Thesis

Cues from structure and meaning–insights from Mandarin sentence processing

An electrophysiological study of argument structure and adjective placement

Max Wolpert

Integrated Program in Neuroscience

McGill University, Montreal

January 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy in Neuroscience

Table of Contents

List of Figures and Tables

Abstract

Mandarin has a combination of features that are relatively unusual across human languages. In this thesis, we took advantage of these features to study sentence processing in native and non-native speakers. To study processing in real time, we used electroencephalography (EEG) data to analyze event-related brain potentials (ERPs). Our behavioral and ERP results together provide new insight into processing of diverse sentence types, as presented in two manuscripts.

In Manuscript I, we considered Mandarin verb-final transitive sentences, which allow both patient-first and agent-first word orders. We asked participants to read sentences and identify the agent, while recording their responses, reaction times, and EEG. Our analysis showed that coverbs are the strongest cue for comprehension and that word order did not impact interpretation in the absence of other cues. We also found that role reversal sentences, where an implausible noun is expressed as the agent of a verb, elicited N400 effects, in line with sentence processing models that predict crosslinguistic differences in core parsing mechanisms.

In Manuscript II, we studied word order in native and non-native processing of adjective phrases. Mandarin adjectives typically precede the nouns they modify but may also occur directly following a noun. Native speakers showed an N400-P600 response for both English and Mandarin adjective placement violations. In contrast to English, Mandarin results showed an additional N400 effect that may be related to the lexical status of adjectives. Non-native ERP results showed an N400 effect with no P600, in line with predictions for intermediate-proficiency second language learners. We further explored individual differences in ERP patterns, showing that participants exhibited a full range of responses that was not reflected in the grand average.

These two manuscripts contribute to our understanding of the diversity of Mandarin sentence structures and their processing. Manuscript I is the first use of an ERP paradigm with a forced-choice agent assignment task and reveals an intricate interplay among competing cues for argument structure processing. Manuscript II replicates prior findings from English, but also presents new data that inform our understanding of Mandarin adjective phrase structure and the extent of individual variability in ERP patterns.

Résumé

Le mandarin présente une combinaison de caractéristiques relativement peu communes parmi les langues humaines. Dans cette thèse, nous avons exploité ces propriétés pour étudier le traitement de phrases chez les locuteurs natifs et non-natifs du mandarin. Afin d'évaluer le traitement en temps réel, nous avons analysé les potentiels évoqués (ERPs) issus des enregistrements électroencéphalographiques (EEG). Nos résultats comportementaux et des ERPs permettent d'obtenir une nouvelle compréhension du traitement de différents types de phrases, tel que présenté dans deux manuscrits.

Dans le Manuscrit I, nous avons considéré des phrases transitives en mandarin ayant leur verbe en position finale, et dont le premier nom pouvait avoir soit le rôle argumental d'Agent, soit celui de Patient. Nous avons demandé aux participants de lire ces phrases et d'identifier l'Agent, alors que nous enregistrions leurs réponses, temps de réaction et EEG. Nos analyses montrent que les coverbes sont l'indice le plus important pour la compréhension de phrases en mandarin, et que l'ordre des mots n'influence pas l'interprétation en l'absence d'autres indices.

Dans le Manuscrit II, nous avons étudié l'effet de l'ordre des mots sur le traitement natif et non-natif de constructions adjectivales. En mandarin, les adjectifs précèdent habituellement les noms qu'ils modifient, mais peuvent également parfois les suivre immédiatement. Les placements agrammaticaux d'adjectifs ont élicité une réponse N400-P600 chez les locuteurs natifs, tant en anglais qu'en mandarin. À la différence de l'anglais, les résultats en mandarin ont montré un effet N400 additionnel qui peut être associé au statut lexical des adjectifs. Les ERPs des locuteurs non-natifs présentaient un effet N400 sans P600. Ces résultats correspondent aux prédictions pour des apprenants d'une langue seconde de niveau intermédiaire. Nous avons aussi exploré les différences individuelles dans les résultats ERPs, montrant que les participants présentaient une gamme complète de réponses qui n'était pas reflétée dans la moyenne générale des ERPs.

Ces deux manuscrits permettent de mieux comprendre la diversité des structures de phrases en mandarin ainsi que leur traitement. Le Manuscript I rapporte la première utilisation d'un paradigme ERP avec une tâche d'identification obligatoire du nom ayant le rôle d'Agent. Il révèle une interaction complexe entre des indices concurrents dans le traitement de la structure argumentale de la phrase. Le Manuscrit II reproduit des résultats antérieurs obtenus en anglais mais présente également de nouvelles données qui éclairent notre compréhension de la structure

des constructions adjectivales en mandarin, ainsi que de l’étendue de la variabilité inter-individuelle dans les données ERP.

Acknowledgments

Firstly, I thank my supervisors, Professor Karsten Steinhauer and Professor Shari Baum. Karsten, I feel incredibly lucky that I managed to convince you to take me on as a student, and I have benefitted academically and personally from your forbearance alongside your high standards for research. Shari, you have been a steady force alongside me from the very beginning of the PhD, always the first to reply and to volunteer your time to give me guidance and feedback. I am fortunate to have trained in a supportive research environment where I could implement my own research project from start to finish.

My next thanks go to Professor Hui Zhang at Nanjing Normal University. 张老师，因为您我在南师大的时候过得又充实又快乐。 I am also grateful for the support and input from my advisory committee members, Professor Lydia White and Professor Debra Titone. Lydia provided the theoretical linguistics expertise that I was eager to have for my project, and I hope that I can continue to do work in the future that linguists find compelling (or at least mildly interesting). Debra brought in experimental know-how and greater depth to my consideration of bilingualism, and never failed to give encouragement.

I owe enormous thanks to Dr. Annie Gilbert, who was my first research mentor as a PhD student. Annie always included me in her research process and shared her knowledge and even introduced me to poutine from Chez Ashton. I owe my entering the field of language sciences to my master's supervisors, Dr. Sendy Caffarra and Dr. Simona Mancini. Not only did they bring me in on an awesome experience with Basque at the interface of pragmatics and morphosyntax, but also kept working with me after the PhD to gently guide me to publish my first first-authored article. I also need to thank the Integrated Program in Neuroscience and the Rotation Program, which was an important steppingstone for me to discover my direction in research.

There is no way I could have gotten to where I am without the Lab Maukis: Alexandre Herbay, Lauren Fromont, and Nastia Glushko. I cherish the memories of shared avocado toasts, road trips to chalets, and dancing around a fire to banish our inner demons. I aspire to be as smart and capable as you all when I grow up. Alex, I could not have made it on my own these last two years of pandemic without you in the lab, and I wish I could carry you around to my future workplaces to hear your laugh and bother you with ERPscope and making figures and presentations beautiful. Lauren, you are the coolest person I know, and I hope I can follow your example and keep social issues at the forefront alongside science. Nastia, as my savior during the pandemic and my go-to partner in crime, thank you for being so much more than my most valuable colleague during this journey. Thanks to you and Polo for giving me a home and my own Montreal family (plus Lego and Ratou of course). If I manage to bring half your ambition, technical know-how, and insight to my future work, I will consider myself wildly successful.

I also owe my making it through the PhD to Alberto Osa García, who somehow despite my shortcomings has always been a steadfast friend and support. De alguna manera, casi diez años después de un recital de poesía portuguesa, los dos nos encontramos en Canadá haciendo un doctorado. Gracias por haberme llevado un amor por el lenguaje, el arte y la vida y por haber soportado siempre mis quejas. Aspiro a tener tu compasión y tu energía por participar en todo que haya en la vida.

Special thanks go to Jiarui Ao, who worked in our lab for three years and made valuable contributions to the work in this thesis. I am humbled that my PhD project has been a part of your early research training. I also thank everyone at Nanjing Normal University who gave input into my experiment design and helped me settle in, including 杨艳琴, 卞京, 纪晓丽, 戴好运, and 刘艳. I also thank John Urban at the Hopkins-Nanjing Center for being enormously generous with his time and energy to recruit participants for me.

Thank you to my McGill classmates and community who have contributed directly and indirectly to my work. Shuyi Zhang, Yufang Ruan, and Xiaoming Jiang all gave valuable input in the early stages of my experimental design. Neelima Vaddadi, I am grateful we got to socialize at the beginning of our PhD journey in PN1 and in GSAN, and now again at the end when we are about to move on to wonderful things! Thank you for being a good thesis buddy. Marcel Farrés Franch, you introduced me to castells and have always expanded my knowledge and my community. Anastasia Sares, you were the first McGill neuroscience student that I met! I admire you immensely and thanks to you I have rekindled my love for fantasy and the fantastic. Dr. Kaija Sander, we have found a common soulmate/deity/idol in Lizzo, and under her grace we shall continue to have wonderful times together.

I also express my heartfelt thanks to those in McGill who have given me valuable supports along the thesis journey. Jos Porter has always listened and offered insight into greater patterns in my thinking. Isabella Scurfield is a superhero in disguise and has helped on multiple occasions to organize my mind and life. Mariève Isabel taught me that writing is a community effort and that there is room for grace in a writing lifestyle. I could not have gotten to this point without the Three Months to Advance Your Thesis program. Donetta Hines has been another invaluable writing support, and thanks to Donetta I now hold an unbreakable association between Spanish-language poetry and writing.

In the greater Montreal community, I need to also thank several people. Océane Apffel Font, you are full of whimsy and care so much about the world. Orlando Hernandez and Denis Croft, you were my first friends in Montreal and have always made me feel I have a community and a home in this city. Claire David, you have such energy and passion and principled thinking in what you do in life. You are one of my role models as a scientist and an educator. Waleed Al-Ahmad, you have helped give me new and deeper perspectives of myself and others, and my work with you has been an integral part of my thesis journey.

Despite my active efforts at vagabonding away from the US, there are several American compatriots that have been a part of my journey here. Jessica Liang and Oishee Sen, I am so impressed at our being functional adults and still staying connected. Rhiannon Anderson, we had a good streak going of being together for New Year's and annoying everyone around us by discussing language all day long. I hope we can find another New Year's to revel in these experiences. Shazeda Ahmed, I am half convinced that one day I will discover you are a bionic cyborg because you are productive and driven beyond normal human capabilities. Thank you for being a source for deep discussions about China, America, life, and love, and I am certain our post-PhD lives hold many wonders ahead. And my most amazing seniors! Volina Noriega, Brieanna Frye, Leanna Frye, and Anais Quetell (honorary senior), my pandemic confinement experience was better for having youse to talk to.

I will finish by thanking Peijian Sun and my family. Peijian, thank you for your faith in me that has transcended time and distance. You are my source of insufferable optimism, and I am humbled to be an important part of your life. I could not have finished this thesis without your direct intervention and gentle but consistent encouragement. Mom and dad, thank you for instilling in me a love of learning. I am grateful we still get to share moments of our experiences and daily life. Sammy, you have been my number one dialed number through the PhD and have always kept me company and reminded me to look after myself. To many more good times ahead!

Acknowledgment of Funding

This thesis work received essential support from multiple sources. The research activity was supported by grants from the Natural Sciences and Engineering Council (NSERC) awarded to Professor Karsten Steinhauer. For testing in Nanjing, lab space and equipment were provided by Professor Hui Zhang and Nanjing Normal University. As a PhD trainee, I received generous support from the Fonds de Recherche du Quebec - Nature et technologies (FRQNT), the US Fulbright Student Program, the McGill Shuk-Tak Liang Fellowship from the Faculty of Medicine, the Mitacs Globalink Research Award, the Centre for Research on Brain, Language and Music, and the Integrated Program in Neuroscience. My attendance at McGill as an international student was also thanks to a generous differential fee waiver for tuition provided by the Integrated Program in Neuroscience.

Contribution to Original Knowledge

The present thesis describes analysis of event-related potential (ERP) and behavioral data from experiments studying the processing of Mandarin sentences. We approached this work from a crosslinguistic perspective, taking advantage of the relatively unusual typological characteristics of Mandarin to advance the limits of our understanding of human sentence processing. There are several firsts in this thesis work, both in the methodology and the results; alongside these firsts, the thesis critically and productively engages with prior work to contribute meaningfully to the field.

Manuscript I details the first use of a forced-choice agent assignment task in an ERP paradigm. We systematically compared four cues influencing argument structure, including a fully balanced design for animacy including animate and inanimate plausible agents. Unlike virtually all prior studies of native processing, we recruited and tested functionally monolingual Mandarin speakers as participants, thereby minimizing the effect of any other languages on our results. At the behavioral level, our results show an updated profile of cue competition in Mandarin monolinguals that challenges and augments existing understanding. We further documented individual differences in our behavioral results, expanding beyond group-level effects. Our ERP results provide evidence for crosslinguistic differences in argument structure processing. Manuscript I's combination of meaningful behavioral and ERP data captures a more complete profile of argument structure processing than the use of either methodology alone and showcases how behavioral and neural measures can be additively interpreted together.

Manuscript II presents an adjective-noun placement experiment that was applied to Mandarin adjectives for the first time. We tested both native and non-native Mandarin speakers to compare to previous findings for processing of English adjectives. For native speakers, our results showed a similar ERP profile to English, but we also observed an earlier, novel effect that may reflect fundamental differences in the lexical nature of Mandarin and English adjectives. For non-native speakers, we observed an ERP profile more consistent with relatively low-proficiency second language learners. This finding was in contrast to previous results with non-native English speakers, and we propose that Mandarin adjective placement has important differences from English, and the nuances of these differences can prove challenging for second language acquisition. We further demonstrated that while ERP component amplitudes may not always correlate when applying stringent measures for comparison, participants may still show a full range of ERP profiles that is not captured by effects at the group-level.

Together, the two manuscripts provide data for real-time sentence processing for two Mandarin syntactic structures. In relation to English, these structures have key similarities and differences that serve as points of crosslinguistic comparison. Not only do the results of the thesis contribute what has been described above, but also lay the foundation for ongoing and future experiments with bilingualism and first language attrition.

Author Contributions

1.1 Manuscript I

MW and KS designed the experiment with input from SB and HZ. KS provided the funding for participants and HZ provided the lab space and testing materials. MW collected the data, conducted the analysis, and wrote the manuscript. JA participated in coding and analyzing questionnaire and reaction time data and assisted in optimizing the protocol for preprocessing EEG data. All authors gave feedback on analysis and the manuscript.

1.2 Manuscript II

MW and KS designed the experiment with input from SB and HZ. KS provided the funding for participants and HZ provided the lab space and testing materials. MW collected the data, conducted the analysis, and wrote the manuscript. All authors gave feedback on analysis and the manuscript.

General Introduction

A word I hate to use in English is I*. It is a melodramatic word. In Chinese, a language less grammatically strict, one can construct a sentence with an implied subject pronoun and skip that embarrassing* I*, or else replace it with* we*. Living is not an original business.*

Li, Yiyun. (2017). *Dear friend, from my life I write to you in your life*. First edition. New York, Random House. (pp. 36-37).

I was once foolish enough to believe knowledge would clarify, but some things are so gauzed behind layers of syntax and semantics…, that simply knowing … does nothing to reveal it.

Vuong, Ocean. (2019). *On earth we're briefly gorgeous: a novel*. New York, Random House. (pp. 100-101).

This thesis presents two studies on sentence processing in Mandarin Chinese (henceforth Mandarin). As an incremental, real-time task, sentence processing is among the most complex of human cognitive feats (Kara D. Federmeier, Jongman, & Szewczyk, 2020). We produce and comprehend intricate utterances, all while following an internal set of rules that constrain how we combine and modify words (Chomsky, 2006; Hyder, Højlund, Jensen, Østergaard, & Shtyrov, 2020). By studying the underlying mechanisms of sentence processing, we move toward understanding the human brain's cognitive architecture (Lewis, 1999).

Neurophysiological methods like electroencephalography (EEG) and event-related potentials (ERPs) can measure the real-time sequence of electrical events underlying this architecture (Kaan, 2007). While language sciences have classically relied on acceptability judgments, where informants directly evaluate a linguistic item (Myers, 2017; Schütze, 2016), this offline approach can be problematic for quantifying linguistic phenomena (Gibson & Fedorenko, 2010) and may not capture the incremental steps that determine sentence processing (Ferreira & Çokal, 2015). ERPs, however, can measure electrophysiological indices of sentence processing with millisecond precision, and have the additional advantage of having well-documented waveforms (called components) that ideally correspond to discrete domains of language processing (Kaan, 2007; Luck, 2014). While the ERP methodology provides these advantages, offline measures are still valuable sources of empirical evidence (Ferreira & Yang, 2019; Sprouse & Almeida, 2017), and behavior is an important basis for understanding the brain (Niv, 2020). In the present thesis, we combine behavioral and ERP methods to characterize the timecourse and decisions involved in Mandarin sentence processing.

In this introduction, we first detail background and motivation for three areas of sentence processing research to which the present thesis contributes: 1) crosslinguistic variation, 2) argument structure, and 3) bilingualism. For each area, we further detail relevant findings from select EEG studies, focusing on the N400 and the P600 components. We then overview the structure of the thesis and the two manuscripts.

1 Crosslinguistic variation in sentence processing

1.1 Overview

This thesis considers Mandarin sentence processing as a basis for crosslinguistic comparison. Languages vary at all levels, from their phonemes to their sentence structure (Evans & Levinson, 2009). There have been influential efforts to characterize the constraints of this variability, including ideas like Universal Grammar (Chomsky, 1957b) and parameter setting (Chomsky, 2014; Chomsky & Lasnik, 1993). Indeed, at the processing level, human language must conform to the architecture of the human brain and the need for efficient communication (Hahn, Jurafsky, & Futrell, 2020). To understand the cognitive confines and the limits (or lack thereof) of the human language faculty, researchers must compare typologically dissimilar languages (Croft, 2007; Haspelmath, 2007).

Mandarin is an ideal language for making some of these comparisons because it is a "typologically hybrid language" (Chappell, Ming, & Peyraube, 2007). In contrast to Indo-European languages that have been the main subject of language sciences research (Croft, 2007), Mandarin has features that can be challenging to explain with classic linguistic terminology. For instance, despite having virtually no morphological inflection (Packard, 2015), Mandarin has relatively flexible word order (Wu & He, 2015), even though flexible word order is usually limited to highly inflected languages (McFadden, 2003). This feature has led some researchers to suggest that the grammatical notion of 'subject' may not exist in Mandarin, and labels like 'topic' and 'comment' may be more appropriate (LaPolla, 1993; Xu, 2015). To this end, some authors have even labeled Mandarin a "semantics-based" language instead of a "syntax-based" one (Su, 2001). Second, despite it being a mostly right-branching, head-initial language like English, Mandarin has left-branching, head-final relative clauses, which has provoked heated debate over whether there is a universal subject relative clause processing advantage (Gibson & Wu, 2013; Jäger, Chen, Li, Lin, & Vasishth, 2015; for review see Lau & Tanaka, 2021). Third, alongside its lack of inflection, Mandarin syntactic categories are also highly flexible, with many words able to serve as adjective, verb, or noun with no change in form (Huang, Chen, & Shen, 2002).

Some researchers have already taken advantage of Mandarin's features to study the extent of language universals. To determine if Mandarin parsers preferentially parse sentence-initial nouns as subjects, Wang, Schlesewsky, Bickel, & Bornkessel-Schlesewsky (2009) tested sentences where the first noun was the plausible subject or object of the subsequent verb. Results suggested that even though Mandarin permits object-initial word orders, subject-initial word orders are still easier to comprehend. Research on Mandarin sentence processing has also focused on verb-final structures (e.g., Bornkessel-Schlesewsky et al., 2011; Chow & Phillips, 2013; Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008; Wang, Schlesewsky, Philipp, & Bornkessel-Schlesewsky, 2012). Languages with primary verb-final word order account for around 40% of documented languages, according to the World Atlas of Language Structures (Dryer, 2013b). Yet other languages, like Mandarin (Li & Thompson, 1976), employ frequent verb-final sentence structures.

For researchers who want to understand incremental, real-time sentence processing, verb final structures are of special interest. Verbs are often described as the driving force for assigning grammatical functions to sentence elements (e.g., Chomsky, 1993; Druks, 2002). For parsing of verb-final structures, however, it would be disadvantageous for comprehenders to wait until the end of a sentence before computing a syntactic structure. Instead, parsers use information from preceding arguments to make predictions about the upcoming verb (Bornkessel & Schlesewsky, 2006a; Chow, Momma, Smith, Lau, & Phillips, 2016). There is experimental evidence that verb-final structures demand a distinct profile of working memory and predictive parsing for comprehension (Friederici & Frisch, 2000; Levy & Keller, 2013), and comprehension strategies for verb-final sentences have been shown to differ crosslinguistically (Vasishth, Suckow, Lewis, & Kern, 2010). Mandarin's verb-final structures can thus provide valuable data for empirical, crosslinguistic comparison.

1.1a The Competition Model

An influential account of crosslinguistic differences and sentence processing is the Competition Model, a functionalist account of language acquisition and processing (first described by Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; extended by MacWhinney, 2005). The Competition Model is an overarching framework of acquisition, production, and impairment for both first and second language (Li & MacWhinney, 2012). The model considers linguistic evidence from sounds to sentence structure and highlights the shared mechanisms of learning throughout the lifespan where second language learning and bilingualism draw on the same processes as first language acquisition (MacWhinney, 2005). Data informing the Competition Model come from over a dozen languages, and each language is considered to have its own features (MacWhinney, 2022a).

At the core of the Competition Model is competition among "cues"[1], units of sensory information that contribute to an understanding of the environment (Ernst & Bülthoff, 2004); just as visual and auditory cues are combined to perceive where a noise is coming from, so are phoneme-, morpheme-, and phrase-level cues integrated to achieve language processing (Martin, 2016). According to the Competition Model architecture, linguistic cues compete for mapping form to function. These cues are assigned different strengths and validities according to the properties of a particular language, and these cues are acquired through experience (MacWhinney, 2005). For bilinguals, each of their languages has different codes that independently govern cue strength for each respective language, and these codes can experience reverse and forward transfer effects (MacWhinney, 1987). The Competition Model has considered extensive experimental evidence from argument structure processing, as will be detailed in Section 2.1.

1.2 EEG data

EEG studies of sentence processing have considered a variety of languages with diverse typological relationships, from Basque (e.g., Carreiras, Duñabeitia, Vergara, de la Cruz-Pavía, & Laka, 2010) to Japanese (e.g, Nakagome et al., 2001) to American Sign Language (e.g., Gutierrez, Williams, Grosvald, & Corina, 2012) and more. Across the range of these studies, many ERP components have been consistently reported, including the two of most relevance for the present study: the N400 and the P600, classically associated with semantic and syntactic processing, respectively. This syntax-semantics divide between the two components is not so simple, however. Neither of these ERP components is specific to language (e.g., N400 elicited by unexpected odors in Grigor, 1999; Sarfarazi, 1999; P600 elicited by arithmetic equations in Núñez-Peña & Honrubia-Serrano, 2004), but the fact that they can be observed across languages is evidence that there are common cognitive underpinnings to human sentence processing. Before turning to their consideration across different languages, we first briefly overview the history and select interpretations of the N400 and P600 components. For visual reference, these components are depicted in Figure 1.

[1] We note that while the Competition Model as an emergentist framework does not see language as separate from other cognitive functions (MacWhinney, 2022b), the notion of cues in processing has similarities to and differences from ideas of parameters in generative linguistics (MacWhinney, 2004; Newmeyer, 2004). In the context of this thesis, we do not take a position as to whether one theoretical framing is better than the other.

