative to Simple, whereas more posterior electrodes seemed to fit the slow prediction hypothesis, with no N400 difference between Resultative and Simple conditions. It remained unclear to us why the Simple patterned differently than the Causative condition, since both of their object nouns were relatively unpredictable based on the offline cloze norming. Such a finding was not consistent with any hypothesis we were aware of.

Although this pattern of data is equally unexpected on both hypotheses, both hypotheses are also consistent with reasonable post-hoc explanations that can inform improvements in the design. If resultative cues can be used rapidly to update predictions, it is possible that we failed to detect a true N400 difference between Simple and Resultative conditions because our cloze probability contrast was not robust enough across items, or that the 1 x 3 design limited power for detecting our effect of interest. In particular, it could be that the N400 to the object in the Causative condition was not reduced for a different reason, perhaps due to properties specific to the Causative construction. In Experiment 2, we worked to mitigate these possibilities by selecting a more tightly controlled subset of Resultative and Simple verb items from Experiment 1, in a different design that compared two different types of compound verbs and their corresponding Simple controls.

Although our sanity check sentences demonstrated a classic N400 effect, showing that participants did engage prediction during the experiment, we note that these items were qualitatively different than experiment materials: the cloze contrast was not closely matched to the experimental materials, target words were placed at sentence final position, and the predictability of target words was driven by multiple sources of contexts, not just subject and a verb. Therefore, in Experiment 2, we also modified the items in the simple predictability contrast to be more comparable to experimental materials.

2.3 Experiment 2

In Experiment 2, we aimed to improve on the design and materials of Experiment 1. We selected a more tightly controlled subset of Resultative and Simple verb items from Experiment 1, and created a 2 x 2 design in which the predictive effect of the resultative was compared against the predictive effect of a different type of compound verb, coordinate verbs. This allowed us to test different sources of online prediction difficulty in complex predicates.

Similar to resultative verbs, coordinate verbs are compound verbs that are composed of two contentful verbal morphemes, V1 and V2. Whereas the verbal morphemes in a Resultative are involved in a causal relation (‘V2 by V1’), the two morphemes of Coordinate are in a coordinate relation (‘V1 and V2’). For example, in the sentence *The store owner hit-scolded the employee*, the interpretation is that the store owner hit and scolded the employee. Although coordinates and resultatives bear a surface similarity in both being composed of two predicates, comprehenders can distinguish them online through cues provided by the meaning of the two verbs and by the subcategorization of V2. For example, given the compound verb *hit-scold*, the V1 *hit* is a transitive verb, which requires an agent and a patient, and so is the V2 *scold.* Since V1 *hit* and V2 *scold* have the same subcategorization, they naturally form a coordinate relation, and both of them are the head of the compound verb. The compound verb *hit-scold* cannot be a resultative verb.

The goal of Experiment 2 was to identify potential sources of online prediction difficulty in complex predicates. As in Experiment 1, the target nouns were more predictable in the complex predicate contexts compared with the simple predicate contexts. If computing a complex predicate is generally hard in a way that causes delays in prediction, then we would expect to observe no N400 effect to the objects in both Resultative and Coordinate contexts. However, if it is resultative predicates specifically that are costly, because of the causal relationship between V1 and V2, then we would expect to obtain an interaction between Set type and

Predictability effect, with a significant N400 contrast in Coordinate context, but not Resultative one.

2.3.1 Participants

The participants were 40 naive young adults (28 females, 18-40 years old, mean: 24) from National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders. Of the 40 participants, 7 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves. The reported results were obtained from the remaining 33 participants (18 females, 19-40 years old, mean: 24). All of them consented to participate in the experiment. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

2.3.2 Materials

Similar to Experiment 1, the materials were sentences of SVO structure, with the verbs varying among different conditions. Two sets of compound verbs were created: Resultative set and Coordinate set. Within each set, in addition to a compound verb condition, we included a simple verb condition as a baseline condition. The verbs in the simple verb conditions were the V1 from the compound verb conditions, followed by an experiential aspect marker *guo* to match the number of characters in the compound verb conditions. In other words, the Resultative set contained Resultative (V1-V2), and R-Simple (V1-asp) conditions whereas the

Coordinate set contained Coordinate (V1-V2) and C-Simple (V1-asp) conditions. Note that resultative compound verbs in Mandarin are usually accomplishment or achievement verbs which describe an event as naturally bounded (Tai, 1984), and they frequently occur with the perfective aspect marker *le*. We thus added a perfective aspect marker *le* at the end of each verb in all experiment conditions.

Although the verbs varied, the subject and object were identical between conditions in the same set. We intended to make the object in Resultative context and Coordinate context more predictable than that their Simple controls. Materials for the Resultative set were 60 Resultative verbs and corresponding Simple verbs selected from Experiment 1. For the Coordinate set, the procedure to finalize the materials was similar to the procedure to Resultative verbs in Experiment 1. Coordinate verbs were chosen from *An Online Revised Mandarin Dictionary by the Ministry of Education, R.O.C.* (http://dict.revised.moe.edu.tw/cbdic/index.html). We did not include Coordinate compound verbs whose V1 and V2 are synonyms. In addition, we excluded Coordinate verbs whose V1 was identical to the V1 of the Resultative verbs, because in this case the baseline condition to the Coordinate condition (C-Simple) and the baseline condition to the Resultative condition (R-Simple) would be identical. We selected 119 coordinate compound verbs and created 119 pair of verbs, with each pair containing one Coordinate verb, and one Simple verb. We added a subject noun phrase in each set, such that the subject noun phrase was the same between conditions. Then, the 119 sets of subject-verb frames (each set contained 2 subject-verb frames, so 238 frames in total), were divided into two lists such. Each list contained 119 sentences. Fifty participants were recruited for the cloze norming (25 participants per list); none of the participants took part in the ERP experiment. Cloze norming data were collected online via Ibex Farm (http://spellout.net/ibexfarm/). We presented the context of a sentence frame all at once and the sentence frame would remain on the screen. Participants were instructed to provide the best continuations for the sentence frames. The presentation order of the sentence stimuli was randomized.

To demonstrate that participants were able to generate predictions based on a minimal sentence context, we also created predictability sentence frames that only contained a subject and a simple verb, which were a better match to the experimental conditions. One hundred subject-verb frames were subject to online cloze norming. Another 25 participants were recruited to perform sentence completion task. None of the participants took part in the ERP experiment. The presentation order of the sentence stimuli was randomized.

Through cloze norming, our goal was to select sentence frames in which a given object noun phrase was highly predicted by Resultative condition and Coordinate condition, but not by their baseline R-Simple and C-Simple conditions. Sentence frames that did not meet this criterion were excluded. The finalized stimuli were 60 items in the Resultative set and 60 items in the Coordinate set. The averaged cloze probability to the target nouns in the Resultative set was 39% for Resultative (range: 16%-80%) and 9% for R-Simple (range: 0%-36%) and in the Coordinate set was 38% for Coordinate (range: 16%-72%) and 10% for C-Simple (range: 0%-44%). The cloze sanity check items were of similar contrast to the experimental materials (High-cloze: 38% vs. Low-cloze: 9%) (See Table 2). After finalizing the target nouns, we added more contexts following the target nouns to make the sentences slightly longer and sound more natural. Each sentence consisted of six to nine words, with each word being one to four characters long.

Condition	Verb	Sentence context	Target	Cloze
Resultative sets				
Resultative	V1-V2	小孩 咬破了 *The kid bit-broke le*	嘴唇 *lip*	High(39%)
		"The kid bit his lip such that his lip was broken."		
R-Simple (Baseline of Resultative)	V1-ASP	小孩 咬過了 *The kid bit-ASP le*	嘴唇 *lip*	Low(9%)
		"The kid had bitten his lip"		
Coordinate sets				
Coordinate	V1-V2	老闆娘 打罵了 *The store owner hit-scolded le*	員工 employee	High(38%)
		"The store owner hit and scolded the employee."		
C-Simple (Baseline of Coordinate)	V1-ASP	老闆娘 打過了 *The store owner hit-ASP le*	員工 employee	Low(10%)
		"The store owner had hit the employee."		

Table 2: Example stimulus in each condition in Experiment 2 (averaged cloze probability in parenthesis)

Two experimental lists were constructed such that no sentence context or target was repeated within the same list. Each list consisted of 240 sentences, including 60 items of Resultative set, 60 items of Coordinate set, 60 items of cloze sanity check items, and additional 60 filler items that were of similar length for an unrelated experiment that will not be described here. The presentation order of the sentence stimuli was randomized within each list. Participants were randomly assigned to one of the two lists.

2.3.3 Procedure

The procedure was identical to Experiment 1.

2.3.4 Data acquisition and analysis

Data acquisition and analysis, including the regions of interest, were identical to Experiment 1. The overall mean rejection rate (including sanity check items) across participants was 24.1 ± 12.4% (mean ± SD); participants with greater than 40% trials rejected were excluded from further analysis. Rejection rates of experimental conditions were summarized as follows: Resultative: 28.6 ± 12.9%, R-Simple: 30.8 ± 16.3%, Coordinate: 27.8 ± 9.4%, and C-Simple: 28.4 ± 12.8%. We carried out a repeated-measure ANOVA on the mean amplitudes in the measurement time windows of 1100-1300 ms, which was 300-500 ms since the onset of the noun, and evaluated effects of Set type (Resultative, Coordinate) and Predictability (High-cloze, Low-cloze). Follow-up paired t-tests were performed when an interaction was observed.

In the sanity check items that were designed to replicate standard N400 effects of cloze probability, we carried out a paired t-test over the same set of electrodes evaluating the effect of predictability (High-cloze, Low-cloze).

2.3.5 Results

2.3.5.1 Behavioral data

The overall accuracy rate to the comprehension questions was 91% (83%-96%), showing that participants were paying attention in the experiment.

2.3.5.2 ERP data

Plotted in Figure 4 shows the grand average ERPs to the verb and object noun in the Resultative and R-Simple conditions and in the Coordinate and C-Simple conditions. Visual inspection suggested that there was no N400 cloze difference to the objects in the Resultative set, but that there was a difference in the Coordinate set. A repeated-measure Type III ANOVA analyses demonstrated a significant Set type by Predictability interaction ($F(1,32) = 4.346$, $p < 0.05$). Follow-up pairwise analyses revealed that that there was a significant difference between Coordinate and its C-Simple baseline ($t(32) = 2.96$, $p < 0.01$), but not between Resultative and its R-Simple baseline ($t(32) = 0.56$, $p = 0.58$).

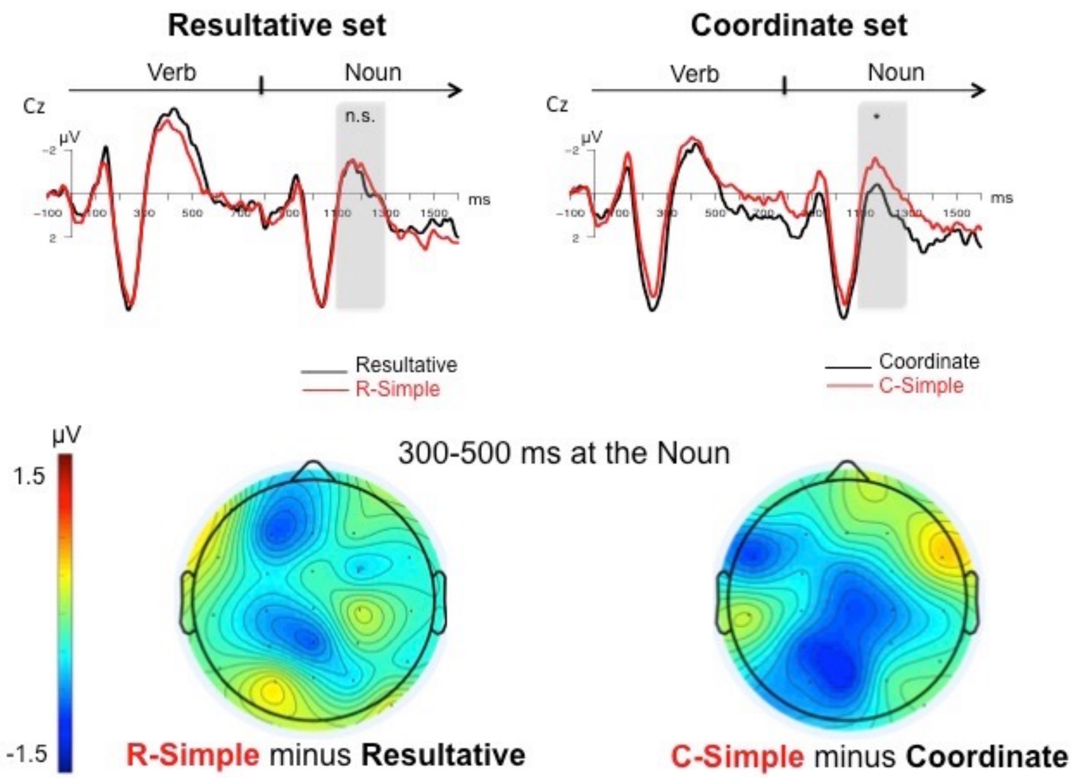

Figure 4: Top: Grand average ERPs from the verb to the noun of the Resultative set (Left) and Coordinate set (Right) at the Cz electrode in Experiment 2. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 2 (Left: R-Simple minus Resultative; Right: C-Simple minus Coordinate).

It is worth noting that visual inspection suggested that the coordinate comparison also showed an earlier increased negativity for the C-Simple condition relative to the Coordinate condition that onset approximately 500 ms into the verb region (more negative for simple verbs than coordinated verbs). Although we did not have any specific hypotheses about what ERP differences might emerge at the verb, this difference might raise the question of whether the N400 difference observed at the object noun in the coordinate conditions might be partly due to ongoing negativity for the C-Simple condition from the verb region. We think this is unlikely as the waveforms appear to come back together prior to the N400 window on the noun, but we will return to this point in examining the results of Experiment 3.

Figure 5 shows the grand average ERPs for the Predictability effect in the sanity check items (High-cloze vs. Low-cloze). Visual inspection suggested that the high-cloze continuations had reduced N400 amplitude than the low-cloze continuations. The results of the pairwise comparison also showed a significant effect of cloze ($t(32) = 4.89$, $p < 0.05$).

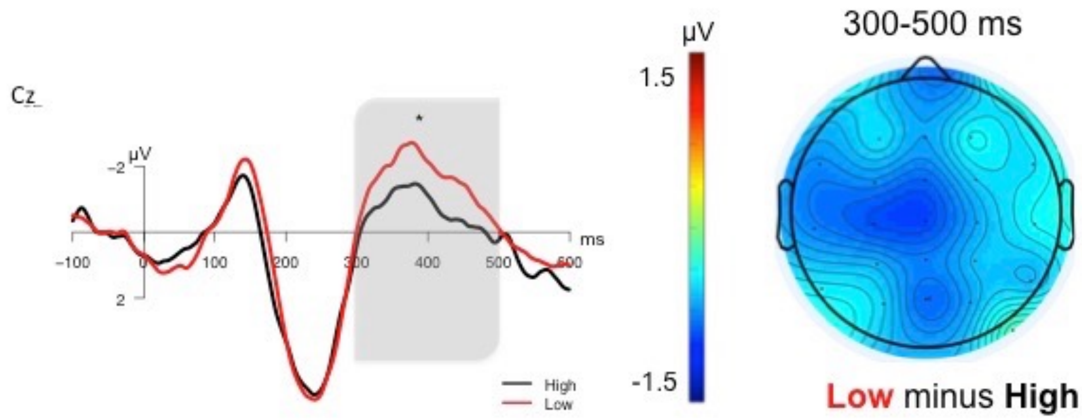

Figure 5: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 2. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 2 (Low cloze minus High cloze).

2.3.6 Discussion

The results of Experiment 2 suggest that prediction for the object noun is not immediately updated by information from Resultative verbs. The interaction between Set type and Predictability indicated that predictability was modulated differentially by the two types of compound verbs we tested: Coordinate verbs and Resultative verbs. Specifically, we found an N400 effect in the Coordinate set, indicating that information encoded in coordinate verbs can impact prediction in time. However, there was no N400 effect in the Resultative set. Based on the significant interaction, we could infer that the computation of Resultative was too slow to impact prediction in time.

Our finding that Coordinate verbs immediately contributed to object predictions, above and beyond V1 alone, is important for ruling out several possible explanations of the failure for Resultative verbs to do so. In Experiment 1, we observed a larger N400 difference between Resultative (V1-V2) and Causative (V2) than between Resultative and Simple (V1). One possible explanation of this pattern could have been simply that predictions were rapidly updated on the basis of V1 only, with V2 contributing little to constrain predictions. However, in Experiment 2, we showed that although both Resultative and Coordinate are compound verbs, comprehenders were only able to quickly incorporate V1 and V2 into their prediction when they form a coordinate relation. Therefore, we would like to argue what slowed down prediction in Resultative is a process that was specific to Resultative verbs. We suggest that it could be the process of computing the causal relationship between V1 and V2 that slowed down prediction, but other alternatives are also possible. We will discuss these alternatives in the General Discussion section.

In Experiment 2, we made the sanity check sentences more comparable to the experimental sentences: The sentence context of sanity check items consisted of a subject and a simple verb, and the cloze contrast was matched to that of experiment conditions. With these manipulations, the N400 effect was still significant, showing that participants were engaged to update their predictions given the minimal contexts. However, it is essential to note that the N400 effect of the sanity check sentences was much smaller than that in Experiment 1. These results suggest that a cloze difference of this magnitude based on subject and a verb corresponds to a relatively small effect size on N400 amplitude.

Based on the results of Experiment 2, we could infer that the computation of Resultative was too slow to impact prediction in time when words are presented with an 800 ms SOA. This hypothesis would predict that with enough time the N400 contrast should emerge. Experiment 3 was designed to test this hypothesis.

2.4 Experiment 3

Experiment 2 showed that participants could quickly update predictions based on Coordinate verbs but not Resultative verbs. In Experiment 3 we asked, could predictions be updated if comprehenders were given several hundred more milliseconds? We used the same materials as in Experiment 2 except that we added a buffer to allow additional processing time by inserting a prenominal modifier with minimal conceptual content, such as a possessive or a quantifier, between the compound verb and its object noun, which resulted in an extra 400 ms of processing time compared to Experiment 2 (see details below). Our hypothesis predicts that the N400 cloze effect should re-emerge in the Resultative set when sufficient processing time is provided.

2.4.1 Participants

The participants were 48 naive young adults (22 females, 18-33 years old, mean: 23) from National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders.

Of the 48 participants, 10 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves. The reported results were obtained from the remaining 38 participants (20 females, 18-33 years old, mean: 23). All of them consented to participate in the experiment. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

2.4.2 Materials

The materials were identical to Experiment 2, except that we inserted a modifier (either a possessive or a quantifier) between a verb and a noun, such that participants might have a little buffer to update their predictions. In half of the sentences, the modifier was a possessive, and the other half was a quantifier.

Although we assumed the predictability of the target noun would remain the same despite the insertion of a modifier, we conducted post-hoc cloze norming to confirm this. Our norming focused on the 60 Resultative sets of subject-verb-modifier frames (each set contained Resultative and R-Simple conditions, so 120 frames were normed in total). They were divided into two lists. Fifty participants were recruited for the cloze norming (25 participants per list); none of the participants took part in the ERP experiment. Cloze norming data were collected online via Ibex Farm (http://spellout.net/ibexfarm/). We presented the context of a sentence frame all at once and the sentence frame would remain on the screen; participants were instructed to provide the best continuations for the sentence frames. The presentation order of the sentence stimuli was randomized. Surprisingly, our norming revealed that the cloze contrast between the Resultative and R-Simple conditions actually became smaller (in Experiment 3, Resultative: 40% vs. R-Simple: 18% whereas in Experiment 2, Resultative: 39% vs. R-Simple: 9%). Fortunately, this difference goes against our hypothesis (a smaller cloze difference in Experiment 3 than 2, although we expect the N400 effect to re-emerge in Experiment 3) and therefore only acts to provide a more conservative test of that hypothesis.

2.4.3 Procedure

The procedure was identical to Experiment 2, except for the presentation rate. With presentation rate of 800 ms in Experiment 2, the EEG recording time was about 40 minutes. As we added a modifier between the verb and the noun in all conditions, the EEG recording time could be even longer. To keep participants from being too tired during the experiment, which could introduce artifacts such as alpha waves, we increased the presentation rate from 800 ms to 600 ms in Experiment 3. Each word appeared on the screen for 500 ms, with a 100 ms inter-stimulus interval (See Figure 6 for details). Given the new SOA, participants had up to 1200 ms (i.e., the duration from a verb to a modifier) to update predictions whereas in Experiment 2, only 800 ms (i.e., the duration of a noun) was available to make predictions.

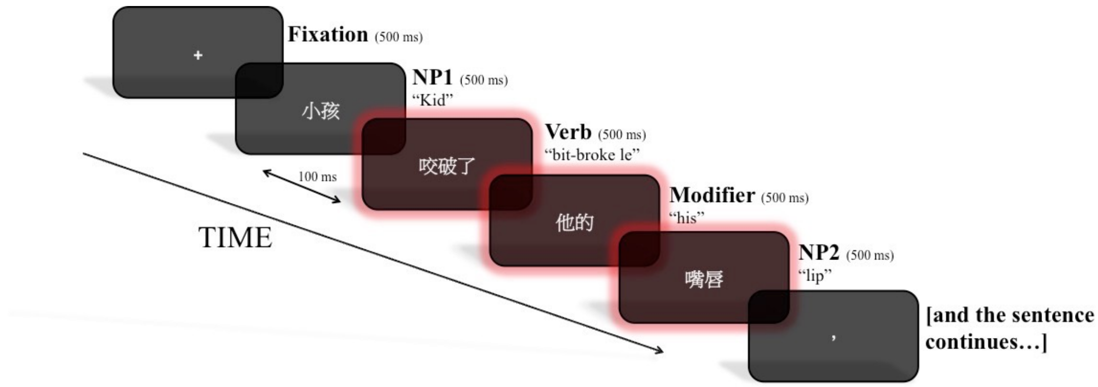

Figure 6: Presentation of stimuli in Experiment 3.

2.4.4 Data acquisition and analysis

Data acquisition and analysis were identical to Experiment 2. We time-locked ERPs to the onset of the verb, where experimental conditions started to differ, with the epoch ranging from -100 ms to 1800 ms, to cover the brainwave responses from the onset of the verb to the end of the object noun (600 ms each for the verb, modifier, and object). As for the sanity check items, the epoch was from -100 to 600 ms. The overall mean rejection rate (including sanity check items) across participants was 23.1 ± 12.7%; participants with greater than 40% trials rejected were excluded from further analysis. Rejection rates of the experimental conditions were summarized as follows: 24.1 ± 11.4%, R-Simple: 24.4 ± 13.2%; Coordinate: 23.8 ± 12.6%, C-Simple: 23.7 ± 11.6%.

2.4.5 Results

2.4.5.1 Behavioral data

The overall accuracy rate to the comprehension questions was 92 % (83%-

98%), showing that participants were paying attention in the experiment.

2.4.5.2 ERP data

Plotted in Figure 7 shows the grand average ERPs to the Resultative and R-Simple conditions and to the Coordinate and C-Simple conditions. Visual inspection suggested that there was an N400 effect to the objects in the Resultative set as well as in the Coordinate set. Statistically, we found a Predictability main effect ($F(1,37) = 10.73$, $p < 0.005$) and no evidence of a Set type by Predictability interaction ($F(1,37) = 0.05$, $p = 0.82$).

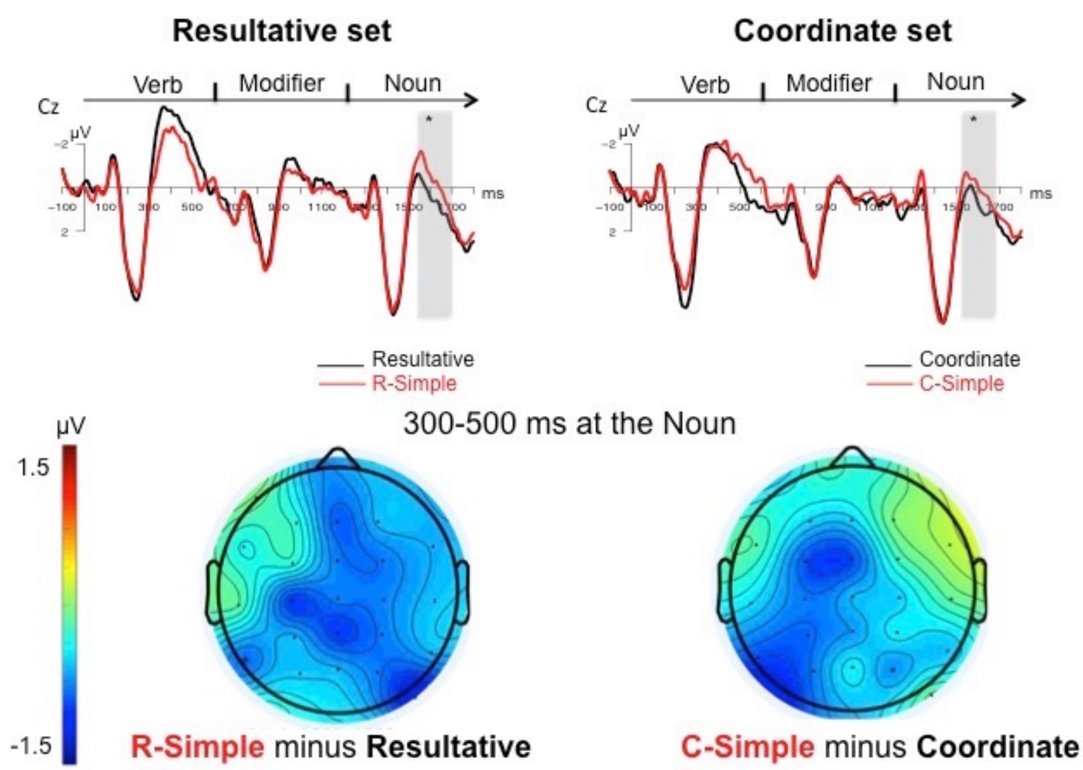

Figure 7: Top: Grand average ERPs from the verb to the noun of the Resultative set (Left) and Coordinate set (Right) at the Cz electrode in Experiment 3. Bottom: Topographic distribution of ERP effects in the 300-500 ms intervals at the noun in Experiment 3 (Left: R-Simple minus Resultative; Right: C-Simple minus Coordinate).

Figure 8 shows the grand average ERPs to High-cloze and Low-cloze sanity check sentences. Visual inspection suggested that the high-cloze continuations elicited a reduced N400 response than the low-cloze continuations. Statistics also showed a significant effect ($t(37) = 6.35$, $p < 0.05$).

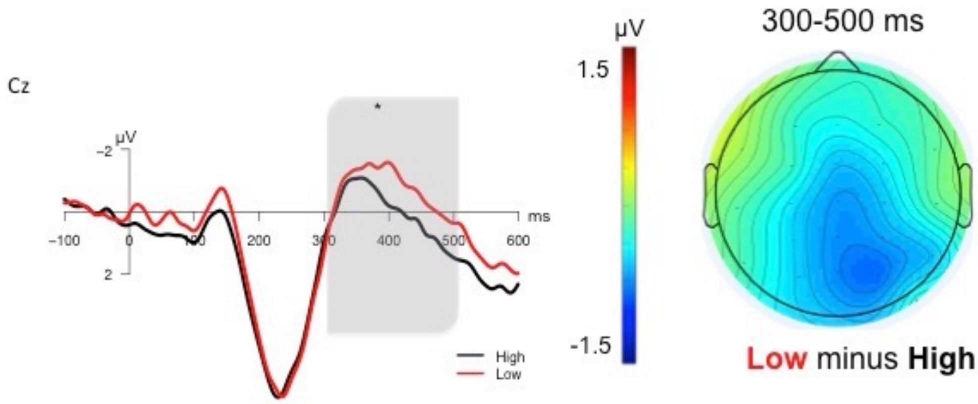

Figure 8: Left: Grand average ERPs to cloze sanity check at Cz in Experiment 3. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 3 (Low cloze minus High cloze).

2.4.6 Discussion

In Experiment 3, we investigated if predictions on the basis of Resultatives can be updated when participants were given several hundred more milliseconds. A modifier was inserted between the verb and the object noun to create a little buffer for participants to update predictions. Under these conditions, we did not obtain an N400 reduction at the object in Resultative set in Experiment 2, but it emerged in Experiment 3. By contrast, the N400 reduction was observed at the object in Coordinate set in both experiments. These effects held even though the addition of the modifier unintentionally made the cloze contrasts slightly smaller than Experiment 2 and Experiment 1 (mean differences of ~20% in Experiment 3 and ~30% in Experiment 2 and Experiment 1). Overall, these results showed that the causal relations between V1 and V2 could constrain predictions, if participants were provided with sufficient time—here, a buffer of 1200 ms between complex verb and object noun.

During the verb region in Experiment 3, we showed the same numerical pattern from Experiment 2 of more negativity for the C-Simple condition than the Coordinate condition at around 500 ms post-verb onset. One concern from Experiment 2 was whether the apparent N400 effect at the noun could rather be due to differences carried over from the verb region. The prenominal modifier in Experiment 3 allowed us to see that the verb-elicited differences appeared to subside by about 800 ms post-verb onset, as illustrated in Figure 7. In order to confirm that there were no reliable differences between Coordinate and C-Simple conditions immediately prior to noun onset, we ran an additional paired t-test on 100 ms before the onset of the noun. Results showed that the Coordinate condition did not differ from the C-Simple condition ($t(37) = 1.29$, $p = 0.20$). Therefore, it unlikely that early differences on the verb are responsible for the significant N400 effect observed on the subsequent object noun for the coordinate comparison in Experiment 2.

2.5 General Discussion

Three ERP experiments were conducted to investigate the predictive mechanism of online sentence comprehension through properties of Mandarin compound verbs. We focused on resultative compound verbs whose V2 predicates the object of V1, featuring that the object is affected by V1. We asked if the causal relationship of a resultative compound verb could rapidly constrain predictions of a subsequent object. The predictive effect of resultative compound verbs was compared against that of coordinate compound verbs, which allowed us to test different sources of online prediction difficulty in complex predicates.

The N400 was used as a neural indicator of what is predicted in the current study. Although results from Experiment 1 were inconclusive, the better-controlled design in Experiment 2 suggested that predictions on the basis of the resultative were not updated in time to impact processing of the object when verb and object onset were separated by 800 ms. This "timing" hypothesis was supported by Experiment 3, where the N400 predictability effect was recovered when participants were provided with up to 1200 ms between verb and object onset to update predictions.

Two classes of explanation for why prediction update is delayed are (1) computing the causal relations expressed by a resultative predicate is slow and/or (2) using the resultative predicate to generate predictions—to retrieve entities/nouns from memory that are likely to complete the message—is slow. We do not have strong evidence to favor one over the other. In fact, as the two classes of explanation target different stages of processing, it is likely that they are not mutually exclusive. In the following, we consider the two accounts in more detail.

2.5.1 Slow prediction due to the computation of a complex resultative predicate

One possibility is that predictions based on the resultative predicate take significant time because computing the causal relations expressed by the predicate takes time. As discussed in earlier sections, complex predicates are different from simple verbs in many aspects. For example, with the combination of two verbal morphemes, comprehenders could be struggling with lexical processing, such as accessing the meaning of V1 and V2, constructing the mental representation of the complex predicate, and decomposing the internal structure of the complex predicate. Any of the above computations could take longer and slow down predictions. To identify sources of online prediction difficulty in complex predicates, we introduced coordinate compound verbs in Experiments 2 and 3, and compare the predictability effect of resultative compound verbs against coordinate ones. Our results revealed that not all complex predicates yield equally slow predictions: with an 800 ms SOA, predictability effects were observed after coordinate verbs but not resultative verbs. These data suggest that what slowed down the predictive mechanism were processes that were specific to resultative compound verbs.

Different theoretical frameworks differ in exactly what kinds of complex predicate representations are computed over compound verbs. Li (1990) proposed that when the two verbal morphemes were merged into a complex predicate, theta roles from V1 and V2 should merge into a composite theta role. For example, in the complex predicate *bit-broke* "broken by biting," whose V1 *bite* required an agent and a patient and V2 *break* required a theme, the theme role from V2 should be merged with the patient role from V1, and then the composite theta role, patient-theme, would be assigned to the object noun phrase. Different from Li, Williams (2014) suggested the semantics does not involve combining the theta roles from V1 and V2. Instead, the entire complex predicate per se has two theta roles, a causer and a causee, and relations to the events of V1 and V2 are inferred. Since our experiments were not set up to test any of the above frameworks, we do not have a stand to argue for one analysis than the other. However, both frameworks feature unique properties in the resultative structure. If some of these properties are particularly costly to compute, we could explain why updating predictions subsequent to a resultative verb took longer than other types of verbs.

2.5.2 Slow prediction due to memory search for an optimal candidate

We also entertained the slow prediction hypothesis proposed by Chow et al. (2018), which would hold that predictions were slow because it takes longer to use the cues from resultatives to retrieve the best fitting word or concept for the context.

Chow, Momma, et al. (2016) specifically propose that lexical prediction can be seen as a two-step memory retrieval process, which involves (1) a fast parallel process that activates all the event schemas associated with the individual context words, and (2) a slow serial search through this initial set for the schemas that match the argument role assignment of the nouns in the context. For example, in *Cop ba thief arrest* (the cop arrest the thief"), it was fast for comprehenders to recognize that *cop* was an agent and *thief* was a patient. Nevertheless, because the information of *cop-as-an-agent* and *thief-as-a-patient* were compound cues and not simplex cues, comprehenders would have to serially search through the semantic space for an item that satisfied all the features, delaying successful prediction. On the other hand, other authors point out the challenges in formulating a principled distinction between simplex and compound cues that captures the semantic retrieval phenomena, and instead suggest that delays in contextual prediction may reflect differential weighting of cue certainty across time (Kuperberg, 2016). What these accounts have in common is that they all place the locus of the timing effects in the process of prediction update, rather than the process of parsing and interpreting the context.

To explain the results of the current study, these kinds of account would posit that the computation of complex predicates, including the configuration of argument structures, was completed rapidly, but what slowed down prediction was the process of retrieving the candidate that best satisfied the context. Consider predictions generated from a coordinate compound verb first. When perceiving the verb *hit-scold,* comprehenders recognized the verb-verb sequence as a coordinate verb. Both verbs then served as retrieval cues at an initial stage, with a set of *hittable* candidates (e.g., ball, employee, etc) and a set of *scold-able* candidates (e.g., employee, politician, etc) being activated. Candidates that matched both cues would be more activated, and comprehenders would just retrieve one of the candidates. In this case, *employee* was retrieved as the best fit to the context *The store owner hit-scolded* _____, because *employee* was both *hittable* and *scold-able*. In other words, a simple summation of the activation elicited by each verb would be likely to yield successful retrieval of the best-fitting candidate[2]. Our ERP results indeed indicated that participants could make use of the cues provided by a coordinate verb to update their prediction promptly.

However, prediction on the basis of a resultative verb could be more complicated. Below we suggest two different possibilities of the memory retrieval process for resultative verbs. Both of them involve a second-step search, which could

[2] Note that this simple summation procedure is not equivalent to the real meaning of a coordinate compound structure, which for example entails that the events denoted by each individual verb should be related as parts of one complex event, not simply that both events involved the object as an argument. However, this procedure would likely be a successful heuristic for retrieving probable objects of a coordinate compound.

explain why prediction based on resultative verbs was slower than coordinate verbs. To begin with, when perceiving the verb *bit-broke* in *The kid bit-broke _____,* comprehenders should recognize the verb-verb sequence as a resultative verb. Then, the initial process could be very similar to that proposed above for the coordinate verbs. Candidates that related to the two verbal morphemes would be more activated. For example, *biting* could activate a set of candidates, such as *toy* and *straw*; *breaking* could activate a set of candidates like *vase* and *glass*. However, since *lip* might not be very highly activated by biting events or breaking events alone, an additional step might be needed for it to be predictively facilitated: the parser had to search through the intersected space of *biting* and *breaking* events to find a candidate that could be *broken by biting.* In this case, although *lip* might not be the predicted target based on a biting event or a breaking event alone, it was the best fit to the context when a causal relationship between the two verbal morphemes was taken into consideration.

Alternatively, it is possible that after identifying *bit-broke* is a resultative verb, the parser pursued a different strategy of memory searching from the beginning than that used with coordinate verbs. On this account, since the parser knew that V1 of a resultative verb is a means predicate, and V2 a result predicate, when searching through the memory space, it might start with V1 as the only retrieval cue. In this case, initial activation would be focused on entities that could participate in a biting event involving a kid, such as *toy*, *corn* and *lip*. Then, at a later stage, these initial candidates were serially searched for one that could be broken by biting, in this case, *lip*.

We do not have evidence to argue for or against these two possibilities of memory retrieval processes discussed above. Whether entities related to V2 are activated at an initial stage, just as we propose for the coordinate verbs, is an empirical question. We will leave the question for future exploration.