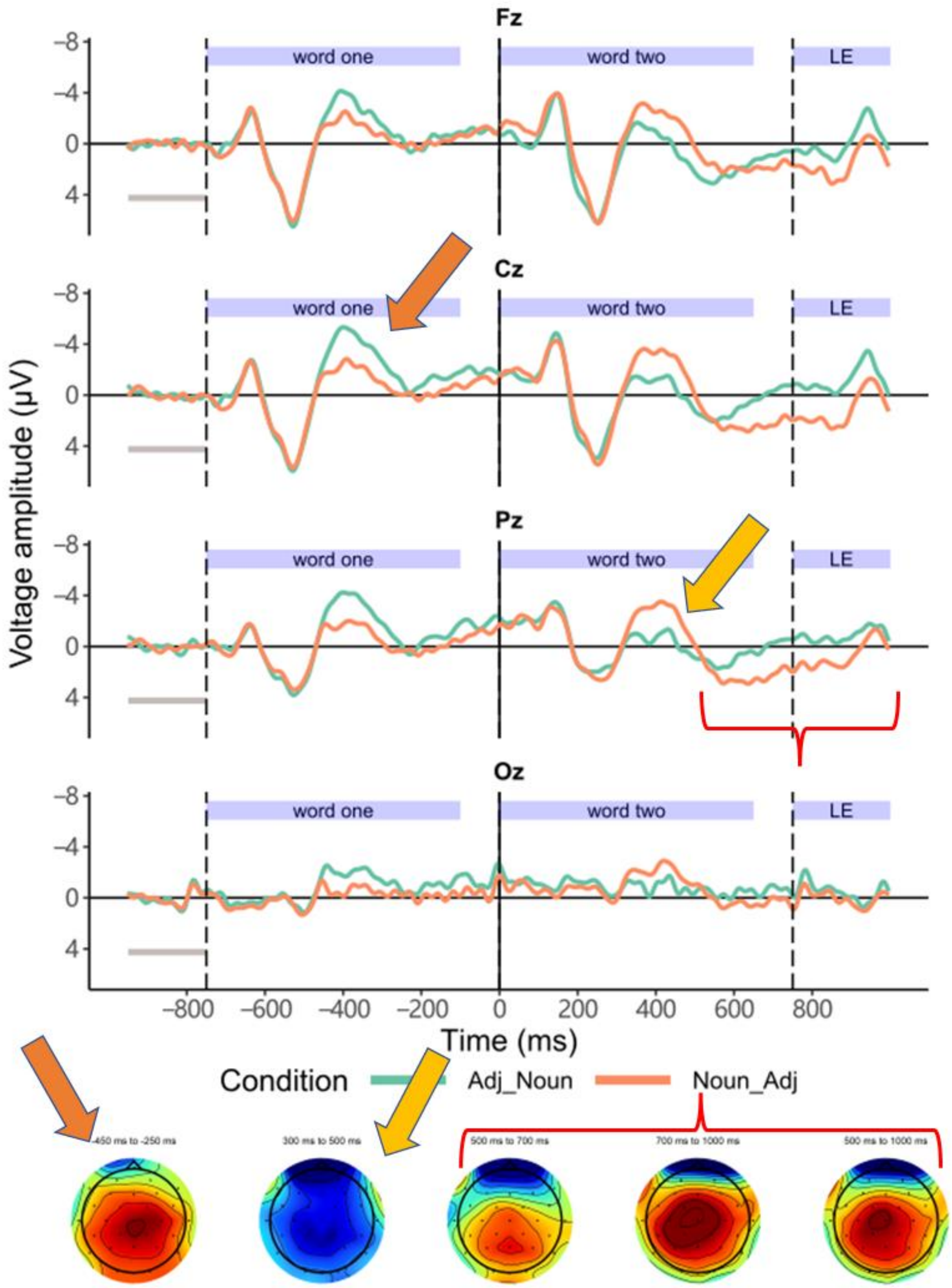

Figure 1. *N400 and P600 effects reproduced from Manuscript II of the present thesis. Arrows indicate two N400 effects and their corresponding scalp maps (note that here the green ERP is subtracted from the orange ERP, so the first N400 appears as a positivity in the scalp map). The bracket represents the P600 effect and its associated scalp maps that show its progression throughout the time window*

The N400 component was first reported over forty years ago in response to semantically anomalous sentences, as in “He spread the warm bread with #socks” (Kutas & Hillyard, 1980).

The N400 component is a negative-going ERP wave peaking around 400 ms elicited by each (content) word in a sentence (Kutas & Federmeier, 2011). First reports characterized the N400 effect as an index of semantic processing, with its amplitude increasing along with semantic processing difficulty. However, the precise functions indexed by the N400 component are still debated. N400 effects have also been reported for morphosyntactic (Guajardo & Wicha, 2014; Weber & Lavric, 2008) and syntactic (Steinhauer, 2014) manipulations, which further complicates the semantic interpretation of the N400. According to some accounts, the N400 is a measure of word retrieval difficulty, and when retrieval of a word's meaning is facilitated by a semantically congruent context, N400 amplitude decreases (Brouwer, Delogu, Venhuizen, & Crocker, 2021; Brouwer, Fitz, & Hoeks, 2012). This retrieval account is closely related to interpretations of the N400 as an index of prediction (DeLong, Urbach, & Kutas, 2005; Lau, Phillips, & Poeppel, 2008; Molinaro, Conrad, Barber, & Carreiras, 2010). Others defend the idea that the N400 is also an index of semantic integration, a process occurring after meaning retrieval and relating meaning to context (Brown & Hagoort, 1993; P. Holcomb, 1993; Nieuwland et al., 2020; Steinhauer, Royle, Drury, & Fromont, 2017).

About a decade after the first report of the N400, several researchers reported a positive-going wave around 600 ms that was dubbed the P600 or the "syntactic positive shift" (Hagoort, Brown, & Groothusen, 1993). The P600 effect was elicited by phrase structure violations such as "Max's *of proof the theorem" (Neville, Nicol, Barss, Forster, & Garrett, 1991) and garden path sentences like "The baker trusted to bake the cake won many awards" (Osterhout & Holcomb, 1992). This effect was first interpreted as indexing syntactic reanalysis and repair, as in ungrammatical sentences or incorrect initial parses where a sequence of words violates syntactic rules (Friederici, 1995). The P600 was later linked explicitly to syntactic integration (Kaan, Harris, Gibson, & Holcomb, 2000), and some accounts have proposed that the N400 and P600 components modularly index retrieval and integration, respectively (Brouwer et al., 2021, 2012). Other accounts interpret the P600 as a member of the P300 family of ERP components; according to this idea, P600 effects do not reflect syntactic structure, but instead reflect a response to unexpected stimuli (Sassenhagen, Schlesewsky, & Bornkessel-Schlesewsky, 2014). The P600 has further been linked to task effects (in line with its interpretation as a member of the P300 family), in contrast to the N400, which may be more task-independent (Schacht, Sommer, Shmuilovich, Martínez, & Martín-Loeches, 2014). Still other researchers have pointed out that reports of P600 effects for distinct linguistic manipulations may actually represent functionally distinct positivities (Leckey & Federmeier, 2020). The P600 effect has also been reported for semantic anomalies, as detailed for the particular case of role reversals in Section 2.2.

In summary, the relationship between the N400 and P600 components is not as simple as once thought, and the two components do not represent discrete, serial stages of syntactic and semantic processing (as suggested in early accounts, like that of Friederici, 1995). There is still another factor that cannot be ignored: variability across languages. Most sentence processing models at least implicitly suppose that their accounts of N400 and P600 effects extend to all languages, with no mechanism for interaction with different linguistic features. Given that languages vary extensively, it is essential that neurocognitive models of sentence processing consider crosslinguistic variability.

1.2a The extended Argument Dependency Model (eADM)

The extended Argument Dependency Model (eADM) is a neurocognitive model centered around crosslinguistic variability in the incremental assignment of structural roles in simple sentences (Bornkessel-Schlesewsky et al., 2011). This model was first proposed as the Argument Dependency Model (Schlesewsky & Bornkessel, 2004), where there was the basic foundation of a parallel, two-stream system with three steps for assigning hierarchical structure. In its extended version, the eADM reiterated the importance of crosslinguistic variation and neural data to its architecture and mapped out finer detail for its proposed three processing steps (Bornkessel-Schlesewsky & Schlesewsky, 2008; Bornkessel & Schlesewsky, 2006a). Here, we briefly outline the three steps (Phases) of the model and their relation to N400 and P600 effects.

Phase 1 encompasses the earliest moments of word recognition when syntactic category is recognized. This step also produces an empty "phrase structure template" where structural roles can be assigned in the subsequent phases. The eADM further divides Phase 2 into two parts. In Phase 2a, the syntactic category identified in Phase 1 determines whether to activate the predicate (i.e., verb) or non-predicate (i.e., noun) processing stream. Phase 2b is proposed to be the source of N400 effects and has different mechanisms for nouns and verbs. For nouns, prominence information (e.g., animacy or case) is used to assign the noun a tentative position in the phrase structure template. For verbs, the representation for argument structure is either linked with nouns that have already been encountered or used to form predictions for upcoming nouns. Additionally for verbs, there is a plausibility check performed on the existing sentence structure.

Finally, Phase 3 encompasses a generalized mapping step where the representations from Phase 2 are integrated into the greater context. P600 effects are proposed to stem from Phase 3.

Crucially, Phase 2b is posited as the processing step where crosslinguistic differences make their impact. Based on the features of a particular language, such as whether there is case marking, strict word order, or verb agreement, the linking, prominence, and plausibility computations will be different. The authors of the eADM themselves compare their proposals for language-specific computations to the notion of cue strength and validity in the Competition Model (Bornkessel & Schlesewsky, 2006a). In a more recent version, the researchers behind the eADM have extended again their model to account for predictive parsing at a more general level (Bornkessel-Schlesewsky & Schlesewsky, 2019). In this update, each language has its own filter for relevant cues for argument assignment, and this language-specific filter is used to generate predictions. We expand on the explanations from the eADM's predictions for argument structure processing in Section 2.2.

2 Argument structure processing

2.1 Overview

Argument structure is the relationship between a verb and its nouns. Verbs are the basic unit of a predicate (Chomsky, 1957a; Croft, 2011) and may take one or more noun arguments to form sentences (Chomsky, 1993). For instance, in the sentence "I write you a thesis," the verb "write" takes the argument "I" as its doer, the argument "you" as its receiver, and the argument "thesis" as the object that undergoes the writing process. The different roles of these arguments are often given precise labels called theta roles. In the same example sentence, the theta roles are agent for "I", recipient for "you", and patient for "thesis." While these labels allow for communication of subtle differences among the types of arguments verbs may take, there is debate over whether and how theta roles exist in a cognitive form (Rissman & Majid, 2019). There is some agreement, however, that parsing mechanisms should require at least an abstract sense of the roles of agent and patient, which have been referred to as proto-agent and proto-patient (Dowty, 1991). In the present thesis, unless otherwise noted, we adopt this view that agent and patient are the primary sentential roles in processing, even when a more precise theta role label may be applied.

Argument structure is a productive line of inquiry for understanding the basic mechanisms of sentence processing. Argument structure processing requires access to knowledge of verbs' logical structure, semantic and event knowledge, and grammatical rules (Friederici & Weissenborn, 2007). Argument structure is thus a complex and fundamental part of language, and it has been shown that detailed argument structure is found in Nicaraguan Sign Language (a new language created organically by deaf children) at the earliest stages of language creation (Senghas, Newport, & Supalla, 1997).

The processes by which the mind manages multiple sources of information for assigning argument structure are the focus of both the Competition Model and the eADM. Here, we first describe argument structure findings from the Competition Model, before turning to the eADM in Section 2.2. As previously discussed, the Competition Model considers that particular languages have individual patterns of cue strength and validity. For argument structure, this has been studied extensively via a binary forced-choice paradigm where participants read transitive sentences with competing cues for indicating which noun is the agent. In the first report of this paradigm, Italian and English sentences were created by orthogonally crossing multiple cues for agent assignment (Bates et al., 1982). For the cue of word order, for instance, a test sentence would be presented as "the horse kicks the carrot", "the horse the carrot kicks", or "kicks the horse the carrot". The cue of word order was then further crossed with animacy, stress, and introducing one of the nouns as a discourse topic with a preceding context. Participants, who were either native Italian or native English speakers, listened to each sentence and chose which of the two nouns was the agent. Results showed a striking difference between the two language groups for which cue participants depended on in the case of conflict: while English speakers relied primarily on the cue of word order (i.e., the noun preceding the verb was preferred as the agent) Italian speakers relied primarily on the cue of animacy (i.e., in the case of an animacy contrast between the two nouns, the animate noun was preferred as the agent). For English, these results showed that word order is the strongest cue and will override animacy and agreement cues in the case of cue conflict. Data from multiple languages have shown that these relative cue strengths are robust among a language's monolingual speakers and vary systematically across languages (MacWhinney, 2022a).

The linguistic features of Mandarin make it especially suited to study argument structure. In the oral register, Mandarin permits multiple word orders and allows both subject and object drop

(Li & Thompson, 1989). This means a sentence like *鸭子吃了* "(the) duck ate" can change meaning depending on context to express "the duck ate (something)" or "(something) ate the duck". Verb-final sentences with two noun arguments are permitted with both object-subject-verb and subject-object-verb word order (with pragmatic restrictions), such that the sentences *孩子苹果吃掉了* "child apple ate" and *苹果孩子吃掉了* "apple child ate" both mean "the child ate the apple" (Li & Thompson, 1976; Li & Thompson, 1989). Evidence from a Competition Model study confirms that plausibility is more important than word order for Mandarin agent assignment (Li, Bates, Liu, & MacWhinney, 1992).

Another important feature for Mandarin argument structure is the coverbs BA and BEI. These coverbs are used in verb-final constructions to explicitly assign agent and patient status. For the noun phrase following a coverb, BA assigns patient status and BEI assigns agent status. To expand on the previous example, the sentence *把鸭子吃了* "BA duck ate" means "(something) ate the duck", while *被鸭子吃了* "BEI duck ate" means "the duck ate (something)". In verb-final sentences with two nouns, both BA and BEI are stronger cues than plausibility (Li et al., 1992), and so allow the expression of role reversal sentences, where an implausible patient receives an action from an implausible agent. Accordingly, the two sentences *苹果把孩子吃掉了* "apple BA child ate" and *孩子被苹果吃掉了* "child BEI apple ate" both mean "the apple ate the child". These role reversal sentences are of special interest for sentence processing research, as will be detailed in Section 2.2.

For the purpose of this thesis, we also point out that multiple researchers have extended the notion of argument structure beyond verbs to also apply to adjectives (Meltzer-Asscher, 2011; Meltzer-Asscher & Thompson, 2014). Just as a verb is described to extend theta roles to its noun arguments, so can an adjective be said to assign a theta role to its own noun arguments, and adjectives differ in the restrictions on their possible arguments and their order in relation to the noun and other adjectives (Loken & Gelman, 2017). In the case of Mandarin, the similarity between adjectives and nouns extends further because there is not a clear word class divide between the syntactic categories of adjective and verb. In fact, some researchers have described Mandarin as a language without adjectives, where descriptive verbs (or "adjectival verbs") instead fill the adjective role (Dixon, 1982; Li & Thompson, 1989). This flexibility of syntactic categories in Mandarin, as shown in Figure 2, makes it a yet more interesting language for the study of argument structure.

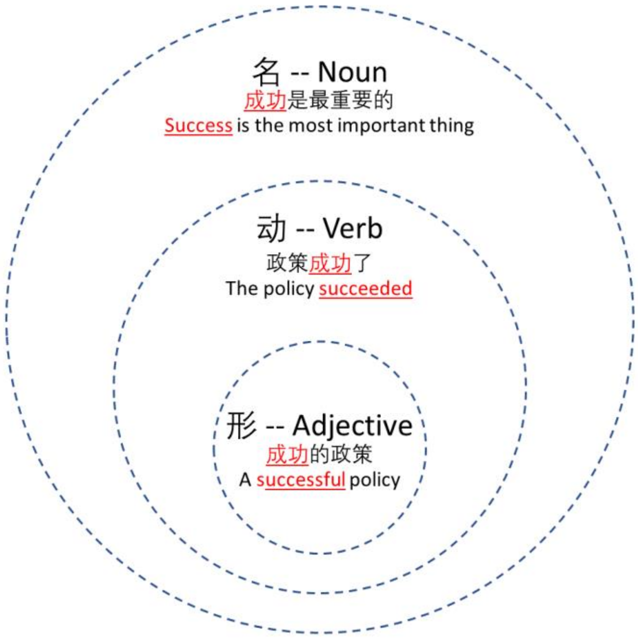

Figure 2*. Model for word class in Mandarin, wherein most adjectives may serve as verbs and most verbs can serve as nouns without inflectional changes (created based on ideas from Chen, 2012; Luuk, 2010).*

2.2 EEG data

In contrast to early characterizations of N400 effects as elicited by semantic anomalies (Kutas & Hillyard, 1980) and P600 effects as elicited by syntactic violations (Neville et al., 1991), there is now ample evidence that these two components index more nuanced processes (Delogu, Brouwer, & Crocker, 2019; Frenzel, Schlesewsky, & Bornkessel-Schlesewsky, 2011). For instance, “syntactic” N400 effects (Guajardo & Wicha, 2014) and “semantic” P600 effects (Hoeks, Stowe, & Doedens, 2004; Kolk, Chwilla, Van Herten, & Oor, 2003) have been reported, demonstrating the limitations of early interpretations of these components.

Many of the advances in our understanding of the N400 and P600 components have been made in the context of studying argument structure processing. The primary focus of this study has been the role reversal, also known as a semantic reversal or a semantic illusion. A role reversal sentence like “The apple eats the child” is semantically anomalous with the stereotypical roles of agent and patient reversed, but otherwise complies with grammatical rules. As such, early interpretations of the N400 and P600 would be consistent with these types of sentences eliciting N400 effects. However, ERP data from almost twenty years ago showed that role reversal sentences elicit P600 effects without modulation of N400 amplitude (Hoeks et al., 2004; Kolk et al., 2003).

These semantic P600 effects have prompted diverse explanations from sentence processing researchers. One such explanation is that select pairings of nouns and verbs are so semantically related that syntactic structure is initially disregarded; in the subsequent steps of comprehension, the parser detects that their semantic interpretation does not match with the verb's assignment of thematic roles, resulting in a syntactic instead of a semantic anomaly (Kim & Osterhout, 2005). Other accounts have proposed parallel but independent processing streams for syntax and semantics, where the semantic stream is initially satisfied because of the high degree of relatedness between the noun and the verb (Kolk & Chwilla, 2007; Kuperberg, 2007). An alternate explanation is that there is a single processing stream where meaning is first retrieved independently from plausibility, as indexed by the N400, and only then integrated into the sentence context, as indexed by the P600 (Brouwer et al., 2021, 2012).

Notably, the majority of these accounts neglect to consider crosslinguistic variation in argument structure processing. Although the semantic P600 effect for role reversal sentences has been reported in English (Kim & Osterhout, 2005), Dutch (Kolk et al., 2003), Japanese (Sakamoto, 2015), and Mandarin (Chow & Phillips, 2013; Chow, Lau, Wang, & Phillips, 2018), there have been different results as well. Languages such as Turkish, Icelandic, and German have shown either N400 or biphasic N400-P600 effects to role reversals, and, in contrast to other reports, Mandarin role reversals have also shown an N400 effect with no semantic P600 (Bornkessel-Schlesewsky et al., 2011). This disparity among ERP effects for role reversals across languages is explicitly addressed by the eADM.

As previously described, the eADM posits three steps for argument structure processing, wherein computations in step two are the source of N400 effects and computations in step three the source of P600 effects (Bornkessel-Schlesewsky & Schlesewsky, 2008; Bornkessel & Schlesewsky, 2006a). When processing non-predicate sentence elements, parsers in step two use language-specific cues to assign a proto-theta role to the noun phrase. These proto-theta roles are similar to those described by Dowty (1991), and in the eADM are dubbed actor (proto-agent) and undergoer (proto-patient)[2]. These actor and undergoer roles are assigned to nouns even before the verb has been encountered. On encountering the verb, the parser links the logical structure of the verb with existing actor and undergoer candidates or waits for an upcoming noun argument to fill the space in the logical structure. Because different languages rely on different cues for expressing argument structure, and the relative weighting of cues in the case of cue conflict is specific to each language, the eADM predicts crosslinguistic differences in how actor and undergoer roles are assigned. Depending on the cue-weighting pattern of a particular language, competition among certain cues may be handled during step two or during step three.

The different computations between step two and step three are the basis for the eADM's explanation of semantic P600 effects (Bornkessel-Schlesewsky & Schlesewsky, 2008). In step two, the formation of actor and undergoer roles and their linking to the verb occur independently of plausibility processing. When encountering a verb, the plausibility of the verb's linking to its arguments is checked in parallel to the linking process. In step three, the plausibility check and the linking of the verb and its arguments are integrated together in a generalized mapping computation. According to the eADM, this generalized mapping process is the source of semantic P600 effects, and the predictions depend on the language. Take for example the English sentence "The hearty meal was devouring the kids" (from Kim & Osterhout, 2005). On encountering the noun "meal," step two assigns "meal" to the actor role based on its being first in the sentence, as word order is the strongest cue for argument assignment in English. On reaching the verb, the linking between "meal" and "devour" is performed such that "meal" is the undergoer of "devour." The plausibility check conducted in parallel to the linking of the verb and its argument is also satisfied because "meal" and "devour" are semantically related. It is then only at stage three in the generalized mapping computation where the mismatch between linking and plausibility is detected, thus eliciting a semantic P600 with no N400 effect.

According to the eADM, languages with a different pattern of cue weighting may show different responses to role reversals (Bornkessel-Schlesewsky et al., 2011). An example is the Mandarin sentence *子弹被侦探击中了* "bullet BEI detective hit." In this sentence, the cues of animacy and BEI both point to "bullet" being the undergoer and "detective" being the actor.

[2] Although proto-agent and proto-patient have been called the only "psychologically real" theta roles, we note that there is neural evidence of processing differences between experiencer- and agent-type verbs (Bourguignon et al., 2012; Kyriaki et al., 2020).

Accordingly, these nouns are assigned their appropriate proto-theta roles during incremental parsing. On reaching the verb, the linking and plausibility check computations are performed. Because linear order of arguments is an unreliable cue in Mandarin, the linking computation must consider additional cues like animacy or likelihood that the noun could serve as an agent to complete the logical structure of the verb. As such, the parser encounters difficulty in stage two of processing the verb, thus yielding an N400 effect. The eADM predicts that the well-formedness check in step three would then result in a P600 effect, but there appears to be crosslinguistic variation at this stage as well. While German, Turkish, Icelandic, and Mandarin have all shown N400 effects to role reversals, only German and Icelandic showed subsequent semantic P600 effects (Bornkessel-Schlesewsky et al., 2011). Given the substantial variability among languages, the eADM accounts for many of the reported semantic P600 phenomena.

The eADM further proposes that the difference between languages that do and do not show an N400 to role reversals is the cue of word order. “Sequence-dependent” languages like English or Dutch, where word order is the strongest cue for assigning argument structure, are predicted to show semantic P600s to role reversals; “sequence-independent” languages like German or Mandarin, where word order is an unreliable cue for expressing argument structure, are predicted to show N400 effects to role reversals, and may or may not show a biphasic N400-P600 effect. This interpretation suggests that flexibility in word order may be one of the fundamental differences among languages that impacts sentence processing.

2.2a Bag of Arguments

We further note an additional account that has considered EEG data from Mandarin to account for the semantic P600 effect and argument structure, which we refer to in this thesis as the Bag of Arguments account (Chow & Phillips, 2013; Chow, Smith, Lau, & Phillips, 2016; Chow et al., 2018, 2016). The researchers behind this account reported semantic P600 effects in Mandarin without an N400 effect (Chow & Phillips, 2013), and later extended this finding with an account of general constraints and limitations on processing of structural roles, without a role for crosslinguistic differences (Chow et al., 2016; Chow et al., 2018). According to the Bag of Arguments account, the N400 effect is best interpreted as an index of how predictable a word is, with more predictable words eliciting smaller N400 components. Critically, there are requirements for this prediction to occur. First, parsers need enough time (at least 800 ms) to predict the structural roles of pre-verb arguments; before this point, parsers can still use the lexical meaning, but without assigning agent or patient roles (Chow et al., 2018). Second, the verb and its nouns must be sufficiently related such that the verb is predictable from its preceding arguments (Chow & Phillips, 2013; Chow et al., 2018). Third, parsers consider only nouns within the same clause as possible arguments for a verb (Chow et al., 2016).

For example, consider the following two role reversal sentences: (1) *小偷把警察抓住了* "thief BA police arrested" and (2) *小偷把警察在上星期抓住了* "thief BA police last week arrested". According to the Bag of Arguments account, both (1) and (2) have nouns and verbs that are combinable such that the verb could be predictable. In (1) the verb follows the second argument directly, while in (2) there is an intervening adverbial phrase. The Bag of Arguments account states that at least 800 ms is needed to predict the structural roles of nouns, so (1) should not elicit an N400 effect but (2) should, assuming the adverbial phrase is sufficiently long. Both sentences should elicit P600 effects as the sentences are implausible. Finally, any arguments in preceding clauses would not be considered in forming a prediction of the verb "arrest".

In contrast to the eADM, the Bag of Arguments account assumes that effects should hold across languages, and its proponents have used both Mandarin and English data to construct their model (W. Chow et al., 2016). The eADM authors have acknowledged results from the Bag of Arguments studies, but attribute the lack of an N400 to the Bag of Arguments account's only considering the coverb BA to form role reversals (Bornkessel-Schlesewsky & Schlesewsky, 2019). Nonetheless, the eADM and Bag of Arguments accounts conflict in whether there is a P600 effect for BA role reversals.

3 Non-native processing

3.1 Overview

Alongside crosslinguistic variability and the mechanisms of argument structure, understanding bilingual[3] language experience is an essential factor for understanding language processing (Wong, Yin, & O'Brien, 2016). There are virtually no complete monolinguals on the

[3] Crucially, we do not use the terms bilingual and bilingualism to exclude multilingualism. We consider that studying processing of two languages sets the framework for extending research to multilingual cases. On a related note, we also find the term second language (or L2) to be helpful in distinguishing between early life acquisition, as in a first language (or L1), and acquisition later in life. Following this logic, a bilingual individual may have multiple second languages and/or multiple first languages. For the experimental study of processing, it is useful to consider a reduced case to interpret the precise mechanisms and interactions at hand.

planet (Edwards, 2006), and all of an individual's linguistic knowledge is contained in the same brain. We know that a bilingual brain is not the sum of two monolinguals' (Grosjean, 1989), but the impacts of multiple languages on brain structure and function remain a line of current research (Wong et al., 2016). Some evidence suggests that all linguistic knowledge shares the same neural representation, and multiple languages largely draw on the same cortical areas as observed in monolinguals (Fabbro, 2001; Paradis, 1990). However, recovery patterns for aphasia in bilinguals are variable, and one language may suffer impairment while the other recovers, suggesting that there may be some non-overlapping representations or processing (Kuzmina, Goral, Norvik, & Weekes, 2019). The field of sentence processing has myriad factors to consider for a single language, so adding bilingualism to the picture is an enormous task.

In addition, researchers must further contend with the fact that bilingualism encompasses a range of language experiences. Some bilinguals learn their two or more languages early in life, either simultaneously from birth or sequentially with early exposure to the second language (Tsimpli, 2014). Some of these bilinguals are heritage speakers, whose dominant language is not their mother tongue (Montrul, 2012), and others are receptive bilinguals, who comprehend their first language or a second language without the ability to produce it (Sherkina-Lieber, 2020). Still other bilinguals learn their second language as adults, and still reach native-like proficiency (Birdsong, 1992; White & Genesee, 1996). Bilinguals also vary in how they use their languages with monolinguals or other bilinguals (Gullifer & Titone, 2020), ranging from a purely monolingual to a fully bilingual mode with language mixing (Grosjean, 2012). All bilinguals further vary in which of their languages they use more proficiently in which contexts (Schmeißer et al., 2015). We acknowledge this diversity of bilingual experience and the complexity of measuring how this diversity impacts processing. In the present thesis, we compare functionally monolingual Mandarin speakers to late English-Mandarin bilinguals. We collected detailed language background information from each of our participants to have a more complete picture of the variability in their language experience.

Just as models of language processing routinely neglect the importance of crosslinguistic variability, so too do models routinely ignore the impact of the diversity of bilingual experience and how two or more languages interact in the same mind (Frank, 2021). Of the models discussed so far, only the Competition Model explicitly accounts for bilingual processing (MacWhinney, 1987). One of the most consistent findings is that features of native language processing transfer to processing of second languages, and this transfer can either facilitate performance in the second language, such as when the ordering of adjectives and nouns is the same between languages, or make learning more difficult, such as when the ordering of adjectives and nouns is different (Bardovi-Harlig & Sprouse, 2017). This transfer may also occur bidirectionally, from the second language to the first (Brown & Gullberg, 2008; Degani, Prior, & Tokowicz, 2011; MacWhinney, 2017; Pavlenko & Jarvis, 2002). The Competition Model provides more nuance to the idea of positive and negative transfer by making the relation to cue strength and validity (MacWhinney, 1987). According to this view, cue weighting from the first language is used directly in the initial stages of second language acquisition but is then fine-tuned through experience to strengthen or weaken particular cues that have been transferred. The Competition Model suggests four scenarios for the relationships between the cue weightings of the first and second language (MacWhinney, 1987). One extreme is that first language strategies are entirely transferred to the second language, even when they are applied inappropriately. The other extreme is that first language strategies are completely abandoned in favor of those of the second language. There are additionally, however, the possibilities of merging the first and second language cue frameworks together as one or achieving at least partially separate representations of each language's cues.

An example of the intricate interplay between bilinguals' two languages is in a Competition Model study of Mandarin-English bilinguals who acquired their second language English at different periods in life (Liu, Bates, & Li, 1992). The study confirmed previous patterns of cue weighting for agent assignment in the monolingual control groups, with English monolinguals almost exclusively relying on word order and Mandarin monolinguals only using word order in the absence of other cues (Li et al., 1992). The study's experimental groups were Mandarin-English bilinguals who were first exposed to English before age four, between ages six and ten, between ages twelve and sixteen, or after the age of twenty. The bilinguals who learned English as adults showed influence from their Mandarin cue-weighting strategies for both languages, while the bilinguals who learned English as infants used English cue-weighting strategies for both languages. This pattern ostensibly suggested a simple story of age of acquisition being the primary factor for cue weighting strategy. The authors noted, however, that Mandarin and English may be "remarkably interpenetrable," meaning that the relative strengths of word order as a cue and both languages' poor or absent inflectional morphology may result in less negative and more positive transfer (Liu et al., 1992). The authors further indicate that while their results showed apparent effects of age of acquisition, these effects were not always linear and showed variability among groups.

This study highlights a central debate in bilingualism research about the relative roles of age of acquisition and proficiency in achieving nativelike processing. Neurologists Penfield & Roberts (1959) and the linguist and psychologist Lenneberg (1967) are credited with the first academic descriptions of a critical period for language acquisition, after which it is harder to learn a new language. Researchers today broadly agree with the notion of a critical period for first language acquisition, after which point a learned language will lack structural complexity (Mayberry & Kluender, 2018). Discussion of critical periods for second language acquisition, however, is often divisive (Johnson & Newport, 1989; Newport, 2018). In a behavioral study of over 600,000 native and non-native English speakers, Hartshorne, Tenenbaum, & Pinker (2018) claimed to demonstrate a critical period for native-like attainment of second language syntax. Their analysis[4] showed that learners' performance on a battery of English syntactic and morphosyntactic measures declined relative to native speakers as a function of their age of acquisition. This was more so the case for learners who acquired their English in a classroom setting, while immersion learners showed higher levels of attainment. In contrast to the idea of a critical period, however, is the presence of multiple individuals in their dataset who do reach native-like performance at late ages. This variability is noted by the authors' softening of their claims in the main text, where they suggest the "optimal" period for second language learning is younger than seven years old, and that their suggested critical period applies only to the rate of learning, not ultimate attainment[5] (Chen & Hartshorne, 2021). Additionally, as the authors note, the plateau effect seen in late second language learners may be due to social effects, where lack of learning is due not to a diminished neurocognitive capacity for language acquisition but merely to a lack of motivation (Hartshorne et al., 2018).

Hartshorne and colleagues' (2018) claim of a critical period for second language learning is specifically for the domain of syntax and morphosyntax. There is strong evidence that

---

[4] There has also been critique of Hartshorne and colleagues' analysis and whether the suggested cutoff ages accurately represent the data (Slik, Schepens, Bongaerts, & Hout, 2022; cf Hartshorne, 2021).

[5] Curiously, results broken down by first language type showed that Mandarin speakers who learned English between the ages of six and ten showed faster learning. This may be relevant to the claim of "interpenetrability" by Liu et al. (1992). See also Chan & Hartshorne (2021) for further discussion of the impact (or lack of impact) of language type on learning rate and attainment.

acquisition of certain linguistic domains, including phonemes (Werker & Tees, 1984), prosodic patterns for speech segmentation (Kuhl, 2004), and lexical tone (Pierce, Klein, Chen, Delcenserie, & Genesee, 2014), involve brain changes that occur within the first year of life. Syntax and morphosyntax, however, may show a different pattern. Data from Hartshorne and colleagues (2018) suggest that the learning rate for grammatical rules decreases for a second language. If this is true, then first language syntax should also be robust and engender brain changes relatively early in life.

Our introduction here of non-native processing and potential limits ultimate attainment of first and second languages also requires consideration of the meaning of the term “native speaker”. In the fields of applied linguistics and education, there has been considerable discussion of the problems underlying the term “native” (Davies, 2004; O’Rourke & Pujolar, 2013), including its implications for psycholinguistics (Cheng et al., 2021). While we note these problems and strive for inclusivity in our research terminology, we find the terms “native” and “non-native” essential for our consideration of core sentence processing mechanisms. For our purposes, the term native speaker implies exposure to a given language early in life, prior to sensitive periods for sound processing as described above. Per Hartshorne and colleagues’ (2018) data, the early life exposure for syntactic rules may extend until approximately seven years old. In the case of our non-native participants, we limited recruitment to those whose first immersion in Mandarin was in mid-adolescence or later.

Lastly, we note there are still open questions about the differences among domains of language for second language acquisition and bilingual processing. For the present thesis, all experimental manipulations relate specifically to word order, as Mandarin has virtually no morphological inflection. We consider that these word order manipulations therefore directly target syntactic processing. Although Mandarin and English have similar canonical word order, word order’s relative cue strength is vastly different between the two languages, as shown in the previously discussed Competition Model studies. This makes word order a special target for investigating English-Mandarin bilingual processing. We also note that the Competition Model may fail to account for potential differences between linguistic domains like phonology, semantics, and syntax, as the model supposes that all domains of languages are subject to the same cue processing demands.

3.2 EEG data

Even when non-native speakers resemble their native counterparts in performance, the underlying neural mechanisms and strategies they use may be distinct. As a non-invasive tool that directly measures real-time electrical activity in the brain, EEG has the potential to uncover processing differences even when behavior is identical. Steinhauer, White, & Drury (2009) suggest an approach for using ERPs to evaluate progress in second language learning, where distinct stages of proficiency are captured in qualitative differences in ERP patterns. For example, a grammatical violation that would elicit a biphasic N400-P600 response in native speakers (e.g., adjective placement in Steinhauer, 2014) would show in second language learners first solely an N400, then a P600, and then the native-like N400-P600 pattern as the learner's proficiency increases. The authors suggest that ERP evidence demonstrates that second language learners can attain native-like processing patterns (Steinhauer, 2014; Steinhauer et al., 2009). Others have suggested that second language learners' processing is for the large part shallower than that of native speakers, even at high levels of proficiency (Clahsen & Felser, 2006b, 2006a, 2018; Roberts, 2012).

These hypotheses have been tested extensively by ERP studies studying non-native grammar. Caffarra, Molinaro, Davidson, & Carreiras (2015) conducted a systematic review of ERP studies on second language speakers' syntactic processing. The authors classified experimental manipulations as falling under two types of syntactic violations: phrase structure[6], where there is an error in syntactic category, word order, or omission of a word, and morphosyntax, where morphological inflection on a word disagrees with the required gender, number, case, person, or tense agreement. Results showed that a P600 effect was the most consistent finding among all reviewed studies, with the next most common finding being an N400 effect. Regression results showed that earlier age of acquisition was associated with greater likelihood of observing N400 effects, while the presence of P600 effects was most predicted by proficiency. The authors interpret the proficiency effect on P600 to support the hypothesis of Steinhauer, White, & Drury (2009). However, earlier ERP components, including the N400, may be less affected by proficiency, suggesting that certain aspects of non-native sentence processing in earlier time windows (i.e., before 600 ms) are constrained by age of acquisition.

---

[6] We note that per Caffarra and colleagues' (2015) classification, the experiments described in the present thesis are all manipulations of phrase structure.

We note a final important element that is relevant to the present thesis: individual differences. Just as there is variability among languages in patterns of cue weighting for argument structure, there is variability among individuals, and this is the case for ERP profiles as well. Individual differences and how to measure them have recently been highlighted as a key area of investigation for the language sciences, including for native and non-native processing (Cunnings & Fujita, 2021; Kidd, Donnelly, & Christiansen, 2018). There is evidence that in addition to N400 and P600 ERP profiles tracking development in proficiency, these two components also track individual differences in sentence processing for both native and non-native speakers (Tanner, Inoue, & Osterhout, 2014; Tanner, McLaughlin, Herschensohn, & Osterhout, 2013; Tanner & Van Hell, 2014). For instance, in one study of early second language learners, a group-level biphasic N400-P600 effect in response to subject-verb agreement violations was revealed to be either an N400 or a P600 pattern at the individual level, while native speakers showed only a P600 effect (Tanner et al., 2013). Additionally, the magnitude of the P600 effect was positively correlated with learners' second language proficiency, while the magnitude of the N400 effect was negatively correlated, indicating that more successful early learners showed greater resemblance to the native ERP pattern (in line with Steinhauer et al., 2009). Another study showed that for non-native speakers, the total magnitude of N400 and P600 responses to grammatical violations was associated with second language proficiency, while age of acquisition and motivation predicted a more native-like P600 response pattern (Tanner et al., 2014). These results demonstrate the potential of the N400 and P600 components to capture individual profiles of non-native processing. Understanding these individual differences in conjunction with bilingual language experience is part of the future of the field of language sciences (Titone & Baum, 2014).

4 Present thesis

The present thesis used ERPs and behavioral tasks to investigate processing of two types of Mandarin sentence structures. The first structure, verb-final sentences, is the topic of Manuscript I, and the second structure, adjective-noun placement, is the topic of Manuscript II. Both manuscripts address crosslinguistic variation by comparing the Mandarin structures to English and other languages. Argument structure is a central focus of Manuscript I, but also applies in a broad sense to adjective placement, as described in 2.1 for the notion of argument structure in adjectives. For bilingualism, only Manuscript II considers bilingual participants, comparing native and non-native processing; however, we have additional bilingual data from the experiment in Manuscript I that do not form part of the thesis that we will briefly touch on in the General Discussion section. Note that as of the time of submission of this thesis, neither manuscript has been submitted to an academic journal for peer review.

The General Discussion expands on the two manuscripts' findings and interpretation, as well as proposing avenues for further investigation. Some data originally planned were not included in the present thesis for reasons of concision and constraints due to the COVID-19 pandemic. For the experiment detailed in Manuscript I, additional data were collected from non-native Mandarin speakers as well as data from English native speakers for verb-final English sentences. For the experiments in both manuscripts, there is ongoing data collection from first language attriters, Chinese immigrants who have resided in an English-speaking country for an extended period of time and report changes in their native Mandarin. Although these data are not part of the present thesis, relevant mentions will be made in the General Discussion for how the present manuscripts can inform predictions for these studies.

Below, we briefly overview the contents of both manuscripts.

4.1 Manuscript I

In the first manuscript, we used a binary agent assignment task, reaction time, and ERPs to study argument structure processing of Mandarin verb-final sentences. Mandarin allows both agent-first and patient-first word order for verb-final sentences, so these sentence types can be ambiguous without context. For instance, in the sentence *气压温度预测了* "temperature pressure predicted", it is ambiguous whether temperature predicts pressure or pressure predicts temperature. If semantic knowledge gives a strong motivation for a certain interpretation, both word orders are interpreted to have the same meaning. For example, the sentences *苹果孩子吃掉了，孩子苹果吃掉了* "apple child ate", "child apple ate" have the same meaning, namely that the child ate the apple. To explicitly assign one of the nouns as an agent in verb-final sentences, it is necessary to use one of the two coverbs BA and BEI. The coverb BA assigns patient status to its following noun, while the coverb BEI assigns patient status to its preceding noun. Using these coverbs, it is possible to create role reversal sentences, such as *苹果把孩子吃掉了* "apple BA child ate" or *孩子被苹果吃掉了* "child BEI apple ate". To create our experimental materials, we manipulated sentence structure by including or excluding coverbs, argument order by alternating which nouns appeared in first and second position, reversibility by constructing sentences with and without clear semantic stereotypes for agent status, and agent animacy by including plausible animate and inanimate agents.

Our experiment addressed the following research questions:

I) How do competing sentence cues impact Mandarin parsers' assignment of argument structure?
   a. Prior Competition Model results suggest that BEI should be stronger than BA and animacy should be stronger than word order. Additionally, word order should be used to assign argument structure in the absence of other cues, with a preference for object-subject-verb word order.

II) How do parsers handle competing cues incrementally throughout a sentence?
   a. Previous results from the eADM research group and others suggest that sentence-initial inanimate nouns elicit larger N400s than sentence-initial animate nouns, due to a subject-first processing heuristic where less ideal subjects (like inanimate nouns) are more costly to process

III) What ERP pattern is elicited by Mandarin role reversals?
   a. According to the eADM, Mandarin role reversals should elicit N400 effects because Mandarin is a sequence-independent language that differs from sequence-dependent languages like English.
   b. According to the Bag of Arguments, Mandarin role reversals should always elicit a P600 effect and only show an N400 effect if the nouns and verb are sufficiently related and if there at least 800 ms between the final argument and the verb.

4.2 Manuscript II

In the second manuscript, we considered native and non-native processing of Mandarin adjective-noun structures. While Mandarin and English adjectives typically precede the nouns the modify, there are grammatical differences as well. Predicate adjectives in Mandarin can be used without a copula, as in *这些书很新* "these books very new." Additionally, most Mandarin attributive adjectives are separated from their noun by the particle DE, as in *很新的书* "very new DE book", and DE is obligatory in the case of adjectives composed of more than one morpheme or when adjectives are modified by an adverb, as in the example of "very new". To maximize comparison with English, we considered the case of monosyllabic adjective-noun pairs (like *新书* "new book") that can occur without the intervening particle DE. We created sentences with constraining syntactic structures that made it ungrammatical for the adjective to follow the noun, as in *父亲最近想看新书了* "father recently wants to read new books." We tested native Mandarin speakers and high-proficiency English-Mandarin bilinguals. For the bilinguals, we also tested their native English processing in a replication of a prior version of the experiment in English.

Manuscript II addressed the following research questions:

I) Can we replicate the biphasic N400-P600 response to English adjective-noun placement violations in English-Mandarin bilinguals?
   a. The previous effect was time-locked to the onset of the second word in the adjective-noun pair.

II) What ERP pattern do Mandarin adjective-noun placement violations elicit?
   a. Given the grammatical differences with English adjectives, and the fact that sentence processing experiments of Mandarin sometimes show different ERP patterns from other languages, this possibility was not trivial.

III) Can non-native Mandarin speakers show a native-like ERP pattern to Mandarin adjective-noun violations?
   a. Previous results showed that both Mandarin and French native speakers could show a native-like pattern for English adjective placement processing. However, French speakers, whose first language has both prenominal and postnominal adjectives, showed an additional effect on the first word in the adjective-noun pair for adjectives that were postnominal in their native French. Because Mandarin and English adjectives typically precede their nouns, non-native speakers should be able to transfer this pattern to help them process the non-native pattern.

5 References

Bardovi-Harlig, K., & Sprouse, R. A. (2017). Negative Versus Positive Transfer. In *The TESOL Encyclopedia of English Language Teaching* (pp. 1–6). https://doi.org/10.1002/9781118784235.eelt0084

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, *11*(3), 245–299. https://doi.org/10.1016/0010-0277(82)90017-8

Birdsong, D. (1992). Ultimate Attainment in Second Language Acquisition. *Language*, *68*(4), 706. https://doi.org/10.2307/416851

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. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(FEB), 1–17. https://doi.org/10.3389/fpsyg.2019.00298

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

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. https://doi.org/10.1016/j.bandl.2012.05.001

Brouwer, H., Delogu, F., Venhuizen, N. J., & Crocker, M. W. (2021). Neurobehavioral Correlates of Surprisal in Language Comprehension: A Neurocomputational Model. *Frontiers in Psychology*, *12*(February), 1–16. https://doi.org/10.3389/fpsyg.2021.615538

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. https://doi.org/10.1016/j.brainres.2012.01.055

Brown, A., & Gullberg, M. (2008). BIDIRECTIONAL CROSSLINGUISTIC INFLUENCE IN L1-L2 ENCODING OF MANNER IN SPEECH AND GESTURE: A Study of Japanese Speakers of English. *Studies in Second Language Acquisition*, *30*(02), 225–251. https://doi.org/10.1017/S0272263108080327

Brown, C., & Hagoort, P. (1993). The Processing Nature of the N400: Evidence from Masked Priming. *Journal of Cognitive Neuroscience*, *5*(1), 34–44. https://doi.org/10.1162/jocn.1993.5.1.34

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews*, *51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Carreiras, M., Duñabeitia, J. A., Vergara, M., de la Cruz-Pavía, I., & Laka, I. (2010). Subject relative clauses are not universally easier to process: Evidence from Basque. *Cognition*, *115*(1), 79–92. https://doi.org/10.1016/j.cognition.2009.11.012

Chan, J., & Hartshorne, J. K. (2021). Is It Easier for Children to Learn English If Their Native Language Is Similar to English？*Proceedings of the 46th Annual Boston University Conference on Language Development*, 103–114.

Chappell, H., Ming, L., & Peyraube, A. (2007). Chinese linguistics and typology: The state of the art. *Linguistic Typology*, *11*(1), 187–211. https://doi.org/10.1515/LINGTY.2007.014

Chen, G. (2012). *A Contrastive Study of Adjectival Class in Chinese and English (汉英形容词对比研究)*. Zhejiang University (浙江大学).