In a more continuous model, the delayed memory access of the predictable candidates in the resultative case might reflect the lower frequency, and thus lower resting activation level, of the complex real-world events that support the prediction (that is, biting events in general will be more frequent than biting events that result in breaking). It could also be that generating a prediction about the likely result of an event does not solely depend on retrieving memories of existing events, but also requires an extra processing step of inference or simulation. All of these explanations predict the dissociation in timing from the coordination contexts, in which identifying predictable candidates can be done with reference to simple events in memory.

2.5.3 ERP responses to verbs

Our results suggest that the computations required to generate predictions following resultative verbs take longer than following coordinate verbs. While our design focused on neural activity during the target noun, these results raise the question of whether traces of those costly computations could be observed during the ERP to the verb itself. To our knowledge, we are the first group that used EEG responses to study the processing of Mandarin resultatives. Since we did not have any a priori hypothesis about the processing of resultative verbs, we plotted the topographic distribution of ERP effects in P200 (150-300 ms), N400 (300-500 ms) and P600 (500-800 ms) intervals at the verb (see Figure 9). If we could observe any pattern across the three experiments, we might get an initial clue about which stage of processing drives slower prediction update in response to resultative verbs, and whether this would be a useful avenue to pursue in future work.

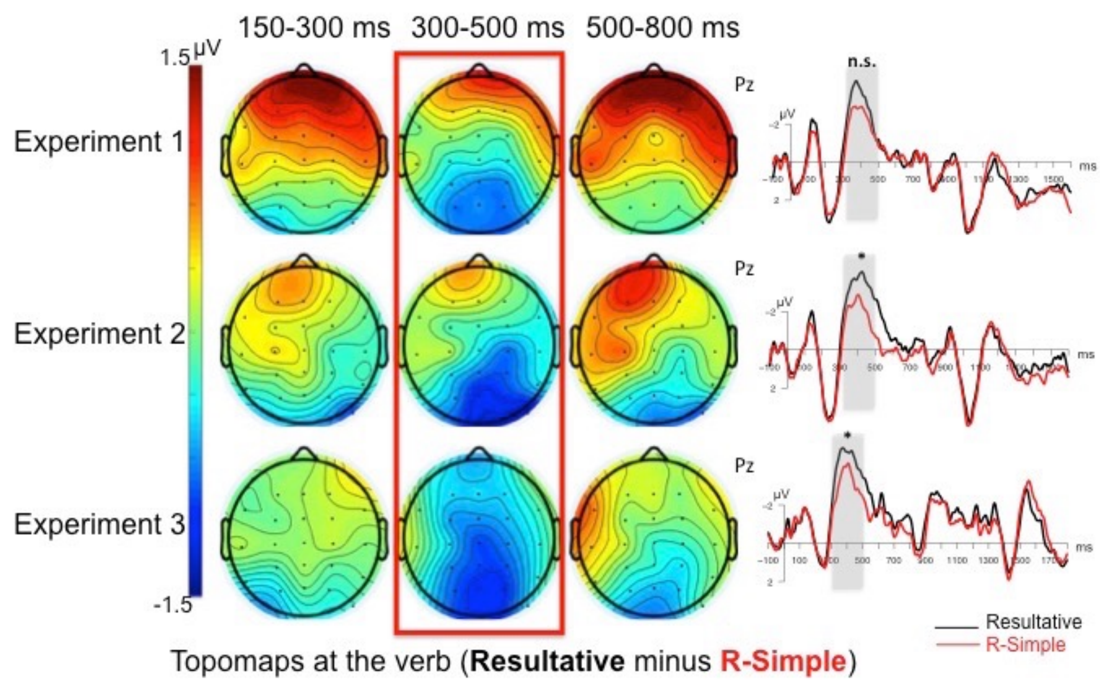

Figure 9: Left: Topographic distribution of Resultative effect across different time windows in Experiments 1-3. Right: Grand average ERPs from the verb to the noun of Resultative and R-Simple conditions at the Pz electrode in Experiments 1-3.

As depicted by Figure 9, relative to the Simple condition, the Resultative condition seemed to elicit a larger negativity over the central-parietal sites in the N400 time window. The topographic distribution and the peak latency resembled an N400 effect. We ran post-hoc paired t-tests to examine the effect, with the same region of interests as what we defined for the N400 effect at the noun. Our analyses showed that the N400 effect was not significant in Experiment 1 ($t(28) = 2.20$, $p =$

0.15), but was significant both in Experiment 2 ($t(32) = 5.47$, $p < 0.05$) and Experiment 3 ($t(37) = 10.63$, $p < 0.001$). We also plotted the topographic maps of the Coordinate set (see Figure 10). However, unlike the Resultative set, we found no effects at the N400 time window when considering Coordinate relative to C-Simple conditions (Experiment 2: $t(32) = 0.26$, $p = 0.62$; Experiment 3: $t(37) = 0.23$, $p = 0.64$).

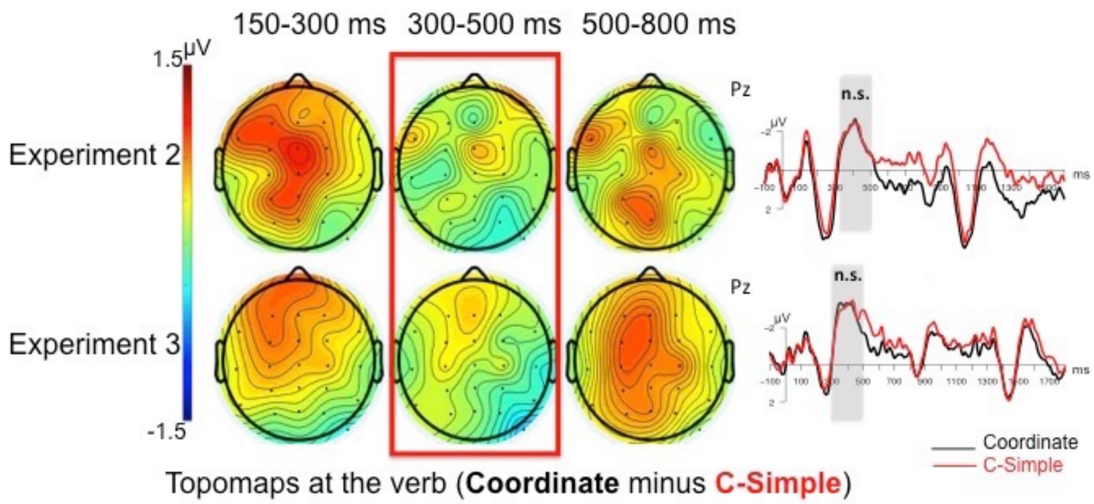

Figure 10: Left: Topographic distribution of Coordinate effect across different time windows in Experiments 1-3. Right: Grand average ERPs from the verb to the noun of Coordinate and C-Simple conditions at the Pz electrode in Experiments 1-3.

Although resultative verbs and coordinate verbs are both compound verbs, the post-hoc analyses reveal that only resultative verbs elicited a larger N400 response. As the N400 is primarily associated with lexical or conceptual processing, these data then tentatively suggest that resultative verbs require additional lexical or conceptual computations that could be tied to the delayed prediction effects at the subsequent noun. However, it could also be the case that these effects on the verb are unrelated to the effects on the object and simply differences between the lexical properties of the resultative and coordinate verbs, such as different degrees of semantic association between V1 and V2, lexical frequency, number of brush strokes, neighborhood density, etc. Since the current study was not aimed at evaluating responses at the verb, we did not attempt to control these properties, and so we leave a more systematic investigation of these differential verb responses for future work.

2.5.4 Implications for L2 acquisition and processing

Finally, as the current study took advantage of language-specific properties by which argument structure is encoded in Mandarin, we would like to briefly discuss potential implications of the current study for L2 acquisition and processing. It is interesting to note that Mandarin resultative compound verbs are notably challenging for L2 learners. Yuan and Zhao (2010) used acceptability rating to study the comprehension of resultative compound verbs in Mandarin native speakers and English L2 learners of Mandarin. Their materials included several types of resultative verbs: not only the type that is called “transitive": in the current study, but also the ones discussed in Section 2.1.2. Note that “transitive” resultative verbs are the type of resultative verbs that are used most frequently in Mandarin; it is also the type most often found in other languages. Yuan and Zhao (2010) showed that advanced L2 learners exhibited mastery only of the transitive type. The authors attributed the benefit of Mandarin transitive resultative verbs to similar thematic configurations in English resultative constructions (i.e. Mom washed the clothes ruined): in both languages, the object noun phrase of a resultative complex predicate received a patient role from V1 and a theme role from V2. Such a transfer effect from learners’ first language could also explain why resultative verbs of other thematic relations were rejected by the L2 learners, although it still remained a puzzle why thematic role reconfiguration was challenging for them. To further explore the mental representation of Mandarin resultatives in L2 learners, we suggest that a better understanding of the computation involved in L1 resultative comprehension should be developed. We believe that the current study constitutes one such step. Although we mainly focused on transitive resultative verbs in this chapter, our method is applicable to other types of resultative verbs and compound verbs of different internal structures in general. We will continue to work on these cases in our future explorations.

2.6 Conclusion

This chapter investigated how quickly two types of complex predicates associated with verb-verb compounds—coordinates and resultatives—could be computed and used to update predictions for the subsequent input, using the N400 response as a measure of online prediction. If processing speed were mainly a function of syntactic complexity, then we would expect both conditions to demonstrate the same temporal dynamics, but if the computations required by certain semantic relations are particularly costly, the two verb types should dissociate. Results from our three experiments indicate that predictions afforded by a resultative verb do not impact processing of the subsequent noun at an effective verb-noun SOA of 800 ms, but that predictive effects emerge with a verb-noun SOA of 1200 ms. This contrasts with the case of coordinate verbs, which impacted predictions at the verb at both SOAs. We discussed two broad families of accounts for the dissociation: (1)

computing the causal relations expressed by a resultative predicate is more taxing and/or (2) retrieving a candidate that fits the resultative context requires more time. Our study shows that evaluating the speed of prediction update with the N400 is an effective approach for dissociating some of the fine-grained subcomputations required for the interpretation of complex verb constructions. Future work using this method, in combination with other tools, can help to lay the groundwork for a detailed time course model of argument structure computation.

Chapter 3: The time course of argument-verb computation in online sentence comprehension: Evidence from the N400

3.1 Introduction

As reviewed in Chapter 1, it is now well established that during online sentence processing, comprehenders engage a predictive mechanism in which expectations about the linguistic form or message are generated early and updated as new input arrives. Since successful prediction depends on finishing computing linguistic information from the sentence context, the timing of prediction can and has been used to investigate the time course of computing different linguistic processes (Chow, Lau, Wang, & Phillips, 2018; Momma, Sakai, & Phillips, 2015). In this chapter, I report a set of experiments that uses this approach in order to map the time course of argument-verb relation computations.

The purpose of this chapter is to investigate the processing stages by which the parser incorporates different pieces of information from the arguments to predict the verb. One recent hypothesis suggests that the processing profiles can be broken into two stages: an earlier stage in which the subset of nouns that denote the verb's arguments are identified to inform verb prediction, and a later stage in which argument role information becomes available to constrain predictions (Chow, Smith, Lau & Phillips, 2016), but the evidence for this idea is limited. The set of experiments described below are designed to test this hypothesis more systematically, with the ultimate goal of mapping the time course of argument-verb relation computations. We will use the N400 response to index successful verb prediction, and successful verb prediction in turn as an indicator that the relevant linguistic information about argument structure in the context must have been computed by that point in time. To foreshadow the results, we will provide new evidence from a novel design showing that the upcoming verb's arguments are identified relatively quickly, and that the identity of the verb's arguments rapidly constrains prediction even when linear order is held constant (*The millionaire*(subject) *the servant*(object) *fired* vs. *The millionaire [the servant*(subject) *fired…*). With data from additional experiments, we will be able to position this argument-identification process as an intermediate stage, after an early stage in which verb prediction appears fully insensitive to argument information, but before the later stage in which argument role information finally impacts the verb prediction (e.g. *The servant*(subject/object) *the millionaire*(object/subject) *fired*).

3.1.1 Slow vs. fast prediction in sentence comprehension

Much existing evidence has showed that comprehenders actively integrate information from the context to predict what is coming next (Federmeier & Kutas, 1999; Federmeier, 2007; Thornhill & Van Petten, 2012). As introduced in Chapters 1 and 2, in such experiments, predictability of a word is often quantified by an offline cloze measure, where participants are asked to provide a continuation to a sentence frame, and the percentage of a word used to complete the sentence frame is defined as the cloze probability of the word (Taylor, 1953). For example, given the sentence frame "He bought her a necklace for her ____," a majority of participants provided "birthday" and only a small proportion provided "collection" as the best continuation to the sentence, "birthday," the high-cloze completion, is defined as a predicted word and "collection," the low-cloze one, as an unpredicted word (Federmeier, Wlotko, Ochoa-Dewald, & Kutas 2007). Since the aim of this chapter is to investigate the computation of verbal argument structure, let's turn to what we know about predictions involving verbs and their arguments.

What are the processes involved in using argument information to predict a verb? Friederici and Frisch (2000) suggest that if arguments are given prior to a verb, the parser would compute the thematic relations of the arguments, and thereby check the computed thematic relations against the argument structure frame of the verb. A considerable amount of studies have now investigated if the thematic relations of argument can be updated quickly to impact predictions of a verb. This line of research reverses the thematic roles assigned to the pre-verbal arguments and tests if the N400, the neural indicator of prediction, is sensitive to the thematic anomaly. Although a few inconsistent results exist—which will be discussed in detail in the Discussion section—a majority of studies show that the N400 is not sensitive to thematic role reversals. In fact, the absence of N400 effect has been replicated among different languages, with various structures. For example, the N400 insensitivity is found in Chow, Smith et al. (2016) with objective relative clause (OSV) in English (e.g. "the customer that the waitress <u>served</u>" vs. "the waitress that the customer <u>served</u>"). It is also observed with simple SOV structure in languages that allow it, such as Mandarin (Chow & Phillips, 2013; Chow et al., 2018). In addition, the pattern still holds even when there is only one pre-verbal argument (Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kim & Osterhout, 2005; Momma et al., 2015). The insensitivity of N400 to role reversal situations appears to be incompatible with existing studies, as the classic N400 observations were that a low-cloze unexpected target word, or a semantically implausible word, would generate a larger N400 response relative to an expected word. However, studies like Chow, Smith et al. (2016) have confirmed that role reversals are an exception—even when cloze probability is collected and shown to differ, there is still no N400 difference to role reversal anomaly.

Various accounts have been proposed to explain the absence of N400 effect to role reversal situations (Kim & Osterhout, 2005; Kuperberg et al., 2007; see Brouwer, Fitz, & Hoeks, 2012 for a review). Different from most of the existing accounts, which question the functional interpretations of the N400 and P600 components, Chow, Momma, Lau and Phillips (2016) propose the slow prediction hypothesis, which suggest that argument roles may impact predictive computations slower than other kinds of information. To support this hypothesis, Momma et al. (2015) manipulated presentation rates with two-word Japanese sentences (*bee-nominative sting* vs. *bee-accusative sting*). Their results show that the N400 is not sensitive to role reversals when the materials are presented at 800 ms presentation rate. However, when the presentation rate is increased to 1200 ms, participants have more time to consider the thematic relations between the argument and the verb, the N400 effect emerges. In a similar spirit, Chow et al. (2018) manipulated the linear distance between arguments and the verb in Mandarin. They find that when the two arguments are adjacent to the verb, the N400 is blind to thematic role reversal situations (*Cop ba thief arrest,* meaning "the cop arrested the thief," vs. *Thief ba cop arrest,* meaning "the thief arrested the cop"). By contrast, when a temporal adverbial is inserted between the second argument and the verb, which creates a little buffer to update predictions on the verb, the N400 effect becomes present (*Yesterday cop ba thief* *arrest,* meaning "the cop arrested the thief yesterday," vs. *Thief ba cop yesterday* *arrest,* meaning "the thief arrested the cop yesterday"). The above findings reveal that argument role information can constrain predictions on the verb within at least one to two seconds, although this is notably longer than many other contextual information sources.

3.1.2 Bag of Words vs. Bag of Arguments hypotheses in argument-verb computation

Prior work has shown that argument role information impacts predictions relatively slowly, but what is happening during this long time window before argument role impacts prediction? How are the necessary computations ordered within this time? Prior to argument roles coming in, do comprehenders just compute basic lexical association, or can some level of structure be playing a role earlier? Chow, Smith et al. (2016) hypothesize that even before argument role impacts prediction, structure is already impacting prediction in the sense that a subset of noun phrases are identified as arguments of the upcoming verb, and this information can constrain the prediction of the verb. They call this the "Bag of Arguments" hypothesis, using a metaphor that elements in the "bag" are relevant information for prediction. In contrast to the "Bag of Arguments" hypothesis is the "Bag of Words" hypothesis, which suggests that all the words in the context can contribute to prediction prior to the late-stage contribution of argument role.

To test these two hypotheses, Chow, Smith et al. (2016) create sentences with three noun phrases in a row. The last two noun phrases are placed in an embedded sentence and the critical verb comes at the end of the embedded sentence. N400 responses are evaluated at the embedded verb. By reversing the order of the first two noun phrases, they introduce different arguments in the embedded sentence ("The *exterminator* inquired which *neighbor* the landlord had evicted" vs. "The *neighbor* inquired which *exterminator* the landlord had evicted"). The Bag of Words hypothesis would predict no difference between conditions, because the three noun phrases, regardless of their order, are all inside the "bag." In other words, with the metaphor that elements in the bag are relevant information for prediction, the scope of the "bag" under the Bag of Words hypothesis is the entire sentence. By contrast, the scope of the "bag" would be smaller under the Bag of Arguments hypothesis, as here the identification of the subset of nouns that are arguments of the upcoming verb is used to predict the upcoming verb; the 'bag' refers to the embedded clause. The Bag of Arguments hypothesis would predict an N400 effect to the sentences described above, since the two conditions involve different arguments in the "bag." In particular, with a *neighbor* and a *landlord* in the bag, the predicted event is evicting. However, evicting would be a less likely event if the arguments are an *exterminator* and a *landlord*. ERP results reveal an N400 effect between conditions, and support the Bag of Arguments hypothesis. The finding shows that comprehenders able to identify noun phrases that could be arguments of a verb and update their predictions based on that.

Note that the Bag of Arguments hypothesis also predicts that argument *roles* would not impact the prediction of an upcoming verb. Metaphorically speaking, these arguments are lumped in the bag, so their argument role information is not distinguishable for prediction. Chow, Smith et al. (2016) include a second experiment where the order of the last two arguments is reversed in the embedded sentence, creating role reversal scenarios ("The restaurant owner forgot which *customer* the *waitress* had served" vs. "The restaurant owner forgot which *waitress* the *customer* had served"). They successfully replicate prior studies by showing a null N400 effect between conditions.

Taken together, Chow, Smith et al. (2016) take their results to support the Bag of Arguments hypothesis, showing that initial verb prediction is constrained by noun phrases that are in the same clause as the target verb. What is implied by this conclusion is that the parser is able to identify which noun phrases could be arguments of the upcoming verb, potentially based on the structure cue provided by the clause boundary. Then, if an additional several hundred milliseconds is provided, argument role could constrain predictions of a verb as well (Chow et al., 2018; Momma et al., 2015). These findings imply that there are two stages of argument-verb computations. First, there exists a time window for the parser to identify if the noun phrases could be arguments of the verb, and to use that information to update predictions. Then, a later stage at which the parser is able to update predictions on the basis of argument roles, and construct detailed representations of a sentence.

However, in Chow, Smith, et al. (2016), the noun phrase outside of the embedded clause is in fact linearly further away from the embedded verb. In other words, with English sentences, whether the noun phrase could be the argument of a verb is confounded with linear distance from the verb. The effects they observed could therefore result from a recency effect or priming. We will revisit the Bag of Arguments hypothesis when we discuss the experiment design of the current study.

3.1.3 The current study

Prior studies have broadly indicated that prediction on the basis of argument roles is slow, and Chow, Smith et al.'s (2016) study provides intriguing initial evidence that this may reflect temporally staged argument structure computations, where initial verb prediction is constrained by the identification of arguments but not their roles—the Bag of Arguments hypothesis. However, as discussed in the section above, only one data point exists so far in support of the Bag of Arguments hypothesis, and in that study whether a noun phrase could be an argument of an upcoming verb is confounded with linear distance between the noun phrase and the verb. Furthermore, the temporal dynamics of the processing stages by which the parser incorporates different pieces of information from the noun phrase to predict the verb, remain relatively vague. In this chapter, our goal is to devise a stronger test of the Bag of Arguments hypothesis, with better control of the linear distance between the noun phrases and a verb. More broadly, our aim is to begin to more comprehensively map the time course of argument-verb computations. We hope that by getting a better understanding of when different pieces information contribute to the prediction of the verb, we can develop a processing model which identifies and maps out the stages comprehenders go through to compute argument-verb relations.

In the three ERP experiments reported here, the basic logic is the following. We used the timing of prediction, as indexed by the N400 effect, to study when different pieces of argument information contribute to predictions of the verb, as successful prediction depended on finishing computation of linguistic information in the context. We varied the presentation rate to investigate the amount of time needed for a particular type of argument information to impact prediction of the verb. All the experiment materials were in Mandarin, which has properties that allow us to keep the linear distance between noun phrases and verbs identical regardless whether the noun phrase could be an argument of the verb (more explanations below). We started our study with a slower presentation rate (800 ms), testing if prediction of the verb was constrained by noun phrases in the same clause as the verb, a situation in which the noun phrases were more likely to be the arguments of the verb (Experiment 4) and comparing the impact of argument role at the same presentation rate (Experiment 5). Thus, the first two experiments established the time frame for the Bag of Arguments hypothesis. Then we tested if a similar effect still held with a faster presentation rate (600 ms) in Experiment 6.

3.2 Experiment 4

In Experiment 4, we tested whether identifying noun phrases as arguments of the verb can be a useful cue to constrain predictions of the verb, when linear distance between the noun phrases and the verb is better controlled. Specifically, the Mandarin *ba* construction places two arguments before the verb (e.g. *Millionaire ba servant fired* meaning "Millionaire fired the servant"). While this sentence is monoclausal, a biclausal sentence could be introduced with the same noun order simply by replacing *ba* with a clausal verb, such as *think* (*Millionaire thought servant fired*...), so that the verb is no longer predicted by the context. The Bag of Arguments hypothesis predicts that comprehenders should be able to identify noun phrases that could be the arguments of the verb relatively quickly. In this example, if both *servant* and *millionaire* are identified as arguments of a verb, it is more likely that the verb is *fire* than if *servant* is the only argument in the "bag." If this is the process used by comprehenders, then we expect to observe a smaller N400 response at the verb in the one-clause *ba* condition compared to the two-clause *think* condition. The Bag of Arguments hypothesis would also predict that there is a time period in processing at which the parser had not committed to the thematic role of the arguments yet; metaphorically speaking, all the relevant arguments are lumped in the bag, with argument roles undefined. We will wait until Experiment 5 to directly test this prediction of the hypothesis.

We relied on previous role reversal studies to determine the presentation rate of Experiment 1. As far as we could tell, 800 ms was the slowest presentation rate where a null effect of argument role on the N400 was observed in role reversal studies (Momma et al., 2015), and thus this rate seemed like a good place to start in narrowing in on the hypothesized time window in which argument(s) of a verb could impact prediction but not the role bounded by the argument.

3.2.1 Participants

The participants were 40 naive young adults (28 females, 18-40 years old, mean: 24) from National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders.

Of the 40 participants, 7 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves. The reported results were obtained from the remaining 33 participants (18 females, 19-40 years old, mean: 24). Informed consent was obtained from all participants. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

3.2.2 Materials

Materials were sentences adapted from Experiment 3 in Chow et al. (2018). We began by selecting 60 sentences, all of which used the SOV *ba* construction in Mandarin. In particular, the construction requires a transitive verb, and the morpheme *ba* always follows an agent argument and is immediately followed by a patient argument. That is, in this construction, unambiguous and reliable cues about the arguments' syntactic roles are available before the presence of the verb. In our experiment setup, the two preverbal arguments were always animate. None of the target verbs were repeated. The 60 sentences were considered the Baseline sentences; we replaced the morpheme *ba* with the verb *think* to create another 60 sentences as the critical Complement sentences. In other words, the two conditions for the experiment were (1) Baseline condition, with the two noun phrases presented in a canonical SOV word order and (2) Complement condition, with the verb *think* separating the two noun phrases into different clauses (see Table 3). Since replacing *ba* with *think* would introduce a clause boundary between two noun phrases, the critical verb, which was then embedded in a subordinate clause, became much less predictable based on the second noun phrase alone. Note that we wrote different post-target verb continuations to the two conditions, because the structure requirements of the two conditions were very different. For the Baseline condition, the two pre-verbal arguments had satisfied the argument structure restrictions of a transitive verb. By contrast, when the transitive verb was embedded in a subordinate clause, such as in the Complement condition, it would introduce another argument into the subordinate clause to make the sentence grammatical. Therefore, depending on the length of the continuations, the length of our sentences ranged between six to nine words long. Despite the length of the sentence varied, the number of words was always identical up to reaching the target verb. Lastly, we adapted the materials to accommodate small lexical differences in language use between Mandarin speakers in China and Taiwan. The 120 sentences were divided into two lists in a latin square design.

Condition	Sentence context	Post target verb continuation
Baseline	富翁 把 僕人 解雇了	之後 立即 請來了 新的 管家
	Millionaire ba servant **fired-ASP**	then immediately hired new housekeeper
	"The millionaire had fired the servant and then immediately hired a new housekeeper."	
Complement	富翁 認為 僕人 解雇了	童工 很 不 應該
	Millionaire thought servant **fired-ASP**	kid very not should
	"The millionaire thought that it was inappropriate for the servant to fire the kid."	

Table 3: Example stimulus in each condition in Experiment 4

To confirm that comprehenders did engage predictive mechanisms during the experiment that modulated N400 amplitude, we also included 30 pairs of sentences instantiating a cloze contrast (High: 38% vs. Low: 9%) as our control items. Different from the experimental conditions, the control sentences were of simple SVO structure, with predictability being examined at the object noun position (e.g. *The tourist had picked up the flowers / the cherries*). Here, prediction was updated based on the information provided by a subject and a verb. The 30 pairs of sentences were counterbalanced between two lists.

Two lists were constructed such that no sentence context or target word was presented twice in one list. Each list consisted of 240 sentences, including 30 sentences in the Baseline condition, 30 sentences in the Complement condition, 30 sentences of high-cloze target in the High condition, 30 sentences of low-cloze target in the Low condition, and an additional 120 filler sentences from an unrelated experiment reported elsewhere (Liao & Lau, 2020). Participants were randomly assigned to one of the two lists. The presentation order of the sentences was randomized.

3.2.3 Procedure

Participants sat in front of a computer screen with their hands on a keyboard. Sentences were segmented into words (see Example in (1)), which were presented one word at a time in a white font (traditional Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 600 ms. Each word appeared on the screen for 600 ms, with a 200 ms inter-stimulus interval, for a stimulus-onset asynchrony (SOA) of 800 ms (See Figure 11 for details). At the end of 20% of the trials, a comprehension question would show up on the screen, and the participant had to answer via a button press in order to proceed to the next trial. Prior to the experimental session, participants were presented with six practice trials with feedback to familiarize themselves with the task. The experimental session was divided into 4 blocks of 60 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 90 minutes.

(1) Sentence segmentation for stimulus presentation:

富翁/把/僕人/解雇了/之後/立即/請來了/新的/管家。

Millionaire/ ba/ servant/ fired-ASP/then/immediately/hired/new/housekeeper

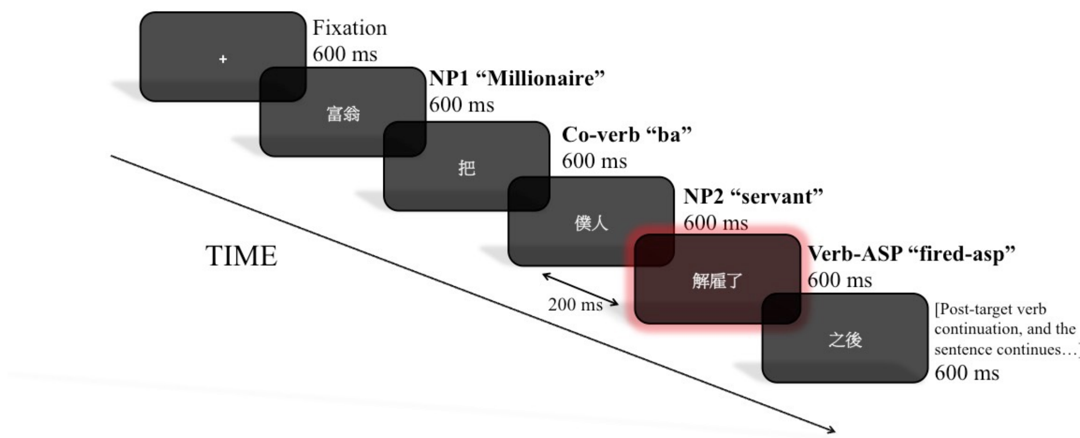

Figure 11: Presentation of stimuli in Experiment 4

3.2.4 Data acquisition and analysis

E-prime 2.0 (Psychology Software Tools Incorporated) was used to present the experimental stimuli, record participants' behavioral data, and send the event codes to the digitization computer. EEG was recorded from 30 electrodes placed according to the 10/20 system (FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T3, C3, CZ, C4, T4, TP7, CP3, CPZ, CP4, TP8, T5, P3, PZ, P4, T6, O1, OZ, O2). Each channel was referenced to an average of the left and right mastoids for both online and off-line analyses. Four additional electrodes (two on the outer canthus of each eye and two on the upper and lower ridge of the left eye) were placed to monitor blinks and horizontal eye movements. The impedance of all the electrodes was kept below 5 kΩ. EEG signals were continuously digitized at 1000 Hz, filtered between DC to 100 Hz (NuAmps, NeuroScan Incorporated).

ERP analyses were time-locked to the onset of the verb for the critical conditions and to the onset of the noun for the sanity check items. The EEG data were processed with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) in Matlab (MathWorks, Inc.). A linear derivation file was first imported to convert the four monopolar eye-movement monitoring channels to two bipolar channels (VEOG and HEOG). We applied a notch filter at 60 Hz and an Infinite Impulse Response (IIR) filter with the band-pass value set between 0.1 Hz to 30 Hz, 12 dB/oct. Then we extracted epochs of length -100 to 800 ms. Baseline correction was applied with the pre-stimulus -100 to 0 ms interval. After baseline correction, artifact rejection was carried out by reviewing the epochs both automatically and manually: At each channel, a 200-ms window was moved across the data (100 ms before and 800 ms after the stimulus) in 100-ms increments and any epoch where the peak-to-peak voltage exceeded 70 μV was rejected. We then reviewed the data, and adjusted the voltage threshold for individual subjects, to ensure that epochs contaminated by excessive blinking, body movements, skin potentials, and amplifier saturation were rejected. The mean rejection rate across participants was 19.2 ± 11.9% (mean ± SD); participants with greater than 40% trials rejected were excluded from further analysis. The following were the rejection rates for each condition: Baseline: 20.0 ± 12.4%; Complement: 17.9 ± 12.7%; High: 21.1 ± 12.1% and Low: 18.0 ± 10.5%.

Our hypotheses centered around the N400 response at the verb for the critical comparisons and at the noun for the sanity check items, so we selected nine electrodes over the central-parietal area (C3, CZ, C4, CP3, CPZ, CP4, P3, PZ, P4), known to show the most prominent N400 effect. We carried out a paired t-test on the mean amplitudes in the measurement time window of 300-500 ms, evaluating effects of Predictability (Baseline, Complement). The sanity check items that were designed to replicate standard N400 effects of cloze probability, we carried out a paired t-test over the same set of electrodes evaluating the effect of Cloze probability (High, Low).

3.2.5 Results

3.2.5.1 Behavioral data

The overall accuracy rate for the comprehension questions was 91 % (79%-100%), showing that participants were paying attention during the experiment.

3.2.5.2 ERP data

Figure 12 below presents the grand average ERPs to N400 effect of Predictability in the critical sentences (Baseline, Complement). Visual inspection suggests that the Complement condition elicited a larger N400 amplitude than the

Baseline condition. The results of the pairwise comparison show a significant effect ($t(32) = 4.31$, $p < 0.05$).

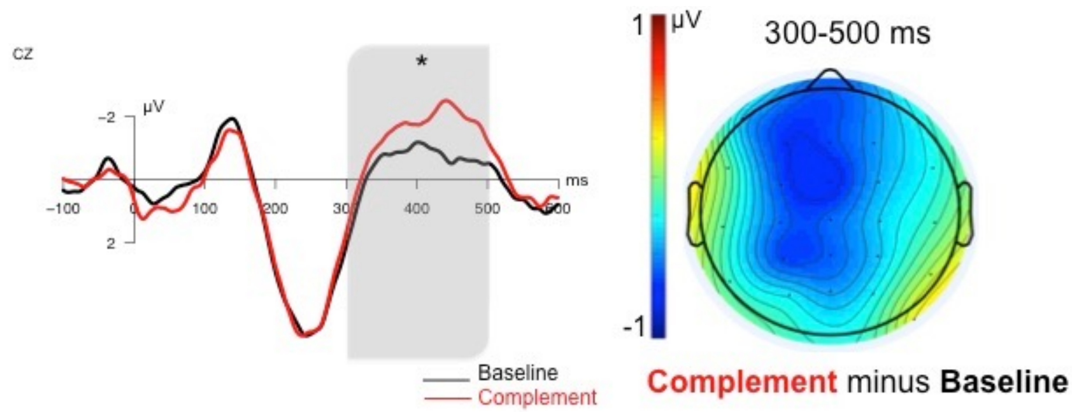

Figure 12: Grand average ERPs to predictability effect of Baseline and Complement at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 4 (Complement minus Baseline).

Figure 13 shows the grand average ERPs for the Cloze probability effect in the control items (High vs. Low). Visual inspection finds that the N400 response to the High condition is reduced relative to the Low condition. The results of the paired t-test show a significant effect ($t(32) = 4.89$, $p < 0.05$). The size of the critical N400 effect was approximately the same as in the control conditions that manipulated cloze probability through other contextual cues.

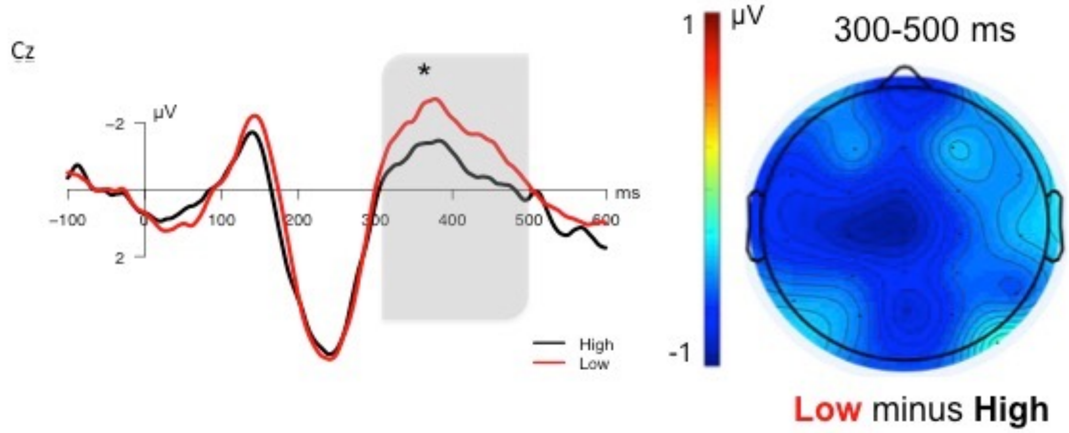

Figure 13: Left: Grand average ERPs to cloze control items at Cz in Experiment 1. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 4 (Low minus High cloze).

3.2.6 Discussion

The Bag of Arguments hypothesis predicts that there is a stage at which the identification of noun phrases that could be arguments of a verb could constrain prediction of that verb, but not the thematic role bound by the noun phrase. In this experiment, we observed a larger N400 to the verb *fired* in the Complement condition, because it was less likely to be an event that would be predicted by the only argument in the embedded clause, *servant.* By contrast, in Baseline condition, *fired* was the predicted event when the two arguments *millionaire* and *servant* were involved, and its N400 amplitude was reduced. This implied that given a slower presentation rate at 800 ms, comprehenders could identify if the noun phrase could be an argument of a verb, and update predictions of the verb based on the that information. The next question is whether argument role could effectively impact predictions of verbs at such a slower presentation rate. We will address this question in Experiment 5.