Chen, T., & Hartshorne, J. K. (2021). More evidence from over 1.1 million subjects that the critical period for syntax closes in late adolescence. *Cognition*, *214*, 104706. https://doi.org/10.1016/j.cognition.2021.104706

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The Problematic Concept of Native Speaker in Psycholinguistics: Replacing Vague and Harmful Terminology With Inclusive and Accurate Measures. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.715843

Chomsky, N. (1957a). 4. Phrase Structure. In *Syntactic Structures* (pp. 26–33). https://doi.org/10.1515/9783112316009-005

Chomsky, N. (1957b). Syntactic Structures. In *Matériaux pour l'histoire de notre temps* (Vol. 27). https://doi.org/10.1515/9783112316009

Chomsky, N. (1993). Chapter 2. Subsystems of core grammar. 2.6. LF-representation and θ-theory (2). In *Lectures on Government and Binding* (pp. 101–117). https://doi.org/10.1515/9783110884166.101

Chomsky, N. (2006). *Language and Mind*. Retrieved from http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=244085

Chomsky, N. (2014). *The Minimalist Program*. https://doi.org/10.7551/mitpress/9780262527347.001.0001

Chomsky, N., & Lasnik, H. (1993). The Theory of Principles and Parameters. In *Syntax* (pp. 506–569). https://doi.org/10.1515/9783110095869.1.9.506

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, *1506*, 76–93. https://doi.org/10.1016/j.brainres.2013.02.016

Chow, W., Smith, C., Lau, E., & Phillips, C. (2016). A " bag-of-arguments " mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/23273798.2015.1066832

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), 803–828. https://doi.org/10.1080/23273798.2018.1427878

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. https://doi.org/10.1080/23273798.2016.1160135

Clahsen, H., & Felser, C. (2006a). Continuity and shallow structures in language processing. *Applied Psycholinguistics*, *27*(1), 107–126. https://doi.org/10.1017/S0142716406060206

Clahsen, H., & Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences*, *10*(12), 564–570. https://doi.org/10.1016/j.tics.2006.10.002

Clahsen, H., & Felser, C. (2018). SOME NOTES ON THE SHALLOW STRUCTURE HYPOTHESIS. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Croft, W. (2007). Typology and linguistic theory in the past decade: A personal view. *Linguistic Typology*, *11*(1), 79–91. https://doi.org/10.1515/LINGTY.2007.007

Croft, W. (2011). Parts of speech as language universals and as language-particular categories. *Approaches to the Typology of Word Classes*, (21), 65–102. https://doi.org/10.1515/9783110806120.65

Cunnings, I., & Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, *42*(3), 579–599. https://doi.org/10.1017/S0142716420000648

Davies, A. (2004). The Native Speaker in Applied Linguistics. In A. Davies & C. Elder (Eds.), *The Handbook of Applied Linguistics* (pp. 431–450). https://doi.org/10.1002/9780470757000.ch17

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, *23*(1), 18–28. https://doi.org/10.1080/20445911.2011.445986

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569. https://doi.org/10.1016/j.bandc.2019.05.007

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121. https://doi.org/10.1038/nn1504

Dixon, R. M. (1982). Where have all the adjectives gone? In R. M. Dixon (Ed.), *Where have All the Adjectives Gone? And Other Essays in Semantics and Syntax* (pp. 1–62). https://doi.org/10.1515/9783110822939.1

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, *67*(3), 547. https://doi.org/10.2307/415037

Druks, J. (2002). Verbs and nouns—a review of the literature. *Journal of Neurolinguistics*, *15*(3–5), 289–315. https://doi.org/10.1016/S0911-6044(01)00029-X

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/81

Edwards, J. (2006). Foundations of Bilingualism. In *The Handbook of Bilingualism* (pp. 7–31). https://doi.org/10.1002/9780470756997.ch1

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fabbro, F. (2001). The bilingual brain: Cerebral representation of languages. *Brain and Language*, *79*(2), 211–222. https://doi.org/10.1006/brln.2001.2481

Federmeier, K. D., Jongman, S. R., & Szewczyk, J. M. (2020). Examining the Role of General Cognitive Skills in Language Processing: A Window Into Complex Cognition. *Current Directions in Psychological Science*, *29*(6), 575–582. https://doi.org/10.1177/0963721420964095

Ferreira, F., & çokal, D. (2015). Sentence Processing. In *Neurobiology of Language*. https://doi.org/10.1016/B978-0-12-407794-2.00022-5

Ferreira, F., & Yang, Z. (2019). The Problem of Comprehension in Psycholinguistics. *Discourse Processes*, *56*(7), 485–495. https://doi.org/10.1080/0163853X.2019.1591885

Frank, S. L. (2021). Toward Computational Models of Multilingual Sentence Processing. *Language Learning*, *71*(S1), 193–218. https://doi.org/10.1111/lang.12406

Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: A new perspective on the N400–P600 distinction. *Neuropsychologia*, *49*(3), 574–579. https://doi.org/10.1016/j.neuropsychologia.2010.12.003

Friederici, A.D. (1995). The Time Course of Syntactic Activation During Language Processing: A Model Based on Neuropsychological and Neurophysiological Data. *Brain and Language*, *50*(3), 259–281.

https://doi.org/10.1006/brln.1995.1048

Friederici, A.D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax–semantic interface. *Brain Research*, *1146*, 50–58. https://doi.org/10.1016/j.brainres.2006.08.038

Friederici, Angela 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. https://doi.org/10.1006/jmla.2000.2709

Gibson, E., & Fedorenko, E. (2010). Weak quantitative standards in linguistics research. *Trends in Cognitive Sciences*, *14*(6), 233–234. https://doi.org/10.1016/j.tics.2010.03.005

Gibson, E., & Wu, H. H. I. (2013). Processing Chinese relative clauses in context. *Language and Cognitive Processes*, *28*(1–2), 125–155. https://doi.org/10.1080/01690965.2010.536656

Grigor, J. (1999). The Effect of Odour Priming on Long Latency Visual Evoked Potentials of Matching and Mismatching Objects. *Chemical Senses*, *24*(2), 137–144. https://doi.org/10.1093/chemse/24.2.137

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934X(89)90048-5

Grosjean, F. (2012). Bilingual and Monolingual Language Modes. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0090

Guajardo, L. F., & Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262–272. https://doi.org/10.1016/j.neuroimage.2013.09.077

Gullifer, J. W., & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, *23*(2), 283–294. https://doi.org/10.1017/S1366728919000026

Gutierrez, E., Williams, D., Grosvald, M., & Corina, D. (2012). Lexical access in American Sign Language: An ERP investigation of effects of semantics and phonology. *Brain Research*, *1468*, 63–83. https://doi.org/10.1016/j.brainres.2012.04.029

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. https://doi.org/10.1080/01690969308407585

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(5), 2347–2353. https://doi.org/10.1073/pnas.1910923117

Hartshorne, J. K. (2021). No evidence that age affects different bilingual learner groups differently : Rebuttal to van der Slik , Schepens , Bongaerts , and van Hout (2021). *PsyArXiv*, 1–20.

https://doi.org/https://doi.org/10.31234/osf.io/bvjys

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*(April), 263–277. https://doi.org/10.1016/j.cognition.2018.04.007

Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, *11*(1), 119–132. https://doi.org/10.1515/LINGTY.2007.011

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. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Holcomb, P. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47–61. https://doi.org/10.1111/j.1469-8986.1993.tb03204.x

Huang, C.-R., Chen, C.-J., & Shen, C. C. C. (2002). The Nature of Categorical Ambiguity and Its Implications for Language Porceesing: A Corpus-based Study of Mandarin Chinese. *Sentence Processing in East Asian Languages*, 53–83.

Hyder, R., Højlund, A., Jensen, M., Østergaard, K., & Shtyrov, Y. (2020). Objective assessment of automatic language comprehension mechanisms in the brain: Novel E/MEG paradigm. *Psychophysiology*, *57*(5), 1–18. https://doi.org/10.1111/psyp.13543

Jäger, L., Chen, Z., Li, Q., Lin, C. J. C., & Vasishth, S. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79–80*, 97–120. https://doi.org/10.1016/j.jml.2014.10.005

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, *1*(6), 571–591. https://doi.org/10.1111/j.1749-818x.2007.00037.x

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159–201. https://doi.org/10.1080/016909600386084

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

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

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*(3), 257–261. https://doi.org/10.1016/j.bandl.2006.07.006

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. https://doi.org/10.1038/nrn1533

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

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. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kuzmina, E., Goral, M., Norvik, M., & Weekes, B. S. (2019). What Influences Language Impairment in Bilingual Aphasia? A Meta-Analytic Review. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00445

Kyriaki, L., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Semantic reversal anomalies under the microscope: Task and modality influences on language-associated event-related potentials. *European Journal of Neuroscience*, *52*(7), 3803–3827. https://doi.org/10.1111/ejn.14862

LaPolla, R. (1993). *Arguments Against ' Subject ' and ' Direct Object ' as Viable Concepts in Chinese Arguments Against ' Subject ' and ' Direct Object ' · as Viable c · o _ ncepts 1n Chinese* *. (January 1993).

Lau, E. F., Phillips, C., & Poeppel, D. (2008, December). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, Vol. 9, pp. 920–933. https://doi.org/10.1038/nrn2532

Lau, E., & Tanaka, N. (2021). The subject advantage in relative clauses: A review. *Glossa*, *6*(1). https://doi.org/10.5334/GJGL.1343

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*(7), 1–15. https://doi.org/10.1111/psyp.13351

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. https://doi.org/10.1016/j.jml.2012.02.005

Lewis, R. L. (1999). Specifying Architectures for Language Processing: Process, Control, and Memory in

Parsing and Interpretation. In *Architectures and Mechanisms for Language Processing* (pp. 56–89). https://doi.org/10.1017/CBO9780511527210.004

Li, C. N., & Thompson, S. A. (1976). On the issue of word order in a synchronic grammar: A case against 'movement transformations.' *Lingua*, *39*(3), 169–181. https://doi.org/10.1016/0024-3841(76)90001-2

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1992). Cues as Functional Constraints on Sentence Processing in Chinese. In H. C. Chen & O. Tzeng (Eds.), *Language Processing in Chinese* (pp. 207–234). https://doi.org/10.1016/S0166-4115(08)61893-2

Li, P., & MacWhinney, B. (2012). Competition Model. In *The Encyclopedia of Applied Linguistics* (pp. 1–5). https://doi.org/10.1002/9781405198431.wbeal0168

Li, & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451–484. https://doi.org/10.1017/S0142716400005762

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis - The assumption that measurement error always reduces effect sizes is fals. *Science*, *355*(6325), 582–584.

Luck, S. J. (Steven J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Luuk, E. (2010). Nouns, verbs and flexibles: implications for typologies of word classes. *Language Sciences*, *32*(3), 349–365. https://doi.org/10.1016/j.langsci.2009.02.001

Macwhinney, B. (2017). Are these approaches incompatible ? *Linguistic Approaches to Bilingualism*, *6*, 730–733. https://doi.org/10.1075/lab.00014.mac

MacWhinney, B. (1987). Applying the Competition Model to bilingualism. *Applied Psycholinguistics*, *8*(4), 315–327. https://doi.org/10.1017/S0142716400000357

MacWhinney, B. (2004). Parameters or cues? *Bilingualism: Language and Cognition*, *7*(1), 35–36. https://doi.org/10.1017/S1366728904001233

MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

MacWhinney, B. (2022a). The Competition Model: Past and Future. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A Life in Cognition* (pp. 3–16). https://doi.org/10.1007/978-3-030-66175-5_1

MacWhinney, B. (2022b). The Competition Model: Past and Future. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 3–16). https://doi.org/10.1007/978-3-030-66175-5_1

Martin, A. E. (2016). Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. *Frontiers in Psychology*, *7*(February), 1–17.

https://doi.org/10.3389/fpsyg.2016.00120

Mayberry, R., & Kluender, R. (2018). Rethinking the critical period for language: New insights into an old question from American Sign Language. *Bilingualism: Language and Cognition*, *21*(5), 886–905. https://doi.org/10.1017/S1366728917000724

McFadden, T. (2003). On Morphological Case and Word-Order Freedom. *Annual Meeting of the Berkeley Linguistics Society*, *29*(1), 295. https://doi.org/10.3765/bls.v29i1.997

Meltzer-Asscher, A. (2011). *Adjectives and Argument Structure* (Tel Aviv University). https://doi.org/10.1075/cilt.197.04ben

Meltzer-Asscher, A., & Thompson, C. K. (2014). The forgotten grammatical category: Adjective use in agrammatic aphasia. *Journal of Neurolinguistics*, *30*, 48–68. https://doi.org/10.1016/j.jneuroling.2014.04.001

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, *1*(1), 1–7. https://doi.org/10.1080/17588920903373952

Montrul, S. (2012). Bilingualism and the Heritage Language Speaker. In *The Handbook of Bilingualism and Multilingualism* (pp. 168–189). https://doi.org/10.1002/9781118332382.ch7

Myers, J. (2017). Acceptability Judgments. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.333

Nakagome, K., Takazawa, S., Kanno, O., Hagiwara, H., Nakajima, H., Itoh, K., & Koshida, I. (2001). A topographical study of ERP correlates of semantic and syntactic violations in the Japanese language using the multichannel EEG system. *Psychophysiology*, *38*(2), 304–315. https://doi.org/10.1111/1469-8986.3820304

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. https://doi.org/10.1162/jocn.1991.3.2.151

Newmeyer, F. J. (2004). Against a parameter-setting approach to typological variation. *Linguistic Variation Yearbook*, *4*(2004), 181–234. https://doi.org/10.1075/livy.4.06new

NEWPORT, E. L. (2018). Is there a critical period for L1 but not L2? *Bilingualism: Language and Cognition*, *21*(5), 928–929. https://doi.org/10.1017/S1366728918000305

Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., … Von Grebmer Zu Wolfsthurn, S. (2020). Dissociable effects of prediction and integration during language comprehension: Evidence from a largescale study using brain potentials. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1791). https://doi.org/10.1098/rstb.2018.0522

Niv, Y. (2020). On the primacy of behavioral research for understanding the brain. *Current Controversies in Philosophy of Cognitive Science*, 134–149. https://doi.org/10.4324/9781003026273-16

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research*, *18*(2), 130–141. https://doi.org/10.1016/j.cogbrainres.2003.09.010

O'Rourke, B., & Pujolar, J. (2013). From native speakers to "new speakers" – problematizing nativeness in language revitalization contexts. *Histoire Épistémologie Langage*, *35*(2), 47–67. Retrieved from https://www.persee.fr/doc/hel_0750-8069_2013_num_35_2_3457

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

Packard, J. L. (2015). Morphology. In *The Oxford Handbook of Chinese Linguistics*. https://doi.org/10.1093/oxfordhb/9780199856336.013.0045

Paradis, M. (1990). Language lateralization in bilinguals: Enough already! *Brain and Language*, *39*(4), 576–586. https://doi.org/10.1016/0093-934X(90)90163-B

Pavlenko, A., & Jarvis, S. (2002). Bidirectional Transfer. *Applied Linguistics*, *23*(2), 190-214+281. https://doi.org/10.1093/applin/23.2.190

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

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. https://doi.org/10.1016/j.bandl.2007.09.005

Pierce, L. J., Klein, D., Chen, J.-K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences*, *111*(48), 17314–17319. https://doi.org/10.1073/pnas.1409411111

Rissman, L., & Majid, A. (2019). Thematic roles: Core knowledge or linguistic construct? *Psychonomic Bulletin and Review*, *26*(6), 1850–1869. https://doi.org/10.3758/s13423-019-01634-5

Roberts, L. (2012). Individual Differences in Second Language Sentence Processing. *Language Learning*, *62*, 172–188. https://doi.org/10.1111/j.1467-9922.2012.00711.x

Sakamoto, T. (2015). 15. Processing of syntactic and semantic information in the human brain: Evidence from ERP studies in Japanese. In *Handbook of Japanese Psycholinguistics* (pp. 457–510). https://doi.org/10.1515/9781614511212-022

Sarfarazi, M. (1999). Visual Event Related Potentials Modulated by Contextually Relevant and Irrelevant Olfactory Primes. *Chemical Senses*, *24*(2), 145–154. https://doi.org/10.1093/chemse/24.2.145

Sassenhagen, J., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2014). The P600-as-P3 hypothesis revisited: Single-trial analyses reveal that the late EEG positivity following linguistically deviant material is reaction time aligned. *Brain and Language*, *137*, 29–39. https://doi.org/10.1016/j.bandl.2014.07.010

Schacht, A., Sommer, W., Shmuilovich, O., Martínez, P. C., & Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE*, *9*(3), 1–7. https://doi.org/10.1371/journal.pone.0091226

Schlesewsky, M., & Bornkessel, I. (2004). On incremental interpretation: Degrees of meaning accessed during sentence comprehension. *Lingua*, *114*(9–10), 1213–1234. https://doi.org/10.1016/j.lingua.2003.07.006

Schmeißer, A., Hager, M., Gil, L. A., Jansen, V., Geveler, J., Eichler, N., … Müller, N. (2015). Related but different: The two concepts of language dominance and language proficiency. In *Language Dominance in Bilinguals* (pp. 36–65). https://doi.org/10.1017/CBO9781107375345.003

Schütze, C. T. (2016). *The empirical base of linguistics*. https://doi.org/10.17169/langsci.b89.101

Senghas, A., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: The emergence of grammatical devices. *In E. Hughes & A. Greenhill (Eds.), Proceedings of the 21st Annual Boston University Conference on Language Development*, 550–561. Cascadilla Press.

Sherkina-Lieber, M. (2020). A classification of receptive bilinguals. *Linguistic Approaches to Bilingualism*, *10*(3), 412–440. https://doi.org/10.1075/lab.17080.she

Slik, F., Schepens, J., Bongaerts, T., & Hout, R. (2022). Critical Period Claim Revisited: Reanalysis of Hartshorne, Tenenbaum, and Pinker (2018) Suggests Steady Decline and Learner-Type Differences. *Language Learning*, *72*(1), 87–112. https://doi.org/10.1111/lang.12470

Sprouse, J., & Almeida, D. (2017). Design sensitivity and statistical power in acceptability judgment experiments. *Glossa*, *2*(1), 1–32. https://doi.org/10.5334/gjgl.236

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Steinhauer, K., Royle, P., Drury, J. E., & Fromont, L. A. (2017). The priming of priming: Evidence that the N400 reflects context-dependent post-retrieval word integration in working memory. *Neuroscience Letters*, *651*, 192–197. https://doi.org/10.1016/j.neulet.2017.05.007

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Su, I. R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112. https://doi.org/10.1017/S0142716401001059

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism*, *17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*(1), 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

Titone, D., & Baum, S. (2014). The future of bilingualism research: Insufferably optimistic and replete with new questions. *Applied Psycholinguistics*, *35*(5), 933–942. https://doi.org/10.1017/S0142716414000289

Tsimpli, I. M. (2014). Early, late or very late? *Linguistic Approaches to Bilingualism*, *4*(3), 283–313. https://doi.org/10.1075/lab.4.3.01tsi

Vasishth, S., Suckow, K., Lewis, R. L., & Kern, S. (2010). Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures. *Language and Cognitive Processes*, *25*(4), 533–567. https://doi.org/10.1080/01690960903310587

Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, *24*(7–8), 1180–1226. https://doi.org/10.1080/01690960802159937

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The Role of Animacy in Online Argument Interpretation in Mandarin Chinese. In *Case, Word Order and Prominence* (Vol. 40, pp. 91–119). https://doi.org/10.1007/978-94-007-1463-2_5

Weber, K., & Lavric, A. (2008). Syntactic anomaly elicits a lexico-semantic (N400) ERP effect in the second language but not the first. *Psychophysiology*, *45*(6), 920–925. https://doi.org/10.1111/j.1469-8986.2008.00691.x

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

White, L., & Genesee, F. (1996). How native is near-native? The issue of ultimate attainment in adult second language acquisition. *Second Language Research*, *12*(3), 233–265. https://doi.org/10.1177/026765839601200301

Wong, B., Yin, B., & O'Brien, B. (2016). Neurolinguistics: Structure, Function, and Connectivity in the Bilingual Brain. *BioMed Research International*, *2016*, 1–22. https://doi.org/10.1155/2016/7069274

Wu, F., & He, Y. (2015). Some Typological Characteristics of Mandarin Chinese Syntax. In *The Oxford*

Handbook of Chinese Linguistics. https://doi.org/10.1093/oxfordhb/9780199856336.013.0020

Xu, L. (2015). Topic Prominence. In *The Oxford Handbook of Chinese Linguistics*. https://doi.org/10.1093/oxfordhb/9780199856336.013.0072

Manuscript I

The child the apple eats – processing of argument structure in Mandarin verb-final sentences

Max Wolpert[1,2], Jiarui Ao[1], Hui Zhang[3], Shari Baum[2,4], Karsten Steinhauer[2,4]

Affiliations:

[1]Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

[2]Centre for Research on Brain, Language and Music, Montreal, QC, Canada

[3]School of Foreign Languages and Cultures, Nanjing Normal University, Nanjing, Jiangsu, China

[4]School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada

1 Abstract

Mandarin Chinese has an unusual combination of linguistic features that have consequences for expression of argument structure. These features include flexible word order with virtually no inflection, as well as the coverbs BA and BEI, which explicitly assign thematic roles. Taking advantage of Mandarin's grammatical properties, we utilized a forced agent-assignment task to characterize native, monolingual processing of verb-final Mandarin sentences with competing cues for agent assignment. In addition to participants' binary agent assignments, we collected reaction time and electroencephalography data to capture real-time processing throughout the sentence. We considered predictions from three models of sentence processing that have considered data from Mandarin: the Competition Model, the extended Argument Dependency Model (eADM), and the Bag of Arguments account. Our behavioral results showed that participants had no inherent preference for agent position, and the coverbs BA and BEI were the strongest cues for agent selection. The interactions between cues were more nuanced than the linear cue strength hierarchy reported by previous Competition Model studies, and we also observed individual differences in reliance on the cues of coverbs and plausibility. In the case of role reversals, where sentence structure contradicts semantic knowledge, we observed an N400 effect without a subsequent semantic P600. This ERP pattern is more consistent with the eADM than the Bag of Arguments account.

KEYWORDS: Mandarin, role reversals, Competition Model, extended Argument Dependency Model, Bag of Arguments, argument structure

2 Introduction

All languages manage to communicate who did what to whom, but languages differ in how they express the doer and receiver of a verb's action. The sentence elements that signal these structural meanings are known as cues, and cues' strength and availability depend on the features of a given language (Li & MacWhinney, 2012; Martin, 2016). Cues such as word order (Dryer, 2013b), case marking (Iggesen, 2013), and verb inflection (Bickel & Nichols, 2013) tend to vary in systematic ways across the worlds' languages (Comrie, Dryer, Gil, & Haspelmath, 2013). Mandarin Chinese (henceforth Mandarin) is notorious for challenging the systematicity of these crosslinguistic patterns because these three cues are either unreliable or absent (Li, Bates, Liu, & MacWhinney, 1992; Li & Thompson, 1989).

Word order is a strong cue in languages like English (MacWhinney, Bates, & Kliegl, 1984), where the linear position of arguments in relation to the verb determines meaning. In most English transitive sentences, arguments directly preceding verbs are agents (the entity that carries out the action of the verb) and arguments directly following verbs are patients (the entity that receives the action of the verb). This means that for sentences (1a) and (1b), the first noun is interpreted as the agent, even when *the apple* is not a plausible agent as in (1b).

(1a) The child eats the apple.

(1b) The apple eats the child.

Case marking, instead of word order, is a stronger cue in languages like German (MacWhinney et al., 1984). A case system marks which argument plays which role through morphological inflection on the nouns or their corresponding articles, as in the German sentence (2).

(2)[7]	*Den*	*Apfel*	*isst*	*der*	*Mann.*
	The_{ACC}	apple	eats	the_{NOM}	man
	"The man eats the apple."				

[7] ACC = accusative case; NOM = nominative case

In (2), case marking on the articles means that *der Mann* (the man) is understood as the agent and *den Apfel* (the apple) the patient, even though German canonical word order works against this interpretation. Yet other languages use number or person agreement between verbs and their arguments to assign agent and patient status, as in the Italian sentence (3) (example from Kasparian & Steinhauer, 2017).

(3)[8] *Il poliziotto che i ladri arresta registra i nomi.*
The policeman that the thieves arrests$_{3rdPersonSing}$ registers the names.
"The policeman that arrests the thieves registers the names."

In (3), the 3rd person singular verb form *arresta* (arrests) agrees with the singular agent *il poliziotto* (*the* policeman) and not the plural patient *i ladri* (the thieves). In this example, verb agreement drives argument structure assignment, even though the patient appears in front of the verb, which is the subject position according to Italian canonical word order. Sentences (2) and (3) also reflect that extensive morphological inflection is associated with free word order (McFadden, 2003).

From these examples, it is clear that different cues have different weights in different languages, and in the case of cue conflict, a stronger cue may override weaker cues. But what about a language that does not have strict word order, a case system, or verb agreement? How can speakers possibly understand who did what to whom when these three cues are not available? In the present study, we considered the case of Mandarin, a language that has no tense, no case, and virtually no morphological inflection, all while having flexible word order (Li & Thompson, 1989).

Mandarin has an uncommon combination of linguistic features that impact argument structure processing. Critically, Mandarin accepts omission of the subject or the object of a sentence (Liu, 2014), so a sentence like (4) changes meaning depending on context. The lack of cue reliability for pre-verbal nouns is in contrast to English, where nouns preceding verbs are usually agents (Bates & MacWhinney, 1989; Li et al., 1992).

---

[8] 3rdPersonSing = 3rd person singular inflection

(4) *鸭子吃了。*

duck eat LE

"The duck ate something." or "Something ate the duck."

Although Mandarin canonical word order is usually understood as subject-verb-object (Li, 1998), object-subject-verb and subject-object-verb orders are also permitted, and these verb-final sentences may be becoming more preferred (Li & Thompson, 1976 but see Sun & Givon, 1985 for opposing evidence). Some linguists also question whether the concept of subject and object is appropriate for Mandarin (LaPolla, 1993), suggesting that 'topic' and 'comment' are more fitting labels. Even the syntactic category of many words is highly flexible, with the same word often able to serve as noun, verb, or adjective based purely on context (Huang, Chen, & Shen, 2002). Such flexible syntax makes it challenging to create an "unambiguously ungrammatical" Mandarin sentence (Lu et al., 2000), and some researchers have labeled Mandarin a "semantics-based" as opposed to a "syntax-based" language (Su, 2001).

Mandarin has another feature important for argument structure: the coverbs BA and BEI. These coverbs can occur in verb-final sentences to disambiguate agent and patient roles. BA assigns agent status to its preceding noun, as in *孩子把苹果吃掉了* "child BA apple ate", resulting in subject-object-verb word order. BEI assigns patient status to its preceding noun, as in *苹果被孩子吃掉了* "apple BEI child ate", resulting in object-subject-verb word order. These word orders are also possible without the coverbs BA and BEI, as already indicated, but may be pragmatically restricted (P. Li et al., 1992). There has been debate about the syntactic categories of BA and BEI (e.g., Bender, 2000; Ting, 1998), but for simplicity we refer to them here as coverbs (Li & Thompson, 1974). Although BA and BEI are both common in verb-final clauses, they do have differences in structure and usage. BA must be followed by a noun phrase and is limited in which verbs it can be used with, and BEI is typically analyzed as a passive construction and can be followed directly by a verb (Deng, Mai, & Yip, 2018; Huang, Li, & Li, 2009). Despite these differences, BA and BEI are powerful cues for argument structure assignment and each assigns a different word order interpretation in noun-noun-verb sentences (although there is discussion of their structural properties and whether they assign theta roles in theoretical analysis, as in Huang et al., 2009).

These grammatical peculiarities of Mandarin make it an ideal language for expanding our understanding of language processing universals. Broadly speaking, there are two possibilities for human sentence processing: it is fundamentally the same regardless of the language or it shows qualitative differences across languages. The case for equivalent processing assumes that because all humans rely on the same cognitive architecture, processing strategies should stem from the same heuristics and mechanisms. Because language processing occurs in real time and memory is limited, parsers must be efficient and rapidly take in semantic information in meaningful syntactic structures (e.g., Christiansen & Chater, 2015; Ferreira & Patson, 2007). However, wide variability across languages (Evans & Levinson, 2009) has the potential to impact processing. The scope of crosslinguistic variability must be appreciated to construct valid models of sentence processing and to understand how this processing is shaped by constraints of the human brain (Hahn et al., 2020).

Argument structure cues are a prime target for investigating basic mechanisms of sentence processing across languages. According to theoretical linguistics accounts, all human languages have verbs, and the verb phrase is a basic unit of the predicate of a sentence (Chomsky, 1957a; Croft, 2011). To construct more complex sentences, languages add noun phrases as arguments to a verb, and the verb assigns theta roles, such as agent and patient, to its arguments (Chomsky, 1993). There is debate about the cognitive reality of theta roles (e.g., Ziegler, Snedeker, & Wittenberg, 2018), but even scientists critical of theta roles acknowledge that for a given verb, parsers must assign the basic notions of doer and receiver, which have been referred to as proto-agent and proto-patient (Dowty, 1991). These assignments may be influenced by parsing heuristics, such as initially assuming that the first noun of a sentence is the agent (Krebs, Malaia, Wilbur, & Roehm, 2018; L. Wang et al., 2009). To focus on the cognitive nature of theta roles at a basic level, we refer only to agents and patients—the doer and receiver of a transitive verb, respectively—unless otherwise specified, even when a given noun may instead be interpreted as theme, experiencer, or other role in theoretical linguistics accounts. For instance, in sentence (1a), "the child", who performs the action of the verb, is given the theta role of agent, while "the apple", which undergoes the action of the verb, is given the theta role of patient. In sentence (1b) in English, these theta roles are reversed by interchanging the order of the first noun and the second noun. In sentences (2) and (3), however, we see that word order is not a crosslinguistic universal for signaling agent and patient. As such, expressing the role-reversed meaning of a sentence may require modifying subject-verb agreement or case marking. These sentences with role reversals, also known as semantic reversals or semantic illusions, have provided significant challenges for sentence processing models, especially for explanations of electrophysiological data.

2.1 Role Reversals

Syntax and semantics have often been conceived as separate cognitive modules. In his 1957 book *Syntactic Structures*, Chomsky emphasized the independence of grammar and meaning (Chomsky, 1957b). His famous example sentence "colorless green ideas sleep furiously" demonstrates that a meaningless sentence can still comply with a language's grammatical rules.

Many psycholinguists have built on this syntax-semantics divide in models of sentence processing, and electroencephalography (EEG) methods have provided important data for these models. The syntax-semantics dichotomy is evident in the initial reports of the N400 and the P600 event-related potential (ERP) components. The amplitude of the N400 component was first proposed as an index of semantic processing, where semantic anomalies elicited larger N400 amplitudes, such as the word *socks* in "#He spread the warm bread with socks"[9] (Kutas & Hillyard, 1980). The P600 component was likewise singularly interpreted as elicited by syntactic anomalies, including phrase structure violations in the sentence "*Max's of proof the theorem" (Neville et al., 1991) and garden path sentences like "The baker trusted to bake the cake won many awards" (Osterhout & Holcomb, 1992). The P600 was even referred to as a "syntactic positive shift" (Hagoort et al., 1993). Some serial models of sentence processing have taken these early interpretations of the N400, P600, and other ERP components to propose discrete, sequential stages of syntactic and semantic processing (e.g., Friederici, 2002).

However, in the early 2000s, reports of semantic anomalies eliciting P600 effects challenged the separation between syntax and semantics in ERP studies. This effect was first reported in Dutch (Hoeks, Stowe, & Doedens, 2004; Kolk, Chwilla, Van Herten, & Oor, 2003) and English (Kim & Osterhout, 2005), where semantically anomalous but grammatically well-formed sentences (e.g., *the javelin has the athletes thrown* and *the hearty meal was devouring the kids*) elicited a P600 and not an N400 effect at the verb. Because these sentences were syntactically correct, this "semantic P600 effect" was incompatible with the idea that the N400 and P600 index separate semantic and syntactic processing modules. Since these initial studies, the

[9] The symbol # indicates a semantic or pragmatic anomaly; the symbol * indicates ungrammaticality

semantic P600 without an N400 effect has also been reported in Japanese (Sakamoto, 2015) and Mandarin (Chow & Phillips, 2013; Chow, Lau, Wang, & Phillips, 2018; Ye & Zhou, 2008).

Several researchers have put forward explanations of semantic P600 effects. Kim & Osterhout (2005) proposed that certain combinations of nouns and verbs are so related that parsers initially disregard syntactic structure and pursue a semantics-based interpretation. This semantic attraction effect occurs independently of syntactic processing, and when parsers try to integrate their interpretation with the verb assigning an implausible thematic role, they detect a syntactic anomaly instead of a semantic one. Later accounts have also proposed parallel, independent syntactic and semantic processing streams (Kolk & Chwilla, 2007; Kuperberg, 2007; Van Herten, Kolk, & Chwilla, 2005). Alternatively, there may be a single processing stream that does not distinguish between semantics and syntax, but instead first retrieves meaning independent of plausibility, as indexed by the N400, and then integrates that meaning into the sentence context, as indexed by the P600 (Brouwer et al., 2021, 2012).

Further complicating matters, several experiments have shown diverse ERP effects for role reversals. For instance, the type of verb can impact processing of role reversals, such that English action verbs (like “to throw”) elicit semantic P600s while English experiencer verbs (like “to love”) have been shown to elicit biphasic N400-P600 effects (Bourguignon, Drury, Valois, & Steinhauer, 2012). ERPs elicited by the nouns of role reversals in sentences like “Fred eats a restaurant in a sandwich” have shown N400 effects with no subsequent semantic P600 (Kos, Vosse, van den Brink, & Hagoort, 2010). Experimental task and presentation modality also vary considerably among experiments on role reversals and both of these factors have been shown to impact ERPs (Kyriaki, Schlesewsky, & Bornkessel-Schlesewsky, 2020).

Among semantic P600 experiments, there is an additional factor that is often ignored: crosslinguistic differences. A tacit assumption across many accounts of role reversals is that parsers across all human languages process arguments via the same underlying mechanisms (e.g., Brouwer et al., 2012; Kuperberg, 2007; van de Meerendonk, Kolk, Chwilla, & Vissers, 2009). These accounts would predict that conditions evoking N400 effects and semantic P600s should be the same across languages. Given its linguistic features, Mandarin serves as an important test case for accounts of argument structure processing and role reversals. We are aware of three models of argument structure processing that explicitly incorporate evidence from Mandarin parsing.

The first two models place crosslinguistic variability at the forefront of their explanations, the Competition Model (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; MacWhinney, 2005) and the extended Argument Dependency Model (eADM, Bornkessel & Schlesewsky, 2006a). These models both center their processing architecture around linguistic cues and their variation across languages, with the Competition Model relying primarily on behavioral data and the eADM extending consideration to ERPs. The third model we refer to as the Bag of Arguments account (Chow, Smith, Lau, & Phillips, 2016; Chow, Momma, Smith, Lau, & Phillips, 2016), which encompasses evidence from multiple studies by the same research group over the past decade. The Bag of Arguments account is based on ERP evidence for prediction of upcoming verbs during real-time sentence processing. This account considers that results from Mandarin should extend to other languages. Below, we introduce each of these three models in more detail.

2.2 Competition Model

The Competition Model (first described in Bates et al., 1982; extended in MacWhinney, 2005) is an emergentist account of language learning and processing which focuses on cues as a central component of sentence processing (MacWhinney, 2022a). Researchers using the framework of the Competition Model have often employed a specific forced-choice task for agent selection in a systematic cue comparison across languages, wherein all possible cues of a given language are combined in all permutations to determine their relative strength and validity in offline judgments. Agent selection differs between languages depending on how often given cues are present–dubbed cue availability–and how often given cues correctly indicate the agent–dubbed cue reliability; in the case of competition between cues, the primary cue driving argument structure assignment is said to have greater cue strength (MacWhinney, 2018).

For example, English possesses the cues of word order, animacy, and agreement. In a sentence such as “The horse kicks the carrots”, all three cues work towards interpreting “the horse” as the agent. If we adjust the word order cue, we have the sentence “The carrots kicks the horse,” where agreement and animacy work towards the interpretation of “the horse” as the agent, while word order works towards the interpretation of “the carrots” as the agent. Multiple experiments have shown that native English speakers consistently rely on word order in these cases of cue competition and that animacy and agreement are equal in strength (Liu, Bates, & Li, 1992; MacWhinney et al., 1984; Su, 2001). In Mandarin, however, parsers have been shown to rely on the following cues in order of decreasing strength: BEI, animacy, word order, and BA, although BA may be as strong as word order in noun-noun-verb sentences (Li et al., 1992). However, it is worth noting that many of these data are based on sentence materials that had no plausible inanimate agents. The lack of plausible inanimate agents may have overestimated the role of animacy in argument assignment, as participants could develop a strategy of excluding inanimate nouns from consideration as agents.

The Competition Model continues to evolve (Li & MacWhinney, 2012) and inform models of language processing (e.g., Bornkessel & Schlesewsky, 2006b; Martin, 2016), and has incorporated evidence from typologically diverse languages (MacWhinney, 2022a). The primary experimental work informing the Competition Model has been offline behavioral judgments of argument assignment after parsing complete sentences. Therefore, although the mechanisms of cue processing occur in real time, the Competition Model has considered limited evidence about cue interpretation as it occurs throughout a sentence in online parsing.

2.3 extended Argument Dependency Model

Drawing on inspiration from the Competition Model, the eADM uses neuroimaging studies of real-time sentence processing to explain the impact of crosslinguistic variability. The eADM posits hierarchical stages of real-time argument structure processing, where nouns are first assigned proto-theta roles and verbs are linked to any existing arguments (Bornkessel-Schlesewsky et al., 2011; Bornkessel-Schlesewsky & Schlesewsky, 2016; Bornkessel & Schlesewsky, 2006a). Central to the model, the relational processing for nouns and verbs is dependent on language-specific cue weighting, where cues such as animacy, word order, and case are encoded and used for argument structure building according to the cues' strength and validity in a given language. This relational processing is accompanied in parallel by a plausibility computation to check that the verb's arguments are nominally plausible. Problems in these relational processing and plausibility computations can lead to N400 effects. P600 effects, however, are limited to subsequent well-formedness and repair computations.

This model architecture leads to an explanation of semantic P600 effects that is grounded in crosslinguistic diversity (Bornkessel-Schlesewsky & Schlesewsky, 2008). The model contends that the N400 effect can be absent for role reversals if language-specific cue weighting satisfies the heuristics of the plausibility computation. This cue weighting is language specific, and the eADM authors suggest that languages span two categories for cue weighting: sequence-dependent, for languages that have strict word order, and sequence-independent, for languages that have flexible word order. For sequence-dependent languages like English or Dutch, parsers are driven almost entirely by word order—neglecting unreliable cues like animacy, case, or verb inflection—so their plausibility check in step two can be nominally satisfied if the verb and its arguments are related. This means that for role reversal sentences like “The hearty meal was devouring the kids” (Kim & Osterhout, 2005), the parser is initially satisfied that the verb *devour* and its argument *meal* are related and only detects the anomaly at the generalized mapping step of stage three (Bornkessel-Schlesewsky & Schlesewsky, 2008). For sequence-independent languages, where word order is not a reliable cue for argument structure, parsers must instead rapidly incorporate information from other cues into their plausibility checks (Bornkessel-Schlesewsky et al., 2011). Since these cues are independent of linear position, processing conflict can already occur in stage two, which produces an N400 effect. The crosslinguistic variability of sequence-independent languages is complicated by findings of role reversals eliciting biphasic N400-P600 effects for German and Icelandic (Bornkessel-Schlesewsky et al., 2011). The eADM proposes that the semantic P600 is elicited by role reversals in languages that have no possible grammatical continuation of the sentence; because Mandarin has prenominal relative clauses, there are possible grammatical continuations of role reversal sentences, and hence only an N400 effect (Bornkessel-Schlesewsky et al., 2011).

Role reversals and ERPs time-locked to verbs are not the only focus of the eADM’s predictions. Pre-verbal nouns also have the potential to impact online processing based on cues like animacy. For instance, it has been proposed that sentence-initial nouns are parsed as subjects by default (Krebs et al., 2018; L. Wang et al., 2009), so sentence-initial inanimate nouns may elicit greater N400 effects than more ideal sentence-initial animate nouns at stage two of the eADM (Bourguignon et al., 2012; Kyriaki et al., 2020; Philipp, Graf, Kretzschmar, & Primus, 2017, but no such effect in Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008). Additionally, in languages that permit verb-final word order, like Mandarin, there should be a behavioral preference and easier online processing for inanimate objects and animate subjects (L. Wang et al., 2012).

The eADM seeks to explain the role of crosslinguistic diversity in sentence processing (Bornkessel & Schlesewsky, 2006a) and accounts for many of the reported semantic P600 effects (Bornkessel-Schlesewsky & Schlesewsky, 2008; Brouwer et al., 2012). However, the model implicitly only considers native speakers and monolinguals. Unlike the Competition Model, the possibility for forward and backward transfer between first and second languages is not addressed, even though many of the supporting data for the eADM are likely from speakers of more than one language.

2.3a Bag of Arguments

The focus of the Bag of Arguments model is on generation of predictions during real-time sentence processing. In this approach, the amplitude of the N400 component is considered to index how much a given word's semantic representation has been preactivated prior to retrieval (Chow et al., 2018). While the Bag of Arguments does not presuppose the specific mechanisms of prediction (Chow et al., 2016; Chow et al., 2018), the model suggests limitations on when and what sources of information can be used. Specifically, parsers chunk a sentence into clauses and rapidly consider nouns' lexical meaning for predicting verbs within the same clause (Chow et al., 2016); however, information about structural roles like agent and patient take longer to impact parsing, at least 800 ms (Chow et al., 2018; Liao & Lau, 2020).

Unlike the eADM and the Competition Model, the Bag of Arguments model does not assume any crosslinguistic differences for processing. All three models consider Mandarin data, but for the Bag of Arguments model, Mandarin serves as a case example of verb-final structures, and results from Mandarin are expected to extend to processing of other languages. Although the authors acknowledge that primary drivers for assigning argument roles are different in different languages, the authors suggest that any cue for argument role assignment can only influence verb prediction if there is at least 800 ms between arguments and the verb (Chow et al., 2018). To our understanding, this 800-ms limit is an experimental observation and not a theoretical limit, so the Bag of Arguments model may be compatible with greater variability at both the group and individual levels for the precise amount of time required.

For the case of role reversals, the Bag of Arguments model explains the semantic P600 effect in terms of the lack of an N400 effect, or that the N400 is "blind" to role reversals (Chow & Phillips, 2013). According to the model, role reversals only elicit an N400 effect if two conditions are met. First, the verb and its arguments are combinable and highly predictable, meaning there exists a plausible interpretation where each noun is either agent or patient and there is strong lexical association between the nouns and the verb (Chow & Phillips, 2013). Second, parsers need at least 800 ms to use structural role information to predict upcoming verbs

(Chow et al., 2018). If these two conditions are met, then role reversals should elicit a biphasic N400-P600 effect; if either of these conditions is false, however, then there will be a semantic P600 effect with no N400 difference (Chow et al., 2018). For the P600 effect in general, the Bag of Arguments model suggests that a posterior P600 will be elicited in the event of plausibility and congruence violations, and the P600 is not modulated by prediction (Chow et al., 2018). If a late positivity is frontal, however, the positivity is supposed to index some element of prediction and to be functionally different from a posterior P600, even if the two effects occur in the same time window (Chow et al., 2018).

With its focus on prediction, the Bag of Arguments has made important contributions towards understanding how parsers handle argument structure information in real-time sentence processing (Chow et al., 2016). In terms of cues for argument structure assignment in Mandarin, the Bag of Arguments has focused on sentences with verb-final word order using the coverb BA; the coverb BEI, however, has not been considered (Chow & Phillips, 2013; Chow et al., 2018). Additionally, like the eADM, studies cited in support of the Bag of Arguments model have not reported controlling for participants' knowledge of other languages.

2.4 Present Study

Given Mandarin's unusual patterns for argument structure assignment, how do Mandarin parsers manage conflicting cues in real-time sentence processing? To address this question, the present study characterized native processing of transitive, verb-final sentences in Mandarin with four competing cues: Structure, Agent Animacy, Reversibility, and Order. To manipulate Structure, we used sentences with one of the coverbs BA and BEI or the plain noun-noun-verb structure with no coverb. For Agent Animacy, we created sentences whose plausible agent was either animate or inanimate (e.g., *仆人镜子擦亮了* "servant mirror polished" or *鸟笼喜鹊困住了* "birdcage magpie trapped"). For Reversibility, we created irreversible sentences that had only one plausible interpretation (e.g., in the sentence *仆人镜子擦亮了* "servant mirror polished", "servant" is the only plausible agent) and reversible sentences that had two equally plausible agents (e.g., *技工老板举报了* "technician boss denounced", both "technician" and "boss" are equally plausible as agent). To strengthen the cue of Reversibility, all reversible sentences had two nouns with shared animacy status, thus maximizing ambiguity, and all irreversible sentences had two nouns with contrasting animacy status, thus maximizing the difference in plausibility. Lastly, we manipulated word order (referred to as Order) so that a given noun appeared in either first or second position in our noun-noun-verb sentences. Note that for irreversible sentences, we expected that participants would choose the plausible agent regardless of Order; for reversible sentences, we expected that the effect of Order may show a preference for object-subject-verb interpretation as per prior findings (P. Li et al., 1992; Su, 2001; Wang et al., 2012; cf. Yu & Tamaoka, 2018).

This orthogonal comparison of cues was inspired by tasks used in previous Competition Model experiments. An example set of sentences is shown in Table 1. Combining the irreversible sentences with BA and BEI coverbs permitted the creation of semantically congruent and semantically anomalous sentences, as shown in Table 2.