3.3 Experiment 5

In Experiment 5, we tested the second prediction of the Bag of Arguments hypothesis with the same slower presentation rate. To recap, the metaphor that the arguments are "lumped in a bag" is meant to express the hypothesis that there is a stage at which identifying the arguments of a verb could constrain prediction but the argument role information bound by the argument does not. In Experiment 4 we showed that at a presentation rate of 800 ms, whether a noun phrase is an argument of the verb did constrain the prediction of the verb. Therefore, in Experiment 5 we asked whether at the same presentation rate, argument role information does or does not impact prediction of the verb. Here, the key property of the design is that across Experiments 4 and 5, we tested the impact of argument identification and argument role with exactly the same timing and tightly matched experimental items. In Experiment 5 we kept the same items for the Baseline condition as Experiment 4. To create the role reversal items, we kept the morpheme *ba* in the Baseline condition, and then reversed the order of the two arguments (*Millionaire ba servant fired* vs. *Servant ba millionaire fired*). One might wonder why we ran a between-subject design as Experiments 4 and 5 rather than a within-subject Clausehood (Baseline, Complement) by Argument role (Baseline, Reversal) design. The main challenge for setting up a within-subject design was we found it difficult to generate a full set of 120 role-reversal sentences without repeating the target verbs. We thus decided to use 60 tightly matched sentences between experiments.

3.3.1 Participants

The participants were 37 naive young adults (24 females, 18-31 years old, mean: 23) from National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders. Of the 37 participants, 10 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves. The reported results were obtained from the remaining 27 participants (18 females, 18-31 years old, mean: 23). Informed consent was obtained from all participants. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

3.3.2 Materials

The experimental materials were 60 pairs of sentences comprising the two conditions: Baseline and (role) Reversal. We began with the same 60 Baseline sentences from Experiment 4. To create the role reversal sentences, we reversed the order of the two arguments and created role reversal sentences, for example: Baseline condition (*Millionaire ba servant fired*, meaning the millionaire fired the servant) and Reversal condition (*Servant ba millionaire fired,* meaning the servant fired the millionaire) (See Table 4). The 60 pairs of items were divided into two lists with latin square method. To show that participants did engage predictive mechanism during the experiment, we included the 30 pairs of cloze items in Experiment 4 as our control in Experiment 5. The same 120 filler sentences used in Experiment 4 were included here as well.

Two lists were constructed such that no sentence context or target word was repeated in the same list. Each list consisted of a total of 240 sentences, including 30 sentences in Baseline condition, 30 sentences in Reversal condition, 30 sentences of high-cloze target in High condition and 30 sentences of low-cloze target in Low condition, and 120 filler sentences. Participants were randomly assigned to one of the two lists.

Condition	Sentence	Post-target continuation
Baseline	富翁 把 僕人 解雇了	之後 立即 請來了 新的 管家
	Millionaire ba servant **fired-ASP**	then immediately hired new housekeeper
	"The millionaire had fired the servant and then immediately hired a new housekeeper."	
Reversal	僕人 把 富翁 解雇了	之後 立即 請來了 新的 管家
	Servant ba millionaire **fired-ASP**	then immediately hired new housekeeper
	"The servant had fired the millionaire and then immediately hired a new housekeeper."	

Table 4: Example stimulus in each condition in Experiment 5.

3.3.3 Procedure

The procedure was identical to Experiment 4. As in Experiment 4, 20% of the sentences would be followed by a comprehension question. Sentences in Reversal condition tended to be semantically implausible, which might be weird to be asked a comprehension question, participants were reminded of answering the questions based on the content they read.

3.3.4 Data acquisition and analysis

Data acquisition and analysis were identical to Experiment 4. The overall mean rejection rate across participants was 19.8 ± 10.3% (mean ± SD). Like Experiment 4, participants with rejection rate greater than 40% were excluded from further analysis. Rejection rates for each condition were summarized below: Baseline: 19.0 ± 11.8%; Reversal: 16.9 ± 8.6%; High: 22.6 ± 14.5% and Low: 20.9 ± 12.1%.

3.3.5 Results

3.3.5.1 Behavioral data

The overall accuracy rate to the comprehension questions was 91 % (75%-100%), showing that participants were paying attention during the experiment.

3.3.5.2 ERP data

Figure 14 shows the grand average ERPs for the Predictability effect to Baseline and Reversal conditions. Visual inspection suggested that there was no N400 difference between the two conditions. The results of the paired t-test similarly showed no significant difference ($t(26) = -0.47$, $p = 0.64$).

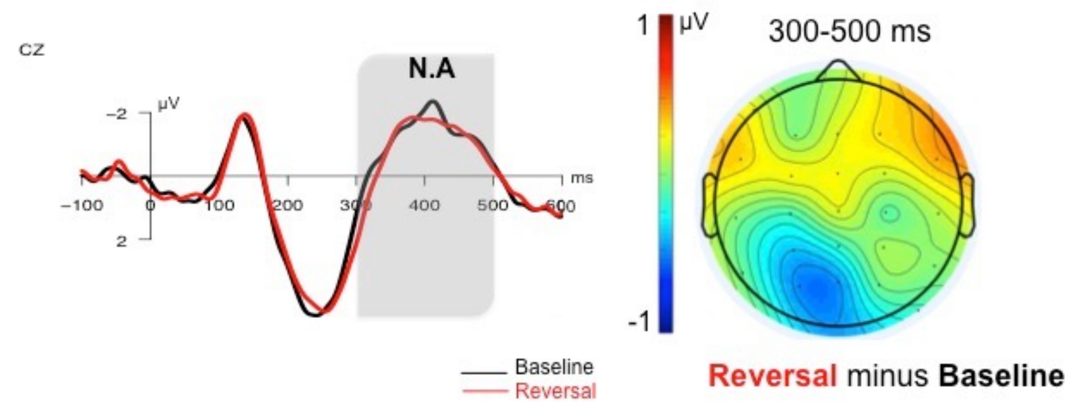

Figure 14: Grand average ERPs to cloze manipulations of Baseline and Reversal at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 5 (Reversal minus Baseline).

By contrast, Figure 15 shows the grand average ERPs to the High and Low conditions. Visual inspection showed that the N400 was reduced to the High relative to the Low condition. The results of the paired t-test showed a significant difference between conditions ($t(26) = -2.32$, $p < 0.05$).

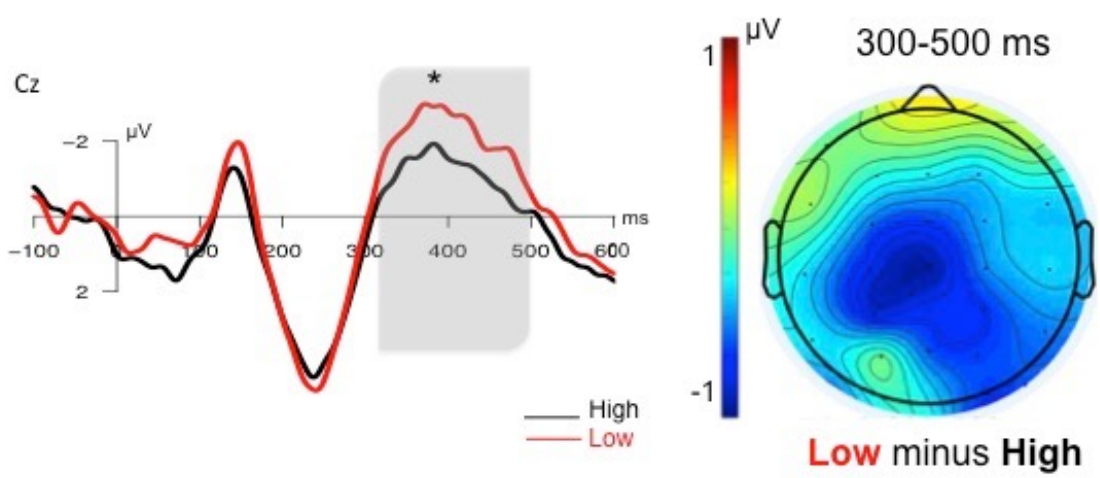

Figure 15: Left: Grand average ERPs to cloze sanity check sentences at Cz in Experiment 5. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 5 (Low minus High cloze).

3.3.6 Discussion

The Bag of Arguments hypothesis predicts there should be a period of time in which identifying the arguments of a verb could exert an effect on prediction but not argument role information bound by the arguments. In Experiment 4 we had observed that with 800 ms presentation rate, comprehenders could tell if the noun phrases could be arguments of a verb. In Experiment 5, we tested if argument role information could impact prediction with the same presentation rate. In particular, given *millionaire-as-an-agent* and *servant-as-a-patient*, the predicted verb would be *fired*, but the role reversal scenario (i.e. *servant-as-an-agent* and *millionaire-as-a-patient*) would not predict the verb *fired*. Interestingly, the N400 was not sensitive to role reversal situations, as if the verb *fired* were a good fit of event for a servant to act on a millionaire. As discussed in Section 3.1.1, the insensitivity of N400 to role reversal situations have been replicated in many languages with various verb final sentence structures (Kim & Osterhout, 2005; Kuperberg et al., 2007; Momma et al., 2015; Chow & Phillips, 2013; Chow, Smith et al., 2016). The null effect could not be attributed to lack of engaging predictive mechanism during the experiment, as we did observe an N400 effect to the cloze manipulation in our control items. A more likely explanation to the null effect of the role reversal situations, as suggested by Chow, Momma et al. (2016), is that it takes longer for prediction to be updated on the basis of argument role. For example, Momma et al. (2015) have found that the N400 effect emerged when the presentation rate was increased to 1200 ms.

In sum, in Experiments 4 and 5, we tested the Bag of Arguments hypothesis, which suggested that there existed a time window where identifying the arguments of a verb could constrain prediction, but not argument roles bound by the argument.

With Mandarin, we were able to manipulate whether noun phrases were arguments of a verb while keep the linear distance between the noun phrases and the verb identical. Results from Experiments 4 and 5 allowed us to narrow down the time window to compute different levels of argument-verb relations. Specifically, given a slower presentation rate at 800 ms, the parser was able to identify noun phrases that were arguments of a verb, and to use that information to update predictions, but not argument roles.

3.4 Experiment 6

The goal of Experiment 6 was to identify if there is a lower time limit for arguments of a verb to be identified to constrain predictions. If there is a time window at which the parser cannot tell if the noun phrases are arguments of a verb, such that only word associative effects are present (i.e. the Bag of Words hypothesis), then we should revise the two-stage model implied by the Bag of Arguments hypothesis into a three-stage model. We tested the same materials as in Experiment 4 (*Millionaire ba servant fired* vs. *Millionaire thought servant fired...*) with a faster presentation rate of 600 ms. Except for the presentation rate, other settings remained identical as Experiment 4.

3.4.1 Participants

The participants were 48 naive young adults (22 females, 18-33 years old, mean: 23) from National Taiwan Normal University. All of them were right-handed native Mandarin speakers, without a history of neurological or psychiatric disorders.

Of the 48 participants, 10 were excluded after pre-processing because of excessive eye-blinking, muscle potentials, sweat artifact and alpha waves. The reported results were obtained from the remaining 38 participants (20 females, 18-33 years old, mean: 23). Informed consent was obtained from all participants. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

3.4.2 Materials

The materials were identical to those in Experiment 4.

3.4.3 Procedure

The procedure was identical to Experiment 4, except for the presentation rate. The presentation rate was increased to 600 ms, with 500 ms stimulus duration and a 100 ms blank interval. See Figure 16 for details.

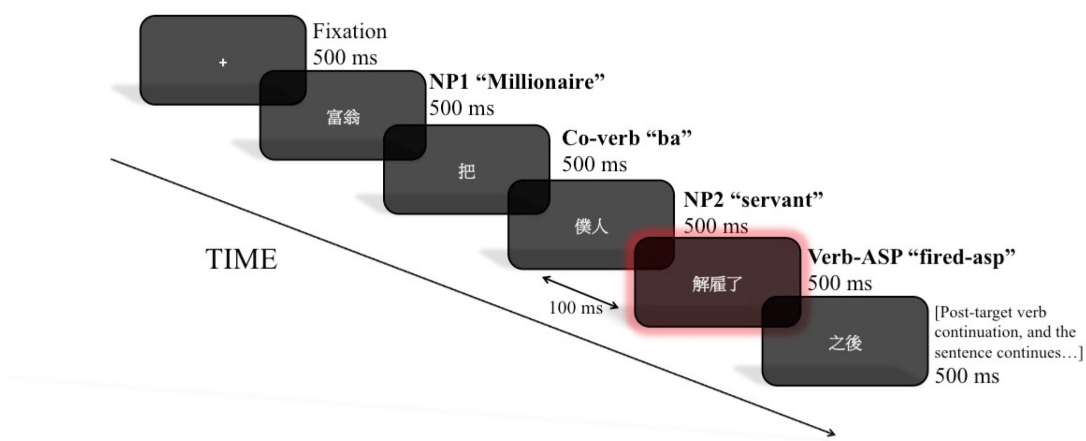

Figure 16: Presentation of stimuli in Experiment 6

3.4.4 Data acquisition and analysis

Data acquisition and analysis were identical to Experiment 4. The mean rejection rate across participants was 23.1 ± 12.7% (mean ± SD); participants with rejection rate greater than 40% were excluded from further analysis. The following were the rejection rates for each condition: Baseline: 22.0 ± 13.5%; Complement: 21.9 ± 12.7%; High: 22.7 ± 13.3% and Low: 22.2 ± 13.2%.

3.4.5 Results

3.4.5.1 Behavioral data

The overall accuracy rate for the comprehension questions was 93% (83%-100%), showing that participants were paying attention during the experiment.

3.4.5.2 ERP data

Figure 17 below is the grand average ERPs illustrating the N400 response in Baseline and Complement sentences. Visual inspection suggested that there was no N400 amplitude difference between the Think condition and the Baseline condition. The results of the pairwise comparison showed a null effect ($t(37) = 0.10$, $p = 0.75$).

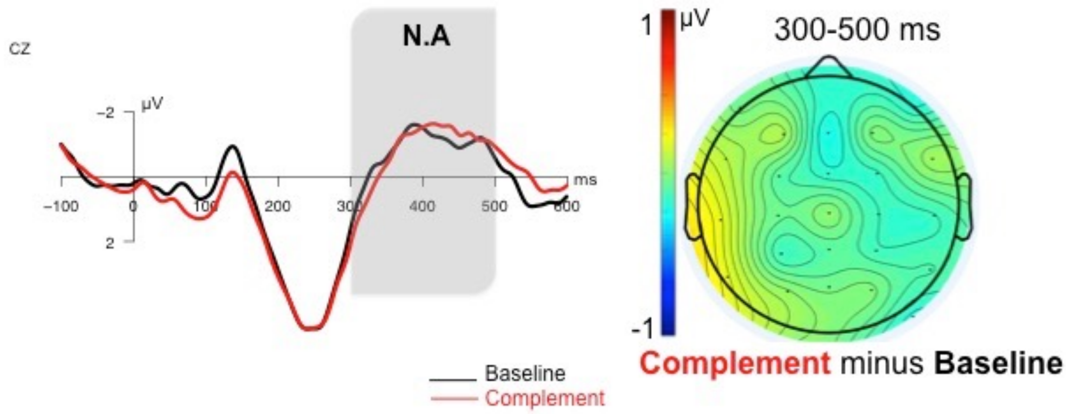

Figure 17: Grand average ERPs to predictability effect of Baseline and Complement at Cz and the topographic distribution of ERP effects in the 300-500 ms interval in Experiment 6 (Complement minus Baseline).

Figure 18 shows the grand average ERPs to High and Low for the control items. Visual inspection suggested that the N400 amplitude was reduced for the High cloze relative to the Low cloze ones. Paired t-test also confirmed the visual inspection ($t(37) = 6.35$, $p < 0.05$).

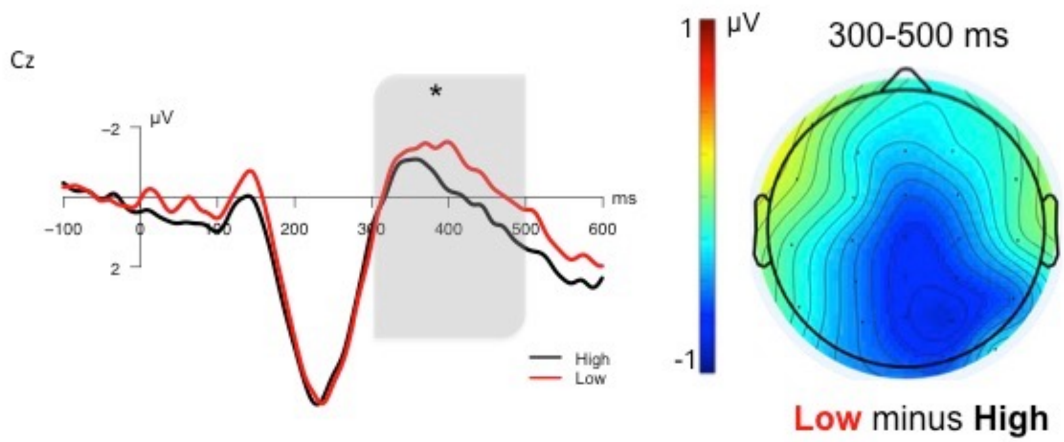

Figure 18: Left: Grand average ERPs to cloze control items at Cz in Experiment 6. Right: The topographic distribution of ERP effects in the 300-500 ms interval in Experiment 6 (Low minus High cloze).

3.4.6 Discussion

In Experiment 6, we aimed at investigating if there is a lower time limit for the parser to detect if noun phrases could be arguments of a verb and use that information to update predictions. We used a slightly faster presentation rate (600 ms) than in Experiment 4 (800 ms). Prior studies have already reported the absence of argument role effects on the N400 at a 600 ms presentation rate (Chow & Phillips, 2013; Kuperberg et al., 2007), as we also showed for the slower 800 ms presentation rate in Experiment 5. Here we rather tested whether the parser could detect if the noun phrases could be arguments of a verb at the 600 ms presentation rate. Different from Experiment 4, where we found an N400 effect in the comparison between Complement and Baseline conditions with the slower presentation rate (800 ms), we found that this effect was absent with the faster presentation rate (600 ms). In other words, under time pressure, prediction of the verb was no longer constrained by arguments of a verb.

The null effect of N400 showed that there is a time window to identify whether a noun phrase is argument of a verb or not; if the time lapse is not long enough, then the parser cannot tell. This was what happened when the presentation rate was increased to 600 ms. The two noun phrases in the Complement condition were parsed as if they were arguments of the verb, which was the case in the Baseline condition. Their N400 responses did not differ from each other. The patterns observed here were compatible with predictions from the Bag of Words hypothesis, which suggests that structure played a limited role in initial verb prediction; word associations were sufficient to account for the effects. One alternative explanation for different results between 600 ms (Experiment 6) and 800 ms (Experiment 4) rates is to suggest that the 600 ms rate was too fast for processing the sentences in general. However, we note that 600 ms/word is in fact on the slower side against most RSVP sentence paradigms in ERP. More importantly, we still obtained an N400 effect of cloze contrast in our control items. This finding is crucial, because it shows that participants did engage predictive mechanism during the experiment with the faster presentation rate.

3.5 General Discussion

Three ERP experiments were conducted to map the time course of argument-verb relation computations. We placed two noun phrases before a verb, and systematically evaluated the timing for different pieces of argument information to impact the prediction of a verb. Results from Experiments 4 and 5 showed that with the slower presentation rate at 800 ms, comprehenders were able to update predictions based on argument(s) of the verb, but prediction based on argument roles was not yet effective. By contrast, when the presentation rate was increased to 600 ms in Experiment 6, comprehenders could no longer detect if the noun phrases could be arguments of an upcoming verb or not. Under time pressure, verb prediction was mainly based on nearby words.

Our work provides important support for the Bag of Arguments hypothesis proposed by Chow, Smith et al. (2016), which suggests that there exists a time window at which argument role information could not inform the prediction of an upcoming verb, but noun phrases that are in the same clause as the verb can. What is implied by their conclusion is the parser is able to identify which noun phrases could be arguments of the upcoming verb, potentially based on the structure cue provided by the clause boundary. To our knowledge, Chow, Smith et al. (2016) is the only study so far that supports the prediction of the Bag of Arguments hypothesis. However, in their experiment setup, whether a noun phrase was an argument of the upcoming verb or not is confounded with linear distance between the noun phrases and the verb. In addition, the temporal dynamics of the processing stages by which the parser incorporates different pieces of information from the arguments to predict the verb remain relatively vague. In our work, we have controlled linear distance between the noun phrases and the verb, and identified a time window that distinguishes arguments identification and argument role computation. In particular, with a presentation rate of 800 ms, the parser was able to identify if noun phrases could be arguments of an upcoming verb, and update predictions based on that. However, prediction on the basis of argument role was not updated at this time point.

Our work goes beyond Chow, Smith et al. (2016), as we show that there is a lower time limit for the parser to detect if a noun phrase could be an argument of the upcoming verb and to use this information to constrain verb predictions. When the time elapsed between the onset of the argument and the verb was 600 ms, we saw no evidence that the parser had identified if the noun phrase was an argument of the verb. Such a finding was in line with the predictions of the Bag of Words hypothesis, which suggested that initial verb prediction was not constrained by structure at all; the mechanism at work was simply word associations. Chow, Smith et al. (2016) do not itself contain data to support a temporal distinction between the Bag of Words and the

Bag of Arguments hypotheses. Our data fills in the gap, as we are able to identify time windows for each of the stages suggested the hypothesis.

In sum, given the findings of our experiments, we would like to propose a model of argument-verb relation computations. We will discuss the model in detail in the section below, but here we would like to point out how our model differs from the proposal of the Bag of Arguments (Chow, Smith et al., 2016). To begin with, the Bag of Arguments hypothesis suggests a two-stage model of argument-verb relation computations. In particular, the Bag of Arguments hypothesis states that there exists a minimum time window for arguments of a verb to impact predictions, but not argument role information. In the current study, we show evidence that there exists a stage before arguments of a verb are identified to inform predictions, in which only word association is at work (the Bag of Words hypothesis). We were able to temporally distinguish stages at which the Bag of Words and the Bag of Arguments hypotheses hold. When such information is incorporated into the model, we ended up with three stages of argument-verb relation computations.

The other aspect in which our model differs from the Bag of Arguments model of Chow, Smith et al. (2016) is at which stage the computation is slow. Chow, Momma et al. (2016) propose that prediction should be seen as a memory retrieval process, and that it is searching through the memory space for a best fit that slowed down predictions—not computing argument role information per se. Take the role reversal situation as an example to illustrate their view, Chow, Momma et al. (2016) argue that comprehenders could compute the thematic relation *bee-as-a-patient* promptly; but prediction is slowed down because comprehenders have to search through their memory space for an event that properly involved *bee-as-a-patient*. In the current study, we did not have evidence to argue for or against Chow, Momma, et al.'s (2016) idea that prediction is slow as a result of memory search, but we tend to rather favor the possibility that parsing itself—that is, the process of assigning argument roles—is slow. We will walk through our rationale in details in the following section.

3.5.1 Toward a processing model of argument-verb relation computations

Based on the results of the three experiments and the findings from prior research (Momma et al., 2015; Chow et al., 2018), we would like to propose a processing model of computing argument-verb relations (see Figure 19). Note that the emphasis should not be placed on the exact time point for each level of computation to occur, but rather levels of argument information the parser computes as time evolves. As depicted in Figure 19, our model suggests that there are three stages for different levels of argument information to be computed in argument-verb relation computations. At an early stage, initial verb prediction is based on word associations. The parser does not differentiate whether these noun phrases are arguments of an upcoming verb; it simply finds an event that can be associated with both noun phrases. For example, as *fire* could be a very plausible event to involve both *a millionaire* and *a servant*, when under time pressure, the parser does not consider other cues in the context in addition to the semantic relatedness between the noun phrases and the verb. This is the time window for the Bag of Words mechanism, or simple word association to work. Then, at an intermediate stage, the parser becomes more sensitive to structural cues. The parser is able to identify whether noun phrases are arguments of the verb and use that information to update predictions (the Bag of Arguments hypothesis). It is only at a later stage that the parser starts to compute argument role information (e.g. servant-as-an-agent and millionaire-as-a-patient) and construct the full structure of a sentence.

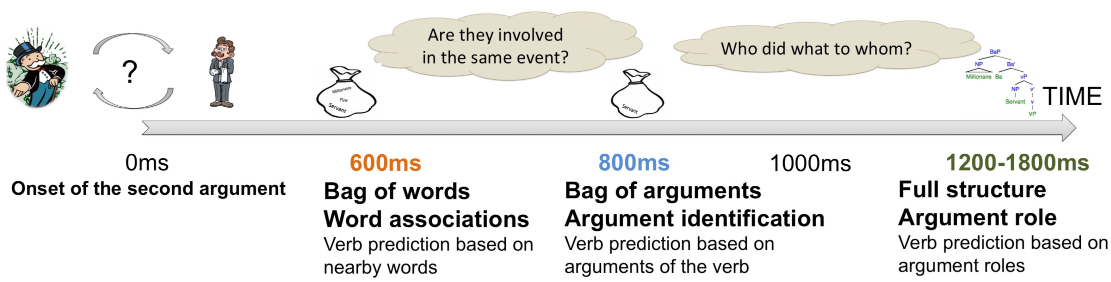

Figure 19: The three-stage processing model of argument-verb computations[3].

Our data from Experiments 4-6 allowed us to identify the time windows for the Bag of Words and the Bag of Arguments stages, and we relied on prior work to identify the time window (between 1200-1800 ms) for argument roles to exert an effect on predictions. In particular, Chow et al. (2018) find that the N400 is not sensitive to role reversal situations when the time lapse between the last argument and the verb is 600 ms, but that the N400 effect emerges when the lapse is increased to 1800 ms, while Momma et al. (2015) with simple Japanese sentences narrow the time window to 1200 ms. We systematically reviewed existing studies on role reversal manipulation, and found that a majority of research that reported a null N400 effect

[3] The time zero is set at the onset of the second argument, and the temporal scale is marked by the time lapse between the presentation of the second argument and the onset of the verb. The idea is to show the amount of time comprehenders have when all the argument information is available in the context for them to predict the verb.

had an lapse window shorter than 1200 ms (Chow & Phillips, 2013; Chow et al., 2015; Hoeks, Stowe, & Doedens, 2004; Kuperberg et al., 2007).

We consider this model a parsing model, which illustrates the time course for different levels of argument-verb information to be computed to feed prediction. Our model implies that parsing is slow, as the parser is only able to compute sophisticated structural information as longer amount of time is granted. Such a slow parsing view is different from a slow prediction view. To be more precise, the slow prediction view holds that computing the relations of an argument and its argument role is not taxing; what slows down prediction is the memory search process to retrieve the best fit of the context (Chow, Momma et al. 2016). In other words, under the slow prediction view, it would not be too challenging to compute *millionaire-as-a-patient* and *servant-as-an-agent*. What slows down prediction is to search for an event that involves them. Momma et al. (2015) examine ERP responses to pre-verbal arguments, coupled with different case markers, such as *bee-accusative* vs. *bee-nominative*. Their results show that the N400 amplitude is larger in arguments with an accusative case relative to a nominative case. They interpret the patterns as showing that the relation between an argument and its argument role could be established very early. To us, it seems not very clear if the N400 effect reflects differences between arguments coupled with different roles (e.g. *bee-accusative* vs. *bee-nominative*), or lexical differences of different case markers (-*accusative* vs. -*nominative*). Therefore, we think the existing evidence is neutral on whether the observed delays reflect slow prediction or slow parsing, and thus for now we prefer to couch the current model in terms of slow parsing. Still, we would like to clarify that it is not our intention to argue against the slow prediction view. In fact, since our experiments were not set up to test the slow prediction view, we do not have direct evidence to argue for or argue against it.

Note that although our model implies that argument role is not committed to initially, we do not suggest that argument role information will never be computed before the presence of a verb. Rather, our suggestion is that there is a minimum time window for different levels of argument information to be computed. Before argument role relation is established, the parser has to identify whether the noun phrases are arguments of the verb.

3.5.2 Reconciling these results with prior role reversal findings

It is important to note that a few studies have reported obtaining an N400 effect in role reversal manipulations with time lags shorter than 1200 ms (Bornkessel-Schlesewsky et al. 2011; Ehrenhofer, 2018). We suggest that there are additional factors that contribute to the shift of the temporal course of argument-verb computations, and that these factors could be further evaluated to extend the scope of the current model. For example, Bornkessel-Schlesewsky et al. (2011), who find an N400 effect in role reversal materials in Turkish and Mandarin, conduct their experiments aurally, in contrast with most other role-reversal studies in the literature. Auditory presentation provides phonological cues, such as coarticulation (and tone sandhi in Mandarin), which are not available in visual presentation. In addition, unlike visual presentation, auditory materials are more difficult to control for the synchronization of the onset and duration of target words across experiment items. In our opinion, the impact from lower level phonetic cues on argument-verb computations might not be significant; without synchronizing the onset and duration of the arguments and the verb, cross-modality comparison does not seem very feasible. How to extend our model to incorporate data in the auditory modality would be a direction for future research.

As the current model was based on data in Mandarin, we would also like to draw special attention to Bornkessel-Schlesewsky et al. (2011), where the authors report an N400 effect in Mandarin role reversal manipulations. In addition to modality differences discussed above, it should be noted that they only found an N400 effect in passive *bei* constructions in Mandarin, not disposable *ba* constructions. Both constructions introduce two preverbal arguments (*ba*: SOV structure; *bei*: OSV structure), but according to Bornkessel-Schlesewsky et al. (2011), only *ba* construction involves structural ambiguity at the verb. The parser might not consider the verb anomalous as it permits a continuation as a relative clause (see Example 2), so the N400 effect is absent at the verb in *ba* constructions. However, we do not find such an interpretation very convincing, as in fact both *ba* and *bei* constructions could take a relative clause continuation after the verb (see Example 3). The absence of N400 effect could not be attributed to the potential structure ambiguity in *ba* constructions. In fact, we believe that the N400 effect in *bei* constructions is more likely to have resulted from a language specific pragmatic principle in Mandarin. Specifically, Mandarin passive *bei* involved a negative connotation. The patient of a passive *bei* sentence always bore a negative consequence of an event, which is reflected as a bigger N400 as early as the presence of the second argument (Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008). What this means is that the pragmatic cue encoded in the passive marker *bei* could facilitate the computation of verb-argument relation, such that the parser was able to detect the role reversal situation more quickly. In the future, we could investigate if the "negative" implication of BEI is not the same kind of information as thematic role information.

(2) 偵探 把 [[子彈 擊中 的] 罐頭] 拿走了。

Detective ba [[bullet hit de] tin] take-away-ASP

"The detective took away the tin which the bullet hit."

(3) 偵探 被 [[子彈 擊中 的] 罐頭] 割到了。

Detective bei [[bullet hit de] tin] cut-ASP

"The detective was cut by the tin which the bullet hit."

Finally, Bourguignon, Drury, Valois, & Steinhauer (2012) show that verb types could modulate the N400 effect in role reversal situations, at least in English. The authors on one hand replicate Kuperberg et al. (2007), showing an absent N400 effect of role reversal with action verbs ("The boys have eaten" vs. "The fries have eaten"); on the other hand, they examine role reversal with psych-verbs, and did obtain an N400 effect at the verb ("The judges have despised" vs. "The movies have despised"). Note that in Bourguignon et al. (2012) the time elapsed between the onset of the argument and the verb is 1000 ms (as they use 500 ms SOA), so the finding for action verbs might still be accommodated by the timing proposed in the current model. However, with the same timing, the N400 effect observed for psych-verbs is more difficult to explain. It is possible that the contrast between the sentient and the nonsentient is psychologically salient, such that given a subject that is nonsentient, the verb is less likely to be a psych verb. By contrast, for the action verbs, the finer distinction (e.g. edible vs. not edible) is not immediately available to the comprehenders; it's not a major division in how comprehenders immediately see the world. Anyway, the intriguing psych-verb data point offers us with a direction to examine the broader question of how verb types interact with argument features identified in the model, such as argument identification and argument roles. We leave this question for future explorations.

3.6 Conclusion

Based on the results of prior studies and our three experiments, we have proposed a model of the time course of argument-verb relation computation. At an initial stage, when the time lapse between the onset of the second argument and of the verb is 600 ms only, the parser does not differentiate structures among the noun phrases. This is the time window for the Bag of Words hypothesis, or simple word association, to work. Then, at an 800 ms lapse, the parser is sensitive to whether the noun phrases are arguments of the upcoming verb, but argument role information does not come into play at this time window (the Bag of Arguments hypothesis). It is only at a later stage (between 1200-1800 ms) that the parser starts to consider argument roles in computing argument-verb relations. Our model thus maps a more detailed time course for online sentence comprehension.

Chapter 4: ERP sensitivity to subcategorization violations in L2

4.1 Introduction

Everyone knows that acquisition of a second language in adulthood is hard. This chapter is part of a broader research endeavor aimed at determining which components of L2 learning are harder than others, and why. Specifically, we will investigate the hypothesis that acquiring subcategorization frames that conflict with their L1 is particularly difficult for L2 speakers.

Only a handful of prior psycholinguistic and electrophysiological studies have investigated the impact of L1 transfer on subcategorization knowledge in L2. Part of the reason may be Clahsen and Felser's influential work (2006) emphasizing the idea that L2 speakers can successfully compute local verb-argument relations in canonical order even when they fail on more complex aspects of the sentence structure. However, as verb-argument computation involves lexical syntax, evidence of native-like computation of verb-argument relations requires more than just getting the interpretation right. To be more specific, while event concepts of common verbs are likely to be the same for speakers of different languages (e.g. eat, sleep), which arguments a verb subcategorizes for is linguistic knowledge, and could vary from language to language. When subcategorization information does not match between L1 and L2, how does this mismatch impact L2 speakers in real time?

In framing the current study, we assume the hypothesis that subcategorization information is encoded in the verb (Chomsky, 1981; 1995). In particular, when a verb is presented before arguments, the incoming arguments are checked against the subcategorization information retrieved from the verb's lexical entry. However, even after the subcategorization knowledge of a verb in L2 has been learned, such that it now constitutes part of the L2 speaker's linguistic knowledge, the process of accessing the subcategorization information from the lexical entry during online sentence comprehension may be more error-prone in L2. This could be particularly problematic in a case in which the L1 verb corresponding to the same event concept has a substantially different subcategorization frame from the L2.

The current study was designed to test this hypothesis, that online verb-argument computation is impacted by L1 transfer. In other words, subcategorization of verbs in L1 will have a significant impact on how learners parse sentences in L2. For example, in English, "bark" takes only one argument whereas it can take two arguments in Mandarin. Therefore, English L1 speakers will reject sentences like "The dog is barking the girl" but those sentences might be accepted by L2 speakers whose L1 is Mandarin, as it means "The dog is barking at the girl" in Mandarin. We will evaluate this hypothesis with EEG, as it has excellent temporal resolution and allows us to track how bilinguals resolve the grammaticality conflict in their two languages in real time.

4.1.1 Prior behavioral studies investigating the processing of subcategorization in L2

A handful of prior studies using behavioral methods have investigated how processing of subcategorization is impacted in L2. Jiang (2007) asked whether L2 speakers could detect subcategorization violations online by examining their reaction time profile in self-paced reading. This study focused on speakers whose L1 was

Mandarin and whose L2 was English. Materials were sentences of a variety of structures, but the ungrammatical ones always involved a complement that the verb did not subcategorize for (e.g. "The mayor promised to offer/*keep the returning advisor a better position soon"; "The teacher wanted/*insisted the student to start all over again"). The results showed that Mandarin L2 speakers of English had a similar processing profile as L1 English speakers. Although the reading time in the L2 was in general slower, both groups showed a slowdown at regions after the subcategorization violation occurred. Jiang (2007) thus argued that L2 speakers were sensitive to verb subcategorization errors in real time. However, Jiang (2007) did not focus on the existence of discrepancies between L1 and L2 subcategorization frames. Therefore, even though L2 speakers appeared to quickly detect subcategorization errors in this study, these results leave the question open of whether native-like detection of subcategorization violations depends on facilitation through their L1.

Although to our knowledge no previous behavioral studies have investigated this question through a violation paradigm, several studies have suggested L1 transfer of subcategorization information by showing that subcategorization *preferences* could be carried over from L1 (Dussias & Cramer Scaltz, 2008; Dussias, Marful, Gerfen, & Bajo, 2010). For example, Dussias, et al. (2010) ran a norming task of 100 English verbs on late Spanish L2 speakers of English. Sentence frames, which contained only a subject and a verb, were given to the bilinguals. The authors looked at the structure of the continuations that the L2 speakers provided, and compared the results with the norming data collected from native English speakers in Garnsey, Pearlmutter, Myers and Lotocky (1997). The cross-study comparison showed that among the 100 verbs, 39 had between-group subcategorization differences. In addition, 10 of the 39 verbs showed a transfer effect from L2 speakers' L1. Although the proportion was not very high, the results suggested that L1 subcategorization could play a role in processing L2.

On the contrary, some researchers have taken the stand that subcategorization preferences are not subject to L1 transfer (Flett, Branigan, & Pickering, 2013; Gries & Wulff, 2005), based on evidence from structural priming manipulations in production. Flett et al. (2013) presented unrelated preposition object (PO) or double object (DO) sentences to participants before they described pictures of a dative event. They compared the sentences produced by German L2 speakers of English, Spanish L2 speakers of English, and L1 English controls. Importantly, while German can take both PO and DO structures, Spanish only allows PO structure. Flett et al. (2013) argued that models in which native structural preferences impact non-native ones should predict several differences between the German and Spanish L2 groups. First, because German speakers encounter PO and DO structures at roughly equal frequencies in their native language, and Spanish speakers only encounter PO structures, such models should predict that the baseline rate of DO structures in L2 English production should be lower for Spanish speakers. Second, because structural priming generally shows an "inverse preference effect" in which less frequent structures are primed more strongly than frequent ones, Spanish speakers should show a relatively larger priming effect for DO structures than German speakers or English native speakers. However, regardless of their linguistic backgrounds, participants of the three groups showed the same baseline rate of DO structures, and showed the same degree of priming in producing a DO structure followed by reading DO sentences or a PO structure followed by reading PO sentences. These results led the authors to downplay the effect of subcategory restrictions from participants' L1. However, we should be cautious about generalizing the results of a priming paradigm. With a brief exposure to a sentence in L2 comprehension, its structural representation may become more activated in speakers' mind to impact immediately subsequent production, washing out L1 impacts. However, this doesn't rule out the possibility that the initial comprehension process is impacted by the L1. For example, it is possible that if we used a different paradigm (such as a violation paradigm), which did not consist of comprehension-to-production priming, L1 transfer could be observed.

4.1.2 ERP studies of subcategorization violation in L1 and L2 speakers

As we have reviewed extensively in prior chapters, EEG has high temporal resolution for tracking task-related computation online, and a number of prior studies have identified ERP responses that appear to be tied to the detection of subcategorization violations in native speakers. Such violations have generally elicited N400 and P600 responses in ERP, although there is some variability.

Friederici and Frisch (2000) was an early landmark study of the response to argument structure and subcategorization violations in ERP, which manipulated two kinds of mismatches between the verb and its arguments in German: number of arguments and type of object arguments. In their first experiment, two arguments preceded the verb (*Anna knows that the inspector*$_{NOM}$ *the banker*$_{ACC}$ *monitored..*).

Number of arguments mismatch was achieved by substituting an intransitive verb for a transitive one, such that the second argument became an unlicensed argument (**Anna knows that the inspector*$_{NOM}$ *the banker*$_{ACC}$ *departed…*). By contrast, type of arguments was achieved by substituting a verb that assigned dative case to the object instead of accusative case (**Anna knows that the inspector*$_{NOM}$ *the banker*$_{ACC}$ *helped…*). ERP responses were time-locked to the onset of the verb. The ERP results showed that compared with canonical sentences, mismatch of number of arguments elicited an N400-P600 biphasic response, and mismatch of type of arguments elicited a LAN-P600 response. As German has relatively free word order, in their Experiment 2, they placed a verb before two arguments (*Today visited the cousin*$_{NOM}$ *the violinist*$_{ACC}$ *in the hospital*), and then manipulated the same kinds of mismatches between the verb and its arguments in Experiment 1. Number of arguments mismatch was achieved by replacing a transitive verb with an intransitive one, such that the second argument became an unlicensed argument (**Today dawdled the cousin*$_{NOM}$ *the violinist*$_{ACC}$ *in the hospital*). Type of arguments was achieved by marking the second argument with a wrong case marker, a dative, when all the critical verbs assigned accusative to their object (*Today visited the cousin*$_{NOM}$ *the violinist*$_{DAT}$ *in the hospital*). ERP responses were time-locked to the onset of the second argument. As in the first experiment, they found that compared with canonical sentences, mismatch of number of arguments elicited both an N400 and a P600 effect, while here mismatch of type of arguments only elicited a P600 effect. Similar results for number of arguments mismatches were observed as part of a Dutch study aimed at investigating the P600 across modalities (Hagoort & Brown 2000), where they reported a P600 effect for these mismatches in both auditory and visual presentation, preceded by a small, fronto-central negativity, larger with auditory presentation. Kielar, Meltzer-Asscher and Thompson (2012) examined similar kinds of subcategorization violations of intransitivity in English ("John visited/*sneezed the doctor"). They reported an N400-P600 effect to the determiner and the noun in both young and healthy elder adults, but only a P600 effect in agrammatic aphasia patients. In a different kind of subcategorization violation paradigm, Osterhout, Holcomb and Swinney (1994) studied the ERP responses to an auxiliary verb in a subordinate clause, which was introduced either by a complement verb or by a transitive verb that did not take a finite complement clause ("The doctor hoped/*forced the patient was lying"). They also found an N400-P600 biphasic response at the auxiliary verb.

How should we understand the ERP responses to subcategorization violations? Studies discussed above all reported a P600 effect, and they all considered it a reflection of syntactic processes that involve reanalysis and repair (Friederici & Frisch, 2000; Hagoort & Brown 2000; Kielar et al., 2012; Osterhout et al., 1994). The negativity occurred before P600 varied in terms of strength and distributions. Hagoort and Brown (2000) noticed that in their data, the distribution of the negativity was more prominent over left frontal cites. They suspected that this effect was a LAN, which reflected word category violations (Friederici, Hahne, & Mecklinger 1996). Different from Hagoort and Brown (2000), other studies found that the distribution of their negativity was more widespread, and was more prominent over the central-parietal cites, they thus interpreted it as an N400 response (Friederici & Frisch, 2000; Kielar et al., 2012; Osterhout et al., 1994). Friederici and Frisch (2000) treated this as an indication of difficulties in integrating lexical information into the context, but for Kielar et al. (2012), it reflected problems in lexical access. From my point of view, if the N400 effect was real, it could also be interpreted as encountering a word whose conceptual and semantic features were not pre-activated. Consider the example “John visited/*sneezed the doctor” from Kielar et al. (2012). Given the intransitive verb “sneezed,” comprehenders would not expect an argument to come next. By contrast, by getting a transitive verb “visited,” the parser may start to expect an argument, and pre-activate features that could be related to that argument. It may be for this reason that the N400 amplitude to the subsequent noun phrase was larger in the intransitive context relative to the transitive context.

To our knowledge, only a handful of studies that have investigated L2 subcategorization violation with EEG. Karatas (2019), in a manipulation of lexical case violations in Turkish, reported a widespread negativity and a P600 effect on the verb at which the violation can be detected in the L1 group, but the P600 effect disappeared in the advanced L2 speakers. Guo, Guo, Yan, Jiang, and Peng (2009) asked a very different question than Karatas (2019). They wondered if L2 speakers relied on a lexical-semantic strategy to process problematic syntactic structures such as subcategorization violations. In a series of prior studies, Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, and Molinaro (2006) showed that L2 speakers often demonstrate N400 effects to syntactic violations such as agreement that would normally elicit P600 effects in L1 speakers. Therefore, Guo et al. (2009) predicted that L2 speakers would show an N400 effect to subcategorization violations, whereas the L1 group a P600 effect. The materials were adapted from Jiang (2007) as discussed in the section above, and sentences of similar structures. In both studies, they compared verbs of different subcategorization frames, with the ungrammatical sentences always involving a complement that the verb did not subcatecorize for (e.g. "The mayor promised to offer/*keep the returning advisor a better position soon"; "The teacher wanted/*insisted the student to start all over again"). ERP responses were time-locked at a critical word at which subcategorization violation can be detected, which collapsed across a variety of parts of speech. The results confirmed their prediction: While L1 speakers showed a P600 effect to subcategorization violations, Mandarin L2 speakers of English showed an N400 effect. However, it is worth noting that, as acknowledged by the authors, the negativity obtained in L2 speakers was only prominent over very lateral electrodes, which was very different than a traditional central-parietal N400 distribution. The results of Guo et al. (2009), did not speak directly to our main question about whether subcategorization restrictions in L1 interfere with L2 verb-argument computation, because like Jiang (2007), Guo et al. (2009) did not examine the difference in subcategorization restrictions between L1 and L2 in constructing the materials and analyzing the results.

4.1.3 The current study

Our goal was to test if L2 verb-argument computation is subject to substantial L1 transfer online. We chose to look at the case of L1 Mandarin and L2 English, because there is a considerable amount of English intransitive verbs whose Mandarin translation could be either transitive or intransitive (i.e. ambitransitive). For example, *My sister listened the music* is acceptable in Mandarin but unacceptable in English. If

L2 speakers are sensitive to English subcategory restriction online, they should show an ERP violation response parallel to native speakers when the unacceptable noun phrase is encountered. We predict that we will obtain a P600 effect to subcategorization violations in L1, because this result has been very reliable across studies (Friederici & Frisch, 2000; Hagoort & Brown 2000; Kielar et al., 2012; Osterhout et al., 1994). By contrast, the negativity observed before P600 in prior studies varied in strength and distribution, and could be dependent on the contextual expectations afforded by the design. Although for this reason this earlier response appears less ideal for testing the transfer hypothesis, we include this "N400" window in our analysis to provide a point of comparison with prior research.

One concern in L2 research aimed at a specific element of processing, is that the L2 speakers may simply show insensitivity or non-native-like responses to all of the experimental manipulations, raising questions about the specificity of the results. Therefore, to show that our L2 speakers of English were able to parse English sentences in real time and recognize grammatical violations that are not subject to L1 transfer, we included a control comparison in the current study. Prior work has showed that even Mandarin L2 speakers who were less proficient in English demonstrated an L1-like ERP response to phrase structure violations like "a proof <u>of</u> the theorem" vs. "Max's <u>of</u> proof the theorem" (Weber-Fox & Neville, 1996). Therefore, we included the same contrast as a control comparison in the current study, where the L2 speakers would be expected to show ERP sensitivity to the violation on any account.

4.2 Experiment 7

4.2.1 Participants

22 native English speakers (10 females, 18-26 years old, mean: 20.4) and 28 English learners whose L1 was Mandarin (18 females, 18-32 years old, mean: 23.5) participated in the study. L1 speakers of English were recruited at the University of Maryland, and none had ever been exposed to Mandarin before. L2 English learners were recruited at National Taiwan Normal University in Taiwan. On average, the L2 speakers reported starting to learn English around the age of 7 (SD = 2). None of them had been exposed to an English-only environment for studying English. All of them were proficient in English, with the following self-reported English proficiency in different skills (1 = not fluent at all; 7 = very fluent): Listening: 5.6 (SD = 0.7); Speaking: 5.2 (SD = 1.0); Reading: 5.6 (SD = 0.9); Writing: 5.0 (SD = 0.8). As a more objective measure of fluency, all participants had passed a standardized English proficiency test beyond Common European Framework of Reference for Languages (CEFR) B2 level. Both groups of participants were right-handed and did not have a history of neurological or psychiatric disorders. All of them consented to participate in the experiment. The experiment protocol was approved by the Institutional Review Board Office at the University of Maryland College Park.

4.2.2 Materials

The critical subcategorization stimuli were sentences of Subject-Verb-Object structure, with the verb being varied between two conditions: (1) Grammatical and

(2) Ungrammatical subcategorization. Verbs in the Grammatical conditions were transitive in both English and Mandarin (e.g. record), whereas verbs in the Ungrammatical conditions usually do not take a direct object in English but are ambitransitive in Mandarin (e.g. listen). Note that although verbs in the Ungrammatical condition were intransitive, they can introduce a subsequent argument with the insertion of a preposition (e.g. listen to the music). The selection of the verbs was based on the intuition of the researcher (Mandarin native speaker with L2 English), and cross-checked with the online Cambridge English Dictionary (https://dictionary.cambridge.org/zht/) and another Mandarin native speaker. We matched the lexical frequency and word length of verbs across the two conditions (lexical frequency: Grammatical: 17352, Ungrammatical: 19382, $t(59) = 0.46$, $p = 0.65$; word length: Grammatical: 6; Ungrammatical: 6, $t(59) = 0.55$, $p = 0.58$). Except for the verbs, the rest of the sentences remained identical. 60 pairs of critical sentences were created, and were proofread by three native English speakers. To ensure that not all sentences with an intransitive English verb were ungrammatical and vice versa, we added two filler conditions with grammatical intransitive verbs and with ungrammatical transitive verbs (Table 5). Each filler condition consisted of 30 sentences.

Condition		Example stimuli
Subcategorization	**Grammatical**	My sister is recording the music.
	Ungrammatical	My sister is listening the music.
Phrase structure (Control items)	**Standard**	The scientists criticized Max's proof of the theorem.
	Anomaly	The scientists criticized Max's of proof the theorem.
Filler	**Grammatical**	The singer sneezed during the concert.
	Ungrammatical	The leader should impose by next week.

Table 5: Example stimuli in each condition in Experiment 7

To show that L2 speakers were able to parse English sentences, we adapted sentences with phrase structure violations from Neville, Nicol, Barss, Forster, and Garrett (1991) as our control comparison items. In particular, Weber-Fox and Neville (1996) reported that L2 speakers, even with delayed exposure to English (with age above 11), were able to show a LAN-P600 effect to phrase structure violation sentences ("The scientists criticized a proof of the theorem" vs. "*The scientists criticized Max's of proof the theorem"). We slightly revised the "standard" sentences by replacing "a proof" with "Max's proof" ("The scientists criticized Max's proof of the theorem"), in order to ensure that not all the proper names occurred in an ungrammatical context. We adapted 30 sentences from Neville et al. (1991), and wrote another 30 sentences of the same structure in order to create 60 pairs of sentences for a controlled comparison.

Two experiment lists were constructed such that no sentence context or critical verb was repeated within the same list. Each list consisted of 60 subcategorization sentences (30 Grammatical and 30 Ungrammatical), 60 filler sentences and 60 phrase structure sentences (30 Standard and 30 Violation). The presentation order was randomized within each list. Participants were randomly assigned to one of the two lists.

4.2.3 Procedure

Participants sat in front of a computer screen and put their hands on a keyboard. Sentences were presented one word at a time in a black font on a white background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 600 ms. Each word appeared on the screen for 400 ms, with a 200 ms inter-stimulus interval, for a stimulus-onset asynchrony (SOA) of 600 ms. At the end of each sentence, participants were asked to judge if the sentence they just read was grammatical or not via button pressing. Prior to the experimental session, participants were presented with six practice trials with feedback to familiarize themselves with the task. Including set-up time, an experimental session lasted around 90 minutes.

After the ERP experiment, L2 speakers were asked to perform an offline paper-pencil task on the same subcategorization violations as a control task to evaluate their ability to recognize the violations when there was no explicit time constraint. During the task, they were asked to correct the grammar of sentences in Subcategorization conditions and Filler conditions that they had read during the ERP experiment. We did not include sentences from Phrase structure conditions, such that the offline task would not last too long. L2 speakers were informed that there was no time limit for them to perform the task. All the L2 speakers finished the task in 40 minutes.

4.2.4 Data acquisition and analysis

The L1 data were collected in the US. EEG was recorded from 29 electrodes placed according to the 10/20 system (FP1, FP2, F7, F3, FZ, F4, F8, FT7, FC3, FCZ, FC4, FT8, T7, C3, CZ, C4, T8, TP7, CP3, CPZ, CP4, TP8, P7, P3, PZ, P4, P8, O1, O2). Each electrode was referenced to the right mastoid online and re-referenced to the average of the left and right mastoids offline. Four additional electrodes (two on the outer canthus of each eye and two on the upper and lower ridge of the left eye) were placed to monitor blinks and horizontal eye movements. The impedance of all the electrodes was kept below 10 kΩ. EEG signals were continuously digitized at 500 Hz, filtered between 0.05 to 100 Hz in US (SynAmps, NeuroScan Incorporated).

The L2 data were collected in Taiwan. EEG was recorded from 30 electrodes placed according to the 10/20 system: the same 29 scalp positions as in the L1 cap, plus OZ. Each electrode was referenced to an average of the left and right mastoids online. Four additional electrodes (two on the outer canthus of each eye and two on the upper and lower ridge of the left eye) were placed to monitor blinks and horizontal eye movements. The impedance of all the electrodes was kept below 10 kΩ. EEG signals were continuously digitized at 500 Hz, filtered between DC to 100 Hz in Taiwan (NuAmps, NeuroScan Incorporated).

The EEG data were processed with EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) in Matlab (MathWorks, Inc.). A linear derivation file was first imported to convert the four monopolar eye-movement monitoring channels to two bipolar channels (VEOG and HEOG). We applied a notch filter at 60 Hz and an Infinite Impulse Response (IIR) filter with the band-pass value set between 0.1 Hz to 30 Hz, 12 dB/oct. Then we extracted epochs of length -100 to 1200 ms, from the onset of the determiner until the end of the noun phrase for the Subcategorization conditions, and from the onset of the preposition to the next word for the control Phrase structure conditions. Baseline correction was applied with the pre-stimulus -100 to 0 ms interval. After baseline correction, artifact rejection was carried out by reviewing the epochs both automatically and manually. At each electrode, a 200-ms window was moved across the data (100 ms before and 1200 ms after the stimulus) in 100-ms increments and any epoch where the peak-to-peak voltage exceeded 70 μV was rejected. We then reviewed the data, and if needed, adjusted the voltage threshold for individual subjects. Epochs contaminated by excessive blinking, body movements, skin potentials, and amplifier saturation were rejected. The overall rejection rates were 80 ± 10% for the L1 group and 88 ± 10% for the L2 group (mean ± SD); participants with greater than 40% trials rejected were excluded from further analysis.

Our hypotheses centered on the N400 and P600 responses at the determiner for the critical Subcategorization conditions and the LAN and P600 responses at the preposition for the Phrase structure conditions. We selected six electrodes over the frontal area (F3, FZ, F4, FC3, FCZ, FC4), six electrodes over the parietal area (CP3, CPZ, CP4, P3, PZ, P4) and averaged them as our clustered region of interest (ROI) for both Subcategorization and Phrase structure comparisons. For Subcategorization conditions, we carried out two repeated-measure Type III ANOVA on the mean amplitudes in the measurement time windows of 300-500 ms and 600-900 ms after the onset of the determiner, evaluating the within-subject factor of Subcategorization (Grammatical, Ungrammatical) and Region (Frontal, Parietal) and between-subject factor Group (L1, L2). For the Phrase structure conditions, we carried out two repeated-measure Type III ANOVAs on the mean amplitudes in the measurement time windows of 300-500 ms and 600-900 ms after the onset of the preposition, evaluating the within-subject factor of Phrase structure (Standard, Violation) and Region (Frontal, Parietal) and between-subject factor Group (L1, L2). When Mauchly's test of Sphericity was violated, Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to adjust the p-values.

4.2.5 Results

4.2.5.1 Behavioral data

The overall accuracy rate of the online grammaticality judgment task was 92% for the L1 group, and 81% for the L2 group, showing that participants were paying attention in the experiment. We further broke down the accuracy rate of the L2 group by conditions, and found that the accuracy rate to the ungrammatical argument structure condition was much lower than the other conditions (Table 6). Although the L2 group was clearly somewhat sensitive to the subcategorization constraint, choosing the "ungrammatical" response much more frequently in the ungrammatical condition than the grammatical condition, they still erroneously chose the "grammatical" response on half of those trials. This cannot be attributed to a generalized bias to judge sentences grammatical, as participants were much more accurate in choosing the "ungrammatical" response in the phrase structure violation condition.

Subcategorization		Phrase structure		Filler	
70.1%		91.1%		82.4%	
Grammatical	Ungrammatical	Standard	Violation	Grammatical	Ungrammatical
90.2%	50.0%	91.6%	90.6%	92.7%	72.0%

Table 6: Accuracy rate of each condition for the L2 group

We had also conducted an offline grammaticality judgment task after the ERP experiment to investigate whether L2 speakers were able to retrieve the argument structure in L2 given unlimited amount of time. The offline responses were coded as "accurate" in the ungrammatical condition if they provided any preposition to repair the intransitive violation sentences, even if they chose the wrong preposition (e.g. listened on the music). However, the accuracy rate to the critical subcategorization conditions was only slightly higher (78.1%) in the offline task than the online task (70.1%).

4.2.5.2 ERP data

4.2.5.2.1 Verb subcategorization violations

Plotted in Figure 20 shows the grand average ERPs to the determiner to the noun to the Subcategorization conditions for the L1 and L2 groups. Visual inspection suggested that there was no N400 difference to the determiner in both groups. As time proceeded, there was a prominent P600 effect in the L1 group, while there was little difference between conditions in the L2 group.

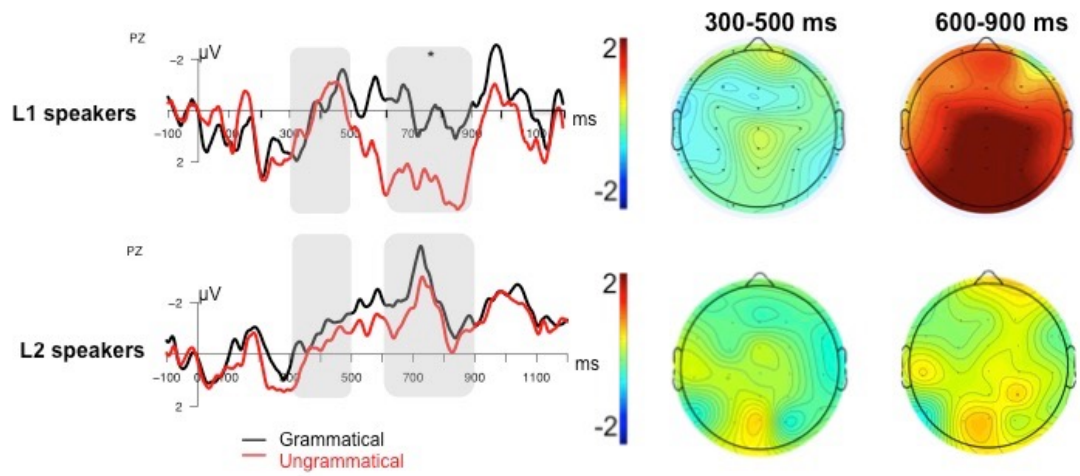

Figure 20: Left: Grand average ERPs of the Subcategorization conditions from the determiner to the nouns at electrode Pz. Right: The topographic distributions in the 300-500 and 600-900 ms intervals at the determiner in L1 and L2 speakers.

During the 300-500 ms time window, we did not obtain any significant group interactions (Subcategorization x Group x Region: F(1,48) = .064, p = .801; Subcategorization x Group: F(1,48) = .446, p = .507), nor did we find a significant main effect (Subcategorization: F =.191, p = .664; Group: F(1,48) = 1.04, p = .313). The statistics confirmed the visual inspection that there was no N400 effect in both groups.

During the 600-900 ms time window, the repeated-measure Type III ANOVA analyses demonstrated a significant Subcategorization by Group interaction (F(1,32)

= 4.346, p < 0.05). Follow-up pairwise analyses revealed that that there was a significant difference between Grammatical and Ungrammatical in the L1 group (t(21) = -2.861, p < 0.05), but not the L2 group (t(27) = -1.095, p = .283). We did not obtain a three-way interaction among Subcategorization x Group x Region (F(1, 48) = 2.261, p = .139), suggesting that the P600 effect in the L1 group was a widespread effect.

4.2.5.2.2 Phrase structure violations

Figure 21 shows the grand average ERPs to the preposition to the next word to the Phrase structure conditions for the L1 and L2 groups. Visual inspection suggested that there was widespread negativity to the preposition during the 300-500 ms time window for both L1 and L2 groups. In the later 600-900 ms time window, only the L1 group showed a P600 deflection. Below we report statistical tests for the two windows.

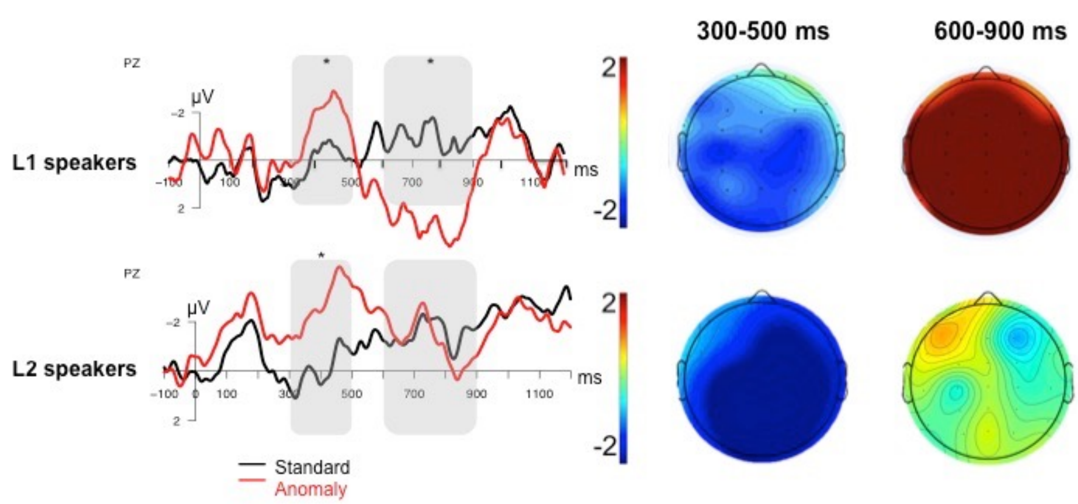

Figure 21: Left: Grand average ERPs of the Phrase structure conditions from the preposition to the following word at electrode Pz. Right: The topographic distributions in the 300-500 and 600-900 ms intervals at the preposition in L1 and L2 speakers.

During 300-500 ms, the repeated-measure Type III ANOVA analyses showed a main effect of Phrase structure ($F = 30.714$, $p < .001$), which was not modulated by Group ($F(1,48) = 2.014$, $p = .162$), nor by Region ($F = 3.388$, $p = .072$). We did not obtain a Phrase structure x Group x Region interaction either ($F(1,48) = 1.172$, $p = .284$). The results confirmed our visual inspection that both group elicited a widespread negativity during this time window.

During 600-900 ms, the repeated-measure Type III ANOVA analyses demonstrated a significant Phrase structure by Group interaction ($F(1,48) = 15.987$, $p < .001$). Follow-up pairwise analyses revealed that that there was a significant difference between Standard and Violation in the L1 group ($t(21) = -4.231$, $p < .001$), but not the L2 group ($t(27) = -.019$, $p = .985$). We did not obtain a three-way interaction among Phrase structure x Group x Region ($F(1,48) = 1.751$, $p = .192$), suggesting that the P600 effect in the L1 group was a widespread effect.

4.3 General Discussion

The current study investigated if L1 subcategorization knowledge impacts verb-argument computation in L2. Although previous studies suggested that L2 speakers could compute verb-argument relations quickly online, we proposed that such computations could be subject to L1 transfer. To test the hypothesis, we selected verbs that were ambitransitive in L1 Mandarin but intransitive in L2 English. We focused on ERP responses to the argument immediately following the verb and predicted that L2 speakers would be insensitive to subcategorization violation at the argument.

4.3.1 Responses to subcategorization violations in L1 and L2

Our ERP results showed that L1 speakers elicited a P600 effect to subcategorization manipulations, but this effect was absent in the L2 group. The L1 response was consistent with the majority of existing studies in L1 subcategorization violations reporting a P600 effect (Osterhout et al., 1994; Friederici & Frisch, 2000; Guo et al., 2009). When the L2 subcategorization was in conflict with the L1, our data revealed that the L2 speakers were not sensitive to the problem online. Specifically, the critical verbs in our experiment were intransitive in English (L2) but could be transitive in Mandarin (L1). The L2 speakers, despite of their high proficiency, appeared to parse it as a transitive verb, such that the ERP responses to the Ungrammatical condition did not differ from the Grammatical baseline.

There are two classes of explanations for the observed L2 insensitivity: non-native knowledge, and/or non-native processing. In particular, it could be that these L2 speakers did not have complete knowledge of L2 verb subcategorization, and thus to compute the verb-argument relations online, they mainly relied on L1 subcategorization knowledge. Alternatively or additionally, it could also be that at least L2 verb subcategorization knowledge existed in these speakers, but during processing, they had difficulties to inhibit L1 subcategorization information in time. Therefore, when the subcategory restrictions of the two languages competed with each other, the effect of L1 would be observed earlier.

To get more insight into which of these explanations was more likely for the current data, we turned to the behavioral responses collected at the end of each trial, where participants were instructed to judge the grammaticality of each sentence. If the absence of ERP sensitivity to the violations were due to lack of grammatical knowledge, we should see insensitivity no matter when or how participants' knowledge was probed. However, if the lack of ERP sensitivity were due to processing deficits, sensitivity might recover with the additional processing time available for making the post-sentence judgment. We found that behavioral responses did show marked sensitivity to the violation: participants said that the Grammatical condition was acceptable 90% of the time, but they said that the Ungrammatical condition was acceptable only 50% of the time. This pattern suggests that both knowledge deficits and processing deficits likely contributed to the absence of the P600 effect. On the one hand, L2 participants were clearly sensitive to the subcategorization violation offline, as their response profile to the Grammatical and Ungrammatical conditions were substantially different—while 90% of Grammatical items were judged grammatical, only 50% of Ungrammatical items were judged ungrammatical. On the other hand, L2 participants often failed to detect the subcategorization violations, falsely accepting those sentences half of the time. Therefore, it could be that they lacked the grammatical knowledge of the subcategorization frame for the verb used in those trials.

As a next step, it would be informative to run a split by performance analysis and see if correct and incorrect rejections of Ungrammatical sentences lead to different ERP responses. If the ERP responses to the correct rejection are different than incorrect one, especially if the correct rejection has a tendency to show a P600 effect, it then would support a more direct contribution of the knowledge account to the ERP results. As long as the L2 has encoded the knowledge of subcategorization correctly, they can make some use of the information online. By contrast, if the ERP responses do not differ between the correct and incorrect rejections, then it might suggest that L1 subcategorization information is difficult to override, even on trials in which speakers have the subcategorization knowledge. Thus this would provide evidence that the processing account is the larger driver of the ERP results. We will run the trial-by-trial analysis to tease apart the relative contribution of the two accounts in the future.

4.3.2 Knowledge and processing accounts

Next we would like to discuss the broader implications of the knowledge account and the processing account. Under the knowledge account, part of the insensitivity to L2 subcategorization online is because such information is not encoded correctly. Once the L2 speaker learns the correct subcategorization information, they can use such information online to some extent. The relatively low accuracy we observed on the ungrammatical subcategorization items in end-of-sentence and end-of-experiment judgments indicates that subcategorization knowledge was certainly far from native-like in the current set of L2 speakers. The question then becomes why the learning of L2 subcategorization is not fully successful. In L2 classrooms, at least those in Taiwan where the L2 data were collected, transitivity of a verb is taught explicitly, and such knowledge is often checked in exams. Therefore, the learning problem here cannot be attributed to lack of negative inputs. We suspect that the difficulty arises because L1 acquisition and L2 learning are different by nature. For example, Ullman (2001) argued that the learning and use of grammar in L1 and L2 tap into different memory systems. In particular, grammatical computations in L1 largely depend upon procedural memory, which is known for learning skills implicitly that involves sequences. By contrast, L2 relies more on declarative memory, a system for learning semantic and episodic knowledge, to study grammatical computations. According to Ullman (2001), the shift to declarative memory for grammar learning can be attributed to late exposure of L2 and/or limited experience to L2. If Ullman (2001) is right, then our finding can have the following pedagogical implications in L2 verb learning: Rather than requesting L2 speakers to memorize the transitivity of a verb, the learning of subcategorization could be more efficient if instructors could increase the amount of L2 inputs, such that L2 speakers might develop the knowledge implicitly, which is subserved by procedure memory.

Because the current dissertation is focused on argument structure processing, I would like to take more time to discuss exactly what kind of L2 processing differences could have been responsible for the absence of ERP sensitivity on those violation trials in which participants did have accurate grammatical knowledge. First, let's consider the steps that L1 speakers go through to process verb-argument relations. When reading the sentence "My sister is listening ____", L1 English speakers can quickly identify the verb "listen" as an intransitive verb, and access its subcategorization information, which is encoded in the verb. At this point, all the open syntactic dependencies in the sentence are completed. Therefore, one possibility is that when the subcategorization violation is encountered at the noun phrase ("the music"), the processing problem that the comprehender is faced with is the fact that there is no attachment site for the noun phrase. They attempt to reanalyze/replay the sentence to see whether they made an error (e.g., was there any alternative subcategorization frame in the verb's lexical entry?) and this reanalysis process generates the P600. An alternative possibility is that L1 speakers take the absence of a period after the intransitive verb as a cue to do further processing: since they know the sentence will continue, they generate an expectation of the most likely (optional) continuation. Although a number of categories are likely (adverb, coordinator, preposition…), the statistics of the experimental items might bias towards the preposition expectation. In this case, the problem in the violation condition is not just that there is no attachment site for the noun phrase, but that the determiner violates the *prediction* for a preposition. L1 English speakers thus could experience difficulties in resolving the conflict of an unexpected lexical category, which will be discussed further in the next section (Section 4.3.3), and the effort to reanalyze the structure of the sentence and/or understand why their expectation was violated would generate the P600.

Now we can turn to the question of why and how for our L2 speakers, L1 Mandarin knowledge could interfere with processing of these violations. One natural place to focus on is the initial processing step for the L1 speaker, the access of subcategorization frame information from the lexical entry of the verb. It is now well established that lexical information in L1 is activated even when the task is exclusively in L2. In a famous study, Thierry and Wu (2007) showed that when reading or listening in L2, L2 speakers automatically and unconsciously translated L2 words into L1. Similarly, according to the Revised Hierarchical Model (the RHM, Kroll & Stewart, 1994), translating from L2 to L1 is much easier because an L2 word is more strongly associated to an L1 equivalent than the other way around. In addition, the RHM proposed that the L2 might not have the privilege to access the concept/meaning of an L2 word directly, because the link between an L2 word and its concept is much weaker than the L1 counterpart. According to this view, the L2 thus often has to be mediated by an L1 translation equivalent in order to access the meaning.

Although the RHM is not uncontroversial (see Dufour & Kroll, 1995; Brysbaert & Duyck, 2010), and we are not committed to it here, it does provide a useful framework for understanding the interference we observe in the current study, as illustrated in Figure 22. According to this account, when encountering the L2 ("listen"), L2 speakers immediately (but unconsciously) translated it into L1 (聽). All the lexical information from the L1 entry thus became activated, including its subcategorization information, and may have erroneously been incorporated into the current parse. For example, instead of accessing only an intransitive frame at 'listen' and concluding that all dependencies had been fulfilled, like the L1 English speaker, the L2 speakers may have distributed syntactic predictions across both the transitive and intransitive possibilities, such that a noun phrase object was predictively projected with some probability. Therefore, when a subsequent noun phrase was presented ("the music"), it would then be slotted into this object position, predicted based on the L1 subcategorization frame, leading to the null ERP effect in the L2 speakers in the current study.

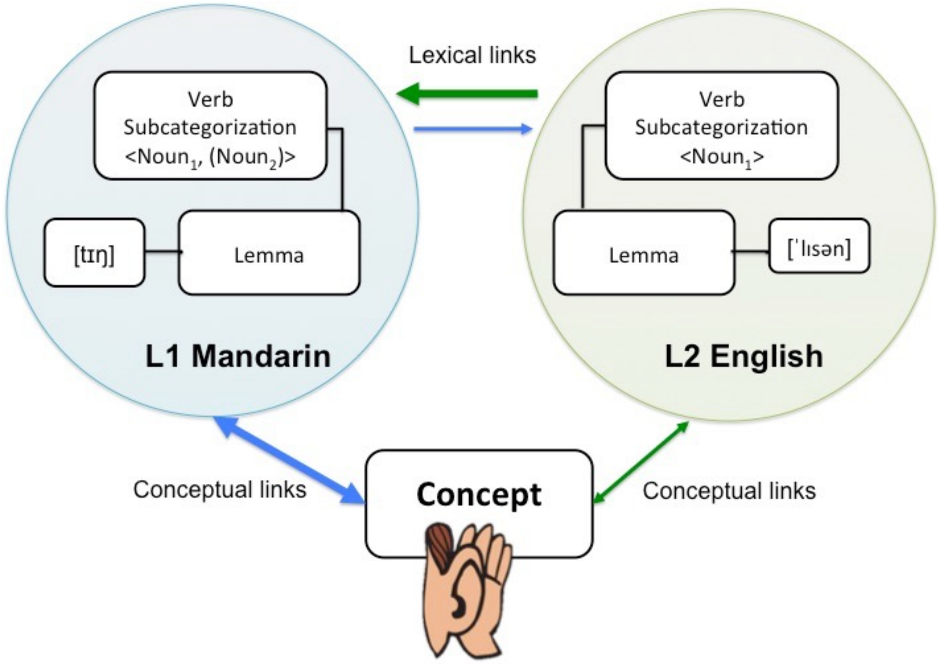

Figure 22: A schematic diagram of online processing upon reading a verb. The figure is adapted from the Revised Hierarchical Model (Kroll & Stewart, 1994).

One question for an account like this one is about what late-stage mechanism allows L2 comprehenders to sometimes detect the violation in the offline behavioral results. Unlike the ERP data, which showed a null effect to subcategorization violation, the behavioral results revealed that L2 speakers were able to reject sentences of subcategorization violations around half of the time. One possibility is that during online processing, lexical association between L2 and L1 words is so strong that L2 speakers are not able to override information from L1. However, post-verbally and through the end of the sentence, where more information becomes available and the cognitive load on L2 speakers is not as heavy, they might be able to revisit the sentence context with more attention to L2 lexical properties / more successful inhibition of L1 lexicon, allowing them to recognize the subcategorization violations.