Table 1. ***Sample sentences for Reversibility and Agent Animacy conditions.*** *Each of the Reversibility and Agent Animacy conditions was crossed with Structure and Word Order so that each noun pair and verb could appear with either noun in first position and with BA, BEI, or plain NNV.*

Reversibility	Agent Animacy	Example sentence
reversible	animate	老板技工举报了。 boss technician denounced
	inanimate	温度气压预测了。 temperature pressure predicted
irreversible	animate	仆人镜子擦亮了。 servant mirror polished
	inanimate	鸟笼喜鹊困住了。 birdcage magpie trapped

*Table 2. **Sample role reversal sentences and their plausible counterparts**. For subject-object-verb word order, BA is the plausible coverb. For object-subject-verb word order, BEI is the plausible coverb.*

Order	Coverb	Example sentence	Plausibility
plausible agent in first position	BA	仆人把镜子擦亮了。 servant BA mirror polished	plausible
	BEI	鸟笼被喜鹊困住了。 birdcage BEI magpie trapped	implausible role reversal
plausible agent in second position	BA	镜子把仆人擦亮了。 mirror BA servant polished	implausible role reversal
	BEI	喜鹊被鸟笼困住了。 magpie BEI birdcage trapped	plausible

Notably, we considered several factors that have not been controlled for in previous experiments. First, to our knowledge, no other ERP study of Mandarin argument structure processing has controlled for participants' knowledge of other languages. There is strong evidence that second language experience can dramatically impact first language argument structure interpretation (Kasparian & Steinhauer, 2017; Liu et al., 1992; Su, 2001), but many studies have tested Mandarin native speakers living in the US or Germany without reporting measures of participants' second language knowledge. Although there are virtually no completely monolingual people on the planet (Edwards, 2006), we recruited participants in Nanjing with minimal knowledge of English and other languages, including non-Mandarin Chinese languages and dialects. Second, many reports of animacy's cue strength have not considered plausible inanimate agents (e.g., Li et al., 1992), despite the potential of inanimate nouns to plausibly carry out the action of many verbs. The absence of inanimate nouns in these studies' stimuli may have overestimated the weight of animacy in cue processing. While Li et al. (1992) included reversible (e.g., *小马踢小牛* "horse kick cow") and irreversible sentences (e.g., *袋鼠洗萝卜* "kangaroo wash carrot"), the agent in irreversible sentences could be chosen based purely on animacy. Third, many Competition Model experiments have relied on stimuli comprised of only several sentences, which could greatly impact the influence of between-item variability. For example, Li et al.'s (Li et al., 1992) study included only six sentence tokens for each of twenty-seven conditions (three levels of word order by three levels of structure by three levels of animacy combinations). Su (2001) improved on this design by adding some plausible inanimate agents to their sentence materials (without manipulating BA and BEI), but their experiment still had few sentence tokens, with only 27 total sentences over 9 conditions (three levels of word order by three levels of animacy). The present experiment builds on these earlier findings by using carefully controlled stimuli with many sentence tokens per condition.

Predictions for both behavioral and ERP results are informed by the three models previously summarized and their previously reported Mandarin results. Studies within the framework of the Competition Model have found that BEI is the strongest cue for argument structure assignment, followed by animacy, BA, and word order in verb-final sentences (Li et al., 1992; Liu et al., 1992; Su, 2001, but note that no effect of word order was found in Miao, 1981). With word order being the weakest cue, it was found to only be used in the absence of all other cues, and the preferred word order interpretation in verb-final sentences was object-subject-verb (Li et al., 1992). According to the eADM, role reversal sentences in Mandarin should elicit an N400 effect, but this may be limited to role reversals with BEI (Bornkessel-Schlesewsky et al., 2011). Aside from effects at the verb, the eADM predicts that sentence-initial inanimate nouns should elicit a greater N400 than sentence-initial animate nouns (Bourguignon et al., 2012; Kyriaki et al., 2020; Philipp et al., 2017). Finally, the Bag of Arguments model predicts a P600 effect for Mandarin role reversal sentences (Chow & Phillips, 2013; Chow et al., 2018), and an N400 should only be present if there is more than 800 ms between the verb and its arguments and high cloze probability (Chow et al., 2016; Chow et al., 2018). These predictions and findings are summarized in Table 3.

Table 3. ***Relevant predictions and prior findings from sentence processing models that have considered Mandarin data.***

Competition Model	• BEI > animacy > BA > word order for cue strength in verb-final sentences (Li et al., 1992) • word order is only used in the absence of all other cues (Li et al., 1992)
eADM	• N400 effect for semantic reversals (Bornkessel-Schlesewsky et al., 2011) ○ specifically, this effect is elicited by role reversals with BEI but not by role reversals with BA • inanimate nouns elicit a larger N400 than animate nouns in sentence-initial position (Bourguignon et al., 2012; Kyriaki et al., 2020; Philipp, Graf, Kretzschmar, & Primus, 2017)
Bag of Arguments	• P600 effect for role reversals (Chow & Phillips, 2013; Chow et al., 2018) ○ evidence so far is only from role reversals with BA, but the prediction should also extend to role reversals with BEI • there will only be an N400 effect if there is high cloze probability and at least 800 ms between the second noun and the verb (Chow et al., 2016; Chow et al., 2018)

3 Methods

3.1 Participants

In total, 39 Mandarin native speakers participated in the study. Of these 39, four were excluded from analysis due to technical problems during the experiment delivery and one was excluded due to failure to stay attentive during the experimental session, resulting in 34 (19 to 25 years old, mean age = 22, SD = 1.9, 19 female) datasets. All participants were recruited via online advertisement and word of mouth in Nanjing and tested at Nanjing Normal University. All participants were right-handed based on the Edinburgh Handedness Inventory (average score = 83, Mandarin version from Yang, Waddington, Adams, & Han, 2018), had normal vision or wore corrective lenses, and did not have any history of neurological disorders. Participants gave written informed consent and were compensated 150 RMB for their time.

To ensure that Mandarin processing was not influenced by other language experience, we limited recruitment to participants who primarily communicated in Mandarin and had limited knowledge of English and other languages, including Chinese languages and dialects. Because

English is a required subject in Chinese primary, secondary, and tertiary schools (J. Yang, 2006), all participants had some previous exposure to English. To minimize the influence of English on processing, we further restricted recruitment to only those who self-reported an English level of 3 or below on a scale from 1 to 6 (1 being no knowledge of English, 6 being nativelike), who did not use English on a regular basis, and who were at or below the College English Test Level 4, which is typically below communicative competence (Yan & Huizhong, 2006). If participants had exposure to a dialect other than standard Mandarin, this was restricted to Northern dialects (e.g., Nanjing, Xuzhou, Nantong, Shandong, Hebei) which are classified as belonging to the Mandarin dialect family and are mutually intelligible (Li & Thompson, 1989). Note that there were exceptionally three participants in the present study who had knowledge of a Chinese language outside of the Mandarin dialect family (Wu, Gan, and Xiang), but they had minimal exposure to these languages in their adult life and primarily used Mandarin.

Participants further completed a detailed language background and usage questionnaire, from which we report summary values in Table 4. To further evaluate their language knowledge, participants also completed a LexTALE lexical decision task in English (Lemhöfer & Broersma, 2012) and Mandarin (I. L. Chan & Chang, 2018). Self-reported proficiency values represent a mean of three separate values for reading, writing, and listening. Exposure percentages represent the self-reported average percent of exposure time from birth to the present. Participants reported percentages in approximately three-year increments throughout their lives, which we then averaged to create an aggregate estimate of lifetime language exposure. Note that the dialect exposure numbers primarily reflect Mandarin dialects (e.g., Nanjing, Nantong, and Xuzhou dialects), which are mutually intelligible with standard Mandarin. Usage percentages represent the average of self-reports of percent of time a language is used in different social contexts, including at school, at the workplace, speaking with friends, and general reading.

Table 4. ***Participants' language experience and proficiency.***

Variable	Mean	Standard deviation	Range
Self-reported Mandarin Proficiency (scale of 1 to 10)	8.6	0.9	6.7-10
Self-reported English Proficiency (scale of 1 to 10)	3.3	2.0	1-7.3
Self-reported Dialect Proficiency (scale of 1 to 10)	6.9	2.5	0-10
English Age of Acquisition (self-reported)	8.9	2.2	5-13
Mandarin LexTALE score (out of 50)	45 out of 50 (90%)	3.8	32-50
English LexTALE score (out of 100)	54 out of 100 (54%)	9.2	40-73
Mandarin exposure percent	46.3	29.0	2.7-98.7
English exposure percent	5.4	5.1	0-17.5
Dialect exposure percent	48.3	30.2	0-97
Mandarin usage percent	92.6	10.2	60.3-100
English usage percent	7.0	9.4	0-35.8
Dialect and other languages usage percent	0.4	1.8	0-9.6

3.2 Materials

We created verb-final sentences with two noun arguments across the two levels of Reversibility (reversible, irreversible) and Agent Animacy (animate, inanimate). Crossing these two factors resulted in four conditions: reversible animate agent, reversible inanimate agent, irreversible animate agent, and irreversible inanimate agent (as summarized in Table 1). To maximize ambiguity in reversible sentences, we chose nouns that shared the same animacy status. We selected 30 transitive verbs for reversible inanimate and irreversible inanimate sentences and 31 transitive verbs for reversible animate and irreversible animate sentences, resulting in 122 unique verbs. To minimize repetitions of sentence materials during the experiment, we selected two noun pairs (noun pair one and noun pair two) for each verb, such that each pair combined with the verb to meet the requirements of the corresponding condition (e.g., reversible with animate agent: *老板技工举报了* "boss technician denounced"; *证人被告举报了* "witness defendant denounced"). These steps resulted in a total of 244 unique noun pairs.

Within these parameters, we further controlled for frequency (using subtitle frequencies from Cai & Brysbaert, 2010) and number of strokes. Frequency values and number of strokes are reported in Table 5. The full sentence materials are reported in the supplementary materials.

Prior to running the EEG experiment, we created an offline questionnaire with our sentence materials to receive information for agent assignment and acceptability ratings from native Mandarin speakers. Note that these questionnaires did not include coverbs, which naturally resulted in decreased acceptability in the absence of a conversational context. Although including BA or BEI in these sentences would increase the naturalness, we wanted to understand how our sentences were comprehended at a purely semantic level and that they would meet a minimum level of acceptability in the NNV structure without a coverb, without systematic differences between conditions in acceptability, as well as ensuring there was a clear semantic direction for our irreversible sentences.

To make our list of stimuli for running the experiment, we next crossed our factors Reversibility and Agent Animacy with Structure (NNV, BA, and BEI) and Order (first and second, representing position of the plausible noun). Note that for reversible sentences, one of the orders was arbitrarily assigned as first so that Order could still be tested and controlled for these items. We assembled ordered lists for presenting sentences to participants. Each of the 122 verbs was used three times, once for each level of Structure, resulting in 366 total sentences. To minimize the effects of repetition, we used the two noun pairs for each verb, so that a given noun only repeated a maximum of once. For example, the two noun pairs *喜鹊鸟笼 / 老鼠箱子* "magpie birdcage / mouse box"and verb *困住* "trap" might appear in the experiment as follows: *喜鹊被鸟笼困住了。* "magpie BEI birdcage trapped"; *喜鹊把鸟笼困住了。* "magpie BA birdcage trapped"; *箱子老鼠困住了。* "box mouse trapped". Note that each sentence ended with the aspect particle LE and a period.

To pseudorandomize our stimuli, we used the program Mix (Van Casteren & Davis, 2006), constraining the randomization such that each level of Structure could repeat a maximum of two times consecutively and a given verb occurred a minimum of 90 trials before or after its previous occurrence. Due to the design of the sentence materials, there was an equal probability of the first or second noun being animate or inanimate and actor or undergoer, so there was no way for participants to develop strategies to predict the role of the nouns until they saw BEI or BA and the final verb. Stimuli were pseudorandomized to maximize distance between repeated verbs (at least 90 items between repetitions) and minimize repetitions of same structure condition to two.

Table 5. ***Controlled variables for sentence materials.***

Reversibility	Agent Animacy	Variable	Pair One -- Mean of noun 1 (upper value) noun 2 (lower value) ± SD	Pair Two -- Mean of noun 1 (upper value) noun 2 (lower value) ± SD	Mean of Verb ± SD
reversible	animate	log(Frequency)	2.36 ± 0.84 2.59 ± 0.58	2.30 ± 0.87 2.66 ± 0.66	2.88 ± 0.51
		Number of strokes	13.0 ± 4.16 14.9 ± 4.35	15.0. ± 4.31 15.4 ± 4.20	16.5 ± 3.08
	inanimate	log(Frequency)	2.73 ± 0.60 2.61 ± 0.67	2.46 ± 0.82 2.59 ± 0.64	2.63 ± 0.57
		Number of strokes	15.6 ± 3.89 16.6 ± 3.96	15.8 ± 3.85 15.4 ± 3.65	16.8 ± 3.88
irreversible	animate	log(Frequency)	2.36 ± 0.80 2.40 ± 0.81	2.48 ± 0.61 2.60 ± 0.67	2.62 ± 0.56
		Number of strokes	14.0 ± 4.30 16.8 ± 4.74	14.1 ± 4.88 17.3 ± 3.59	17.4 ± 3.94
	inanimate	log(Frequency)	2.29 ± 0.90 2.50 ± 0.78	2.36 ± 0.67 2.38 ± 0.83	2.49 ± 0.72
		Number of strokes	15.5 ± 3.99 15.2 ± 4.38	14.7 ± 4.39 14.7 ± 5.08	17.67 ± 4.44

Table 6. ***Pretest results for sentence materials.*** *Acceptability Rating is based on a scale from 1 to 5, where 1 indicated "completely unacceptable" and 5 indicated "completely acceptable". Agent Preference is an average value where each sentence item was presented to participants in both possible orders. A value closer to 1 or 0 indicates a strong preference for one of the nouns, while a value closer to 0.5 indicates no preference for either noun.*

Reversibility	Agent Animacy	Acceptability Rating (Mean ± SD)	Agent Preference (Mean ± SD)
reversible	animate	2.12 ± 1.40	0.44 ± 0.50
	inanimate	2.50 ± 1.58	0.50 ± 0.50
irreversible	animate	2.21 ± 1.48	0.83 ± 0.38
	inanimate	2.13 ± 1.45	0.83 ± 0.38

3.3 Procedure

All parts of the experiment were approved by the McGill Faculty of Medicine Institutional Review Board following the guidelines of the Canadian Tri-Council Policy Statement and by the School of Foreign Languages and Cultures at Nanjing Normal University (南京师范大学外国语学院). After reviewing and signing the consent form, participants sat in a sound-attenuated booth. All stimuli were presented with Presentation® software (Version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) using the Windows XP operating system.

Sentences were presented visually word-by-word, for 650 ms per word and each word followed by a 100 ms inter-stimulus interval (i.e., stimulus onset asynchrony (SOA) = 750 ms). Each trial began with a cue for the participants to blink ("(--)") for 2000 ms, then a fixation cross ("+") for 500 ms. Each sentence ended with the particle LE appearing for 650, before displaying a prompt for the agent assignment task. For the task prompt, the two nouns from the preceding sentence appeared on the screen, with the first noun on the left and the second noun on the right, with "   !!!!!   " between them. Participants were asked to choose which of the two nouns was the agent (*施事*) by pressing the A or L key, corresponding to the left or right noun, respectively. The experiment was divided into four blocks of approximately 90 sentences each, with a scheduled break between each block. Participants could also pause at any response prompt to rest before continuing. The total experiment time, including preparation of the cap and cleanup, lasted from two to three hours.

3.4 EEG Recording and Preprocessing

Participants' EEG was recorded from 32 Ag/AgCl electrodes mounted on an elastic cap according to the international 10-20 system (EASYCAP, Herrsching, Germany). To monitor vertical and horizontal eye movement, one electrode was positioned below the right eye and another to the left of the left eye. Electrode impedance was kept below 10 kΩ. The recordings were amplified online with a bandpass filter of 0.05-100 Hz, referenced online to electrode FCz, and digitized at a sampling rate of 500 Hz.

We used EEGLAB (v2019.1) and ERPLAB (v7.0.0) to preprocess the data. The EEG signal was downsampled to 250 Hz and re-referenced to the average of the linked mastoids (TP9 and TP10). We then used a high-pass filter at a cutoff of 0.1 Hz (FIR filter, Kaiser window, Kaiser beta = 4.89856, transition bandwidth = 0.2 , filter order = 3934) and a low-pass filter at a cutoff of 30 Hz (FIR filter, Kaiser window, transition bandwidth = 10, Kaiser beta = 4.89856, filter order = 80). To correct eye movement artifacts, we decomposed the data using independent component analysis (ICA, runica algorithm in EEGLAB, with the option 'extended', 1). Note that exclusively for the ICA decomposition, we used data that was high-pass filtered at a cutoff of 0.5 Hz (FIR filter, Kaiser window, Kaiser beta = 4.89856, transition bandwidth = 0.2, filter order = 3934) because using a higher high-pass filter on data for ICA decomposition improves the signal-to-noise ratio (Winkler, Debener, Muller, & Tangermann, 2015). The data were also cleaned automatically prior to ICA using the pop_rejcont function (epochlength 2, overlap 1, freqlimit 1-25, threshold 10, taper hamming). The final ICs were then copied to the data filtered at 0.1 and 30 Hz for analysis. We removed a maximum of two ICs per participant, one IC each for vertical and horizontal eye movement.

The signal was then segmented into epochs from -200 to 1000 ms around each critical word (first noun, coverb, second noun, and verb), with pre-stimulus 200 ms baseline correction. To appreciate the ERP changes across the entire sentence, we additionally created whole-sentence epochs with 200 ms pre-onset baselines; including the baseline interval, these epochs spanned 3100 ms for NNV sentences and 3850 ms for BA and BEI sentences. Based on visual inspection of role-reversal sentences in the whole-sentence epochs, we determined that there were important ERP differences occurring before verb onset, which make a pre-onset baseline problematic (Steinhauer & Drury, 2012). To minimize the impact of baseline differences at the verb, we re-epoched the verb time window with a post-stimulus-onset 200 ms baseline correction (baseline interval of 0 to 200 ms), with the assumption that early components in this interval should have minimal difference between conditions. Results from this post-stimulus-onset 200 ms baseline correction are reported in the text because we believe this more accurately reflects verb-linked activity; results from the original pre-stimulus 200 ms baseline correction are reported in supplementary materials for transparency.

Artifact rejection was performed across epochs with a moving window threshold of 80 µV (window size = 500 ms). During review of the artifact rejection process, we determined that electrodes Fp1 and Fp2 were exceptionally noisy across participants and excluded them from analysis. For select subjects whose automatic rejection resulted in greater than 15% of trials being rejected, epoched data were manually inspected to include additional trials and the overall quality of individual datasets. This review resulted in all 34 subjects being included for final analysis (i.e., all individuals had fewer than 15% rejected trials).

3.5 Data Analysis

We analyzed responses and reaction times for the agent assignment task and ERPs time-locked to the onset of target words in sentences. The four factors manipulated in the sentence materials were included in the analysis of each of these measures. The factor Structure was comprised of three levels: NNV, BA, and BEI. Structure was treatment coded such that NNV was the reference level to evaluate the effect of coverbs in relation to sentences with no coverb. The factors Reversibility (reversible and irreversible), Agent Animacy (animate and inanimate) and Order (first and second, denoting position of the plausible noun in irreversible sentences) were sum coded. Unless otherwise noted, these were the factors and contrast coding.

All data analysis was done using *R* version 4.02 (R Core Team, 2017). To account for variability across items and within participants, we computed mixed effects models using the *glmer* function from package *lme4* version 1.1-23 (Bates, Mächler, Bolker, & Walker, 2015), including the optimizer = ‘bobyqa’ parameter. Model coefficients were calculated by maximum likelihood estimates using the Laplace approximation. In the case of the binary data from the agent assignment task, we added the argument family = ‘binomial’ to fit a logistic mixed effects model. To ensure that model effects were interpretable, we limited fixed and random effects to a maximum of three-way interactions, even if there were possible higher order interactions. For random effects structures, all factors with possible variability within items or participants were included in the maximal possible structure. Note that because Agent Animacy and Reversibility did not vary across individual items, these factors were not included in the random structure for item.

All *p*-values were calculated with the Satterthwaite approximation calculated based on Wald *Z*-scores in *lmerTest* package version 3.1-2 (Kuznetsova, Brockhoff, & Christensen, 2017). To construct the maximum possible model and random structure, we used the *buildmer* package version 1.8 (Voeten, 2021) using the direction = 'order' parameter, which adds effects to the model in order of their contribution to log-likelihood. We then again used *buildmer* to do stepwise removal of model variables with the direction = 'backward' parameter to maximize log-likelihood score. These optimized models are reported in the text to maximize power and minimize overfitting (per Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), but the maximal models are reported for reference in the supplementary materials (per Barr, Levy, Scheepers, & Tily, 2013).

Significant interactions were followed up with post-hoc tests for pairwise comparisons using the *emmeans* package version 1.4.8 (Russell, 2019) and Tukey method for adjustment of *p*-values to correct for multiple comparisons. Interactions were visualized with the emmip function from the *emmeans* package. Note that while *p*-values are reported for model results, our inferences and interpretations were not limited to significance testing; instead, we further considered our hypotheses and predictions, effect sizes, and the limitations of data quantity and quality (Loken & Gelman, 2017).

For ease of interpretability, model results are reported with graphical depictions of coefficients and confidence intervals generated by the plot_model function from the *sjPlot* package version 2.8.9 (Lüdecke, 2021); full model outputs, including random effects, are reported in the supplementary materials in tables generated from the tab_model function from the *sjPlot* package. For simplicity, model results reported in the text are limited to significant effects or effects that were related to initial predictions. For data arrangement and general plotting, we used the *tidyverse* package (Wickham et al., 2019), with final figure adjustment performed using the software Inkscape version 0.92 (Inkscape Project, 2020).

3.5a Agent Assignment

Binary agent assignment responses were analyzed with logistic mixed effects models. Because irreversible sentences had a single plausible interpretation, while reversible sentences had two plausible interpretations, we analyzed reversible and irreversible sentences separately.

This allowed us to better understand the effect of plausibility on agent assignment, while limiting model coefficients to a maximum of three-way interactions. For both reversible and irreversible sentences, the maximum specified model included the fixed effects of Structure, Agent Animacy, and Order, random slopes and intercepts for Structure, Agent Animacy, and Order by participant, and random slopes and intercepts for Structure and Order by item. We report coefficients, confidence intervals, and *p*-values on the odds ratios scale, but original tests were performed on the log odds scale. For interpretability, interactions are illustrated on the probability scale.

3.5b Reaction Times

Reaction times for agent assignment responses were analyzed with linear mixed effects models. Reaction times were first cleaned to exclude response times above 10 seconds or below 100 milliseconds. We then cleaned reaction times by condition, limiting to those values within 1.5 standard deviations for each subcondition of Structure, Reversibility, and Animacy (Ratcliff, 1993).These steps resulted in excluding 12.4% of trials from further analysis; we note that some of the excluded trials included instances when participants took breaks before responding. Reaction times were then natural log transformed to ensure that we met assumptions of distribution normality for analysis (Gelman & Hill, 2006). Note that we also analyzed the raw reaction time values (Lo & Andrews, 2015) and results were similar to those found for the log-transformed data; these results are reported in supplementary materials for transparency (Morís Fernández & Vadillo, 2020). The maximum specified model included the fixed effects of Structure, Reversibility, Agent Animacy, and Order, random slopes and intercepts for Structure, Reversibility, Agent Animacy, and Order by participants, and random slopes and intercepts for Structure and Order by item. As an additional step, we also ran a model with the additional factor of Difference Score (the difference between a participant's reliance on plausibility cues and their reliance on coverb cues), which is introduced in the section Individual Differences in Cue Weighting for Agent Assignment. Recent work has demonstrated the importance of individual differences in psychology and language research (Cunnings & Fujita, 2021; Kidd et al., 2018), and including Difference Score in the model explained additional variability in the data. As such, the model including Difference Score is reported in the Results section, with the model excluding Difference Score reported in supplementary materials. We report coefficients, confidence intervals, and *p*-values on the log-transformed scale, but model predictions were back-transformed to milliseconds for interpretability.

3.5c ERPs

As noted above, ERPs were analyzed at the first noun, coverb, second noun, and verb position of the sentence. Condition averages were calculated for each subject and then grand average ERPs were calculated for each condition. These grand average ERPs by condition were used for visual inspection and are represented in all ERP figures in the present study. Statistical models, however, were all based on average amplitudes for specific time windows in single trial epochs.