Assuming this account, one prediction that we make is that the effect should go ‘both ways’, across Mandarin and English L1/L2. Here our ERP data showed that Mandarin speakers accepted transitive uses of “listen” in L2 English, by hypothesis because of interference from accessing L1 Mandarin lexical syntax. We therefore also predict that English speakers who learn Mandarin as their L2 may show temporary processing difficulty with the transitive use of “listen” in Mandarin, due to L1 transfer from English. Specifically, this would predict that they might show a P600 effect when processing the grammatical “listen the music” in Mandarin, similar to what we found for true subcategorization violations of English in this chapter. This would presumably contrast with L1 Mandarin speakers, for whom none of the sentences involve subcategorization violations, predicting no differences between “listen the music” and a control sentence like “record the music” in Mandarin.

A secondary question for the L2 processing account is whether L2 speakers would find the actual English sentence ("listen to the music") unacceptable initially, as there is no PP subcategorization frame available in the L1 lexical representation for ‘listen’ in Mandarin. This is an empirical question, and we do not have data points in this chapter to address this question. However, if the translation from L2 to L1 is automatic and unconscious, as suggested by Thierry and Wu (2007) and Kroll & Stewart (1994), we can predict that L2 speakers might indeed find it unexpected when “listen” is immediately followed by “to.” At the point when the preposition is presented, L1 subcategorization information is still more activated in L2 speakers’ mind and it is not overridden yet. It is possible that at the preposition, L2 speakers are reminded that English “listen” is an intransitive verb, and it requires a preposition to introduce a noun phrase into the sentence context.

Tanner, McLaughlin, Herschensohn, and Osterhout (2013) argued that L2 speakers progressed through different stages of learning: At an early stage, L2 speakers tended to focus more on lexical semantics during sentence processing. As their proficiency improved, they were able to compute more complicated grammatical rules online. If we assume this model, one possibility is that the L2 participants in the current study were still at the stage where influences from L1 were difficult to override. According to this model, however, if their proficiency improved, they would be able to access the L2 subcategorizations more quickly, and perform more native-like computations. However, note that it has been debated to what degree L2 ultimate attainment can be native-like. Clahsen and Felser (2006) took a more pessimistic stand. They proposed that syntactic representations constructed by L2 speakers were shallower and contained less structural details. Interestingly, although Clahsen and Felser (2006) argued that it was challenging for L2 speakers to construct hierarchical details and more abstract elements of sentence structures, they believed that L2 speakers could compute verb-argument relations with ease. Our finding suggested that verb-argument relation might not be as straightforward as the authors thought. It involved complex lexical syntax and thus was prone to L1 transfer when processing in real time.

Our findings of a null effect in L2 subcategorization violation seemed to conflict with Guo et al. (2009), who reported an N400 effect in their L2 group. One important difference between these studies is that Guo et al. (2009) did not manipulation L1-L2 verb subcategorization discrepancies. If many of the L2 items used in their study shared exactly the same subcategorization restrictions between L1 and L2, then this overlap could have allowed the immediate detection of subcategorization violations in L2. In addition, as had been reviewed in the Introduction section, the N400 effect obtained in Guo et al. (2009) did not have a typical distribution. An N400 effect was usually more prominent over central-parietal sites, but in Guo et al. (2009), the effect was very lateralized and it was in fact not significant in the midline. The authors attributed the atypical N400 distribution to a sentence comprehension task imposed in their experiments. However, many other studies that adopted a sentence comprehension task were able to observe a canonical N400 effect in L2 (Foucart, Martin, Moreno, & Costa 2014; Foucart et al., 2015). Based on the "N400 effect" they found, the authors concluded that L2 speakers used a semantic strategy to process sentences online. In the current study, we had taken these concerns into considerations and would like to propose a different account than Guo et al. (2009). We argued that rather than adopting a non-native like semantic strategy in processing subcategorization violations, the challenge that L2 speakers faced was to override interferences from L1 in parsing.

It is also worth noting that while a number of previous L1 studies of subcategorization violations showed a biphasic N400-P600 response (Friederici & Frisch, 2000; Kielar et al., 2012; Osterhout et al., 1994), we only found a P600 effect in our L1 control. Among these studies, Kielar et al.'s study (2012) is the most similar to our own. We both looked at the ERP responses at an unlicensed noun phrase (a determiner and a noun) following an intransitive verb. However, by taking a closer look at Kielar et al., (2012), we noticed there were some problems with their analyses, which might undermine their argument of observing an N400-P600 effect at the determiner and the noun. To begin with, instead of running a long epoch analysis that covered ERP responses from the determiner to the noun, Kielar et al. (2012) did two short-epoch analyses, one for the determiner and the other one for the noun. Such analyses might be problematic. In particular, any effect observed at the subsequent epoch (i.e. the noun) could be resulted from cancelling out a late effect from the previous epoch (i.e. the determiner). In addition, Kielar et al. (2012) did not use the same time window to evaluate the N400 effect at the two epochs (300-400 ms for the determiner and 300-500 ms for the noun). Therefore, it is not clear to us whether an N400 effect should be observed in L1 for violations of intransitivity in English.

Lastly, we suggested what drove the effect in the chapter was the mismatch of subcategorization restrictions between L1 and L2, but one might question if conceptual differences between languages could contribute to the effect, at least partially. Do the differences in subcategorization frames between the English verbs and their Mandarin translations correlate with systematic differences in how speakers of those different languages perceive the denoted events, such that they have different concepts? Although this seems like a reasonable possibility in some cases, in this chapter, the only difference between languages is the presence or absence of a preposition, most of which seem to be of the relatively arbitrary kind that vary a great deal from language to language. For this reason it is our belief that the event concepts for the common verbs used in the current study are unlikely to have been systematically different as a function of preposition use for the speakers of these two languages.

4.3.3 An ELAN response to subcategorization violations in L1 (?)

In addition to the P600 difference between our L1 and L2 groups, as we examined the data closely, we noticed that the Ungrammatical condition elicited an early negativity peaked around 150 ms relative to the grammatical condition only in the L1 group. The early negativity looked more prominent over the left frontal sites (See Figure 23). We suspected that this effect could be an ELAN. We conducted post-hoc tests to evaluate this possibility by comparing the mean amplitude between 100-200 ms after the onset of the determiner with the same ROIs defined for the N400 and P600 time windows. A repeated-measure Type III ANOVA analyses was carried out, evaluating the with-subject factors of Grammaticality (Grammatical, Ungrammatical) and ROI (Frontal, Posterior), and between-subject factor of Group (L1, L2). Results showed a significant Grammaticality x ROI x Group interaction. Follow-up paired t-tests revealed that only the frontal region of the L1 group showed a significant effect (Frontal: $t(21) = 3.184$, $p < .01$; Posterior: $t(21) = 1.537$, $p = .139$). The L2 group did not have a significant interaction between Grammaticality and ROI ($F(1,27) = .316$, $p = .579$). Statistic results confirmed our visual inspection that subcategorization violations lead to an ELAN apparent effect in the L1, but not the L2 group.

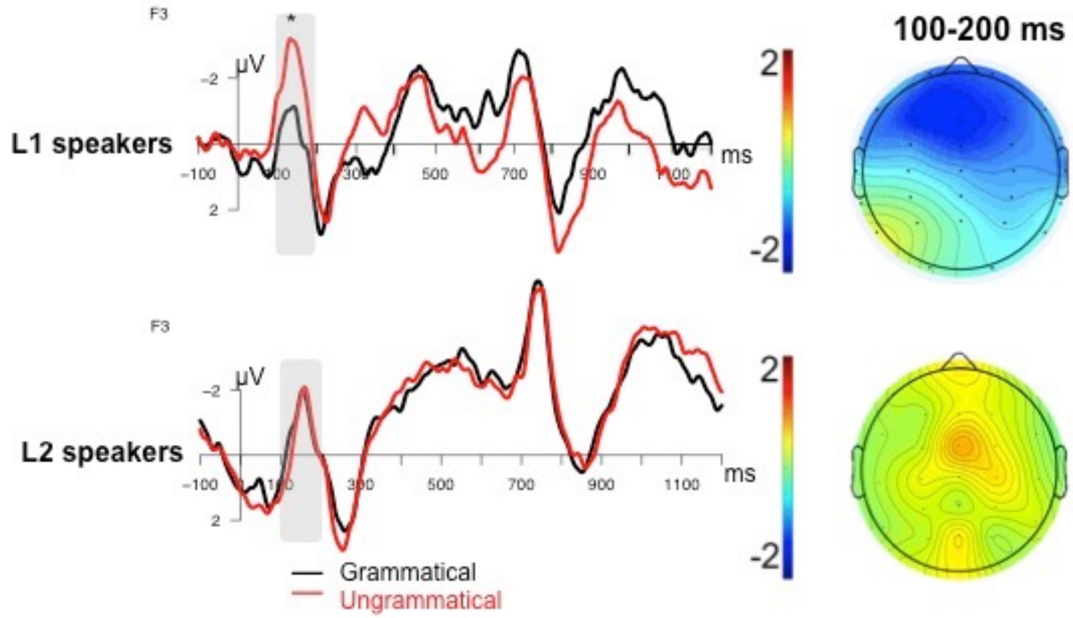

Figure 23: Left: Grand average ERPs of the Phrase structure conditions from the preposition to the following word at electrode F3. Right: The topographic distributions in the 100-200 ms interval at the preposition in L1 and L2 speakers.

If the ELAN effect in the L1 group is real, what could it reflect when L1 speakers processed sentences involved subcategorization violations? Here we would like to argue that it might reflect a prediction mismatch of parts of speech (Friederici, 2000); this would show that L1 speakers actively built up sentence structures with subcategorization provided in the verb. In our experiment setup, critical ungrammatical sentences always contained an intransitive verb, which was immediately followed by a determiner and a noun (e.g. "listened the music"). When L1 speakers reached the intransitive verb, they knew that it could not be the end of a sentence, as there was no period after the verb. They might thus expect an upcoming word to be a preposition, since we did include filler sentences that had a preposition immediately follow an intransitive verb (e.g. "sneezed during the concert"). However, as the sentences proceeded, what was presented to L1 speakers was a determiner, rather than a preposition. The presence of a determiner was in conflict with the structural representation L1 speakers constructed based on the subcategorization information encoded in the verb. Therefore, an ELAN effect might be expected.

The ELAN effect in the L1 group, even though had never been reported from prior empirical studies, appears to be predicted by the Extended Argument Dependency Model (eADM, Bornkessel & Schlesewsky, 2006). The eADM is a language comprehension model, which aimed to account for cross-linguistic unity and diversity in the processing of verb-argument relations. According to the eADM, subcategorization information is immediately used for structure building, and then the verb-argument relations could be computed afterwards. Our ELAN-P600 effect in the L1 group showed that when the subcategorization constraint was violated, comprehenders experienced difficulties in initial structure building and conflict repairing at a later stage. Such a biphasic response, either ELAN-P600 effect or LAN-P600 alike, was common in studies that involved structure violations (Neville et al., 1991; Kaan & Swaab, 2003, but see Osterhout et al., 2004 for an opposing view).

4.3.4 Responses to phrase structure violations in L1 and L2

We included phrase structure violations as our control comparison. The materials were adapted from Neville et al. (1991), and we expected to replicate the findings in Neville et al. (1991) and Weber-Fox and Neville (1996): A biphasic LAN-P600 effect in both L1 and L2 groups. Our results showed a widespread negativity in the LAN time window (300-500 ms) in both groups but a P600 effect only in the L1 group. The fact that both groups showed the same early response to the phrase structure violations allows us to rule out several possible alternative explanations for the null effect of subcategorization violations in the L2 group; it is not the case that the L2 group were insensitive to all kinds of grammatical violations online, nor that we were unable to record grammatically-sensitive ERP responses from that group.

Beyond the above main point, it is interesting to note several aspects of the phrase structure violation results, which were slightly different from the previous results reported by the Neville group (1991). First, the topographic distribution of the negativity in the 300-500 ms time window for both L1 and L2 groups appeared more prominent over the central-parietal area than left-frontal, as classically associated with the LAN. We revisited the effect reported in the two papers, and found that while the topographic distribution in Neville et al. (1991) was indeed left-frontally oriented, the effect in Weber-Fox and Neville (1996) was widespread, with slightly larger signal over the left hemisphere.

Another difference between our results and the previous findings is that unlike Weber-Fox and Neville (1996), we did not observe a P600 effect in the L2 group. However, in Weber-Fox and Neville (1996), they showed that the size of P600 effect was modulated by age of exposure. For those who were exposed to English by age 10, there was a P600 effect to phrase structure violation. For those who learned English afterwards, such an effect became attenuated. Although all of our L2 participated started to learn English before the age of 10, we still did not obtain a significant effect. We speculate that there may be other differences between the subject populations that contributed to the different results. One potential factor was whether the L2 were immersed in a natural target language setting. Weber-Fox and Neville (1996) recruited L2 participants in the US, who had been stayed in an English environment for at least five years. By contrast, the L2 participants in the current study were recruited in Taiwan, where English was taught through formal education, and chances to use English on a daily basis were more restricted. Although the L2 participants were proficient in English, as they were all above the CEFR B2 level, they may have had less certainty about their knowledge of the grammar, and thus may not have attempted to recover the ungrammatical sentences as those who had more exposure to the language did.

It is also worth noting that the phrase structure violation ("Max's proof <u>of</u> the theorem" vs. "Max's <u>of</u> proof the theorem") in Neville et al. (1991) has been criticized for confounding sentence position and different pre-target word baseline, in a way that has called into question whether the LAN effect might be an artifact (Steinhauer & Drury, 2012). We did conduct a supplementary analysis with a 100 ms post-target word baseline, and found that the patterns of the two groups remained the same. However, in future work, to avoid concerns about the baseline problem, we could adopt a different control comparison that fully crosses critical words and sentence context ("They wanted to leave/*about yesterday"; "She was thinking about/*leave you yesterday").

Despite our ERP responses to phrase structure violations were slightly different than prior studies, we would like to bring in behavioral performances of the L2: the accuracy rate of Phrase structure conditions in the L2 was pretty high (above 90%). Although behavioral performance was an offline measure, it showed that the

L2 did have phrase structure knowledge in English. When compared the behavioral results of the two types of violations in L2, the sharp contrast of accuracy rate between Phrase structure conditions (91.1%) and Subcategorization conditions (70.1%) revealed that subcategorization was indeed more challenging for the L2.

4.4 Conclusion

We hypothesized that native language subcategorization knowledge is particularly difficult for L2 speakers to override in online processing. The results of the ERP experiment reported here supported our hypothesis: the L1 group showed a prominent P600 effect to subcategorization violations, but the L2 group was insensitive to the violation such that they showed a null effect, even as they showed sensitivity to phrase structure violations in our control comparison. Both deficits in L2 knowledge (not having the right information encoded in the lexicon), and deficits in L2 processing (not being able to override L1 subcategorization information online) likely contribute jointly to the insensitivity observed here. We hypothesize that the processing deficit may reflect interference associated with automatic access of conflicting L1 lexical information at the verb, which may take time for L2 speakers to override. Together, our data serve as a reminder that computing verb-argument relations, although a seemingly simple task, in fact requires accessing lexical syntax which may be vulnerable to L1 interference in L2.

Chapter 5: Conclusion

5.1 Summary

As we saw in Chapter 1, understanding how verbs are related to their arguments in real time is critical to building a theory of online language comprehension. In this dissertation I have investigated the incremental processing of verb-argument relations. My goals were to temporally disassociate the subcomputations required for complex verb-argument relations, to map out the temporal stages by which argument relations are computed, from clausehood to thematic roles, and to begin to extend the model to online processing in second language/bilingual populations. As all of my research questions are about the time course of online sentence comprehension, ERP has been my primary methodology of choice throughout.

In Chapter 2, I investigated how quickly complex verb-argument relations can be computed to impact the prediction of a subsequent argument. I took advantage of the substantial differences in verb-argument structure provided by Mandarin, whose compound verbs encode complex event relations, such as resultatives (*Kid bit-broke lip*: the kid bit his lip such that it broke) and coordinates (*Store owner hit-scolded employee*: the store owner hit and scolded an employee). I tested sentences in which the object noun could be predicted on the basis of the preceding compound verb, and used N400 responses to the noun to index successful prediction. Results from my three experiments indicated that predictions afforded by a resultative verb did not impact processing of the subsequent noun at an effective verb-noun SOA of 800 ms, but such effects emerged with a verb-noun SOA of 1200 ms. This contrasted with the case of coordinate verbs, which impacted predictions at the verb at both SOAs. I discussed two broad families of accounts for the dissociation: (1) computing the causal relations expressed by a resultative predicate was more taxing and/or (2) retrieving a candidate that fits the resultative context required longer time. These results presented a first step towards temporally dissociating the fine-grained subcomputations required to parse and interpret verb-argument relations.

In Chapter 3, I examined what the temporal stages are by which argument relations are computed, from argument identification to thematic roles. Chow, Smith, Lau, and Phillips (2016) showed that initial verb prediction was driven by arguments in the same clause as the verb, but not argument roles (the Bag of Arguments hypothesis). Here I focused on mapping the time course of identifying which subset of noun phrases are arguments of the verb. I did this by extending the standard paradigm to include structures containing a clause boundary like *The millionaire thought that the servant fired…,* and evaluating sensitivity to this boundary on the N400 to the verb in two experiments (Experiments 4 and 6) that varied in SOA. A control experiment on role reversal sentences (Experiment 5) was conducted to replicate prior studies. These experiments showed that when the time lapse between the onset of the second noun phrase and of the verb is 600 ms only, no effect of argument identification is observed on the N400 response. However, at an 800 ms lapse, N400 amplitude is modulated by arguments of the verb in the expected direction (a smaller N400 response when the clause boundary results in a context that more strongly predicts the critical noun). As expected from prior literature, I show that at the 800 ms lapse there is still no effect of argument role reversal on the N400. Based on these results, I proposed a model of the time course of argument-verb relation computation. Specifically, at an initial stage, the parser does not differentiate if the noun phrases are arguments of a verb (the “Bag of Words” stage). Then, at the second stage (the “Bag of Arguments” stage), the parser could use the clause boundary as a cue to differentiate argument(s) of the verb, but not their argument roles. Based on prior work (Chow, Lau, Wang, & Phillips, 2018; Momma, Sakai, & Phillips, 2015), it is suggested that only at a later stage (between 1200-1800 ms) does the parser start to consider argument roles in computing argument-verb relations. In sum, this model maps a more detailed time course for verb-argument relation computations in online sentence comprehension.

In Chapter 4, I began to investigate the extent to which L1 argument structure knowledge can interfere with L2 processing. I hypothesized that native language argument structure knowledge is particularly difficult for L2 speakers to override in online processing. I constructed sentences with verbs of mismatched argument structures in L1 Mandarin and L2 English, like “My sister listened the music.” The results of the ERP experiment supported my hypothesis: the L1 group showed a prominent P600 effect to argument structure violations, but the L2 group was insensitive to the violation, such that they showed a null effect, even as they showed sensitivity to phrase structure violations in the control comparison. Both deficits in L2 knowledge (not having the right information encoded in the lexicon), and deficits in L2 processing (being in general slower to access this information and/or to compute verb-argument relations online) likely contribute jointly to the insensitivity observed in this study. These data indicate that computing verb-argument relations requires accessing lexical syntax, which is vulnerable to L1 interference in L2.

5.2 Outstanding questions

Although the studies from Chapters 2 to 4 can stand on their own, here I would like to discuss how their conclusions can be linked up together and to broader work in the processing of verb-argument relations. By taking a bird's-eye view, we can obtain a more comprehensive understanding on the incremental process of verb-argument computations. In the sections below, I will walk through how the findings from each chapter can (or cannot) inform one another. Then, I will discuss the extent to which the conjoined findings are compatible with prior work reviewed in Chapter 1.

5.2.1 Connecting the dots, across chapters

In Chapter 2, I was able to temporally dissociate the fine-grained subcomputations required for resultative and coordinate compound verbs. If the computation of resultative verbs is associated with processing delay, then a follow-up question is whether such timing differences in forming argument relations when different kinds of verbs are encountered, can be mapped onto the processing model I proposed in Chapter 3 for the temporal steps involved in forming argument relations when only the arguments have been encountered. Although none of the current designs were aimed at this kind of question, some of my data could potentially provide some clues. In fact, the materials in Chapter 3 involved a variety of different verbs, including resultative compound verbs, coordinate compound verbs and simple verbs of different word length (as some of the simple verbs contained only one character and some had two), and therefore one could conduct an itemized analysis investigating the response at the verb. However, since the experiments in Chapter 3 were not set up to differentiate different types of verbs, I suspect that our statistical power would be too limited.

The model proposed in Chapter 3 can also be cross-checked to see if the three-stage time course of computations still hold in L2 speakers, as it has been debated whether L2 speakers can construct detailed structure representations online (Clahsen & Felser, 2006). If the L2 parser does not consider structural cues carefully to compute argument-verb relations as time evolves, then the L2 speakers would show a very different processing profile than the L1 group. The three-stage model might be collapsed into two stages, with word association on one end, and full structure representation on the other end when given unlimited amount of time. However, I would rather suspect that the L2 speakers have gone through the same three stages as L1 speakers do when computing argument-verb relations, but the time course is possibly delayed. Prior work has shown that advanced L2 speakers are able to apply island constraints online, which require the parser to be aware of a clause/phrase boundary, such that it knows whether an argument can be extracted or not (Omaki & Schulz, 2011; Perpinan, 2020). For example, it is ungrammatical to form a dependency across an island such as a relative clause (*"[*What* did the reporter meet the politician [$_{\text{Relative Clause}}$who supported ___ at the congress]]"). Although it remains unclear whether an argument role has been committed to the argument when the filler-gap dependency has been constructed, L2 evidence on island constraints at least show that the L2 parser is able to construct complicated structural representations, and it is sensitive to clause boundaries online.

More broadly, in Chapters 2 and 3, I relied on the timing of prediction to study the delayed computations in L1; one might wonder if such a paradigm could be used to study L2 processing. In particular, when L1 computations are slow, such processes are expected to be even slower in L2. Although ecological validity is a concern, as it would not be very natural to present materials with an extremely slow SOA, there are alternatives by which a buffer can still be created while maintaining the SOA. As an example, in Chow et al. (2018), they achieved this kind of design by inserting a temporal adjunct between an argument and the target verb (*(Last week), the cop ba (last week) arrested*). In my view, a systematic evaluation of temporal dynamics in which L1 and L2 differ in sentence comprehension can enhance our understanding of language and cognition more broadly.

5.2.2 Broader connections to prior work on verb-argument relation processing

In this dissertation I used the delay of prediction to infer that some aspects of verb-argument computations were slower than others, but what might lead to the slowdown in the cases I observed? In Chapter 2, I suggested that the slowdown resulted from the computation of representation of a resultative verb (i.e. parsing) and/or the search for an argument that fitted the context (i.e. memory retrieval). However, in Chapter 3, where I proposed the time course for different pieces of argument information to be computed to impact the computation of argument-verb relations, I am more inclined to suggest that this model is a parsing model, and thus aspects of the parse itself could be slower to compute. This view is different from the slow prediction hypothesis proposed by Chow, Momma, Smith, Lau and Phillips (2016), which argues that the computation of argument relations, such as computing millionaire-as-an-agent*,* is not taxing. However, I do not find the current data incompatible with a slow memory retrieval view suggested by Chow, Momma et al. (2016). It seems likely that in constructing the detailed structure representation of a sentence, certain aspects of parsing and memory retrieval are measurably slower than others.

As reviewed in Chapter 1, the model proposed by MacDonald, Pearlmutter, and Seidenberg (1994) emphasizes rapid access of argument structure information. The finding in Chapter 4 was in line with this view, because when argument structure information is violated ("*My sister listened the music"), L1 speakers could notice it very quickly. However, to me it is less clear whether the finding in Chapter 2 fits with MacDonald et al (1994). On one hand, the model of MacDonald et al (1994) does not spell out how they would analyze compound verbs. On the other hand, I do not have evidence to differentiate if the temporal delay associated with resultatives verbs resulted from computation of the verbs and/or searching for an argument that fits the context. If the delay was simply because of the latter, and that the computation of a complex event can be completed instantly, then the finding of Chapter 2 can still be accounted for by MacDonald et al's model (1994).

In sentence comprehension, psycholinguists have debated if language processing is always driven completely by bottom-up input, which involves building up inputs from words to phrases in a hierarchical-organized structure (MacDonald et al., 1994; Lewis & Phillips, 2014). Ferreira, Bailey and Ferraro (2002) take a more pessimistic view, and suggest that "the representation is just "good enough" to provide an interpretation that satisfied comprehenders," and that the representation is not detailed enough to distinguish important details such as who is doing what to whom. From my point of view, results from Chapters 2 and 3 in this dissertation provide a different perspective to approach the debate. To be more precise, sentence comprehension is a dynamic process, and it is really just a matter of time to generate detailed and accurate representations of the linguistic inputs. For example, in the model proposed in Chapter 3, structural information does contribute very little in computing argument-verb relations at an early stage. Thus, I suspect that the intermediate Bag of Arguments stage, where clause information could serve as a cue but not argument role, could still be a "good enough" stage. That is because under "good enough" parsing, structural representations are not computed to the full degree, but certainly it does not mean that structures play no role at all. Comprehenders with probabilistic heuristics could still be sensitive to a clause boundary in sentences like [*Millionaire thought* [*servant fired…*]]. Concerning the argument *servant* alone, *fire* is less likely to be a relevant event, such that its N400 amplitude is not reduced. Presumably, if comprehenders are processing in a hurry or under bad conditions, they might never make it all the way to the argument role stage.

Prior work has suggested that there are distinctive processes between using argument information to predict a verb and using a verb to predict an argument (Friederici & Frisch, 2000; Bornkessel & Schlesewsky, 2006). In particular, when a verb is presented before arguments, the incoming arguments are checked against the argument structure information provided by the verb. By contrast, when arguments are presented before the verb, thematic relations among the arguments should be computed, such that argument relations could be checked against the requirements of a verb. The work in this dissertation implies that the time courses for the two types of computations (argument-verb relations and verb-argument relations) are very different. For argument-verb computations, a model is proposed in Chapter 3, where I suggest that the stages for different argument information to be computed to impact argument-verb computations. In respect to verb-argument computations, a separate model could be suggested based on the findings in Chapters 2 and 4. Chapter 4 shows that argument structure information is immediately accessible (at least in L1 speakers) when a verb is present in the context. Specifically, when the lapse between the onset of a verb and the onset of the next word is 600 ms, the parser is able to detect argument structure violations. Results from Chapter 2 show that the fine-grained subcomputations required for different complex verbs can be temporally disassociated. Specifically, while an 800 ms lapse between the complex verb and the argument is sufficient for the computation of coordinate verbs, it takes up to 1200 ms for resultatives. Future work should try to map out more details along the time scale in both argument-verb and verb-argument computations (see Section 5.3 below).

5.3 Future work

Argument structure is an interface between syntax and semantics, and it provides us with a window to evaluate how syntax and semantics interact during online sentence comprehension. The findings in this dissertation suggest many promising routes for future investigations, and I will discuss some of them in this section.

5.3.1 Temporal dynamics of computing complex events

In Chapter 2 I have temporally disassociated the subcomputations required for resultative and coordinate complex verbs. With the same method, I shall be able to temporally disassociate other types of compound verbs. For example, one common type of compound verbs in Mandarin is subordinate complex verbs (or sometimes called modifier-head complex verbs). Subordinate complex verbs involve a modification relation, with V1 as a modifier and V2 as a head. Take the verb *raw-ate* as an example. While V2 *eating* is the head of the event, the modifier *raw* constrains the features of an upcoming argument. It will be awkward if the patient of the raw-eating event is something that is normally eaten raw, since the modifier is then uninformative (*She raw-ate-asp some #fruits/scallops*). By setting up cloze contrasts and manipulating presentation rates among experiments, we should be able to narrow down the time course of computing modification relations.

Now that we have demonstrated that complex verbs involve different event relations, which might lead to different processing profiles along the time scale, an intriguing follow-up direction would be to investigate if we can temporally disassociate complex verbs that involve the same event relations, but of different predicate types. Mandarin resultatives might provide a window for this question. As introduced in Section 2.1.2 in Chapter 2, there are at least three types of resultative verbs in Mandarin: transitive resultatives (*bit-broke*, whose V1 is transitive and V2 predicates the object), unergative resultatives (*washed-tired*, whose V1 is still transitive and V2 predicates the subject), and ergative (*upset-cried*, whose V1 and V2 are both intransitive but V1 predicate the subject and V2 predicate the object). With the same method as in Chapter 2, we can keep cloze contrast constant among different types of resultatives, and see if we can obtain an N400 reduction to predictable object defined through offline cloze norming. Such an inquiry could map out the complex argument roles computations along the time scale.

5.3.2 Types of argument information that feed into parsing along the temporal scale

The processing model proposed in Chapter 3 is based on empirical data with two pre-verbal arguments. Our data show that it is not until the third stage that the parser starts to consider thematic relations of the two arguments. A good follow-up question is how adding or dropping the number of pre-verbal arguments would affect the temporal scale of our model. We can start to evaluate the question from Mandarin, the language of the data that motivate this model. In Mandarin, subject relative clauses have only one argument coupled with a structural cue (either *ba* or *bei*) before the verb. The surface structure is identical to the *ba* constructions examined in Chapter 3, with the first argument being moved downwards. Consider the following sentences:

(1) 富翁 把 僕人 <u>解僱了</u>(Canonical *Ba* construction)

Millionaire ba servant <u>fired-ASP</u>

"The millionaire fired the servant."

(2) 把 僕人 解僱 的 富翁 很 小氣 (Subject relative clause, Canonical)

Ba servant fired de millionaire very stingy

"The millionaire that fired the servant was very stingy."

(3) 把 富翁 解僱 的 僕人 很 小氣 (Subject relative clause, Role reversal)

Ba servant fired de millionaire very stingy

"The servant that fired the millionaire was very stingy."

If reducing the number of pre-verbal arguments facilitates the computation, we would expect the parser to compute the argument-verb relations earlier, and obtain an N400 effect earlier on the time scale. In fact, prior studies have shown that with two pre-verbal arguments in Mandarin, the parser becomes sensitive to thematic relations by 1800 ms (Chow et al., 2018). By contrast, when there is only one pre-verbal argument, although the data is in Japanese, the timing reported is reduced to 1200 ms (Momma et al., 2015). However, Mandarin and Japanese differ in many ways, such that it is not very convincing to make a direct comparison. The Mandarin subject relative clause serves to fill in the gap. No matter what, we predict that the exact timing can be shifted along the scale, but the parser is getting more and more sensitive to structure information as time evolves. This model should be applicable to any language that permits a verb final sentence structure.

In the current study, both of the pre-verbal arguments were animate entities. There is much prior research investigating the role of argument animacy in sentence processing. In particular, prior studies suggested that the linear order of arguments is closely related to animacy hierarchy (Oh, Sum, & Sim, 2016; Paczynski & Kuperberg, 2011). That is, in production there is a tendency to organize entities along an animacy continuum: Human ranks the highest, followed by animals, plants, object entities and then abstract thoughts; an argument of a higher animacy ranking tends to occur at a structurally higher position (Paczynski & Kuperberg, 2011). Corpus studies showed that subject nouns are much more likely to be animate, although the proportions vary across languages (Øvrelid, 2004; Snider & Zaenen, 2006; Minkoff, 2000). However, measures of comprehension difficulty from EEG do not yield a consistent pattern of processing cost for sentences that deviate from these tendencies. Experiments in some languages have shown a larger N400 to an inanimate argument in sentence initial position relative to an animate one (Korean: Oh, et al., 2016; English: Weckerly & Kutas, 1999; Bourguignon, Drury, Valois, Steinhauer, 2012), while experiments in other languages do not (Mandarin: Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008; Li, Zhao, Zheng, & Yang, 2014; German: Bornkessel & Schlesewsky, 2006; Turkish: Dimeral, Schlesewsky, Bornkessel-Schlesewsky, 2008). Either way, our model makes the prediction that interactions between animacy and sentence position should not be observed until our "second stage," where the parser starts to be aware of a clause boundary.

Previous studies, along with the results in Chapter 3, show that morpho-syntactic cues to argument role are relatively slower to impact verb predictions. For example, there is no N400 effect to role reversal situations marked by case marker at a normal presentation rate (*Bee-Nominative* sting vs. **Bee-Accusative* sting). One possibility is that this reflects a strict limitation on the speed of argument role computations. However, could prediction on the basis of argument roles be facilitated by a different kind of cue to argument role? That is, rather than marking the argument role with a (morpho)syntactic cue, could a semantic cue like a volitional adverb facilitate the processes? Kuperberg, Kreher, Sitnikova, Caplan, and Holcomb (2007) have reported an absence of N400 effect to role reversal situations in English ("For breakfast, the boy /*egg must eat"), and we can adapt their sentences to evaluate if a semantic cue allows the computation to be faster ("For breakfast, the egg intentionally/occasionally ate"). In our setup, only volitional adverbs (e.g. intentionally) imply an agent in the context; frequency adverbs (e.g. occasionally) do not have such a connotation. If a semantic cue can rapidly facilitate argument role-verb prediction, it would indicate that the processing speed limitation is not on argument role computation per se, but the process of combining contextual cues to determine the role in the first place. If this is the case, it then raises broader questions about why some cues are more efficient than others.

5.3.3 Computations of verb-argument relations in different populations

The scope of my research can be extended by studying the verb-argument computations in different populations, such as L2 speakers and heritage speakers. In this way we can gain better insights not only how the grammars of the two languages interact with each other, but also how the language system interacts with other cognitive abilities.

In Chapter 4, we show that when argument structures do not match between L1 and L2, verb-argument computation in L2 is likely subject to L1 transfer. Still, it remains unclear how the L2 argument structure is represented. It is possible that the L2 speakers do not have the right information encoded in the lexicon at all. Another possibility for accounting for the L1 transfer effect online is that the L2 speakers are in general slower to compute verb-argument relations. If so, a follow-up question will be what contributes to the delay. Accessing less-activated L2 argument structure information? Inhibiting the influence from L1? These are open questions that deserve further explorations.

Heritage languages, despite often being considered not acquired in full, in fact do show systematic rules. For example, Polinsky (2018) suggests that some areas of heritage grammars are relatively more stable across languages, such as the distinction between nouns and verbs, whereas other domains are more vulnerable, such as word order and morphology. To our knowledge, it remains less clear if heritage speakers possess the "correct" argument structure knowledge of either language, not to mention to what degree argument structures of their two languages interfere during online sentence processing. Admittedly, the bi-directional influence between heritage and dominant language in heritage speakers might not be the same as the L1-L2 interaction in L2 speakers. However, similarities and discrepancies may be a starting point to narrow in on the big questions: How different aspects of language are mentally represented and how they interact with general cognitive abilities.

Appendix (Experiment materials)

1. Experiment 1 (in Chapter 2)

	Condition	Subject	Verb	Target	Post-target continuation				
1	Resultative	丈夫	翻亂了	抽屜	還是	找不到	印章		
1	Causative	丈夫	弄亂了	抽屜	還是	找不到	印章		
1	Simple	丈夫	翻過了	抽屜	還是	找不到	印章		
2	Resultative	士兵	刺死了	敵人	是	為了	自保		
2	Causative	士兵	弄死了	敵人	是	為了	自保		
2	Simple	士兵	刺過了	敵人	才能	脫身			
3	Resultative	女工	染黑了	頭髮	，	看起來	年輕	許多	
3	Causative	女工	弄黑了	頭髮	，	看起來	年輕	許多	
3	Simple	女工	染過了	頭髮	，	看起來	年輕	許多	
4	Resultative	女主人	拌涼了	沙拉	，	等待	客人	到來	
4	Causative	女主人	弄涼了	沙拉	，	等待	客人	到來	
4	Simple	女主人	拌過了	沙拉	，	等待	客人	到來	
5	Resultative	女兒	畫正了	方形	，	贏得	老師	讚賞	
5	Causative	女兒	弄正了	方形	，	贏得	老師	讚賞	
5	Simple	女兒	畫過了	方形	，	贏得	老師	讚賞	
6	Resultative	女朋友	織寬了	毛衣	，	有點	懊惱		
6	Causative	女朋友	弄寬了	毛衣	以	遮掩	身材		
6	Simple	女朋友	織過了	毛衣	還	想	織	圍巾	
7	Resultative	小狗	舔濕了	臉頰	，	他	用	手帕	擦掉
7	Causative	小狗	弄濕了	臉頰	他	用	手帕	擦掉	
7	Simple	小狗	舔過了	臉頰	，	他	用	手帕	擦掉
8	Resultative	小孩	咬破了	嘴唇	，	痛得	哇哇叫		
8	Causative	小孩	弄破了	嘴唇	，	痛得	哇哇叫		
8	Simple	小孩	咬過了	嘴唇	，	若有所思	的	樣子	
9	Resultative	小學生	踢倒了	椅子	被	老師	處罰		
9	Causative	小學生	弄倒了	椅子	被	老師	處罰		
9	Simple	小學生	踢過了	椅子	又	踢	桌子		
10	Resultative	工人	填平了	坑洞	用路	品質	改善	很多	
10	Causative	工人	弄平了	坑洞	用路	品質	改善	很多	
10	Simple	工人	填過了	坑洞	改善了	用路	品質		
11	Resultative	少女	剪斜了	瀏海	，	換了	新	造型	
11	Causative	少女	弄斜了	瀏海	，	換了	新	造型	
11	Simple	少女	剪過了	瀏海	，	換了	新	造型	
12	Resultative	少年	吹響了	號角	，	典禮	開始了		
12	Causative	少年	弄響了	號角	，	典禮	開始了		
12	Simple	少年	吹過了	號角	，	典禮	開始了		
13	Resultative	木匠	釘歪了	釘子	，	結構	就	不穩	

13	Causative	木匠	弄歪了	釘子	，	結構	就	不穩	
13	Simple	木匠	釘過了	釘子	才	接著	確認	設計圖	
14	Resultative	歹徒	刺傷了	警察	才	從	窗戶	逃逸	
14	Causative	歹徒	弄傷了	警察	才	從	窗戶	逃逸	
14	Simple	歹徒	刺過了	警察	才	從	窗戶	逃逸	
15	Resultative	主人	打跑了	小偷	所以	財務上	沒有	損失	
15	Causative	主人	弄跑了	小偷	所以	財務上	沒有	損失	
15	Simple	主人	鋪過了	床鋪	讓	客人	好好	休息	
16	Resultative	主人	鋪厚了	床鋪	讓	客人	好好	休息	
16	Causative	主人	弄厚了	床鋪	讓	客人	好好	休息	
16	Simple	主持人	揉過了	麵團	想要	做	麵包		
17	Resultative	主持人	揉軟了	麵團	想要	做	麵包		
17	Causative	主人	弄跑了	小偷	所以	財務上	沒有	損失	
17	Simple	外公	寫過了	毛筆	想	休息	一下		
18	Resultative	外公	寫壞了	毛筆	，	心裡	很	懊惱	
18	Causative	主持人	弄軟了	麵團	想要	做	麵包		
18	Simple	外婆	剁過了	豬肉	才能	煮	肉燥		
19	Resultative	外婆	剁碎了	豬肉	煮成	肉燥			
19	Causative	外公	弄壞了	毛筆	，	心裡	很	懊惱	
19	Simple	奶奶	戴過了	帽子	又	把	帽子	放	回去
20	Resultative	奶奶	戴反了	帽子	卻	一直	沒	發現	
20	Causative	外婆	弄碎了	豬肉	煮成	肉燥			
20	Simple	奶奶	熬過了	稀飯	，	準備	開飯		
21	Resultative	奶奶	熬稠了	稀飯	比較	有	飽足感		
21	Causative	奶奶	弄反了	帽子	，	一直	戴	不好	
21	Simple	市長	想過了	計畫	卻	還是	束手無策		
22	Resultative	市長	想定了	計畫	要	完成	環保	工程	
22	Causative	奶奶	弄稠了	稀飯	，	鍋子	卻	差點	燒焦
22	Simple	母親	擰過了	毛巾	可是	毛巾	還是	濕濕的	
23	Resultative	母親	擰乾了	毛巾	晾在	窗台上			
23	Causative	市長	弄定了	計畫	要	完成	環保	工程	
23	Simple	申請者	填過了	表單	才	繳交	申請費		
24	Resultative	申請者	填明了	表單	，	等待	通知		
24	Causative	母親	弄乾了	毛巾	然後	晾在	窗台上		
24	Simple	同事	踩過了	地板	所以	要	記得	拖地	
25	Resultative	同事	踩髒了	地板	，	可是	卻	不	處理
25	Causative	申請者	弄明了	表單	才	繳交	申請費		
25	Simple	名醫	治過了	疾病	卻	治不好	她	的	傷心
26	Resultative	名醫	治癒了	疾病	，	他	心中	非常	感激
26	Causative	同事	弄髒了	地板	，	可是	卻	不	處理
26	Simple	老闆	抹過了	奶油	卻	忘了	加	果醬	

27	Resultative	老闆	抹勻了	奶油	卻	忘了	加	果醬		
27	Causative	名醫	弄癒了	疾病	，	他	心中	非常	感激	
27	Simple	冷水	澆過了	大火	，	火勢	已	獲得	控制	
28	Resultative	冷水	澆滅了	大火	還好	沒有	什麼	災情		
28	Causative	老闆	弄勻了	奶油	卻	忘了	加	果醬		
28	Simple	助教	收過了	作業	就	不再	收了			
29	Resultative	助教	收齊了	作業	拿回去	批改				
29	Causative	冷水	弄滅了	大火	還好	沒有	什麼	災情		
29	Simple	助理	裁過了	紙張	接著	發給	大家	使用		
30	Resultative	助理	裁齊了	紙張	接著	發給	大家	使用		
30	Causative	助教	弄齊了	作業	拿回去	批改				
30	Simple	弟弟	磨過了	鞋底	才	進	家門			
31	Resultative	弟弟	磨薄了	鞋底	不知道	要	做	什麼		
31	Causative	助理	弄齊了	紙張	接著	發給	大家	使用		
31	Simple	弟弟	磨過了	鞋底	才	進	家門			
32	Resultative	村民	打昏了	強盜	再	把	他	送到	警局	
32	Causative	村民	弄昏了	強盜	再	把	他	送到	警局	
32	Simple	村民	打過了	強盜	再	送到	警局			
33	Resultative	村長	扭傷了	腳踝	好險	沒有	大礙			
33	Causative	村長	弄傷了	腳踝	好險	沒有	大礙			
33	Simple	村長	扭過了	腳踝	，	舊傷	一直	沒有	好	
34	Resultative	汽車	撞歪了	電線桿	才	成功	停下來			
34	Causative	汽車	弄歪了	電線桿	才	成功	停下來			
34	Simple	汽車	撞過了	電線桿	才	停下來				
35	Resultative	男子	砍倒了	樹木	拿來	當	柴火			
35	Causative	男子	弄倒了	樹木	拿來	當	柴火			
35	Simple	男子	砍過了	樹木	拿來	當	柴火			
36	Resultative	男主人	倒滿了	酒	一口氣	灌下去				
36	Causative	男主人	弄滿了	酒	一口氣	灌下去				
36	Simple	男主人	倒過了	酒	，	一口氣	灌下去			
37	Resultative	男孩	摔瘸了	右腿	只能	跛腳	走路			
37	Causative	男孩	弄瘸了	右腿	只能	跛腳	走路			
37	Simple	男孩	穿過了	鞋子	一雙	又	一雙	，	很	
38	Resultative	男孩	穿舊了	鞋子	，	想	存錢	買	新的	
38	Causative	兒子	弄睡了	妹妹	因為	他	說了	很多	故事	
38	Simple	男孩	摔過了	右腿	只能	跛腳	走路			
39	Resultative	兒子	哄睡了	妹妹	因為	他	說了	很多	故事	
39	Causative	妻子	弄鹹了	菜	只好	加	水	稀釋	味道	
39	Simple	男孩	穿過了	鞋子	才	決定	購買			
40	Resultative	妻子	炒鹹了	菜	只好	加水	稀釋	味道		
40	Causative	姊姊	弄熱了	雙手	用來	取暖				

40	Simple	兒子	哄過了	妹妹	，	她	就	不	哭了
41	Resultative	姊姊	搓熱了	雙手	用來	取暖			
41	Causative	姑姑	弄厚了	牛排	，	可能	不容易	料理	
41	Simple	妻子	炒過了	菜	，	要	我	試試看	味道
42	Resultative	姑姑	切厚了	牛排	，	可能	不容易	料理	
42	Causative	店員	弄涼了	飲料	才	端給	客人		
42	Simple	姊姊	搓過了	雙手	用來	取暖			
43	Resultative	店員	冰涼了	飲料	才	端給	客人		
43	Causative	朋友	弄響了	吉他	，	聲音	傳得	很遠	
43	Simple	姑姑	切過了	牛排	，	令人	食指大動		
44	Resultative	朋友	彈響了	吉他	，	聲音	傳得	很遠	
44	Causative	法官	弄輕了	罪刑	原告	決定	上訴		
44	Simple	店員	冰過了	飲料	才	端給	客人		
45	Resultative	法官	判輕了	罪刑	，	原告	仍然	決定	上訴
45	Causative	牧羊人	弄散了	狼群	以	守護	羊群		
45	Simple	朋友	彈過了	吉他	，	聲音	傳得	很遠	
46	Resultative	牧羊人	驅散了	狼群	來	守護	羊群		
46	Causative	表哥	弄臭了	襪子	只好	自己	洗		
46	Simple	法官	判過了	罪刑	，	原告	仍然	決定	上訴
47	Resultative	表哥	泡臭了	襪子	感覺	很	噁心		
47	Causative	阿姨	弄紅了	螃蟹	所以	賣相	很好		
47	Simple	牧羊人	驅過了	狼群	，	守護	羊群		
48	Resultative	阿姨	蒸紅了	螃蟹	，	難怪	香氣	逼人	
48	Causative	孩子	弄開了	房門	發現	裡面	藏著	驚喜	
48	Simple	表哥	泡過了	襪子	結果	褪色了			
49	Resultative	孩子	推開了	房門	發現	裡面	藏著	驚喜	
49	Causative	客人	弄髒了	桌巾	所以	要求	換	新的	
49	Simple	阿姨	蒸過了	螃蟹	難怪	香氣	逼人		
50	Resultative	客人	翻髒	桌巾	所以	要求	換	新的	
50	Causative	室友	弄熟了	雞蛋	然後	分給	我們	吃	
50	Simple	孩子	推過了	房門	可是	推不動			
51	Resultative	室友	煮熟了	雞蛋	然後	分給	我們	吃	
51	Causative	政府	弄寬了	法規	可是	民眾	反應	兩極	
51	Simple	客人	翻過了	桌巾	，	看起來	似乎	很	挑惕
52	Resultative	政府	修寬了	法規	可是	民眾	反應	兩極	
52	Causative	施工單位	弄通了	隧道	預計	明年	通車		
52	Simple	室友	煮過了	雞蛋	然後	分給	我們	吃	
53	Resultative	施工單位	鑿通了	隧道	預計	明年	通車		
53	Causative	洪水	弄垮了	堤防	引發	災情			
53	Simple	政府	修過了	法規	可是	民眾	反應	兩極	
54	Resultative	洪水	沖垮了	堤防	引發	災情			

54	Causative	炸彈	弄飛了	小鳥	，	戰爭	彷彿	迫在眉睫	
54	Simple	施工單位	鑿過了	隧道	預計	明年	通車		
55	Resultative	炸彈	嚇飛了	小鳥	，	戰爭	開始了		
55	Causative	風	弄散了	落葉	，	勾起	鄉愁		
55	Simple	洪水	沖過了	堤防	需要	好好	整修		
56	Resultative	風	吹散了	落葉	，	引起	鄉愁		
56	Causative	哥哥	弄扁了	蟑螂	所以	引來	一陣	尖叫	
56	Simple	炸彈	嚇過了	小鳥	，	人們	也	嚇壞了	
57	Resultative	哥哥	踩扁了	蟑螂	所以	引來	一陣	尖叫	
57	Causative	孫女	弄粗了	眉毛	決定	重畫	一遍		
57	Simple	風	吹過了	落葉	，	勾起	鄉愁		
58	Resultative	孫女	描粗了	眉毛	決定	重畫	一遍		
58	Causative	師傅	弄白了	牆壁	，	美化了	空間		
58	Simple	哥哥	踩過了	蟑螂	因此	引來	一陣	尖叫	
59	Resultative	師傅	刷白了	牆壁	，	美化了	空間		
59	Causative	校長	弄定了	接班人	可是	背後	傳言	很多	
59	Simple	孫女	描過了	眉毛	讓	妝容	更	完整	
60	Resultative	校長	選定了	接班人	可是	背後	傳言	很多	
60	Causative	秘書	弄甜了	茶	，	招待	來訪	的	客戶
60	Simple	師傅	刷過了	牆壁	，	美化了	空間		
61	Resultative	秘書	泡甜了	茶	難怪	被	董事長	嫌棄	
61	Causative	高溫	弄黑了	鍋子	，	變得	很	難	清洗
61	Simple	校長	選過了	接班人	可是	沒有	中意	人選	
62	Resultative	高溫	燒黑了	鍋子	，	變得	很	難	清洗
62	Causative	高溫	弄黑了	鍋子	，	變得	很	難	清洗
62	Simple	秘書	泡過了	茶	，	招待	來訪	的	客戶
63	Resultative	張奶奶	醃辣了	泡菜	非常	開胃			
63	Causative	張奶奶	弄辣了	泡菜	非常	開胃			
63	Simple	張奶奶	醃過了	泡菜	想	接著	醃	蘿蔔	
64	Resultative	強盜	摔破了	花瓶	而且	摔得	粉碎		
64	Causative	強盜	弄破了	花瓶	，	花瓶	馬上	變成	碎片
64	Simple	強盜	摔過了	花瓶	花瓶	馬上	變成	碎片	
65	Resultative	設計師	挖空了	牆壁	以	加強	採光		
65	Causative	設計師	弄空了	牆壁	想	節省	經費		
65	Simple	設計師	挖過了	牆壁	，	想	營造	不同	感覺
66	Resultative	貨車	撞翻了	攤販	，	現場	一片	混亂	
66	Causative	貨車	弄翻了	攤販	，	現場	一片	混亂	
66	Simple	貨車	撞過了	攤販	再	撞上	護欄		
67	Resultative	隊員	擊敗了	對手	，	贏得	勝利		
67	Causative	隊員	弄敗了	對手	，	贏得	勝利		
67	Simple	隊員	擊過了	對手	，	贏得	勝利		

68	Resultative	媽媽	燙平了	襯衫	，	明天	面試	可以	穿
68	Causative	媽媽	弄平了	襯衫	，	明天	面試	可以	穿
68	Simple	媽媽	燙過了	襯衫	，	明天	面試	可以	穿
69	Resultative	嫂嫂	炸糊了	年糕	導致	口感	變差了		
69	Causative	嫂嫂	弄糊了	年糕	導致	口感	變差了		
69	Simple	嫂嫂	炸過了	年糕	導致	口感	變好了		
70	Resultative	新郎	穿皺了	西裝	，	很	尷尬		
70	Causative	新郎	弄皺了	西裝	，	很	尷尬		
70	Simple	新郎	穿過了	西裝	卻	覺得	不	合身	
71	Resultative	爺爺	磨利了	菜刀	，	使用	起來	更	上手
71	Causative	爺爺	弄利了	菜刀	，	使用	起來	更	上手
71	Simple	爺爺	磨過了	菜刀	，	使用	起來	更	上手
72	Resultative	經理	灌醉了	秘書	，	想	一親芳澤		
72	Causative	經理	弄醉了	秘書	，	想	一親芳澤		
72	Simple	經理	灌過了	秘書	可是	秘書	沒醉		
73	Resultative	舅舅	折斷了	免洗筷	只好	丟到	垃圾桶		
73	Causative	舅舅	弄斷了	免洗筷	只好	丟到	垃圾桶		
73	Simple	舅舅	折過了	免洗筷	然後	開始	大快朵頤		
74	Resultative	農夫	捲高了	袖子	繼續	工作			
74	Causative	農夫	弄高了	圍牆	防止	作物	遭竊		
74	Simple	農夫	砌過了	圍牆	防止	作物	遭竊		
75	Resultative	農夫	砌高了	圍牆	防止	作物	遭竊		
75	Causative	農夫	弄高了	袖子	繼續	工作			
75	Simple	農夫	捲過了	袖子	繼續	工作			
76	Resultative	遊客	摸亮了	銅像	還	打卡	拍照		
76	Causative	遊客	弄亮了	銅像	還	打卡	拍照		
76	Simple	遊客	摸過了	銅像	還	打卡	拍照		
77	Resultative	雷聲	嚇醒了	寶寶	，	害	寶寶	大哭	
77	Causative	雷聲	弄醒了	寶寶	，	害	他	大哭	
77	Simple	雷聲	嚇過了	寶寶	，	害	寶寶	大哭	
78	Resultative	僕人	折疊了	棉被	因此	房間	變得	很	整齊
78	Causative	僕人	弄疊了	棉被	，	房間	變	整齊	
78	Simple	僕人	折過了	棉被	因此	房間	變	整齊	
79	Resultative	演員	撕破了	劇本	拒絕	演出			
79	Causative	演員	弄破了	劇本	實在	是	無心之過		
79	Simple	演員	磕過了	額頭	讓	導演	補	鏡頭	
80	Resultative	演員	磕腫了	額頭	，	非常	敬業		
80	Causative	演員	弄腫了	額頭	好像	是	被	蜜蜂	叮
80	Simple	演員	撕過了	劇本	還	和	導演	吵架	
81	Resultative	廠商	蓋矮了	房子	只好	向	客戶	道歉	
81	Causative	廠商	弄矮了	房子	以	配合	設計師	的	要求

81	Simple	廠商	蓋過了	房子	才 蓋 其他 設施
82	Resultative	敵軍	炸翻了	戰艦	， 已經 取得 先機
82	Causative	敵軍	弄翻了	戰艦	， 已經 取得 先機
82	Simple	敵軍	炸過了	戰艦	， 已經 取得 先機
83	Resultative	窯爐	燒紅了	磚頭	小心 不要 靠近
83	Causative	窯爐	弄紅了	磚頭	小心 不要 靠近
83	Simple	窯爐	燒過了	磚頭	， 小心 不要 靠近
84	Resultative	學生	擦亮了	皮鞋	為了 參加 畢業典禮
84	Causative	學生	弄難了	題目	所以 表現 不如 預期
84	Simple	學生	擦過了	皮鞋	為了 參加 畢業典禮
85	Resultative	學生	想難了	題目	所以 表現 不如 預期
85	Causative	學生	弄亮了	皮鞋	為了 參加 畢業典禮
85	Simple	學生	想過了	題目	可是 無法 解題
86	Resultative	學徒	拌勻了	水泥	以 抹到 牆上
86	Causative	學徒	弄勻了	水泥	以 抹到 牆上
86	Simple	學徒	拌過了	水泥	以 抹到 牆上
87	Resultative	選手	練粗了	手臂	想 為 國家 爭光
87	Causative	選手	弄粗了	手臂	想 嚇唬 對手
87	Simple	選手	練過了	手臂	想 為 國家 爭光
88	Resultative	嬸嬸	貼反了	春聯	因此 被 鄰居 取笑
88	Causative	嬸嬸	弄反了	春聯	被 鄰居 取笑
88	Simple	嬸嬸	貼過了	春聯	很 有 過年 氣氛
89	Resultative	藝術家	拉直了	線條	做成 美麗 的 藝術品
89	Causative	藝術家	弄直了	線條	做成 美麗的 藝術品
89	Simple	藝術家	拉過了	線條	做成 美麗的 藝術品
90	Resultative	藝術家	剪斷了	棉線	準備 用來 創作
90	Causative	藝術家	弄斷了	棉線	準備 用來 創作
90	Simple	藝術家	剪過了	棉線	準備 用來 創作

2. Experiment 2 (in Chapter 2)

	Condition	Subject	Verb	Object	Post-target continuation
Resultative set					
1	Resultative	丈夫	翻亂了	抽屜	還是 找不到 印章
1	R-Simple	丈夫	翻過了	抽屜	還是 找不到 印章
2	Resultative	士兵	刺死了	敵人	是 為了 自保
2	R-Simple	士兵	刺過了	敵人	才能 脫身
3	Resultative	女工	染黑了	頭髮	， 看起來 年輕 許多
3	R-Simple	女工	染過了	頭髮	， 看起來 年輕 許多
4	Resultative	女主人	拌涼了	沙拉	等待 客人 到來
4	R-Simple	女主人	拌過了	沙拉	等待 客人 到來
5	Resultative	女朋友	織寬了	毛衣	有點 懊惱

5	R-Simple	女朋友	織過了	毛衣	還	想	織	圍巾	
6	Resultative	小狗	舔濕了	臉頰	，	他	用	手帕	擦掉
6	R-Simple	小狗	舔過了	臉頰	，	他	用	手帕	擦掉
7	Resultative	小孩	咬破了	嘴唇	，	痛得	哇哇叫		
7	R-Simple	小孩	咬過了	嘴唇	若有所思	的	樣子		
8	Resultative	小學生	踢倒了	椅子	被	老師	處罰		
8	R-Simple	小學生	踢過了	椅子	又	踢	桌子		
9	Resultative	工人	填平了	坑洞	，	用路	品質	改善	很多
9	R-Simple	工人	填過了	坑洞	，	改善了	用路	品質	
10	Resultative	少女	剪斜了	瀏海	，	換了	新造型		
10	R-Simple	少女	剪過了	瀏海	，	換了	新	造型	
11	Resultative	少年	吹響了	號角	，	典禮	開始了		
11	R-Simple	少年	吹過了	號角	，	典禮	開始了		
12	Resultative	木匠	釘歪了	釘子	，	結構	就	不穩	
12	R-Simple	木匠	釘過了	釘子	才	確認	設計圖		
13	Resultative	歹徒	刺傷了	警察	才	從	窗戶	逃逸	
13	R-Simple	歹徒	刺過了	警察	才	從	窗戶	逃逸	
14	Resultative	主人	鋪厚了	床鋪	，	讓	客人	好好	休息
14	R-Simple	主人	鋪過了	床鋪	，	讓	客人	好好	休息
15	Resultative	主持人	揉軟了	麵團	想要	做	麵包		
15	R-Simple	主持人	揉過了	麵團	想要	做	麵包		
16	Resultative	外婆	剁碎了	豬肉	煮成	肉燥			
16	R-Simple	外婆	剁過了	豬肉	才能	煮	肉臊		
17	Resultative	奶奶	熬稠了	稀飯	，	比較	有	飽足感	
17	R-Simple	奶奶	熬過了	稀飯	準備	開飯			
18	Resultative	奶奶	戴反了	帽子	一直	沒	發現		
18	R-Simple	奶奶	戴過了	帽子	又	把	帽子	放回去	
19	Resultative	母親	擰乾了	毛巾	晾在	窗台上			
19	R-Simple	母親	擰過了	毛巾	可是	毛巾	還是	濕濕的	
20	Resultative	同事	踩髒了	地板	，	自己	卻	不處理	
20	R-Simple	同事	踩過了	地板	所以	要	記得	拖地	
21	Resultative	名醫	治癒了	疾病	，	他	心中	非常	感激
21	R-Simple	名醫	治過了	疾病	卻	治不好	她的	傷心	
22	Resultative	老闆	抹勻了	奶油	卻	忘了	加	果醬	
22	R-Simple	老闆	抹過了	奶油	卻	忘了	加	果醬	
23	Resultative	弟弟	磨薄了	鞋底	因為	走	太多	路	
23	R-Simple	弟弟	磨過了	鞋底	才	進	家門		
24	Resultative	村長	扭傷了	腳踝	好險	沒有	大礙		
24	R-Simple	村長	扭過了	腳踝	，	舊傷	一直	沒有	好
25	Resultative	汽車	撞歪了	電線桿	才	成功	停下來		
25	R-Simple	汽車	撞過了	電線桿	才	停下來			

26	Resultative	男子	砍倒了	樹木	拿來	當	柴火		
26	R-Simple	男子	砍過了	樹木	拿來	當	柴火		
27	Resultative	男孩	穿舊了	鞋子	想	存錢	買	新的	
27	R-Simple	男孩	穿過了	鞋子	一雙	又	一雙	才	離開
28	Resultative	姊姊	搓熱了	雙手	用來	取暖			
28	R-Simple	姊姊	搓過了	雙手	用來	取暖			
29	Resultative	姑姑	切厚了	牛排	可能	不容易	料理		
29	R-Simple	姑姑	切過了	牛排	令人	食指	大動		
30	Resultative	店員	冰涼了	飲料	才	端給	客人		
30	R-Simple	店員	冰過了	飲料	才	端給	客人		
31	Resultative	朋友	彈響了	吉他	，	聲音	傳得	很	遠
31	R-Simple	朋友	彈過了	吉他	，	聲音	傳得	很	遠
32	Resultative	法官	判輕了	罪刑	，	原告	決定	上訴	
32	R-Simple	法官	判過了	罪刑	，	原告	仍然	決定	上訴
33	Resultative	表哥	泡臭了	襪子	，	感覺	很	噁心	
33	R-Simple	表哥	泡過了	襪子	結果	褪色了			
34	Resultative	阿姨	蒸紅了	螃蟹	難怪	香氣	逼人		
34	R-Simple	阿姨	蒸過了	螃蟹	難怪	香氣	逼人		
35	Resultative	客人	翻髒了	桌巾	，	要求	換	新的	
35	R-Simple	客人	翻過了	桌巾	，	要求	換	新的	
36	Resultative	室友	煮熟了	雞蛋	分給	我們	吃		
36	R-Simple	室友	煮過了	雞蛋	分給	我們	吃		
37	Resultative	施工單位	鑿通了	隧道	，	預計	明年	通車	
37	R-Simple	施工單位	鑿過了	隧道	預計	明年	通車		
38	Resultative	炸彈	嚇飛了	小鳥	，	戰爭	開始了		
38	R-Simple	炸彈	嚇過了	小鳥	，	人們	也	嚇壞了	
39	Resultative	風	吹散了	落葉	勾起	鄉愁			
39	R-Simple	風	吹過了	落葉	勾起	鄉愁			
40	Resultative	哥哥	踩扁了	蟑螂	，	引來	一陣	尖叫	
40	R-Simple	哥哥	踩過了	蟑螂	，	引來	一陣	尖叫	
41	Resultative	孫女	描粗了	眉毛	決定	重畫	一遍		
41	R-Simple	孫女	描過了	眉毛	讓	妝容	更	完整	
42	Resultative	秘書	泡甜了	茶	被	董事長	嫌棄		
42	R-Simple	秘書	泡過了	茶	招待	來訪	的	客戶	
43	Resultative	張奶奶	醃辣了	泡菜	，	非常	開胃		
43	R-Simple	張奶奶	醃過了	泡菜	想	接著	醃	蘿蔔	
44	Resultative	設計師	挖空了	牆壁	以	加強	採光		
44	R-Simple	設計師	挖過了	牆壁	想	營造	不同	感覺	
45	Resultative	貨車	撞翻了	攤販	，	現場	一片	混亂	
45	R-Simple	貨車	撞過了	攤販	再	撞上	護欄		
46	Resultative	媽媽	燙平了	襯衫	明天	面試	可以	穿	

46	R-Simple	媽媽	燙過了	襯衫	明天	面試	可以	穿		
47	Resultative	爺爺	磨利了	菜刀	，	使用	起來	更	上手	
47	R-Simple	爺爺	磨過了	菜刀	，	使用	起來	更	上手	
48	Resultative	舅舅	折斷了	免洗筷	只好	丟到	垃圾桶			
48	R-Simple	舅舅	折過了	免洗筷	只好	丟到	垃圾桶			
49	Resultative	農夫	捲高了	袖子	繼續	工作				
49	R-Simple	農夫	捲過了	袖子	繼續	工作				
50	Resultative	農夫	砌高了	圍牆	防止	作物	遭竊			
50	R-Simple	農夫	砌過了	圍牆	防止	作物	遭竊			
51	Resultative	遊客	摸亮了	銅像	還	打卡	拍照			
51	R-Simple	遊客	摸過了	銅像	還	打卡	拍照			
52	Resultative	僕人	折疊了	棉被	，	房間	變得	很	整齊	
52	R-Simple	僕人	折過了	棉被	，	房間	因此	變	整齊	
53	Resultative	廠商	蓋矮了	房子	只好	向	客戶	道歉		
53	R-Simple	廠商	蓋過了	房子	才	蓋	其他	設施		
54	Resultative	敵軍	炸翻了	戰艦	，	已經	取得	先機		
54	R-Simple	敵軍	炸過了	戰艦	，	已經	取得	先機		
55	Resultative	窯爐	燒紅了	磚頭	，	小心	不要	靠近		
55	R-Simple	窯爐	燒過了	磚頭	，	小心	不要	靠近		
56	Resultative	學生	擦亮了	皮鞋	為了	參加	畢業典禮			
56	R-Simple	學生	擦過了	皮鞋	為了	參加	畢業典禮			
57	Resultative	學徒	拌勻了	水泥	以	抹到	牆上			
57	R-Simple	學徒	拌過了	水泥	以	抹到	牆上			
58	Resultative	嬸嬸	貼反了	春聯	被	鄰居	取笑			
58	R-Simple	嬸嬸	貼過了	春聯	，	很	有	過年	氣氛	
59	Resultative	藝術家	剪斷了	棉線	準備	用來	創作			
59	R-Simple	藝術家	剪過了	棉線	準備	用來	創作			
60	Resultative	藝術家	拉直了	線條	做成	美麗	的	藝術品		
60	R-Simple	藝術家	拉過了	線條	做成	美麗	的	藝術品		
Coordinate set										
1	Coordinate	大學生	選購了	電腦	帶回	宿舍				
1	C-Simple	大學生	選過了	電腦	帶回	實驗室				
2	Coordinate	女子	刷洗了	廁所	才	上床	休息			
2	C-Simple	女子	刷過了	廁所	才	上床	休息			
3	Coordinate	女孩	拆用了	化妝品	並且	拍	影片	介紹		
3	C-Simple	女孩	拆過了	化妝品	試用	，	沒有	馬上	購買	
4	Coordinate	女學生	塗改了	答案	才	交卷				
4	C-Simple	女學生	塗過了	答案	卻	沒	寫上	新的		
5	Coordinate	少年	拆解了	炸彈	，	被	譽為	天才	少年	
5	C-Simple	少年	拆過了	炸彈	，	被	譽為	天才	少年	
6	Coordinate	王妃	攔截了	信件	救了	王爺	一命			

6	C-Simple	王妃	攔過了	信件	救了	王爺	一命		
7	Coordinate	主唱	錄製了	唱片	不過	仍	需要	後製	
7	C-Simple	主唱	錄過了	唱片	不過	仍	需要	後製	
8	Coordinate	主編	刪改了	文章	讓	排版	更	美觀	
8	C-Simple	主編	刪過了	文章	因為	稿件	太多了		
9	Coordinate	市政府	核發了	執照	商家	才	開始	營業	
9	C-Simple	市政府	核過了	執照	商家	才	開始	營業	
10	Coordinate	犯人	追打了	被害人	，	路人	發現	趕緊	報案
10	C-Simple	犯人	追過了	被害人	，	犯案	動機	可疑	
11	Coordinate	同學們	傳看了	漫畫	結果	被	老師	沒收	
11	C-Simple	同學們	傳過了	漫畫	結果	被	老師	沒收	
12	Coordinate	名媛	退換了	包包	，	指定	要	最新款	
12	C-Simple	名媛	退過了	包包	，	指定	要	最新款	
13	Coordinate	收銀員	換算了	匯率	公告	在	網站上		
13	C-Simple	收銀員	換過了	匯率	公告	在	網站上		
14	Coordinate	老奶奶	借用了	雨傘	才	沒有	淋濕		
14	C-Simple	老奶奶	借過了	雨傘	可是	沒有	歸還		
15	Coordinate	老闆	存放了	現金	在	保險箱	裡面		
15	C-Simple	老闆	存過了	現金	又	存	美金		
16	Coordinate	老闆娘	打罵了	員工	後來	被	檢舉		
16	C-Simple	老闆娘	打過了	員工	後來	被	檢舉		
17	Coordinate	考生	選填了	志願	可是	個個	沒	把握	
17	C-Simple	考生	選過了	志願	可是	個個	沒	把握	
18	Coordinate	助理	遞交了	辭呈	，	等待	長官	批准	
18	C-Simple	助理	遞過了	辭呈	，	等待	長官	批准	
19	Coordinate	抗議人士	編印了	文宣	準備	沿路	發放		
19	C-Simple	抗議人士	編過了	文宣	準備	沿路	發放		
20	Coordinate	李先生	拐賣了	小女孩	很	沒有	良心		
20	C-Simple	李先生	拐過了	小女孩	卻	沒有	成功		
21	Coordinate	男朋友	改訂了	餐廳	作為	求婚	地點		
21	C-Simple	男朋友	改過了	餐廳	因為	前一家	菜色	不好	
22	Coordinate	男童	組裝了	機器人	，	玩得	很	開心	
22	C-Simple	男童	組過了	機器人	又	組好	小汽車		
23	Coordinate	岳母	催迫了	女婿	趕快	與	女兒	和好	
23	C-Simple	岳母	催過了	女婿	想要	抱	孫子		
24	Coordinate	法官	審讀了	判決書	，	被告	不發一語		
24	C-Simple	法官	審過了	判決書	，	被告	不發一語		
25	Coordinate	表妹	彈唱了	一首歌	，	很	動聽		
25	C-Simple	表妹	彈過了	一首歌	，	很	動聽		
26	Coordinate	保姆	沖泡了	牛奶	給	寶寶	喝		
26	C-Simple	保姆	沖過了	牛奶	給	寶寶	喝		

27	Coordinate	前女友	調閱了	監視器	才	掌握	他的	行蹤	
27	C-Simple	前女友	調過了	監視器	才	掌握	他的	行蹤	
28	Coordinate	政府	修護了	古蹟	才	開放	參觀		
28	C-Simple	政府	修過了	古蹟	才	開放	參觀		
29	Coordinate	科學家	訂造了	儀器	以	進行	最新的	實驗	
29	C-Simple	科學家	訂過了	儀器	以	進行	最新的	實驗	
30	Coordinate	美容師	按壓了	穴道	讓	我	覺得	很	放鬆
30	C-Simple	美容師	按過了	穴道	讓	我	覺得	很	放鬆
31	Coordinate	英文系	徵聘了	教授	可是	沒有	人	應徵	
31	C-Simple	英文系	徵過了	教授	可是	沒有	人	應徵	
32	Coordinate	軍方	埋設了	地雷	等待	敵軍	經過		
32	C-Simple	軍方	埋過了	地雷	等待	敵軍	經過		
33	Coordinate	軍隊	攻取了	城池	，	贏得	勝利		
33	C-Simple	軍隊	攻過了	城池	卻	無法	攻下來		
34	Coordinate	家庭主婦	挑取了	青菜	準備	做	晚餐		
34	C-Simple	家庭主婦	挑過了	青菜	準備	做	晚餐		
35	Coordinate	旅客	看管了	行李箱	輪流	去	使用	洗手間	
35	C-Simple	旅客	看過了	行李箱	覺得	不需要	購買	新的	
36	Coordinate	留學生	搬運了	行李	終於	抵達	新家		
36	C-Simple	留學生	搬過了	行李	又	搬	傢俱		
37	Coordinate	粉絲	想慕了	偶像	天天	做	白日夢		
37	C-Simple	粉絲	想過了	偶像	很	擔心	於是	送上	補品
38	Coordinate	逃難者	埋藏了	財物	避免	被	強盜	洗劫而空	
38	C-Simple	逃難者	埋過了	財物	避免	被	強盜	洗劫而空	
39	Coordinate	參選人	分送了	面紙	，	想	拉攏	選民	
39	C-Simple	參選人	分過了	面紙	，	想	拉攏	選民	
40	Coordinate	婆婆	縫補了	衣服	給	公公	穿		
40	C-Simple	婆婆	縫過了	衣服	給	公公	穿		
41	Coordinate	將軍	聽信了	讒言	決定	不再	效忠	朝廷	
41	C-Simple	將軍	聽過了	讒言	仍然	決定	效忠	朝廷	
42	Coordinate	教授	核銷了	經費	才	發現	經費	短缺	
42	C-Simple	教授	核過了	經費	才	發現	經費	短缺	
43	Coordinate	船員	捕殺了	鯨魚	，	民間團體	因此	出面	抗議
43	C-Simple	船員	捕過了	鯨魚	，	民間團體	因此	出面	抗議
44	Coordinate	陳小姐	編譯了	新書	交給	出版社	出版		
44	C-Simple	陳小姐	編過了	新書	交給	出版社	出版		
45	Coordinate	園丁	噴灑了	農藥	才	下班			
45	C-Simple	園丁	噴過了	農藥	才	下班			
46	Coordinate	嫌犯	改裝了	車子	以	逃避	警方	追緝	
46	C-Simple	嫌犯	改過了	車子	以	逃避	警方	追緝	
47	Coordinate	業者	停用了	網路	，	店面	也	出售了	