For the first noun of the sentence, referred to hereafter as noun one, we analyzed average amplitude in the N400 time window from 300-500 ms. This time window analysis was planned a priori based on reports of greater N400 effects for inanimate nouns than for animate nouns (Bourguignon et al., 2012; Kyriaki et al., 2020). At the second-word position of the sentence, there was either another noun (noun two, in the case of NNV sentences), the coverb BA, or the coverb BEI. At this sentence position, we analyzed average amplitudes in the P200 (100 to 300 ms) and N400 (300 to 500 ms) time windows. The P200 time window was selected for analysis after visual observation of large differences in the ERPs between sentence structure types. The N400 time window was analyzed as a validation step to confirm expectations that nouns elicited larger N400 amplitudes than coverbs, thus giving more weight to the unexpected differences in P200 amplitude. At the verb position of the sentence, we analyzed the N400 (300 to 500 ms) and P600 (700 to 900 ms).

For each time window analyzed, we used single trial average amplitude to calculate linear mixed effects models. We first ran models on midline electrodes, including the factor Electrode (Fz, Cz, Pz, Oz), to confirm the presence of effects, and then over all other electrodes on the scalp excluding the midline, with the additional levels of Anteriority (frontal, central, and posterior) and Laterality (right, left). N400 and P600 effects, the primary components investigated in the present study, typically present with a posterior distribution on the scalp (Kaan, 2007); with this in mind, we treatment coded the factors Electrode and Anteriority with the reference levels of Pz and posterior, respectively. In contrast, P200 effects typically have a frontal distribution (Potts & Tucker, 2001), so for models in the P200 time window, we exceptionally used the reference levels of Fz and frontal for Electrode and Anteriority, respectively. Note that for the model specifications below, the factor Electrode was substituted by Anteriority and Laterality for the models over non-midline electrodes.

For noun one, the maximum specified model included the fixed effects of noun one Animacy (animate, inanimate) and Electrode, with random slopes and intercepts included for noun one Animacy and Electrode by participant and by item. For noun two and coverb, the maximum specified model included the fixed effects of Structure (NNV, BA, BEI) and Electrode, with random slopes and intercepts included for Structure and Electrode by participant and by item.

At the verb, only unambiguous role reversal sentences were analyzed (contrasting plausible vs implausible sentences), which limited trials to irreversible BA and BEI sentences (see Table 2). For clarity with respect to predictions about role reversal effects, the factor Order was recoded in terms of Plausibility; BA sentences with the plausible noun in first position were coded as plausible, while BEI sentences with the plausible noun in first position were coded as implausible, with the same logic applied for sentences with the plausible noun in second position. Plausibility was treatment coded with plausible as the reference level. Because there were only two levels of Structure (BA and BEI) with neither level more suited as a reference, Structure was sum coded for this analysis. The maximum specified model for the verb included the fixed effects of Structure, Agent Animacy, Plausibility, and Electrode. Random slopes and intercepts were included for Structure, Agent Animacy, Plausibility, and Electrode by participant, and random slopes and intercepts for Structure, Plausibility, and Electrode by item. Additionally at the verb, we ran models at individual midline electrodes as confirmation for the effects across midline electrodes. These models are reported in supplementary materials.

Because the components of interest in the present study (N400, P600, P200) are typically maximal at or near midline electrodes (Kaan, 2007), we primarily report in the text results from models at the midline; results from lateral electrodes are reported in the text if they show effects beyond the models at midline electrodes. Note that full models for lateral electrodes excluding the midline are reported in the supplementary materials. Additionally, simple effects of the topographical factors Electrode, Anteriority, and Laterality, or interactions involving only these factors, are not reported or discussed in the text because they are not related to the experimental manipulations. Lastly, for models at the verb, we discuss in the text only those effects that included Plausibility because this is the only factor for which we had predictions.

All final analyses were performed on single trial average amplitudes, but average ERPs were calculated by condition for plotting purposes. All figures showing ERP voltage against time and scalp maps reflect these average ERPs and were plotted using the R package *ERPscope* (Herbay, 2022).

4 Results

4.1 Agent Assignment

4.1a Reversible Sentences

Figure 1 shows the coefficients for the logistic mixed effects model. The model intercept was not significant, showing that the likelihood of first-noun choice for NNV sentences was not different from chance ($\beta = 1.25$, SE = 0.24, $Z = 1.20$, $p = 0.232$). The strongest effects were the simple effects of BA and BEI, where sentences with BA were much more likely to elicit first noun agent selection than NNV sentences without a coverb, and sentences with BEI were much less likely to elicit first noun agent selection than NNV sentences with no coverb (BA: $\beta = 16.96$, SE = 5.83, $Z = 8.23$, $p < 0.001$; BEI: $\beta = 0.06$, SE = 0.02, $Z = -7.73$, $p < 0.001$). The lack of any contribution from Order in the optimized model indicates that reversible sentence items showed equal likelihood of either noun serving as a plausible agent in NNV sentences without a coverb. This result was somewhat surprising, as it was inconsistent with our prediction that participants would rely on word order in the absence of other cues.

While there was not a significant simple effect of Agent Animacy, there was an interaction between Agent Animacy and Structure for both BA and BEI sentences (BA: $\beta = 1.24$, SE = 0.12, $Z = 2.18$, $p = 0.03$; BEI: $\beta = 0.65$, SE = 0.06, $Z = -4.66$, $p < 0.001$). Post-hoc pairwise comparisons showed that for BA sentences, plausible animate agents were more likely to elicit first noun agent choice than plausible inanimate agents, while for BEI sentences, plausible animate agents were significantly less likely to elicit first noun agent choice than plausible inanimate agents (BA: $\beta = 1.68$, SE = 0.31, $Z = 2.8$, $p = 0.05$; BEI: $\beta = 0.47$, SE = 0.08, $Z = 4.5$, $p < 0.001$). Additionally, the effect of coverb type on animate agent sentences was stronger than the effect of coverb type on inanimate agent sentences (animate: $\beta = 532.70$, SE = 313.82, $Z = 10.7$, $p < 0.001$; inanimate: $\beta = 148.23$, SE = 138.24, $Z = 8.6$, $p < 0.001$). Overall, this interaction indicates that each coverb was stronger for animate agents than for inanimate agents. This interaction is illustrated on the probability scale in Figure , where animate agent sentences with coverbs were more likely to elicit agent assignments consistent with the coverb cue.

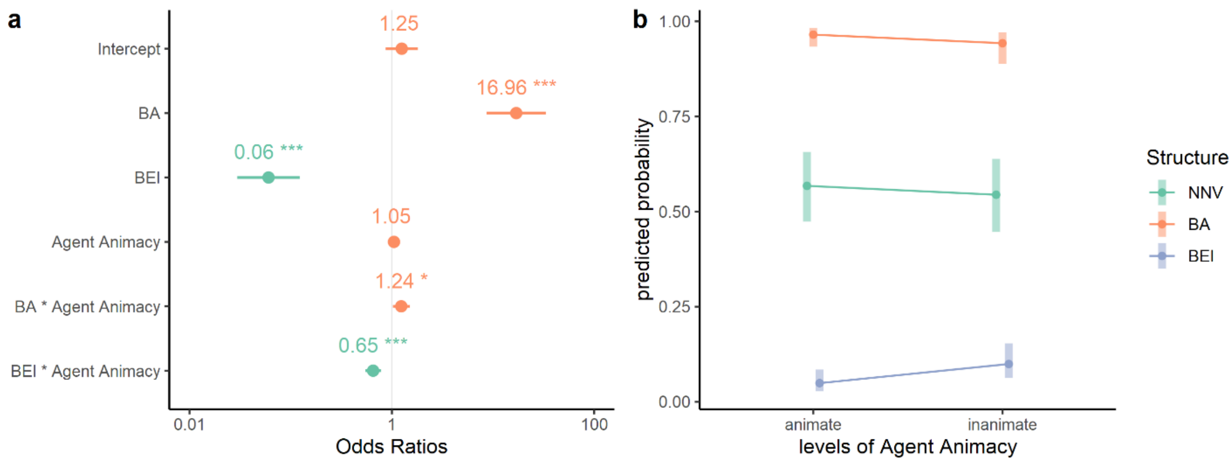

Figure 1. *a) Model coefficients of logistic mixed effects model for first noun agent selection in* ***reversible sentences****. Coefficients are shown as odds ratios. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001; b) Interaction between Agent Animacy and Structure for first agent noun selection in* ***reversible sentences****. Error bars show 95% confidence intervals.*

4.1b Irreversible Sentences

For irreversible sentences, the logistic mixed effects model yielded the coefficients shown in Figure 2. The results for irreversible sentences showed several similarities to those for reversible sentences. First, the model intercept was not significant, showing that the likelihood of first-noun choice for NNV sentences was not different from chance (β = 1.24, SE = 0.31, Z = 0.87, p = 0.382). Second, just as for the reversible sentences, the simple effects of BA and BEI were large, with BA sentences much more likely and BEI sentences much less likely to elicit first noun agent selection than NNV sentences (BA: β = 18.21, SE = 7.00, Z = 7.55, p < 0.001; BEI: β = 0.07, SE = 0.03, Z = -6.21, p < 0.001). Unlike for reversible sentences, the model for irreversible sentences showed a strong simple effect of Order, indicating that the plausible agent was much more likely to be chosen than the implausible agent in NNV sentences (β = 11.16, SE = 2.67, Z = 10.09, p < 0.001). This effect demonstrates that without a coverb the irreversible sentence items had only one plausible interpretation.

The model also revealed a two-way interaction between Structure and Order, which was marginally significant for BA sentences and significant for BEI sentences (BA: β = 0.72, SE = 0.13, Z = -1.86, p = 0.063; BEI: β = 0.42, SE = 0.06, Z = -6.52, p < 0.001). Because NNV is the reference level for the factor of Structure, these two coefficients being less than one indicates that the effect of Order was smaller for BA and BEI than it was for NNV sentences. As seen in the predicted probabilities,, when Order and the coverbs BA or BEI were congruent (e.g., *镜子*

被仆人擦亮了 "mirror BEI servant polished" or *鸟笼把喜鹊困住了* "birdcage BA magpie trapped") the effect on agent assignment choice was similar to that of NNV. However, when Order and coverbs BA and BEI were incongruent (e.g., ,*镜子把仆人擦亮了* "mirror BA servant polished" or *鸟笼被喜鹊困住了* "birdcage BEI magpie trapped"), the coverb cue was stronger than Order, and participants were more likely to interpret the role reversal meaning. This difference was confirmed by post-hoc pairwise comparisons for the difference between congruent and incongruent arrangements of Order with BA and BEI (BA: β = 64.44, SE = 34.90, $Z = 7.69$, $p < 0.001$; BEI: β = 0.05, SE = 0.09, $Z = 6.24$, $p < 0.001$).

There was not a significant main effect of Agent Animacy or a significant two-way interaction between Agent Animacy and Structure; however, there was a significant two-way interaction between Order and Agent Animacy (β = 1.69, SE = 0.13, $Z = 6.73$, $p < 0.001$). This coefficient shows that for NNV sentences without a coverb, the effect of Order was stronger for animate sentences than for inanimate sentences. Post-hoc pairwise comparisons further showed that when the plausible noun was in first position, plausible animate agents (such as *仆人镜子擦亮了* "servant mirror polish") were twice as likely to be chosen as plausible inanimate agents (such as *鸟笼喜鹊困住了* "birdcage magpie trap") (β = 2.24, SE = 0.40, $Z = 4.52$, $p < 0.001$). When the plausible noun was in second position, plausible animate agents (such as "mirror servant polish") were more likely to be chosen than plausible inanimate agents (such as "magpie birdcage trap"), as indicated by a smaller likelihood of first noun agent assignment (β = 0.434, SE = 0.064, $Z = -5.63$, $p < 0.001$). Additionally, the effect of Order for plausible animate agents was almost five times greater (β = 126.71, SE = 62.51, $Z = 9.82$, $p < 0.001$) than for plausible inanimate agents (β = 24.57, SE = 11.58, $Z = 6.79$, $p < 0.001$). These effects demonstrate that plausible animate agents were more preferred for agent assignment than plausible inanimate agents.

Lastly, there was a three-way interaction among Order, Agent Animacy, and Structure only for BEI sentences. This indicates that BEI sentences showed a different pattern for Order and Agent Animacy effects than NNV sentences, while BA and NNV showed approximately the same pattern. We determined that this interaction stems primarily from the effect of Order for BEI sentences between animate and inanimate plausible agents. Namely, for sentences with NNV and BA structures, the effect of Order for plausible animate agents was greater by an order of magnitude (NNV: $\beta = 354$, SE = 183, $Z = 11.34$, $p < 0.001$; BA: $\beta = 195$, SE = 126, $Z = 8.15$, $p < 0.001$) than the effect of Order for plausible inanimate agents (NNV: $\beta = 43.7$, SE = 21.3, $Z = 7.76$, $p < 0.001$; BA: $\beta = 21.3$, SE = 11.7, $Z = 5.59$, $p < 0.001$). For sentences with BEI structure, however, there was relatively minimal difference between the effects of Order for plausible animate agents and plausible inanimate agents (animate: $\beta = 29.5$, SE = 15.8, $Z = 6.30$, $p < 0.001$; inanimate: $\beta = 15.9$, SE = 8.13, $Z = 5.41$, $p < 0.001$). These interactions are illustrated in Figure 2 on the probability scale. Interpreting this interaction in terms of role reversals for sentences with BA or BEI indicates a difference between the two coverbs. BA sentences have a role reversal when the plausible agent is in second position, and it can be seen in Figure 2 that the strength of BA's cue for assigning agent status to the first, implausible noun is stronger when the plausible agent is inanimate (e.g., *喜鹊把鸟笼困住了* "magpie BA birdcage trapped"), where the first noun is predicted to be chosen 80% (SE = 8.4%) of the time, than when the plausible agent is animate (e.g., *镜子把仆人擦亮了* "mirror BA servant polished"), where the first noun is predicted to be chosen 67% (SE = 11.4%) of the time. This difference shows that the cue of Order in BA sentences is stronger when the plausible agent is animate than when the plausible agent is inanimate. In contrast for BEI sentences, there is minimal difference for role reversals with a plausible inanimate agent (e.g., *鸟笼被喜鹊困住了* "birdcage BEI magpie trapped"), where the first noun is predicted to be chosen 31% (SE = 8.3%) of the time, and role reversals with a plausible animate agent (e.g., *仆人被镜子擦亮了* "servant BEI mirror polished"), where the first noun is predicted to be chosen 29% (SE = 7.9%) of the time. For role reversals overall, it appears that BEI reversals are unaffected by the animacy status of the plausible agent, while BA reversals are more plausible (i.e., more likely to elicit first noun agent selection) when the plausible agent is inanimate.

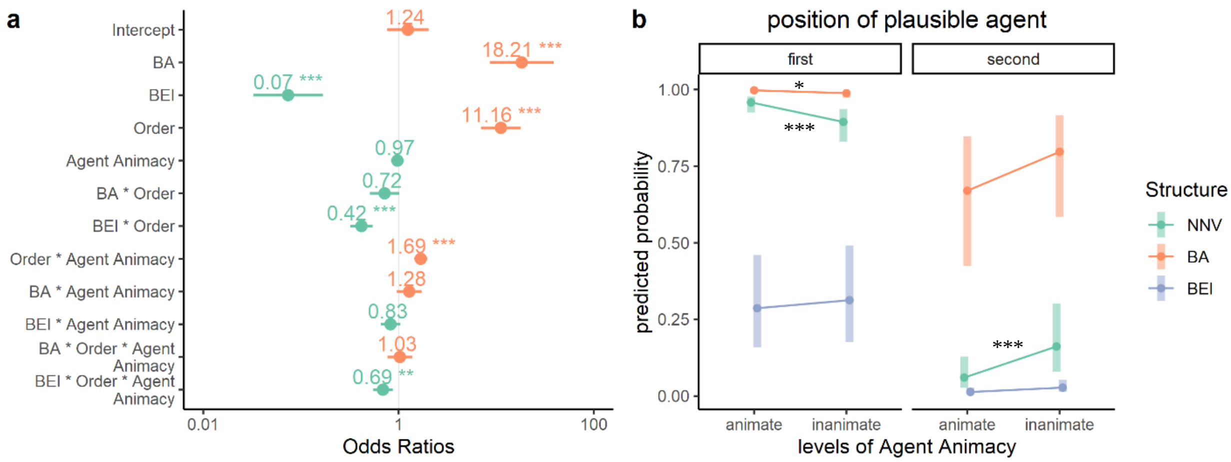

Figure 2. *a) Model coefficients of logistic mixed effects model for first noun agent selection in* ***irreversible sentences****. Coefficients are shown as odds ratios. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001; b) Interaction among Order, Agent Animacy, and Structure for first agent noun selection in* ***irreversible sentences****. Error bars show 95% confidence intervals.*

4.1c Individual Differences in Cue Weighting for Agent Assignment

To better understand the variability in agent assignments, we compared aggregate and individual response patterns by condition (shown in supplementary figures). We observed different strategies in several participants' responses, such as some participants who always followed the cue of plausibility for irreversible sentences and other participants who always followed the cue of coverb when available. To quantify these differences, we computed two scores for each participant. The Coverb Score represented how often participants used coverbs for agent assignment when the coverb was the only cue, i.e., in reversible sentences with BA or BEI. This score was calculated by subtracting the average proportion of first-noun agent selections for all BEI reversible sentences from the average proportion of first-noun agent selections for all BA reversible sentences. A Coverb Score of 1 indicates that a participant always used the coverb cue, whereas a score of 0 indicates that a participant always disregarded the coverb cue. Observed Coverb Scores ranged from 0.2 to 1.

The Plausibility Score represented how often participants use Plausibility as a cue when there was no conflict, i.e., in irreversible sentences without a coverb. Plausibility Score was calculated by subtracting the average proportion of first-noun agent selections for irreversible sentences with the plausible agent as the second noun from the average proportion of first-noun agent selections for irreversible sentences with the plausible agent as the first noun. A Plausibility Score of 1 indicates that a participant always followed the cue of plausibility in the absence of a coverb, while a score of 0 indicates that a participant always disregarded the cue of plausibility. Observed Plausibility Scores ranged from 0 to 1. Note that it is possible for a participant to have a Coverb Score and a Plausibility Score of 1; this would mean that they always followed the coverb cue for reversible sentences and always relied on plausibility in the absence of a coverb.

Examining the distribution of these scores showed variation among participants in their strategies for assigning agents. To capture this variation in a single number, we calculated a Difference Score by subtracting the Coverb Score from the Plausibility Score. Difference Score ranged from -0.98 (for the most coverb-driven) to 0.79 (for the most plausibility-driven). For visualization purposes, we labeled participants with a Difference Score between -0.33 and 0.33 as having a balanced strategy (i.e., the middle third of the values range), where they gave approximately equal weight to the two cues of plausibility and coverb. For participants having a Difference Score of less than -0.33, we labeled them as having a coverb-driven strategy. Lastly, for participants with a Difference Score of more than 0.33, we labeled them as having a plausibility-driven strategy. The distribution of Coverb and Plausibility Scores and the corresponding agent assignment strategy are shown in Figure 3.

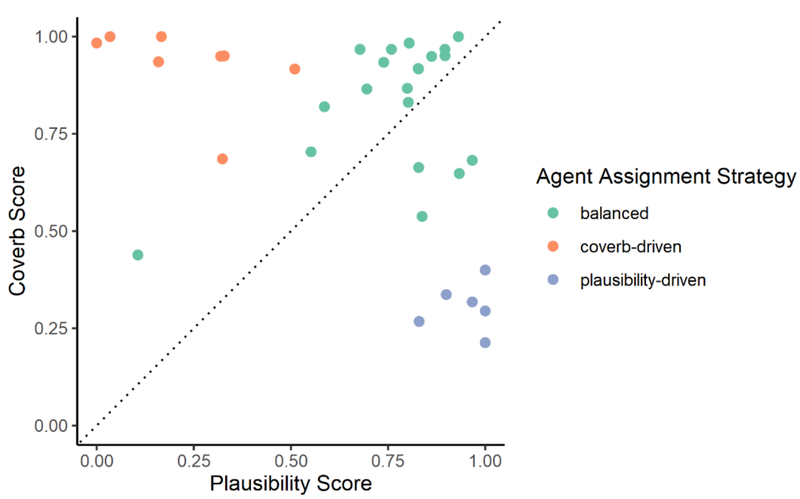

Figure 3. *The dotted line shows the line y = x where Coverb and Plausibility Scores are equal in value. Discrete labels for agent assignment strategy are shown here, but subsequent use of scores for analysis was done with Difference Score as a continuous variable.*

4.2 Reaction Time

Given the range of individual differences captured by the Difference Score measure, we included Difference Score in the model for reaction time as an exploratory step. We found that including this variable yielded additional insights, and the model including Difference Score is reported here. The simpler model with only the factors originally manipulated in the experiment is included in supplementary materials.

The linear mixed effects model of natural log-transformed reaction times optimized for log-likelihood score yielded the coefficients shown in Figure 4. The intercept value (β = 7.31, SE = 0.10, $Z = 75.31$, $p < 0.001$) represents the reaction time for NNV sentences averaged across all other variables (due to other variables being sum coded). Other coefficients' values indicate that response time for those sentences were either faster (negative coefficient value) or slower (positive coefficient value) than the average response time for agent assignment of NNV sentences. Note that the intercept is not depicted in Figure 4 to conserve space.

The simple effect of Difference Score was marginally significant (β = 0.36, SE = 0.20, $Z = 1.85$, $p = 0.065$), indicating that participants who were more plausibility-driven and less coverb-driven had longer reaction times for NNV sentences. As seen for the agent assignment task, the effect of Structure was one of the largest effects, with BA sentences (β = -0.36, SE = 0.06, $Z = -6.06$, $p < 0.001$) and BEI sentences (β = -0.30, SE = 0.06, $Z = -5.39$, $p < 0.001$) both associated with faster reaction times than NNV sentences with no coverb. There was also a simple effect of Reversibility (β = 0.14, SE = 0.02, $Z = 9.16$, $p < 0.001$), where reversible NNV sentences had slower reaction times than irreversible NNV sentences. The final simple effect was Agent Animacy (β = -0.03, SE = 0.01, $Z = -2.48$, $p = 0.013$), where NNV sentences with inanimate agents had longer reaction times than NNV sentences with animate agents.

The model showed a two-way interaction between Difference Score and Reversibility (β = 0.17, SE = 0.03, $Z = 5.30$, $p < 0.001$). To evaluate this interaction in a similar way as for interactions between categorical variables, we compared the predicted values for reversible and irreversible sentences at the extreme values of Difference Score (-0.98 and 0.84). Pairwise comparisons showed that for participants with a high Difference Score, who relied on the cue of plausibility over coverb, were slower for reversible sentences than for irreversible sentences (β = 1.46, SE = 0.08, $Z = 7.00$, $p < 0.001$). Participants with a low Difference Score, who relied on the cue of coverb over plausibility, were also slower for reversible sentences than for irreversible sentences (β = 0.853, SE = 0.05, $Z = -2.74$, $p = 0.03$), but the difference in speed was not as large as for participants with a high Difference Score.

There was also a two-way interaction between Structure and Reversibility for both BA and BEI sentences. Reversible sentences with BA and BEI required less time for response than NNV sentences (BA: β = 1.57, SE = 0.09, $Z = 7.50$, $p < 0.001$; BEI: β = 1.549, SE = 0.088, $Z = 7.69$, $p < 0.001$). Reversible NNV sentences yielded significantly longer reaction times than irreversible NNV sentences (β = 1.32, SE = 0.04, $Z = 8.81$, $p < 0.001$).