47	C-Simple	業者	停過了	網路	，	店面	也	出售了	
48	Coordinate	準新娘	訂製了	禮服	明天	要	試穿		
48	C-Simple	準新娘	訂過了	禮服	又	訂了	喜餅		
49	Coordinate	管理員	點收了	管理費	才	給	我	收據	
49	C-Simple	管理員	點過了	管理費	才	給	我	收據	
50	Coordinate	舞者	搶購了	舞鞋	導致	舞鞋	缺貨		
50	C-Simple	舞者	搶過了	舞鞋	可惜	特製版	只有	一雙	
51	Coordinate	鄰居	鏟除了	雜草	所以	院子	煥然一新		
51	C-Simple	鄰居	鏟過了	雜草	所以	院子	煥然一新		
52	Coordinate	學弟	敲打了	鍵盤	，	沈迷	線上	遊戲	
52	C-Simple	學弟	敲過了	鍵盤	想	找	女孩	聊天	
53	Coordinate	學長	愛慕了	學妹	卻	不敢	表白		
53	C-Simple	學長	愛過了	學妹	卻	不敢	表白		
54	Coordinate	機器	壓製了	模型	才能	量產			
54	C-Simple	機器	壓過了	模型	才能	量產			
55	Coordinate	縣長	減省了	預算	，	工程	仍然	能	進行
55	C-Simple	縣長	減過了	預算	，	工程	仍然	能	進行
56	Coordinate	檢察官	簽發了	搜索票	，	限期	逮捕	嫌犯	
56	C-Simple	檢察官	簽過了	搜索票	，	限期	逮捕	嫌犯	
57	Coordinate	獵人	射殺了	鹿	又	捕到	野兔		
57	C-Simple	獵人	射過了	鹿	又	射到	野兔		
58	Coordinate	醫師	拔除了	智齒	只	花	十分鐘		
58	C-Simple	醫師	拔過了	智齒	才	體會到	病人	的	感覺
59	Coordinate	警方	查扣了	毒品	並	調查	毒品	流向	
59	C-Simple	警方	查過了	毒品	沒有	發現	其他	證據	
60	Coordinate	護理師	哄騙了	病人	還	唱歌	給	病人	聽
60	C-Simple	護理師	哄過了	病人	還	唱歌	給	病人	聽

3. Experiment 3 (in Chapter 2)

	Condition	Subject	Verb	Modifier	Target	Post-target continuation			
Resultative set									
1	Resultative	丈夫	翻亂了	我的	抽屜	還是	找不到	印章	
1	R-Simple	丈夫	翻過了	我的	抽屜	還是	找不到	印章	
2	Resultative	士兵	刺死了	很多	敵人	是	為了	自保	
2	R-Simple	士兵	刺過了	很多	敵人	才能	脫身		
3	Resultative	女工	染黑了	她的	頭髮	，	看起來	年輕	許多
3	R-Simple	女工	染過了	她的	頭髮	，	看起來	年輕	許多
4	Resultative	女主人	拌涼了	所有的	沙拉	等待	客人	到來	
4	R-Simple	女主人	拌過了	所有的	沙拉	等待	客人	到來	
5	Resultative	女朋友	織寬了	你的	毛衣	有點	懊惱		
5	R-Simple	女朋友	織過了	你的	毛衣	還	想	織	圍巾

6	Resultative	小狗	舔濕了	他的	臉頰	，	他	用	手帕	擦掉
6	R-Simple	小狗	舔過了	他的	臉頰	，	他	用	手帕	擦掉
7	Resultative	小孩	咬破了	他的	嘴唇	，	痛得	哇哇叫		
7	R-Simple	小孩	咬過了	他的	嘴唇	若有所思	的	樣子		
8	Resultative	小學生	踢倒了	他的	椅子	被	老師	處罰		
8	R-Simple	小學生	踢過了	他的	椅子	又	踢	桌子		
9	Resultative	工人	填平了	所有的	坑洞	，	用路	品質	改善	很多
9	R-Simple	工人	填過了	所有的	坑洞	，	改善了	用路	品質	
10	Resultative	少女	剪斜了	她的	瀏海	，	換了	新造型		
10	R-Simple	少女	剪過了	她的	瀏海	，	換了	新	造型	
11	Resultative	少年	吹響了	他的	號角	，	典禮	開始了		
11	R-Simple	少年	吹過了	他的	號角	，	典禮	開始了		
12	Resultative	木匠	釘歪了	不少	釘子	，	結構	就	不穩	
12	R-Simple	木匠	釘過了	不少	釘子	才	確認	設計圖		
13	Resultative	歹徒	刺傷了	所有的	警察	才	從	窗戶	逃逸	
13	R-Simple	歹徒	刺過了	所有的	警察	才	從	窗戶	逃逸	
14	Resultative	主人	鋪厚了	那裡的	床鋪	，	讓	客人	好好	休息
14	R-Simple	主人	鋪過了	那裡的	床鋪	，	讓	客人	好好	休息
15	Resultative	主持人	揉軟了	一些	麵團	想要	做	麵包		
15	R-Simple	主持人	揉過了	一些	麵團	想要	做	麵包		
16	Resultative	外婆	剁碎了	不少	豬肉	煮成	肉燥			
16	R-Simple	外婆	剁過了	不少	豬肉	才能	煮	肉臊		
17	Resultative	奶奶	熬稠了	不少	稀飯	，	比較	有	飽足感	
17	R-Simple	奶奶	熬過了	不少	稀飯	想要	當作	早餐		
18	Resultative	奶奶	戴反了	她的	帽子	一直	沒	發現		
18	R-Simple	奶奶	戴過了	她的	帽子	又	把	帽子	放回去	
19	Resultative	母親	擰乾了	許多	毛巾	晾在	窗台上			
19	R-Simple	母親	擰過了	許多	毛巾	再	把	毛巾	拿去晒	
20	Resultative	同事	踩髒了	這裡的	地板	，	自己	卻	不處理	
20	R-Simple	同事	踩過了	這裡的	地板	所以	要	記得	拖地	
21	Resultative	名醫	治癒了	她的	疾病	，	她	心中	非常	感激
21	R-Simple	名醫	治過了	她的	疾病	卻	治不好	她的	傷心	
22	Resultative	老闆	抹勻了	許多	奶油	卻	忘了	加	果醬	
22	R-Simple	老闆	抹過了	許多	奶油	卻	忘了	加	果醬	
23	Resultative	弟弟	磨薄了	自己的	鞋底	因為	走	太多	路	
23	R-Simple	弟弟	磨過了	自己的	鞋底	才	進	家門		
24	Resultative	村長	扭傷了	自己的	腳踝	好險	沒有	大礙		
24	R-Simple	村長	扭過了	自己的	腳踝	，	舊傷	一直	沒有	好
25	Resultative	汽車	撞歪了	這裡的	電線桿	才	成功	停下來		
25	R-Simple	汽車	撞過了	這裡的	電線桿	才	停下來			
26	Resultative	男子	砍倒了	不少	樹木	拿來	當	柴火		

26	R-Simple	男子	砍過了	不少	樹木	拿來	當	柴火		
27	Resultative	男孩	穿舊了	不少	鞋子	想	存錢	買	新的	
27	R-Simple	男孩	穿過了	不少	鞋子	可是	滿意的	沒幾雙		
28	Resultative	姊姊	搓熱了	她的	雙手	用來	取暖			
28	R-Simple	姊姊	搓過了	她的	雙手	用來	取暖			
29	Resultative	姑姑	切厚了	一些	牛排	可能	不容易	料理		
29	R-Simple	姑姑	切過了	一些	牛排	令人	食指	大動		
30	Resultative	店員	冰涼了	所有的	飲料	才	端給	客人		
30	R-Simple	店員	冰過了	所有的	飲料	才	端給	客人		
31	Resultative	朋友	彈響了	你的	吉他	，	聲音	傳得	很	遠
31	R-Simple	朋友	彈過了	你的	吉他	，	聲音	傳得	很	遠
32	Resultative	法官	判輕了	你的	罪刑	，	原告	決定	上訴	
32	R-Simple	法官	判過了	你的	罪刑	，	原告	仍然	決定	上訴
33	Resultative	表哥	泡臭了	他的	襪子	，	感覺	很	噁心	
33	R-Simple	表哥	泡過了	他的	襪子	結果	褪色了			
34	Resultative	阿姨	蒸紅了	不少	螃蟹	難怪	香氣	逼人		
34	R-Simple	阿姨	蒸過了	不少	螃蟹	難怪	香氣	逼人		
35	Resultative	客人	翻髒了	這裡的	桌巾	，	要求	換	新的	
35	R-Simple	客人	翻過了	這裡的	桌巾	，	要求	換	新的	
36	Resultative	室友	煮熟了	許多	雞蛋	分給	我們	吃		
36	R-Simple	室友	煮過了	許多	雞蛋	分給	我們	吃		
37	Resultative	施工單位	鑿通了	那裡的	隧道	，	預計	明年	通車	
37	R-Simple	施工單位	鑿過了	那裡的	隧道	預計	明年	通車		
38	Resultative	炸彈	嚇飛了	很多	小鳥	，	戰爭	開始了		
38	R-Simple	炸彈	嚇過了	很多	小鳥	，	人們	也	嚇壞了	
39	Resultative	風	吹散了	那裡的	落葉	勾起	鄉愁			
39	R-Simple	風	吹過了	那裡的	落葉	勾起	鄉愁			
40	Resultative	哥哥	踩扁了	那裡的	蟑螂	，	引來	一陣	尖叫	
40	R-Simple	哥哥	踩過了	那裡的	蟑螂	，	引來	一陣	尖叫	
41	Resultative	孫女	描粗了	她的	眉毛	決定	重畫	一遍		
41	R-Simple	孫女	描過了	她的	眉毛	讓	妝容	更	完整	
42	Resultative	秘書	泡甜了	一些	茶	被	董事長	嫌棄		
42	R-Simple	秘書	泡過了	一些	茶	招待	來訪	的	客戶	
43	Resultative	張奶奶	醃辣了	一些	泡菜	，	非常	開胃		
43	R-Simple	張奶奶	醃過了	一些	泡菜	想	接著	醃	蘿蔔	
44	Resultative	設計師	挖空了	這裡的	牆壁	以	加強	採光		
44	R-Simple	設計師	挖過了	這裡的	牆壁	想	營造	不同	感覺	
45	Resultative	貨車	撞翻了	許多	攤販	，	現場	一片	混亂	
45	R-Simple	貨車	撞過了	許多	攤販	再	撞上	護欄		
46	Resultative	媽媽	燙平了	我的	襯衫	明天	面試	可以	穿	
46	R-Simple	媽媽	燙過了	我的	襯衫	明天	面試	可以	穿	

47	Resultative	爺爺	磨利了	他的	菜刀	，	使用	起來	更	上手
47	R-Simple	爺爺	磨過了	他的	菜刀	，	使用	起來	更	上手
48	Resultative	舅舅	折斷了	他的	免洗筷	只好	丟到	垃圾桶		
48	R-Simple	舅舅	折過了	他的	免洗筷	只好	丟到	垃圾桶		
49	Resultative	農夫	捲高了	他的	袖子	繼續	工作			
49	R-Simple	農夫	捲過了	他的	袖子	繼續	工作			
50	Resultative	農夫	砌高了	這裡的	圍牆	防止	作物	遭竊		
50	R-Simple	農夫	砌過了	這裡的	圍牆	防止	作物	遭竊		
51	Resultative	遊客	摸亮了	那裡的	銅像	還	打卡	拍照		
51	R-Simple	遊客	摸過了	那裡的	銅像	還	打卡	拍照		
52	Resultative	僕人	折疊了	很多	棉被	，	房間	變得	很	整齊
52	R-Simple	僕人	折過了	很多	棉被	，	房間	因此	變	整齊
53	Resultative	廠商	蓋矮了	很多	房子	只好	向	客戶	道歉	
53	R-Simple	廠商	蓋過了	很多	房子	才	蓋	其他	設施	
54	Resultative	敵軍	炸翻了	很多	戰艦	，	已經	取得	先機	
54	R-Simple	敵軍	炸過了	很多	戰艦	，	已經	取得	先機	
55	Resultative	窯爐	燒紅了	那裡的	磚頭	，	小心	不要	靠近	
55	R-Simple	窯爐	燒過了	那裡的	磚頭	，	小心	不要	靠近	
56	Resultative	學生	擦亮了	自己的	皮鞋	為了	參加	畢業典禮		
56	R-Simple	學生	擦過了	自己的	皮鞋	為了	參加	畢業典禮		
57	Resultative	學徒	拌勻了	很多	水泥	以	抹到	牆上		
57	R-Simple	學徒	拌過了	很多	水泥	以	抹到	牆上		
58	Resultative	嬸嬸	貼反了	這裡的	春聯	被	鄰居	取笑		
58	R-Simple	嬸嬸	貼過了	這裡的	春聯	再	貼	其他	地方	
59	Resultative	藝術家	剪斷了	一些	棉線	準備	用來	創作		
59	R-Simple	藝術家	拉過了	很多	線條	做成	美麗	的	藝術品	
60	Resultative	藝術家	拉直了	很多	線條	做成	美麗	的	藝術品	
60	R-Simple	藝術家	剪過了	一些	棉線	準備	用來	創作		
Coordinate set										
1	Coordinate	大學生	選購了	他們的	電腦	帶回	宿舍			
1	C-Simple	大學生	選過了	他們的	電腦	帶回	實驗室			
2	Coordinate	女子	刷洗了	她的	廁所	才	上床	休息		
2	C-Simple	女子	刷過了	她的	廁所	才	上床	休息		
3	Coordinate	女孩	拆用了	很多	化妝品	並且	拍	影片	介紹	
3	C-Simple	女孩	拆過了	很多	化妝品	試用	，	沒有	馬上	購買
4	Coordinate	女學生	塗改了	她的	答案	才	交卷			
4	C-Simple	女學生	塗過了	她的	答案	卻	沒	寫上	新的	
5	Coordinate	少年	拆解了	不少	炸彈	，	被	譽為	天才	少年
5	C-Simple	少年	拆過了	不少	炸彈	，	被	譽為	天才	少年
6	Coordinate	王妃	攔截了	你的	信件	救了	王爺	一命		
6	C-Simple	王妃	攔過了	你的	信件	救了	王爺	一命		

7	Coordinate	主唱	錄製了	很多	唱片	不過	仍	需要	後製	
7	C-Simple	主唱	錄過了	很多	唱片	不過	仍	需要	後製	
8	Coordinate	主編	刪改了	你的	文章	讓	排版	更	美觀	
8	C-Simple	主編	刪過了	你的	文章	因為	稿件	太多了		
9	Coordinate	市政府	核發了	所有的	執照	商家	才	開始	營業	
9	C-Simple	市政府	核過了	所有的	執照	商家	才能	開始	營業	
10	Coordinate	犯人	追打了	很多	被害人	，	路人	發現	趕緊	報案
10	C-Simple	犯人	追過了	很多	被害人	，	犯案	動機	可疑	
11	Coordinate	同學們	傳看了	我的	漫畫	結果	被	老師	沒收	
11	C-Simple	同學們	傳過了	我的	漫畫	結果	被	老師	沒收	
12	Coordinate	名媛	退換了	很多	包包	，	指定	要	最新款	
12	C-Simple	名媛	退過了	很多	包包	，	指定	要	最新款	
13	Coordinate	收銀員	換算了	那裡的	匯率	公告	在	網站上		
13	C-Simple	收銀員	換過了	那裡的	匯率	公告	在	網站上		
14	Coordinate	老奶奶	借用了	你的	雨傘	才	沒有	淋濕		
14	C-Simple	老奶奶	借過了	你的	雨傘	可是	沒有	歸還		
15	Coordinate	老闆	存放了	不少	現金	在	保險箱	裡面		
15	C-Simple	老闆	存過了	不少	現金	又	存	美金		
16	Coordinate	老闆娘	打罵了	不少	員工	後來	被	檢舉		
16	C-Simple	老闆娘	打過了	不少	員工	後來	被	檢舉		
17	Coordinate	考生	選填了	許多	志願	可是	個個	沒	把握	
17	C-Simple	考生	選過了	許多	志願	可是	個個	沒	把握	
18	Coordinate	助理	遞交了	他的	辭呈	，	等待	長官	批准	
18	C-Simple	助理	遞過了	他的	辭呈	，	等待	長官	批准	
19	Coordinate	抗議人士	編印了	他們的	文宣	準備	沿路	發放		
19	C-Simple	抗議人士	編過了	他們的	文宣	準備	沿路	發放		
20	Coordinate	李先生	拐賣了	不少	小女孩	很	沒有	良心		
20	C-Simple	李先生	拐過了	不少	小女孩	卻	沒有	成功		
21	Coordinate	男朋友	改訂了	自己的	餐廳	作為	求婚	地點		
21	C-Simple	男朋友	改過了	自己的	餐廳	讓	氣氛	更好		
22	Coordinate	男童	組裝了	很多	機器人	，	玩得	很	開心	
22	C-Simple	男童	組過了	很多	機器人	又	組好	小汽車		
23	Coordinate	岳母	催迫了	她的	女婿	趕快	與	女兒	和好	
23	C-Simple	岳母	催過了	她的	女婿	想要	抱	孫子		
24	Coordinate	法官	審讀了	他們的	判決書	，	被告	不發一語		
24	C-Simple	法官	審過了	他們的	判決書	，	盡量	保持	公正	
25	Coordinate	表妹	彈唱了	你的	一首歌	，	很	動聽		
25	C-Simple	表妹	彈過了	你的	一首歌	，	很	動聽		
26	Coordinate	保姆	沖泡了	一些	牛奶	給	寶寶	喝		
26	C-Simple	保姆	沖過了	一些	牛奶	給	寶寶	喝		
27	Coordinate	前女友	調閱了	這裡的	監視器	才	掌握	他的	行蹤	

27	C-Simple	前女友	調過了	這裡的	監視器	才	掌握	他的	行蹤	
28	Coordinate	政府	修護了	所有的	古蹟	才	開放	參觀		
28	C-Simple	政府	修過了	所有的	古蹟	才	開放	參觀		
29	Coordinate	科學家	訂造了	一些	儀器	以	進行	最新的	實驗	
29	C-Simple	科學家	訂過了	一些	儀器	以	進行	最新的	實驗	
30	Coordinate	美容師	按壓了	我的	穴道	讓	我	覺得	很	放鬆
30	C-Simple	美容師	按過了	我的	穴道	讓	我	覺得	很	放鬆
31	Coordinate	英文系	徵聘了	一些	教授	可是	沒有	人	應徵	
31	C-Simple	英文系	徵過了	一些	教授	可是	沒有	人	應徵	
32	Coordinate	軍方	埋設了	很多	地雷	等待	敵軍	經過		
32	C-Simple	軍方	埋過了	很多	地雷	等待	敵軍	經過		
33	Coordinate	軍隊	攻取了	那裡的	城池	，	贏得	勝利		
33	C-Simple	軍隊	攻過了	那裡的	城池	卻	無法	攻下來		
34	Coordinate	家庭主婦	挑取了	許多	青菜	準備	做	晚餐		
34	C-Simple	家庭主婦	挑過了	許多	青菜	準備	做	晚餐		
35	Coordinate	旅客	看管了	他們的	行李箱	輪流	去	使用	洗手間	
35	C-Simple	旅客	看過了	他們的	行李箱	覺得	不需要	購買	新的	
36	Coordinate	留學生	搬運了	他們的	行李	終於	抵達	新家		
36	C-Simple	留學生	搬過了	他們的	行李	又	搬	傢俱		
37	Coordinate	粉絲	想慕了	他們的	偶像	天天	做	白日夢		
37	C-Simple	粉絲	想過了	他們的	偶像	很	擔心	於是	送上	補品
38	Coordinate	逃難者	埋藏了	所有的	財物	避免	被	強盜	洗劫而空	
38	C-Simple	逃難者	埋過了	所有的	財物	避免	被	強盜	洗劫而空	
39	Coordinate	參選人	分送了	很多	面紙	，	想	拉攏	選民	
39	C-Simple	參選人	分過了	很多	面紙	，	想	拉攏	選民	
40	Coordinate	婆婆	縫補了	一些	衣服	給	公公	穿		
40	C-Simple	婆婆	縫過了	一些	衣服	給	公公	穿		
41	Coordinate	將軍	聽信了	不少	讒言	決定	不再	效忠	朝廷	
41	C-Simple	將軍	聽過了	不少	讒言	仍然	決定	效忠	朝廷	
42	Coordinate	教授	核銷了	很多	經費	才	發現	經費	短缺	
42	C-Simple	教授	核過了	很多	經費	才	發現	經費	短缺	
43	Coordinate	船員	捕殺了	很多	鯨魚	，	民間團體	因此	出面	抗議
43	C-Simple	船員	捕過了	很多	鯨魚	，	民間團體	因此	出面	抗議
44	Coordinate	陳小姐	編譯了	我的	新書	交給	出版社	出版		
44	C-Simple	陳小姐	編過了	我的	新書	交給	出版社	出版		
45	Coordinate	園丁	噴灑了	很多	農藥	才	下班			
45	C-Simple	園丁	噴過了	很多	農藥	才	下班			
46	Coordinate	嫌犯	改裝了	他的	車子	以	逃避	警方	追緝	
46	C-Simple	嫌犯	改過了	他的	車子	以	逃避	警方	追緝	
47	Coordinate	業者	停用了	他們的	網路	，	店面	也	出售了	
47	C-Simple	業者	停過了	他們的	網路	，	店面	也	出售了	

48	Coordinate	準新娘	訂製了	一些	禮服	明天	要	試穿		
48	C-Simple	準新娘	訂過了	一些	禮服	又	訂了	喜餅		
49	Coordinate	管理員	點收了	我的	管理費	才	給	我	收據	
49	C-Simple	管理員	點過了	我的	管理費	才	給	我	收據	
50	Coordinate	舞者	搶購了	多數的	舞鞋	導致	舞鞋	缺貨		
50	C-Simple	舞者	搶過了	多數的	舞鞋	剩下的	都是	瑕疵品		
51	Coordinate	鄰居	鏟除了	很多	雜草	所以	院子	煥然一新		
51	C-Simple	鄰居	鏟過了	很多	雜草	所以	院子	煥然一新		
52	Coordinate	學弟	敲打了	自己的	鍵盤	，	沈迷	線上	遊戲	
52	C-Simple	學弟	敲過了	自己的	鍵盤	想	找	女孩	聊天	
53	Coordinate	學長	愛慕了	我的	學妹	卻	不敢	表白		
53	C-Simple	學長	愛過了	我的	學妹	卻	不敢	表白		
54	Coordinate	機器	壓製了	一些	模型	才能	量產			
54	C-Simple	機器	壓過了	一些	模型	才能	量產			
55	Coordinate	縣長	減省了	這裡的	預算	，	工程	仍然	能	進行
55	C-Simple	縣長	減過了	這裡的	預算	，	工程	仍然	能	進行
56	Coordinate	檢察官	簽發了	一些	搜索票	，	限期	逮捕	嫌犯	
56	C-Simple	檢察官	簽過了	一些	搜索票	，	限期	逮捕	嫌犯	
57	Coordinate	獵人	射殺了	那裡的	鹿	又	捕到	野兔		
57	C-Simple	獵人	射過了	那裡的	鹿	又	射到	野兔		
58	Coordinate	醫師	拔除了	他的	智齒	只	花	十分鐘		
58	C-Simple	醫師	拔過了	他的	智齒	才	體會到	病人	的	感覺
59	Coordinate	警方	查扣了	他的	毒品	並	調查	毒品	流向	
59	C-Simple	警方	查過了	他的	毒品	沒有	發現	其他	證據	
60	Coordinate	護理師	哄騙了	那裡的	病人	還	唱歌	給	病人	聽
60	C-Simple	護理師	哄過了	那裡的	病人	還	唱歌	給	病人	聽

4. Experiment 4 and Experiment 6 (in Chapter 3)

Condition	NP1	(co)verb	NP2	Critical verb	Post-target continuation				
1-Complement	女歌手	認為	狗仔隊	甩了	經紀人	是	有	其他	原因
1-Baseline	女歌手	把	狗仔隊	甩了	，	躲開了	他們	的	跟拍
2-Complement	父母	認為	孩子	撫養了	生病的	小孫女	，	很辛苦	
2-Baseline	父母	把	孩子	撫養了	長大	，	很辛苦		
3-Complement	民眾	認為	暴君	推翻了	前政權	，	人民	還是	受苦的
3-Baseline	民眾	把	暴君	推翻了	之後	決定	推選	新的	君王
4-Complement	刑警	認為	嫌犯	抓住了	人質	想要	談判		
4-Baseline	刑警	把	嫌犯	抓住了	並且	立即	帶回	警局	
5-Complement	奸商	認為	投資者	騙了	其他人	，	覺得	不可思議	
5-Baseline	奸商	把	投資者	騙了	，	轉眼間	逃去無蹤		
6-Complement	老陳	認為	河豚	放生了	青蛙				
6-Baseline	老陳	把	河豚	放生了	，	並	沒有	告訴	任何人

7-Complement	老顧客	認為	按摩師	誇獎了	女客人	是	別有居心	
7-Baseline	老顧客	把	按摩師	誇獎了	一番	，	按摩師	笑得	合不攏嘴
8-Complement	吳管家	認為	流浪狗	收留了	貴賓狗	是	因為	發情了	
8-Baseline	吳管家	把	流浪狗	收留了	下來	並且	用心地	照顧	牠
9-Complement	吳管家	認為	乞丐	趕走了	主人的	愛犬	，	很著急	
9-Baseline	吳管家	把	乞丐	趕走了	然後	趕緊	把	門	關上
10-Complement	私家偵探	認為	那個政客	調查了	對手	是	另有	隱情	
10-Baseline	私家偵探	把	那個政客	調查了	一個月，	終於	查到	他的	婚外情
11-Complement	那位商人	認為	高官	賄賂了	里長	得到	不少	好處	
11-Baseline	那位商人	把	高官	賄賂了	以後	得到	不少	好處	
12-Complement	那位顧客	認為	售貨員	罵了	司機	一頓	，	很過分	
12-Baseline	那位顧客	把	售貨員	罵了	一頓	，	要求	他	道歉
13-Complement	那個叛徒	認為	群眾	出賣了	他	，	決定	不再	上當
13-Baseline	那個叛徒	把	群眾	出賣了	，	大家	決定	不再	上當
14-Complement	那個病人	認為	董大夫	感謝了	家屬	很貼心			
14-Baseline	那個病人	把	董大夫	感謝了	一番	，	場面	十分	感人
15-Complement	那個商人	認為	工人	辭退了	外勞	是	躲避	警方	調查
15-Baseline	那個商人	把	工人	辭退了	以後	便	破產了		
16-Complement	那個惡霸	認為	小弟弟	打了	學長	一頓	，	很好奇	
16-Baseline	那個惡霸	把	小弟弟	打了	一頓	，	很過分		
17-Complement	那群駿馬	認為	那個畫家	吸引了	主人的	注意力	，	很開心	
17-Baseline	那群駿馬	把	那個畫家	吸引了	，	他	總算	找到	靈感了
18-Complement	那模特兒	認為	總編輯	迷倒了	設計師	，	有點	吃味	
18-Baseline	那模特兒	把	總編輯	迷倒了	以後	便	立即	成為	封面女郎
19-Complement	來賓	認為	相聲演員	稱讚了	主持人	是	很大的	殊榮	
19-Baseline	來賓	把	相聲演員	稱讚了	一番	，	才	拿到	簽名
20-Complement	周老闆	認為	那個畢業生	錄用了	其他	助理	是	想要	偷懶
20-Baseline	周老闆	把	那個畢業生	錄用了	以後	，	生意	竟然	變好了
21-Complement	牧羊人	認為	羊群	趕了	牧羊犬	回家	很好笑		
21-Baseline	牧羊人	把	羊群	趕了	到	草原上	吃草		
22-Complement	保姆	認為	小孩	照顧了	寵物	可以	培養	責任心	
22-Baseline	保姆	把	小孩	照顧了	又照顧	，	特別地	寵愛	
23-Complement	皇后	認為	白雪公主	毒害了	小動物	，	心地	根本	不善良
23-Baseline	皇后	把	白雪公主	毒害了	以後	更	滿意	自己的	美貌
24-Complement	皇帝	認為	貪官	殺了	親信	之後	就會	倒戈	
24-Baseline	皇帝	把	貪官	殺了	之後	，	任命了	新的	官員
25-Complement	皇帝	認為	佟妃	冷落了	宮女	，	下令	責罰	佟妃
25-Baseline	皇帝	把	佟妃	冷落了	以後	繼續	花天酒地		
26-Complement	軍人	認為	囚犯	關了	心房	所以	無法	問到	細節
26-Baseline	軍人	把	囚犯	關了	很久	，	避免	機密	外洩
27-Complement	恐怖分子	認為	那群旅客	劫持了	他們的	領導	當	人質	

27-Baseline	恐怖分子	把	那群旅客	劫持了	三天	當	人質		
28-Complement	海洋學家	認為	鯨魚	研究了	廢棄	沈船	，	很可愛	
28-Baseline	海洋學家	把	鯨魚	研究了	一年多	卻	沒有	任何	發現
29-Complement	消防員	認為	那個傷者	救了	其他	受困的	人		
29-Baseline	消防員	把	那個傷者	救了	出來	，	自己	卻	重傷
30-Complement	鬥牛士	認為	蠻牛	制服了	小牛	就	很心疼	小牛	
30-Baseline	鬥牛勇士	把	蠻牛	制服了	之後	便	拔刀	殺了	牠
31-Complement	專家	認為	傳染病人	隔離了	彼此	可以	防止	病菌	傳播
31-Baseline	專家	把	傳染病人	隔離了	起來	，	防止	病菌	傳播
32-Complement	球迷	認為	這個球星	圍了	髮帶	有點	不好看		
32-Baseline	球迷	把	這個球星	圍了	一圈	，	想要	索取	合照
33-Complement	這匹野馬	認為	騎師	摔了	鞭子	是	生氣的	表現	
33-Baseline	這匹野馬	把	騎師	摔了	，	開始	狂奔		
34-Complement	這個學生	認為	鄧老師	氣壞了	校長	其實	是	情有可原	
34-Baseline	這個學生	把	鄧老師	氣壞了	卻	不肯	道歉		
35-Complement	這隻小狗	認為	那位貴婦	咬了	漢堡	，	一直	汪汪叫	
35-Baseline	這隻小狗	把	那位貴婦	咬了	一口	還	一直	汪汪叫	
36-Complement	這隻花貓	認為	那隻麻雀	吃了	飼料	以後	會	飛走	
36-Baseline	這隻花貓	把	那隻麻雀	吃了	，	連	羽毛	都	沒留下
37-Complement	富翁	認為	傭人	解雇了	小童工	很	不應該		
37-Baseline	富翁	把	傭人	解雇了	之後	立即	請來了	新的	管家
38-Complement	童工	認為	地主	戲弄了	女僕	所以	向	老夫人	報告
38-Baseline	童工	把	地主	戲弄了	一番	，	連累	其他人	受罰了
39-Complement	裁判	認為	那個參賽者	處罰了	聽眾	因為	唱得	很難聽	
39-Baseline	裁判	把	那個參賽者	處罰了	，	宣佈	比賽	暫停	
40-Complement	評審	認為	得獎者	表揚了	幕後	工作人員	，	非常	飲水思源
40-Baseline	評審	把	得獎者	表揚了	一番	然後	頒發	獎狀	
41-Complement	黑道老大	認為	那位候選人	殺掉了	他的	對手			
41-Baseline	黑道老大	把	那位候選人	殺掉了	並且	製造	自殺的	景象	
42-Complement	媽媽	認為	新生兒	抱了	陌生人	會	很容易	被	騙走
42-Baseline	媽媽	把	新生兒	抱了	又抱	，	臉上	露出了	笑容
43-Complement	慈善家	認為	孤兒	領養了	流浪貓	會	好好地	疼愛	牠
43-Baseline	慈善家	把	孤兒	領養了	，	並且	供	他	上學
44-Complement	滅蟲專家	認為	那些老鼠	消滅了	人類的	存糧			
44-Baseline	滅蟲專家	把	那些老鼠	消滅了	以後	才	向	住戶	收費
45-Complement	爺爺	認為	孫子	抱緊了	流浪貓	沒有	關係		
45-Baseline	爺爺	把	孫子	抱緊了	，	害怕	他	又	走失了
46-Complement	獅子	認為	小弟弟	嚇了	他的	媽媽	一跳		
46-Baseline	獅子	把	小弟弟	嚇了	一跳				
47-Complement	經理人	認為	新樂隊	解散了	高音部	，	粉絲	也會	流失
47-Baseline	經理人	把	新樂隊	解散了	以後	讓	他們	單獨	發展

48-Complement	馴獸師	認為	獅子	馴服了	其他	小動物	，	很得意	
48-Baseline	馴獸師	把	獅子	馴服了	，	變得	非常	乖巧	
49-Complement	演員	認為	國家元首	逗笑了	小寶寶	，	表現	很親民	
49-Baseline	演員	把	國家元首	逗笑了	而且	獲得了	不少	讚美	
50-Complement	模特兒	認為	那個畫家	迷住了	賭博	就	畫不出	佳作了	
50-Baseline	模特兒	把	那個畫家	迷住了	並且	欺騙了	他的	感情	
51-Complement	窮書生	認為	仙女	娶了	兒媳婦	之後	就	失去	法力了
51-Baseline	窮書生	把	仙女	娶了	之後	，	便	開始	奮發向上
52-Complement	緝毒犬	認為	那個毒販	找到了	同伴	，	十分	警戒	
52-Baseline	緝毒犬	把	那個毒販	找到了	，	顯得	十分	神氣	
53-Complement	戰友	認為	烈士	埋葬了	親人	並	在	墳前	燒香
53-Baseline	戰友	把	烈士	埋葬了	，	並	在	墳前	燒香
54-Complement	選手	認為	教練	激怒了	家長	導致	氣氛	很緊張	
54-Baseline	選手	把	教練	激怒了	，	氣氛	很緊張		
55-Complement	總司令	認為	部隊	調遣	班長	離開	營區	很理想	
55-Baseline	總司令	把	部隊	調遣	到	邊境	以茲	防守	
56-Complement	壞學生	認為	小弟弟	帶壞了	手下	於是	找人	教訓	他
56-Baseline	壞學生	把	小弟弟	帶壞了	，	小弟弟	從此	一直	翹課
57-Complement	警察	認為	災民	安置了	那個	年幼的	小孩	很有	義氣
57-Baseline	警察	把	災民	安置了	，	才	交班		
58-Complement	鐘老闆	認為	這群員工	開除了	工讀生	是	狐假虎威的	表現	
58-Baseline	鐘老闆	把	這群員工	開除了	，	拒絕	給付	任何	津貼
59-Complement	魔術師	認為	白鴿	變走了	玫瑰				
59-Baseline	魔術師	把	白鴿	變走了	以後	觀眾	都	很興奮	
60-Complement	鸚鵡	認為	老劉	逗了	另一隻	鸚鵡	是	因為	偏心
60-Baseline	鸚鵡	把	老劉	逗了	一番	使	他	哈哈大笑	

5. Experiment 5 (in Chapter 3)

Condition	NP1	Ba	NP2	Critical verb	Post-verb continuation				
1-Baseline	富翁	把	傭人	解雇了	之後	立即	請來了	新的	管家
1-Reversal	傭人	把	富翁	解雇了	之後	立即	請來了	新的	管家
2-Baseline	軍人	把	囚犯	關了	很久	，	避免	機密	外洩
2-Reversal	囚犯	把	軍人	關了	很久	，	避免	機密	外洩
3-Baseline	這隻小狗	把	那位貴婦	咬了	一口	還	一直	汪汪叫	
3-Reversal	那位貴婦	把	這隻小狗	咬了	一口	還	一直	汪汪叫	
4-Baseline	奸商	把	投資者	騙了	，	轉眼間	逃去無蹤		
4-Reversal	投資者	把	奸商	騙了	，	轉眼間	逃去無蹤		
5-Baseline	演員	把	國家元首	逗笑了	而且	獲得了	不少	讚美	
5-Reversal	國家元首	把	演員	逗笑了	而且	獲得了	不少	讚美	
6-Baseline	這個學生	把	鄧老師	氣壞了	卻	不肯	道歉		
6-Reversal	鄧老師	把	這個學生	氣壞了	卻	不肯	道歉		

7-Baseline	黑道老大	把	那位候選人	殺掉了	並且	製造	自殺的	景象	
7-Reversal	那位候選人	把	黑道老大	殺掉了	並且	製造	自殺的	景象	
8-Baseline	鬥牛勇士	把	蠻牛	制服了	之後	便	拔刀	殺了	牠
8-Reversal	蠻牛	把	鬥牛勇士	制服了	之後	便	拔刀	殺了	牠
9-Baseline	老顧客	把	按摩師	誇獎了	一番	，	按摩師	笑得	合不攏嘴
9-Reversal	按摩師	把	老顧客	誇獎了	一番	，	按摩師	笑得	合不攏嘴
10-Baseline	牧羊人	把	羊群	趕了	到	草原上	吃草		
10-Reversal	羊群	把	牧羊人	趕了	到	草原上	吃草		
11-Baseline	刑警	把	嫌犯	抓住了	並且	立即	帶回	警局	
11-Reversal	嫌犯	把	刑警	抓住了	並且	立即	帶回	警局	
12-Baseline	這隻花貓	把	那隻麻雀	吃了	，	連	羽毛	都	沒留下
12-Reversal	那隻麻雀	把	這隻花貓	吃了	，	連	羽毛	都	沒留下
13-Baseline	女歌手	把	狗仔隊	甩了	，	躲開了	他們	的	跟拍
13-Reversal	狗仔隊	把	女歌手	甩了	，	躲開了	他們	的	跟拍
14-Baseline	壞學生	把	小弟弟	帶壞了	，	小弟弟	從此	一直	翹課
14-Reversal	小弟弟	把	壞學生	帶壞了	，	小弟弟	從此	一直	翹課
15-Baseline	那位顧客	把	售貨員	罵了	一頓	，	要求	他	道歉
15-Reversal	售貨員	把	那位顧客	罵了	一頓	，	要求	他	道歉
16-Baseline	父母	把	孩子	撫養了	長大	，	很辛苦		
16-Reversal	孩子	把	父母	撫養了	長大	，	很辛苦		
17-Baseline	吳管家	把	流浪狗	收留了	下來	並且	用心地	照顧	牠
17-Reversal	流浪狗	把	吳管家	收留了	下來	並且	用心地	照顧	牠
18-Baseline	鸚鵡	把	老劉	逗了	一番	使	他	哈哈大笑	
18-Reversal	老劉	把	鸚鵡	逗了	一番	使	他	哈哈大笑	
19-Baseline	周老闆	把	那個畢業生	錄用了	以後	，	生意	竟然	變好了
19-Reversal	那個畢業生	把	周老闆	錄用了	以後	，	生意	竟然	變好了
20-Baseline	童工	把	地主	戲弄了	一番	，	連累	其他人	受罰了
20-Reversal	地主	把	童工	戲弄了	一番	，	連累	其他人	受罰了
21-Baseline	皇帝	把	貪官	殺了	之後	，	任命了	新的	官員
21-Reversal	貪官	把	皇帝	殺了	之後	，	任命了	新的	官員
22-Baseline	滅蟲專家	把	那些老鼠	消滅了	以後	才	向	住戶	收費
22-Reversal	那些老鼠	把	滅蟲專家	消滅了	以後	才	向	住戶	收費
23-Baseline	警察	把	災民	安置了	，	才	交班		
23-Reversal	災民	把	警察	安置了	，	才	交班		
24-Baseline	模特兒	把	那個畫家	迷住了	並且	欺騙了	他的	感情	
24-Reversal	那個畫家	把	模特兒	迷住了	並且	欺騙了	他的	感情	
25-Baseline	民眾	把	暴君	推翻了	之後	決定	推選	新的	君王
25-Reversal	暴君	把	民眾	推翻了	之後	決定	推選	新的	君王
26-Baseline	魔術師	把	白鴿	變走了	以後	觀眾	都	很興奮	
26-Reversal	白鴿	把	魔術師	變走了	以後	觀眾	都	很興奮	
27-Baseline	皇后	把	白雪公主	毒害了	以後	更	滿意	自己的	美貌

27-Reversal	白雪公主	把	皇后	毒害了	以後	更	滿意	自己的	美貌
28-Baseline	球迷	把	這個球星	圍了	一圈	，	想要	索取	合照
28-Reversal	這個球星	把	球迷	圍了	一圈	，	想要	索取	合照
29-Baseline	媽媽	把	新生兒	抱了	又抱	，	臉上	露出了	笑容
29-Reversal	新生兒	把	媽媽	抱了	又抱	，	臉上	露出了	笑容
30-Baseline	馴獸師	把	獅子	馴服了	，	變得	非常	乖巧	
30-Reversal	獅子	把	馴獸師	馴服了	，	變得	非常	乖巧	
31-Baseline	戰友	把	烈士	埋葬了	，	並	在	墳前	燒香
31-Reversal	烈士	把	戰友	埋葬了	，	並	在	墳前	燒香
32-Baseline	恐怖分子	把	那群旅客	劫持了	三天	當	人質		
32-Reversal	那群旅客	把	恐怖分子	劫持了	三天	當	人質		
33-Baseline	慈善家	把	孤兒	領養了	，	並且	供	他	上學
33-Reversal	孤兒	把	慈善家	領養了	，	並且	供	他	上學
34-Baseline	那群駿馬	把	那個畫家	吸引了	，	他	總算	找到	靈感了
34-Reversal	那個畫家	把	那群駿馬	吸引了	，	他	總算	找到	靈感了
35-Baseline	來賓	把	相聲演員	稱讚了	一番	，	才	拿到	簽名
35-Reversal	相聲演員	把	來賓	稱讚了	一番	，	才	拿到	簽名
36-Baseline	那模特兒	把	總編輯	迷倒了	以後	便	立即	成為	封面女郎
36-Reversal	總編輯	把	那模特兒	迷倒了	以後	便	立即	成為	封面女郎
37-Baseline	窮書生	把	仙女	娶了	之後	，	便	開始	奮發向上
37-Reversal	仙女	把	窮書生	娶了	之後	，	便	開始	奮發向上
38-Baseline	那個惡霸	把	小弟弟	打了	一頓	，	很過分		
38-Reversal	小弟弟	把	那個惡霸	打了	一頓	，	很過分		
39-Baseline	海洋學家	把	鯨魚	研究了	一年多	卻	沒有	任何	發現
39-Reversal	鯨魚	把	海洋學家	研究了	一年多	卻	沒有	任何	發現
40-Baseline	爺爺	把	孫子	抱緊了	，	害怕	他	又	走失了
40-Reversal	孫子	把	爺爺	抱緊了	，	害怕	他	又	走失了
41-Baseline	這匹野馬	把	騎師	摔了	，	開始	狂奔		
41-Reversal	騎師	把	這匹野馬	摔了	，	開始	狂奔		
42-Baseline	那位商人	把	高官	賄賂了	以後	得到	不少	好處	
42-Reversal	高官	把	那位商人	賄賂了	以後	得到	不少	好處	
43-Baseline	獅子	把	小弟弟	嚇了	一跳				
43-Reversal	小弟弟	把	獅子	嚇了	一跳				
44-Baseline	評審	把	得獎者	表揚了	一番	然後	頒發	獎狀	
44-Reversal	得獎者	把	評審	表揚了	一番	然後	頒發	獎狀	
45-Baseline	私家偵探	把	那個政客	調查了	一個月，	終於	查到	他的	婚外情
45-Reversal	那個政客	把	私家偵探	調查了	一個月，	終於	查到	他的	婚外情
46-Baseline	鐘老闆	把	這群員工	開除了	，	拒絕	給付	任何	津貼
46-Reversal	這群員工	把	鐘老闆	開除了	，	拒絕	給付	任何	津貼
47-Baseline	吳管家	把	乞丐	趕走了	然後	趕緊	把	門	關上
47-Reversal	乞丐	把	吳管家	趕走了	然後	趕緊	把	門	關上

48-Baseline	那個叛徒	把	群眾	出賣了	，	大家	決定	不再	上當
48-Reversal	群眾	把	那個叛徒	出賣了	，	大家	決定	不再	上當
49-Baseline	那個商人	把	工人	辭退了	以後	便	破產了		
49-Reversal	工人	把	那個商人	辭退了	以後	便	破產了		
50-Baseline	保姆	把	小孩	照顧了	又照顧	，	特別地	寵愛	
50-Reversal	小孩	把	保姆	照顧了	又照顧	，	特別地	寵愛	
51-Baseline	消防員	把	那個傷者	救了	出來	，	自己	卻	重傷
51-Reversal	那個傷者	把	消防員	救了	出來	，	自己	卻	重傷
52-Baseline	緝毒犬	把	那個毒販	找到了	，	顯得	十分	神氣	
52-Reversal	那個毒販	把	緝毒犬	找到了	，	顯得	十分	神氣	
53-Baseline	裁判	把	那個參賽者	處罰了	，	宣佈	比賽	暫停	
53-Reversal	那個參賽者	把	裁判	處罰了	，	宣佈	比賽	暫停	
54-Baseline	那個病人	把	董大夫	感謝了	一番	，	場面	十分	感人
54-Reversal	董大夫	把	那個病人	感謝了	一番	，	場面	十分	感人
55-Baseline	皇帝	把	佟妃	冷落了	以後	繼續	花天酒地		
55-Reversal	佟妃	把	皇帝	冷落了	以後	繼續	花天酒地		
56-Baseline	總司令	把	部隊	調遣	到	邊境	以茲	防守	
56-Reversal	部隊	把	總司令	調遣	到	邊境	以茲	防守	
57-Baseline	老陳	把	河豚	放生了	，	並	沒有	告訴	任何人
57-Reversal	河豚	把	老陳	放生了	，	並	沒有	告訴	任何人
58-Baseline	經理人	把	新樂隊	解散了	以後	讓	他們	單獨	發展
58-Reversal	新樂隊	把	經理人	解散了	以後	讓	他們	單獨	發展
59-Baseline	選手	把	教練	激怒了	，	氣氛	很緊張		
59-Reversal	教練	把	選手	激怒了	，	氣氛	很緊張		
60-Baseline	專家	把	傳染病人	隔離了	起來	，	防止	病菌	傳播
60-Reversal	傳染病人	把	專家	隔離了	起來	，	防止	病菌	傳播

6. Experiment 7 (in Chapter 4)

	Condition	Sentence
Subcategorization		
1	Grammatical	My sister recorded the music I played.
1	Ungrammatical	My sister listened the music I played.
2	Grammatical	Linda should not deceive the girl on the street.
2	Ungrammatical	Linda should not laugh the girl on the street.
3	Grammatical	The judges will encourage the contestants fairly.
3	Ungrammatical	The judges will comment the contestants equally.
4	Grammatical	The fans will neglect his affairs for a while.
4	Ungrammatical	The fans will gossip his affairs for a while.
5	Grammatical	The villagers might minimize the pollution from the factory.
5	Ungrammatical	The villagers might complain the pollution from the factory.
6	Grammatical	Drivers should obey the laws for safety.

6	Ungrammatical	Drivers should conform the laws for safety.
7	Grammatical	The inspector should identify the motivation for the murder.
7	Ungrammatical	The inspector should delve the motivation for the murder.
8	Grammatical	The new faculty criticized the study of language.
8	Ungrammatical	The new faculty obsessed the study of language.
9	Grammatical	Rebecca accepted my decision on moving out.
9	Ungrammatical	Rebecca interfered my decision on moving out.
10	Grammatical	The teacher punished the student angrily.
10	Ungrammatical	The teacher glared the student angrily.
11	Grammatical	Her boyfriend might host a party tonight.
11	Ungrammatical	Her boyfriend might participate a party tonight.
12	Grammatical	The housewife vacuumed the room quickly.
12	Ungrammatical	The housewife glanced the room quickly.
13	Grammatical	His parents comforted the families of the victims.
13	Ungrammatical	His parents sympathized the families of the victims.
14	Grammatical	Joseph declined the support from the government.
14	Ungrammatical	Joseph relied the support from the government.
15	Grammatical	Debbie really appreciated the cuisine at the restaurant.
15	Ungrammatical	Debbie really cared the cuisine at the restaurant.
16	Grammatical	The grandma has handled the tragedy successfully.
16	Ungrammatical	The grandma has coped the tragedy successfully.
17	Grammatical	The president may reject the proposal completely.
17	Ungrammatical	The president may object the proposal completely.
18	Grammatical	The girl grabbed the tail of the dog.
18	Ungrammatical	The girl stepped the tail of the dog.
19	Grammatical	The designer allowed the customer to modify her pattern.
19	Ungrammatical	The designer agreed the customer to modify her pattern.
20	Grammatical	Nick and Joshua missed the bus to school.
20	Ungrammatical	Nick and Joshua waited the bus to school.
21	Grammatical	The director should finalize the plans for the project.
21	Ungrammatical	The director should talk the plans for the project.
22	Grammatical	English learners should consult the dictionary when learning vocabulary.
22	Ungrammatical	English learners should refer the dictionary when learning vocabulary.
23	Grammatical	The guest speaker praised the campus yesterday.
23	Ungrammatical	The guest speaker arrived the campus yesterday.
24	Grammatical	The manager forwarded the email promptly.
24	Ungrammatical	The manager replied the email promptly.
25	Grammatical	The mother hugged her son closely.
25	Ungrammatical	The mother gazed her son admiringly.
26	Grammatical	He should have toured the country earlier.
26	Ungrammatical	He should have immigrated the country earlier.

27	Grammatical	Her grandson planted a tree by the road.
27	Ungrammatical	Her grandson collided a tree by the road.
28	Grammatical	Her father would undergo the treatment for his disease.
28	Ungrammatical	Her father would consent the treatment for his disease.
29	Grammatical	You should ride the lifeboat securely.
29	Ungrammatical	You should cling the lifeboat securely.
30	Grammatical	The elder should resolve the dispute wisely.
30	Ungrammatical	The elder should intervene the dispute wisely.
31	Grammatical	My colleagues edited the film in the office.
31	Ungrammatical	My colleagues chatted the film in the office.
32	Grammatical	The weatherman refuted the possibility of rain.
32	Ungrammatical	The weatherman hinted the possibility of rain.
33	Grammatical	The coach might assess his performance tomorrow.
33	Ungrammatical	The coach might worry his performance tomorrow.
34	Grammatical	The dog might lick the visitor at the door.
34	Ungrammatical	The dog might bark the visitor at the door.
35	Grammatical	The couple built a house of their own.
35	Ungrammatical	The couple yearned a house of their own.
36	Grammatical	He shouldn't fake his age to get the discount.
36	Ungrammatical	He shouldn't lie his age to get the discount.
37	Grammatical	The athlete achieved his goal of getting a medal.
37	Ungrammatical	The athlete persisted his goal of getting a medal.
38	Grammatical	The mayor might raise the price of the tickets.
38	Ungrammatical	The mayor might inquire the price of the tickets.
39	Grammatical	The landlord violated his privacy by mistake.
39	Ungrammatical	The landlord intruded his privacy by mistake.
40	Grammatical	The candidate introduced his policy on education.
40	Ungrammatical	The candidate alluded his policy on education.
41	Grammatical	The project's success requires their contributions.
41	Ungrammatical	The project's success depends their contributions.
42	Grammatical	The chef adopted the suggestion of adding honey.
42	Ungrammatical	The chef hesitated the suggestion of adding honey.
43	Grammatical	The professor sponsored the development of the program.
43	Ungrammatical	The professor lectured the development of the program.
44	Grammatical	The biologist conducted his experiment on insects.
44	Ungrammatical	The biologist persevered his experiment on insects.
45	Grammatical	The people killed the tyrant in the end.
45	Ungrammatical	The people rebelled the tyrant in the end.
46	Grammatical	The birds inhabited the south of the island.
46	Ungrammatical	The birds migrated the south of the island.
47	Grammatical	The tourists admired the view along the seashore.

47	Ungrammatical	The tourists marveled the view along the seashore.
48	Grammatical	The vine covered the fence of the farm.
48	Ungrammatical	The vine creeped the fence of the farm.
49	Grammatical	Lily collected the leaves for decoration.
49	Ungrammatical	Lily disposed the leaves on the grass.
50	Grammatical	My uncle visited the city in his thirties.
50	Ungrammatical	My uncle dwelled the city in his thirties.
51	Grammatical	The ladies fostered the pet happily.
51	Ungrammatical	The ladies chattered the pet happily.
52	Grammatical	Ghosts haunted that castle on the hilltop.
52	Ungrammatical	Ghosts existed that castle on the hilltop.
53	Grammatical	Ruby has maintained a balance between motherhood and work.
53	Ungrammatical	Ruby has strived a balance between motherhood and work.
54	Grammatical	The strangers invaded his garden by the river.
54	Ungrammatical	The strangers proceeded his garden by the river.
55	Grammatical	The results confirmed my hypothesis so far.
55	Ungrammatical	The results disagreed my hypothesis so far.
56	Grammatical	His remark revealed the truth in a subtle way.
56	Ungrammatical	His remark diverged the truth in a subtle way.
57	Grammatical	The master hired a servant yesterday.
57	Ungrammatical	The master peered a servant questioningly.
58	Grammatical	The fire destroyed the forest last month.
58	Ungrammatical	The fire raged the forest last month.
59	Grammatical	Hazel did not welcome any criticism about the show.
59	Ungrammatical	Hazel did not respond any criticism about the show.
60	Grammatical	The crowd interrupted the speaker on the podium.
60	Ungrammatical	The crowd hissed the speaker on the podium.
Phrase structure		
1	Standard	The scientist scrutinized Max's proof of the theorem.
1	Anomaly	The scientist scrutinized Max's of proof the theorem.
2	Standard	Hanna recalled Bruce's warning about the rain.
2	Anomaly	Hanna recalled Bruce's about warning the rain.
3	Standard	The man donated Larry's painting of the ocean.
3	Anomaly	The man donated Larry's of painting the ocean.
4	Standard	Tyler purchased Kyle's gift at the store.
4	Anomaly	Tyler purchased Kyle's at gift the store.
5	Standard	Angela used Karen's fork for vegetables.
5	Anomaly	Angela used Karen's for fork vegetables.
6	Standard	The fiction aroused Olivia's interest in dinosaurs.
6	Anomaly	The fiction aroused Olivia's in interest dinosaurs.
7	Standard	Simon threw Kate's umbrella on the sofa.

7	Anomaly	Simon threw Kate's on umbrella the sofa.
8	Standard	The listeners discussed Frank's speech on migrants.
8	Anomaly	The listeners discussed Frank's on speech migrants.
9	Standard	John discovered Bob's pictures of the suspect.
9	Anomaly	John discovered Bob's of pictures the suspect.
10	Standard	The artist despised Nina's sketch of the landscape.
10	Anomaly	The artist despised Nina's of sketch the landscape.
11	Standard	Anthony remembered Monica's slogans about peace.
11	Anomaly	Anthony remembered Monica's about slogans peace.
12	Standard	The women overlooked John's complaints about the noise.
12	Anomaly	The women overlooked John's about complaints the noise.
13	Standard	Helen decorated Alice's treehouse in the summer.
13	Anomaly	Helen decorated Alice's in treehouse the summer.
14	Standard	The journal published Harry's paper about drugs.
14	Anomaly	The journal published Harry's about paper drugs.
15	Standard	The observers followed Lauren's guide on birds.
15	Anomaly	The observers followed Lauren's on guide birds.
16	Standard	The lady sold Mary's portrait of her grandfather.
16	Anomaly	The lady sold Mary's of portrait her grandfather.
17	Standard	Doris read Scott's novel about magic.
17	Anomaly	Doris read Scott's about novel magic.
18	Standard	Winston retrieved Stephen's list of supplies.
18	Anomaly	Winston retrieved Stephen's of list supplies.
19	Standard	The chorus sang Lisa's songs about freedom.
19	Anomaly	The chorus sang Lisa's about songs freedom.
20	Standard	Nate tuned Dylan's piano for the concert.
20	Anomaly	Nate tuned Dylan's for piano the concert.
21	Standard	The instructor challenged Alan's poem about the moon.
21	Anomaly	The instructor challenged Alan's about poem the moon.
22	Standard	The firm stole Mike's ideas about marketing.
22	Anomaly	The firm stole Mike's about ideas marketing.
23	Standard	The citizens disliked Fred's jokes about the Prince.
23	Anomaly	The citizens disliked Fred's about jokes the Prince.
24	Standard	Martha played Robert's movie about dolphins.
24	Anomaly	Martha played Robert's about movie dolphins.
25	Standard	Brian opened Gary's box in the closet.
25	Anomaly	Brian opened Gary's in box the closet.
26	Standard	The staff cancelled Andrew's presentation on solar energy.
26	Anomaly	The staff cancelled Andrew's on presentation solar energy.
27	Standard	Jack wanted Daisy's instructions on feeding cats.
27	Anomaly	Jack wanted Daisy's on instructions feeding cats.

28	Standard	The newspaper quoted Ian's depiction of the accident.
28	Anomaly	The newspaper quoted Ian's of depiction the accident.
29	Standard	The agency rejected Maggie's application for the position.
29	Anomaly	The agency rejected Maggie's for application the position.
30	Standard	The housekeeper placed Erin's pillow on the bed.
30	Anomaly	The housekeeper placed Erin's on pillow the bed.
31	Standard	Joyce questioned Colin's forecast of the weather.
31	Anomaly	Joyce questioned Colin's of forecast the weather.
32	Standard	The network broadcast Kevin's findings about planets.
32	Anomaly	The network broadcast Kevin's about findings planets.
33	Standard	The team believed Jim's statement about the past.
33	Anomaly	The team believed Jim's about statement the past.
34	Standard	Gloria located Carol's notebook behind the bookcase.
34	Anomaly	Gloria located Carol's behind notebook the bookcase.
35	Standard	Emily copied Terry's cartoons about animals.
35	Anomaly	Emily copied Terry's about cartoons animals.
36	Standard	The children ate Eddie's chocolate in the bag.
36	Anomaly	The children ate Eddie's in chocolate the bag.
37	Standard	Wendy saw Jean's photos of her friends.
37	Anomaly	Wendy saw Jean's of photos her friends.
38	Standard	The widow needed Sammy's advice about taxes.
38	Anomaly	The widow needed Sammy's about advice taxes.
39	Standard	The judge skimmed Lucy's article about crime.
39	Anomaly	The judge skimmed Lucy's about article crime.
40	Standard	The boys heard Zoe's stories about Africa.
40	Anomaly	The boys heard Zoe's about stories Africa.
41	Standard	Jeff requested Julia's help on the project.
41	Anomaly	Jeff requested Julia's on help the project.
42	Standard	The chemist cited Howard's formulas about reactions.
42	Anomaly	The chemist cited Howard's about formulas reactions.
43	Standard	Luke surveyed consumers' opinions on plastic bags.
43	Anomaly	Luke surveyed consumers' on opinions plastic bags.
44	Standard	The police circulated Ruth's description of the thief.
44	Anomaly	The police circulated Ruth's of description the thief.
45	Standard	Morgan shredded Carter's documents about military secrets.
45	Anomaly	Morgan shredded Carter's about documents military secrets.
46	Standard	Oliver felt Jane's fear of heights.
46	Anomaly	Oliver felt Jane's of fear heights.
47	Standard	George lost Daniel's textbook on engineering.
47	Anomaly	George lost Daniel's on textbook engineering.
48	Standard	The policeman submitted Peter's report of the case.

48	Anomaly	The policeman submitted Peter's of report the case.
49	Standard	Mina connected Amy's keyboard to the computer.
49	Anomaly	Mina connected Amy's to keyboard the computer.
50	Standard	Anne resented Tom's remarks on her looks.
50	Anomaly	Anne resented Tom's on remarks her looks.
51	Standard	The reader analyzed Bill's review of the play.
51	Anomaly	The reader analyzed Bill's of review the play.
52	Standard	Sarah dropped Leo's mug on the floor.
52	Anomaly	Sarah dropped Leo's on mug the floor.
53	Standard	The gardener watered Maria's roses in the yard.
53	Anomaly	The gardener watered Maria's in roses the yard.
54	Standard	The guard received Sue's note about the ransom.
54	Anomaly	The guard received Sue's about note the ransom.
55	Standard	Jill enjoyed Richard's films about love.
55	Anomaly	Jill enjoyed Richard's about films love.
56	Standard	Ellen joined Roger's protest for change.
56	Anomaly	Ellen joined Roger's for protest change.
57	Standard	Alex cleaned Tony's container for the crabs.
57	Anomaly	Alex cleaned Tony's for container the crabs.
58	Standard	The carpenter fixed Zach's stove in the kitchen.
58	Anomaly	The carpenter fixed Zach's in stove the kitchen.
59	Standard	Bella drank Luke's milk in the fridge.
59	Anomaly	Bella drank Luke's in milk the fridge.
60	Standard	Eric translated Ted's books about America.
60	Anomaly	Eric translated Ted's about books America.
Filler		
1	Grammatical	The singer sneezed during the concert.
2	Grammatical	The waitress smiled at the gentleman.
3	Grammatical	That patient struggled with the recovery process.
4	Grammatical	The musician daydreamed about getting the prize.
5	Grammatical	The jogger fainted in the heat.
6	Grammatical	The apples rotted in the orchard.
7	Grammatical	Justice will certainly prevail over injustice.
8	Grammatical	The technician might retire from the company.
9	Grammatical	The intern snored at the front desk.
10	Grammatical	That plant can thrive in the pond.
11	Grammatical	The assistant should apologize for the mistakes.
12	Grammatical	The audience flocked to the concert.
13	Grammatical	The two nations cooperate on the issue.
14	Grammatical	These flowers will bloom throughout the summer.
15	Grammatical	Norbert should not flirt with his secretary.

16	Grammatical	The farmers sweat under the sun.
17	Grammatical	Peggy has been coughing since midnight.
18	Grammatical	The actor swims in the pool.
19	Grammatical	The Queen prayed for her health.
20	Grammatical	Her baby cried in the cradle.
21	Grammatical	Rachel may travel to the North.
22	Grammatical	The teenager knelt beside his brother.
23	Grammatical	Aaron should not frown at his girlfriend.
24	Grammatical	Grandpa used to fish in the park.
25	Grammatical	The agent winked at my roommate.
26	Grammatical	The girls were jumping on the sofa.
27	Grammatical	The doctor nodded for Cindy to come in.
28	Grammatical	The beautiful necklace belonged to my daughter.
29	Grammatical	The passenger shouted at the driver.
30	Grammatical	Dozens of foreigners died in the earthquake.
1	Ungrammatical	The nanny soothed in the playground.
2	Ungrammatical	We have to preserve for future generations.
3	Ungrammatical	The princess should avoid in social media.
4	Ungrammatical	The farm produced for the community.
5	Ungrammatical	The seller might overstate during the meeting.
6	Ungrammatical	The babysitter must receive for her work.
7	Ungrammatical	A resident repaired for the neighbors.
8	Ungrammatical	The lawyer fulfilled for his client.
9	Ungrammatical	The spy tentatively installed on that computer.
10	Ungrammatical	The instructor abandoned in the afternoon.
11	Ungrammatical	The reporter deliberately humiliated on the spot.
12	Ungrammatical	Nancy's injury might ruin in her life.
13	Ungrammatical	The surgeon should sharpen for next Monday.
14	Ungrammatical	The students displayed in the exhibition.
15	Ungrammatical	The governor will impose by next week.
16	Ungrammatical	Their hunter detected in the woods.
17	Ungrammatical	The dealer shouldn't betray in this case.
18	Ungrammatical	The scholar cannot tolerate on the website.
19	Ungrammatical	The banker will evaluate on his team.
20	Ungrammatical	Ben really cherished in the past.
21	Ungrammatical	The champion finally defeated in the contest.
22	Ungrammatical	His aunt blamed before her leaving.
23	Ungrammatical	The baseball player bought for his teammate.
24	Ungrammatical	The cook should put on the shelf.
25	Ungrammatical	The marketing department promoted for next season.
26	Ungrammatical	Her husband has betrayed for several years.

27	Ungrammatical	The kids insulted at the station.
28	Ungrammatical	The writer inserted into the passage.
29	Ungrammatical	The bartender wiped from that table.
30	Ungrammatical	The sailor might injure in that storm.

Bibliography

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological review*, *113*(4), 787.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., ... & Schlesewsky, M. (2011). Think globally: Cross-linguistic variation in electrophysiological activity during sentence comprehension. *Brain and language*, *117*(3), 133-152.

Bourguignon, N., Drury, J. E., Valois, D., & Steinhauer, K. (2012). Decomposing animacy reversals between agents and experiencers: an ERP study. Brain and language, 122(3), 179-189.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135-149.

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.

Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service?. *Bilingualism: Language and cognition*, *13*(3), 359-371.

Chomsky, N. (1981). Lectures on government and binding, foris, dordrecht. *ChomskyLectures on Government and Binding1981*.

Chomsky, N. (1995). Categories and transformations. *The minimalist program*, *219*, 394.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain research*, *1506*, 76-93.

Chow, W. Y., Lau, E., Wang, S., & Phillips, C. (2018). Wait a second! Delayed impact of argument roles on on-line verb prediction. *Language, Cognition and Neuroscience, 33*(7)*,* 1-26.

Chow, W. Y., Momma, S., Smith, C., Lau, E., & Phillips, C. (2016). Prediction as memory retrieval: timing and mechanisms. *Language, Cognition and Neuroscience, 31*(5), 617-627.

Chow, W. Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience, 31*(5), 577-596.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, *42*(4), 368-407.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics*, *27*(1), 3-42.

Craig, C. (1977). *The structure of Jaceltec*. Austin/London: Texas Press.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Demiral, Ş. B., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2008). On the universality of language comprehension strategies: Evidence from Turkish. *Cognition*, *106*(1), 484-500.

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*(2), 166-180.

Dussias, P. E., & Scaltz, T. R. C. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta psychologica*, *128*(3), 501-513.

Dussias, P. E., Marful, A., Gerfen, C., & Molina, M. T. B. (2010). Usage frequencies of complement-taking verbs in Spanish and English: Data from Spanish monolinguals and Spanish—English bilinguals. *Behavior Research Methods*, *42*(4), 1004-1011.

Ehrenhofer, L. (2018). *Argument roles in adult and child comprehension*. [Unpublished doctoral dissertation]. University of Maryland.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491-505.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.

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*(3), 149-161.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain research*, *1146*, 75-84.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science*, *11*(1), 11-15.

Flett, S., Branigan, H. P., & Pickering, M. J. (2013). Are non-native structural preferences affected by native language preferences?. *Bilingualism: Language and Cognition*, *16*(4), 751-760.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465-495.

Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C., & Costa, A. (2015). Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. *Neuropsychologia*, *75*, 291-303.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, *6*(2), 78-84.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and language*, *43*(3), 476-507.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1219.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58-93.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. University of Chicago Press.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112.

Gries, S. T., & Wulff, S. (2005). Do foreign language learners also have constructions?. *Annual Review of Cognitive Linguistics*, *3*(1), 182-200.

Guo, J., Guo, T., Yan, Y., Jiang, N., & Peng, D. (2009). ERP evidence for different strategies employed by native speakers and L2 learners in sentence processing. *Journal of Neurolinguistics*, *22*(2), 123-134.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531-1549.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes*, *8*(4), 439-483.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive brain research*, *19*(1), 59-73.

Huang, C. T. J. (1998). *Logical relations in Chinese and the theory of grammar*. Taylor & Francis.

Jackendoff, R., & Jackendoff, R. S. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, USA.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language learning*, *57*(1), 1-33.

Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: An electrophysiological differentiation. *Journal of cognitive neuroscience*, *15*(1), 98-110.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

Karatas, N. B. (2019). *The comparison of L1 and L2 case processing: ERP evidence from Turkish. [Unpublished doctoral dissertation]. University of Maryland.*

Kielar, A., Meltzer-Asscher, A., & Thompson, C. K. (2012). Electrophysiological responses to argument structure violations in healthy adults and individuals with agrammatic aphasia. *Neuropsychologia*, *50*(14), 3320-3337.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*(2), 205-225.

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, *13*(3), 373-381.

Kuo, G. (2006, January). *Processing Chinese resultative compounds: a study on its morphological headedness.* Paper presented at the 4th Conference of the European Association of Chinese Linguistics (EACL-4), Budapest, Hungary.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience, 31*(5), 602-616.

Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience, 22*(12), 2685-2701.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. *Brain and Language, 100*(3), 223-237.

Kuperberg, G. R., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of cognitive neuroscience*, *23*(5), 1230-1246.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 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. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In Bar, M. (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). Oxford University Press.

Laszlo, S., & Federmeier, K. D. (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*(3), 326-338.

Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920.

Lewis, S., & Phillips, C. (2015). Aligning grammatical theories and language processing models. *Journal of Psycholinguistic Research*, *44*(1), 27-46.

Li, X. Q., Zhao, H. Y., Zheng, Y. Y., & Yang, Y. F. (2015). Two-stage interaction between word order and noun animacy during online thematic processing of sentences in Mandarin Chinese. *Language, Cognition and Neuroscience*, *30*(5), 555-573.

Li, Y. (1990). On VV compounds in Chinese. *Natural language & linguistic theory, 8*(2), 177-207.

Liao, C-H & Lau E. (2020). Enough time to get results? An ERP investigation of prediction with complex events. *Language, Cognition, and Neuroscience.* 1-21.

Liao, C-H., & Chan, S-H. (2016). Direction matters: Event-related brain potentials reflect extra processing costs in switching from the dominant to the less dominant language. *Journal of Neurolinguistics, 40*, 79-97.

Lin C-J., & Jäger, L. (2014, March). *Reading resultative verb compounds in Chinese sentences: An eye-tracking study.* Poster presented at the 2nd East Asian Psycholinguistics Colloquium (EAPC2), Chicago, IL. University of Chicago.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological review*, *101*(4), 676.

Minkoff, S. (2000). Animacy hierarchies and sentence processing. In A. Carnie & E. Guilfoyle (Eds.), The syntax of verb initial languages (pp. 201–212). Oxford: Oxford University Press.

Momma, S., Sakai, H., & Phillips, C. (2015, March). *Give me several hundred more milliseconds: the temporal dynamics of verb prediction.* Paper presented at the 28th annual CUNY Conference on Human Sentence Processing, Los Angeles, CA.

Myers, J., (2006). Processing Chinese compounds: a survey of the literature. In G., Libben & G., Jarema (Eds.), *Representation and Processing of Compound Words* (pp.169-196). Oxford University Press, Oxford.

Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of cognitive Neuroscience*, *3*(2), 151-165.

Oh, S. J., Sung, J. E., Sim, H. S., Oh, S. J., Sung, J. E., & Sim, H. S. (2016). Age-related differences in animacy effects as a function of word-order canonicity in a verb-final language: evidence from ERP. *Communication Sciences & Disorders*, *21*(4), 653-667.

Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563-588.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, *31*(6), 785-806.

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*(4), 786.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*, 271-308.

Osterhout, L., McLaughlin, J., Pitkänen, I., Frenck‐Mestre, C., & Molinaro, N. (2006). Novice learners, longitudinal designs, and event‐related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, *56*, 199-230.

Øvrelid, L. (2004). Disambiguation of syntactic functions in Norwegian: modeling variation in word order interpretations conditioned by animacy and definiteness. In *Proceedings of the 20th Scandinavian Conference of Linguistics* (pp. 1-17). Helsinki: University of Helsinki.

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.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of cognitive neuroscience*, *10*(6), 717-733.

Perpiñán, S. (2020). Wh-Movement, Islands, and Resumption in L1 and L2 Spanish: Is (Un) Grammaticality the Relevant Property?. *Frontiers in Psychology*, *11*.

Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., & Schlesewsky, M. (2008). The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language*, *105*(2), 112-133.

Polinsky, M. (2018). *Heritage languages and their speakers* (Vol. 159). Cambridge University Press.

Schegloff, E. (2000). Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, *29*(1), 1-63.

Shen, Y., & Mochizuki, K. (2010, May). Inheritance of argument structure and compounding constraints of resultative compound verbs in Chinese and Japanese. In L.E. Clemens & C-M. L. Liu (Eds.), *Proceedings of the 22 North American Conference on Chinese Linguistics (NACCL-22) and the 18th the International Association of Chinese Linguistics (IACL-18)* (pp. 341-355). Harvard University, Cambridge, MA.

Snider, N., & Zaenen, A. (2006). Animacy and syntactic structure: Fronted NPs in English. In M. Butt, M. Dalrymple, & T. H. King (Eds.), Intelligent linguistic architectures: Variations on themes by Ronald M. Kaplan. Stanford: CSLI Publications.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and language*, *120*(2), 135-162.

Sturt, P. (2007). Semantic re-interpretation and garden path recovery. *Cognition*, *105*(2), 477-488.

Su, J. J., Molinaro, N., Gillon-Dowens, M., Tsai, P. S., Wu, D. H., & Carreiras, M. (2016). When "he" can also be "she": An ERP study of reflexive pronoun resolution in written mandarin Chinese. *Frontiers in Psychology, 7*, 151.

Tai, J. H. (1984). Verbs and times in Chinese: Vendler's four categories. *Parasession on Lexical Semantics, 20*, 289-296.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*(2), 367-382.

Taylor, W. L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism Bulletin, 30*(4), 415-433.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530-12535.

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392.

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and cognition*, *4*(2), 105-122.

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*(7), 1181-1197.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176-190.

Wang, Y., Jiao, Q., & Pang, Y. (1987). *A dictionary of Chinese verb-resultative complement phrases.* Beijing: Beijing Language Institute Press.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of cognitive neuroscience*, *8*(3), 231-256.

Weckerly, J., & Kutas, M. (1999). An electrophysiological analysis of animacy effects in the processing of object relative sentences. *Psychophysiology*, *36*(5), 559-570.

Williams, A. (2014). Causal VVs in Mandarin. In C. T. J. Huang, Y-H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (pp. 311-341). Chichester: Wiley Blackwell.

Williams, A. (2015). *Arguments in syntax and semantics*. Cambridge University Press.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., & Kuperberg, G. (2014). The difference between "giving a rose" and "giving a kiss": Sustained neural activity to the light verb construction. *Journal of Memory and Language, 73*, 31-42.

Yuan, B., & Zhao, Y. (2011). Asymmetric syntactic and thematic reconfigurations in English speakers' L2 Chinese resultative compound constructions. *International Journal of Bilingualism, 15*(1), 38-55.

Zhang, B., & Peng, D. (1992). Decomposed storage in the Chinese lexicon. In G.E. Stelmach & P.A. Vroon (Eds.), *Advances in Psychology* (Vol. 90, pp. 131-149). North-Holland.

Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia*, *48*(6), 1551-1562.