Reversibility showed a final two-way interaction with Agent Animacy. Post-hoc pairwise comparisons showed that sentences with plausible inanimate agents elicited longer reaction times for reversible sentences but not for irreversible sentences (reversible: β = 0.90, SE = 0.02, $Z = -5.27$, $p < 0.0001$; irreversible: β = 1.01, SE = 0.02, $Z = 0.27$, $p = 0.99$). Additionally, reversible sentences showed longer reaction times than irreversible sentences for inanimate plausible agents but not for animate plausible agents (inanimate: β = 1.19, SE = 0.03, $Z = 6.05$, $p < 0.0001$; animate: β = 1.07, SE = 0.03, $Z = 2.23$, $p = 0.12$).

Difference Score showed another two-way interaction with Agent Animacy. Post-hoc pairwise comparisons showed that participants with a high Difference Score, who relied more on plausibility, had faster reaction times for sentences with animate agents than for sentences with inanimate agents (β = 0.90, SE = 0.03, $Z = -3.28$, $p = 0.0058$).

The final two-way interaction was between Agent Animacy and Order. Post-hoc pairwise comparisons showed that when plausible agents were in first position, plausible animate agents were associated with faster reaction times than plausible inanimate agents (β = 0.919, SE = 0.019, $Z = -4.17$, $p < 0.001$).

Finally, there were three three-way interactions, as shown in Figure 4. The first was among Structure, Reversibility, and Order, for both BA and BEI sentences (BA: β = 0.03, SE = 0.01, $Z = 2.39$, $p = 0.017$; BEI: β = -0.03, SE = 0.01, $Z = -2.26$, $p = 0.024$). As can be seen in Figure 4, reversible sentences were predicted to elicit slower reaction times in all conditions but BEI congruent sentences when the plausible agent was in second position. Post-hoc pairwise comparisons showed that role reversal sentences elicited longer reaction times than congruent sentences for both BA and BEI, although this effect was only significant for BEI (BA: β = 0.94, SE = 0.03, $Z = 3.07$, $p = 0.09$; BEI: β = 1.14, SE = 0.07, $Z = 2.20$, $p = 0.55$).

The second three-way interaction was between Structure, Reversibility, and Agent Animacy. Post-hoc pairwise comparisons showed that for reversible sentences, NNV sentences with inanimate agents elicited significantly longer reaction times than NNV sentences with animate agents ($\beta = 0.85$, SE = 0.03, $Z = -5.28$, $p < 0.001$), while this difference was not significant for BA or BEI sentences (BA: $\beta = 0.93$, SE = 0.03, $Z = -2.38$, $p = 0.42$; BEI: $\beta = 0.92$, SE = 0.03, $Z = -2.67$, $p = 0.24$). For irreversible sentences, NNV sentences had slower reaction times than BA sentences for both plausible animate and inanimate agents (animate: $\beta = 1.36$, SE = 0.09, $Z = 4.81$, $p = 0.0001$; inanimate: $\beta = 1.24$, SE = 0.08, $Z = 3.41$, $p = 0.03$). BEI irreversible sentences, however, were not significantly factor than NNV sentences regardless of Agent Animacy (animate: $\beta = 1.20$, SE = 0.07, $Z = 3.07$, $p = 0.09$; inanimate: $\beta = 1.14$, SE = 0.07, $Z = 2.20$, $p = 0.55$).

The final three-way interaction was between Difference Score, Structure, and Reversibility, but only for BEI sentences ($\beta = -0.07$, SE = 0.03, $Z = -2.33$, $p = 0.02$). Post-hoc pairwise comparisons showed that for participants who relied more on plausibility (high Difference Score), the largest difference in reaction times was between reversible and irreversible NNV sentences ($\beta = 1.774$, SE = 0.11, $Z = 8.98$, $p < 0.0001$). These same participants also tended to have a greater difference in reaction times between reversible and irreversible BA sentences, but no significant difference for Reversibility in BEI sentences (BA: $\beta = 1.48$, SE = 0.09, $Z = 6.12$, $p < 0.0001$; BEI: $\beta = 1.19$, SE = 0.08, $Z = 2.74$, $p = 0.21$).

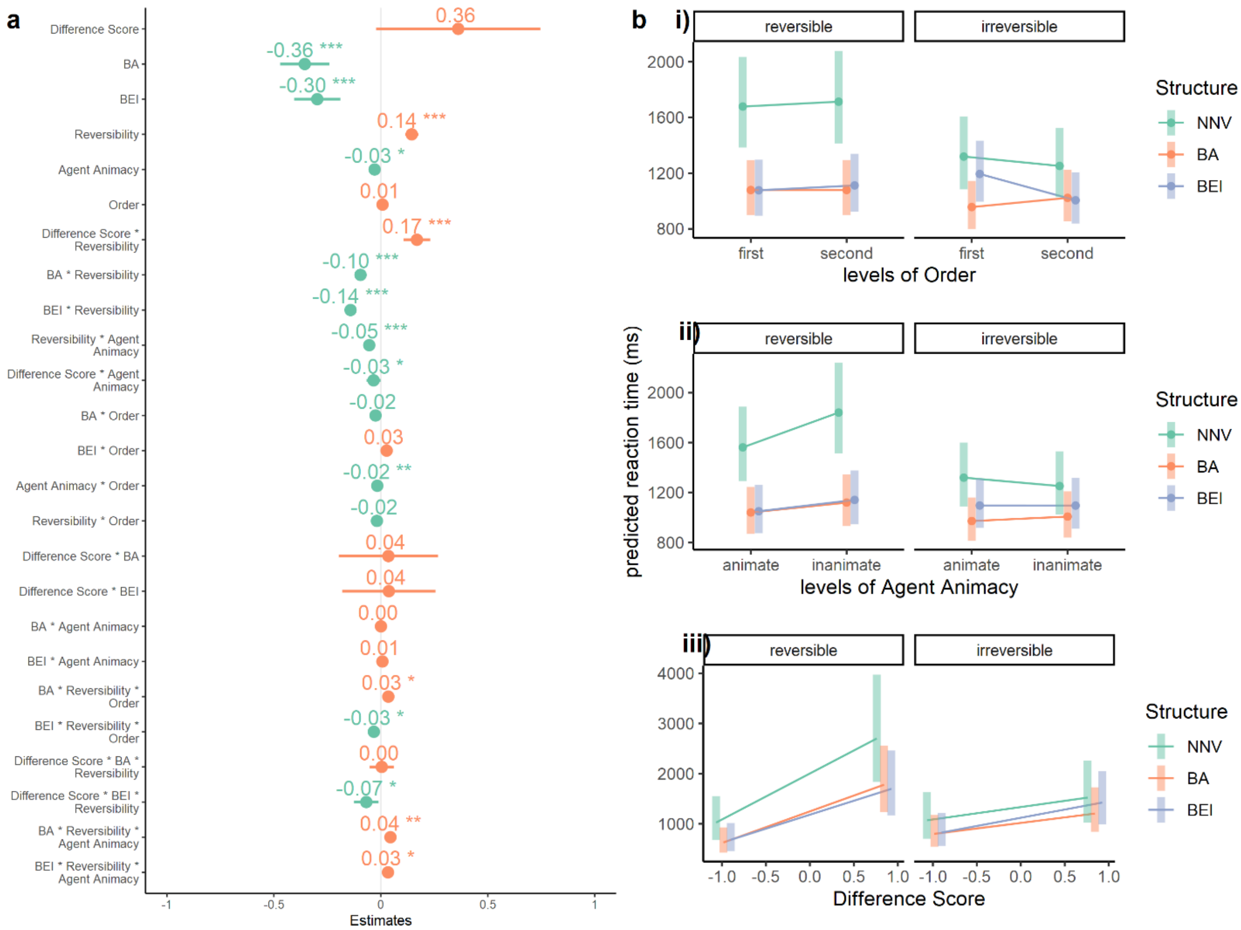

Figure 4. a) Model coefficients of linear mixed effects model for natural log-transformed reaction time of agent assignment task. Intercept ($\beta = 7.31$, $SE = 0.10$, $Z = 75.31$, $p < 0.001$) not depicted to conserve space. Error bars represent 95% confidence intervals. * indicates $p < 0.05$, ** indicates $p < 0.01$, *** indicates $p < 0.001$; b) Three significant three-way interactions for reaction times, back-transformed to the millisecond timescale for ease of interpretability. Error bars show 95% confidence intervals.

4.3 ERPs

4.3a Noun One

Visual inspection of ERPs between animate and inanimate nouns in initial sentence position did not reveal a substantial difference in N400 amplitude, although there was a tendency for inanimate nouns to elicit a greater negativity across the scalp, as seen in the scalp map in Figure 5. A mixed effects model for midline electrodes showed no significant effect of noun one animacy ($\beta = 0.08$ μV, SE = 0.20, $Z = 0.40$, $p = 0.69$, model results reported in supplementary materials).

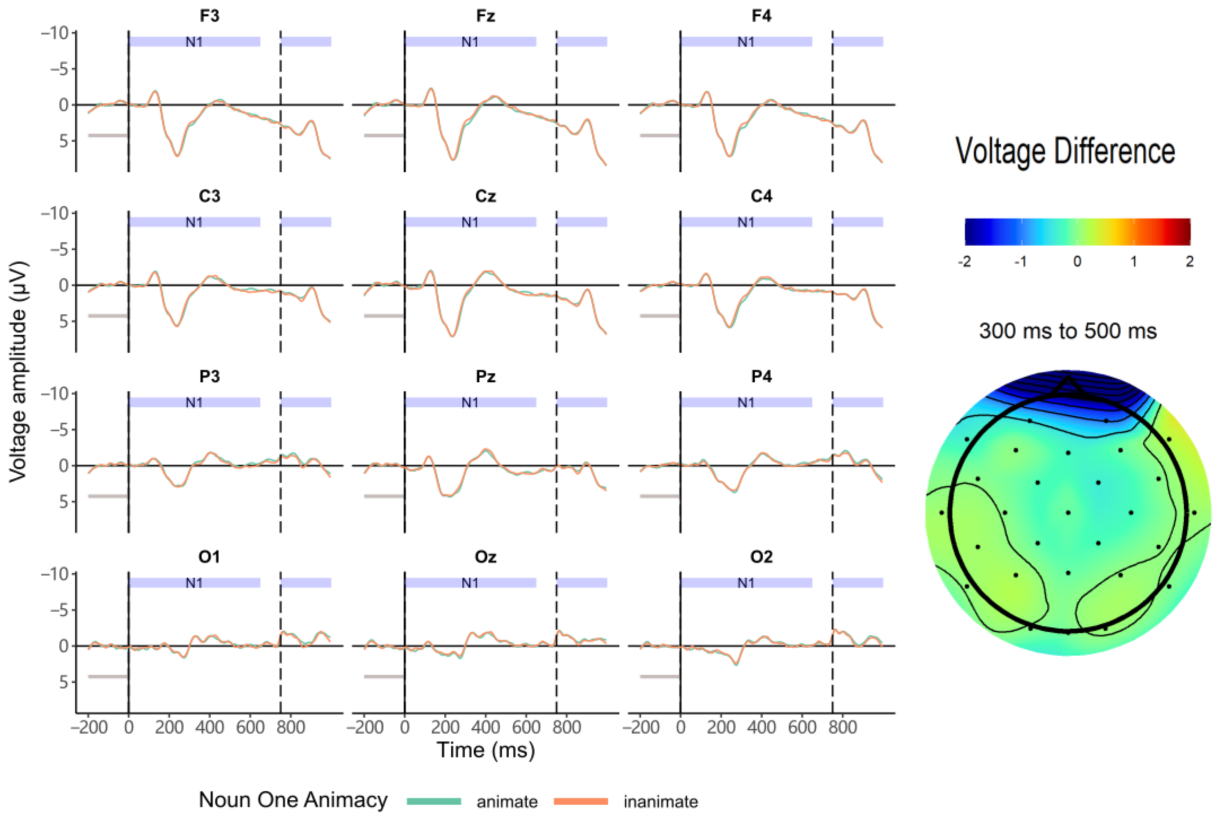

***Figure** 5. ERPs for the effect of noun one animacy, with a scalp map showing inanimate minus animate for the N400 time window from 300 to 500 ms. The 200-ms pre-onset baseline interval is indicated with a gray rectangle.*

4.3b Coverb and Noun Two

Visual inspection for the respective second word in NNV, BA and BEI sentences suggested a dramatic and unexpected in the P200 amplitude linked to the onset of the coverbs BA and BEI. Direct comparison between ERPs locked to second word onset showed that BA elicited a smaller P200 than BEI and Noun Two in NNV sentences, as can be appreciated in Figure 6. Noun two further elicited a sizeable N400 component, consistent with word class effects (Steinhauer, Pancheva, Newman, Gennari, & Ullman, 2001). The smaller P200 for BA was not predicted a priori but given the dramatic differences visible in the ERP waveforms, we ran linear mixed effects models on single trial amplitudes in the P200 and N400 time windows.

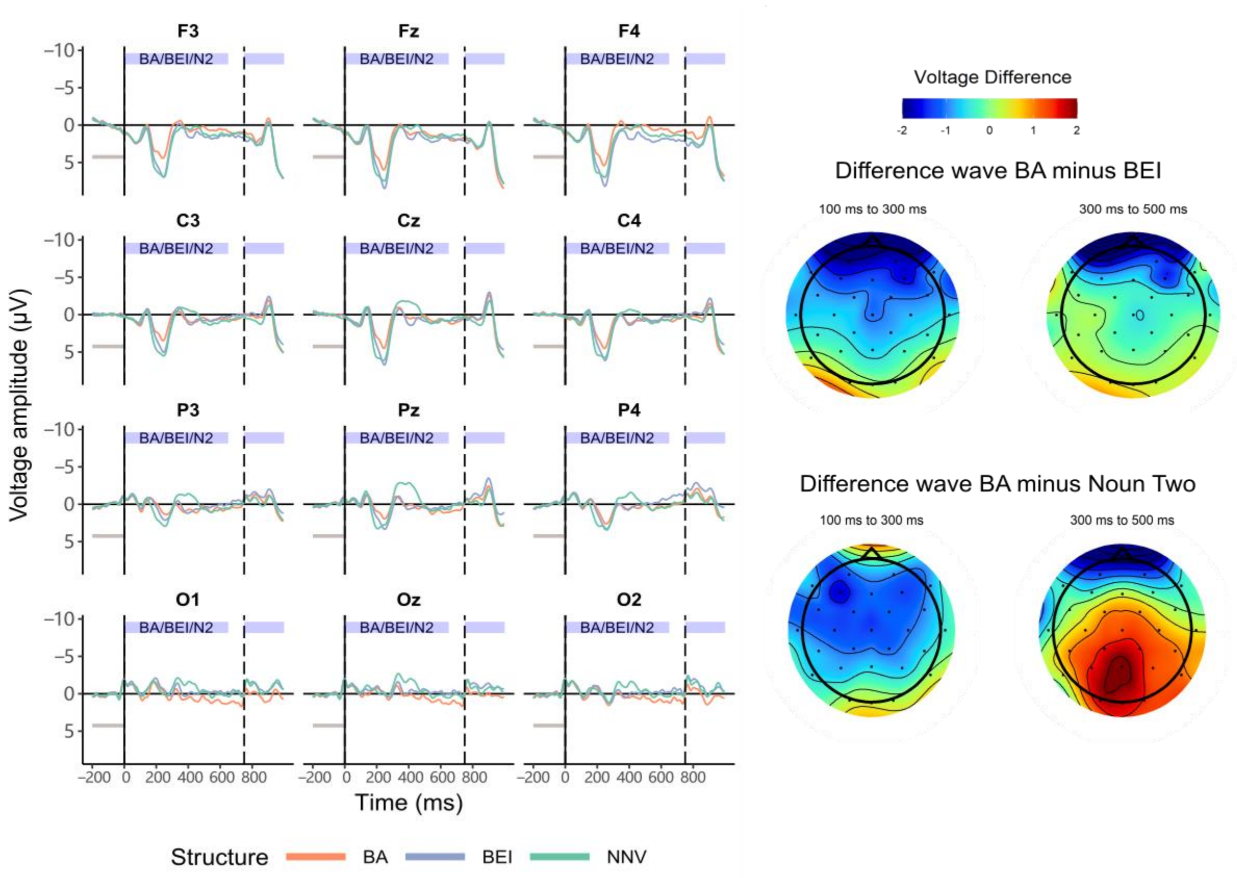

Figure 6. *ERPs for the effect of Structure at the second word position. The 200-ms pre-onset baseline interval is indicated with a gray rectangle. Scalp maps show BA minus BEI (top) and BA minus noun two (bottom) for the P200 and N400 time windows.*

4.3b.i P200

In the P200 time window, linear mixed effects models showed that the average amplitude for BA sentences was significantly smaller than for NNV sentences. There was a simple effect of BA (β = -1.31 μV, SE = 0.38, Z = -3.43, p = 0.001), showing that P200 amplitude for BA at Fz was smaller than for Noun Two at NNV. There was also a two-way interaction between Electrode and Structure for Oz and BA (β = 1.28 μV, SE = 0.35, Z = 3.62, $p < 0.001$). Following this up with post-hoc comparisons showed that P200 amplitude at BA was significantly smaller than at Noun Two in NNV sentences at the electrodes Cz and Fz (Cz: β = 1.40 μV, SE = 0.38, Z = 3.69, p = 0.01; Fz: β = 1.31 μV, SE = 0.38, Z = 3.43, p = 0.030). With correction for multiple comparisons, P200 amplitude for BEI did not differ significantly from that for BA or Noun Two in NNV sentences. Model coefficients and predictions for midline electrodes are shown in Figure 7.

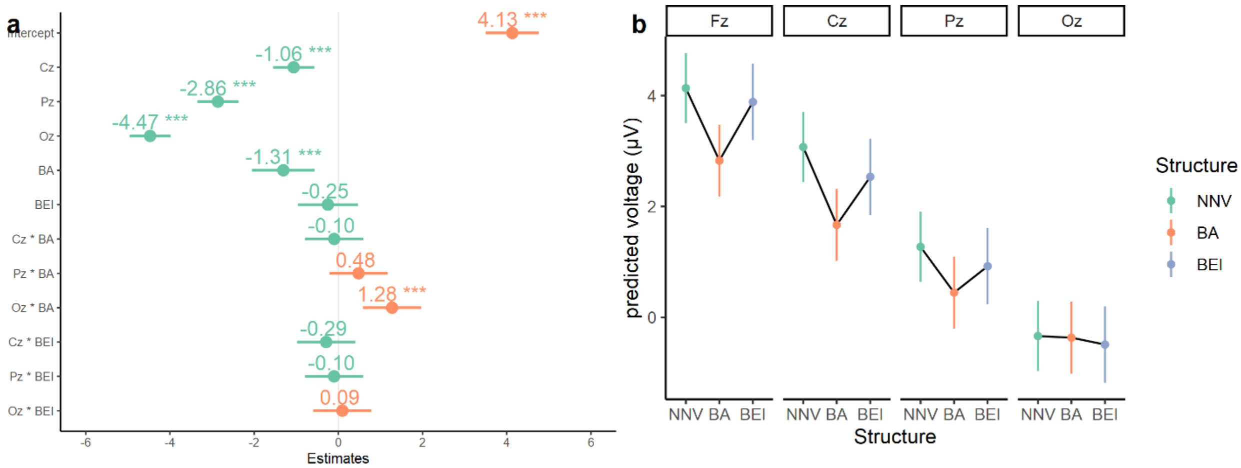

Figure 7. *(a) Model coefficients of linear mixed effects model for P200 amplitude at the second word. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001 (b) Model predictions for two-way interaction between Electrode and Structure. Error bars show 95% confidence intervals.*

4.3b.ii N400

In the N400 time window, linear mixed effects models confirmed that the average amplitude for BA and BEI sentences was smaller than for NNV sentences. There were simple effects of Structure for both BA and BEI sentences, where the magnitude of N400 amplitude was less than for Noun Two in NNV sentences (BA: $\beta = 1.90$ µV, SE = 0.52, $Z = 3.67$, $p < 0.001$; BEI: $\beta = 1.69$ µV, SE = 0.52, $Z = 3.25$, $p = 0.001$). There were also two-way interactions between Structure and Electrode for Fz with BA and BEI (BA: $\beta = -1.66$ µV, SE = 0.50, $Z = -3.32$, $p = 0.001$; BEI: $\beta = -1.24$ µV, SE = 0.50, $Z = -2.47$, $p = 0.013$). Post-hoc comparisons showed that the magnitude of N400 amplitude for BA and BEI was smaller than for Noun Two in NNV at Electrode Pz, although only marginally for BEI (BA: $\beta = -1.90$ µV, SE = 0.52, $Z = -3.67$, $p = 0.013$; BEI: $\beta = -1.69$ µV, SE = 0.52, $Z = -3.25$, $p = 0.053$). Model coefficients and predictions are shown in Figure 8.

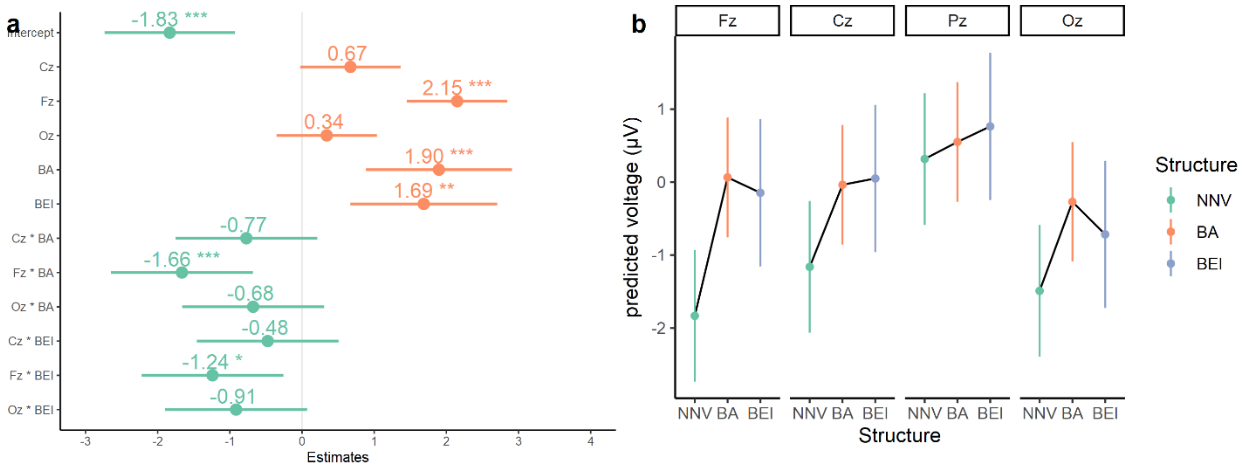

Figure 8. *(a) Model coefficients of linear mixed effects model for N400 amplitude at the second word. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001 (b) Model predictions for two-way interaction between Electrode and Structure. Error bars show 95% confidence intervals.*

4.3c Verb

Visual inspection of ERPs for role reversal sentences at the verb appeared to show a broadly distributed, centro-parietal N400 between 300 and 500 ms and a sustained, localized frontal positivity at Fz around 800 ms, as can be observed in the ERP plots and scalp maps in Figure 9. We further show ERPs broken down by Agent Animacy and Structure in Figure 10. Dividing role reversal sentences separately into BA and BEI appeared to largely show the same pattern, but with some key differences. BA role reversals showed a larger and broader N400 effect with a sustained frontal positivity, and a later, broad positivity beginning around 700 ms in both posterior and frontal locations. BEI role reversals showed a smaller, more localized N400 effect with a sustained frontal negativity, and a central-posterior right-lateralized positivity in the late P600 time window also beginning around 700 ms. For Agent Animacy, role reversals with plausible animate agents appeared to show a broad N400 effect, and role reversals with plausible inanimate agents appeared to show a localized N400 effect followed by a broadly distributed positivity.

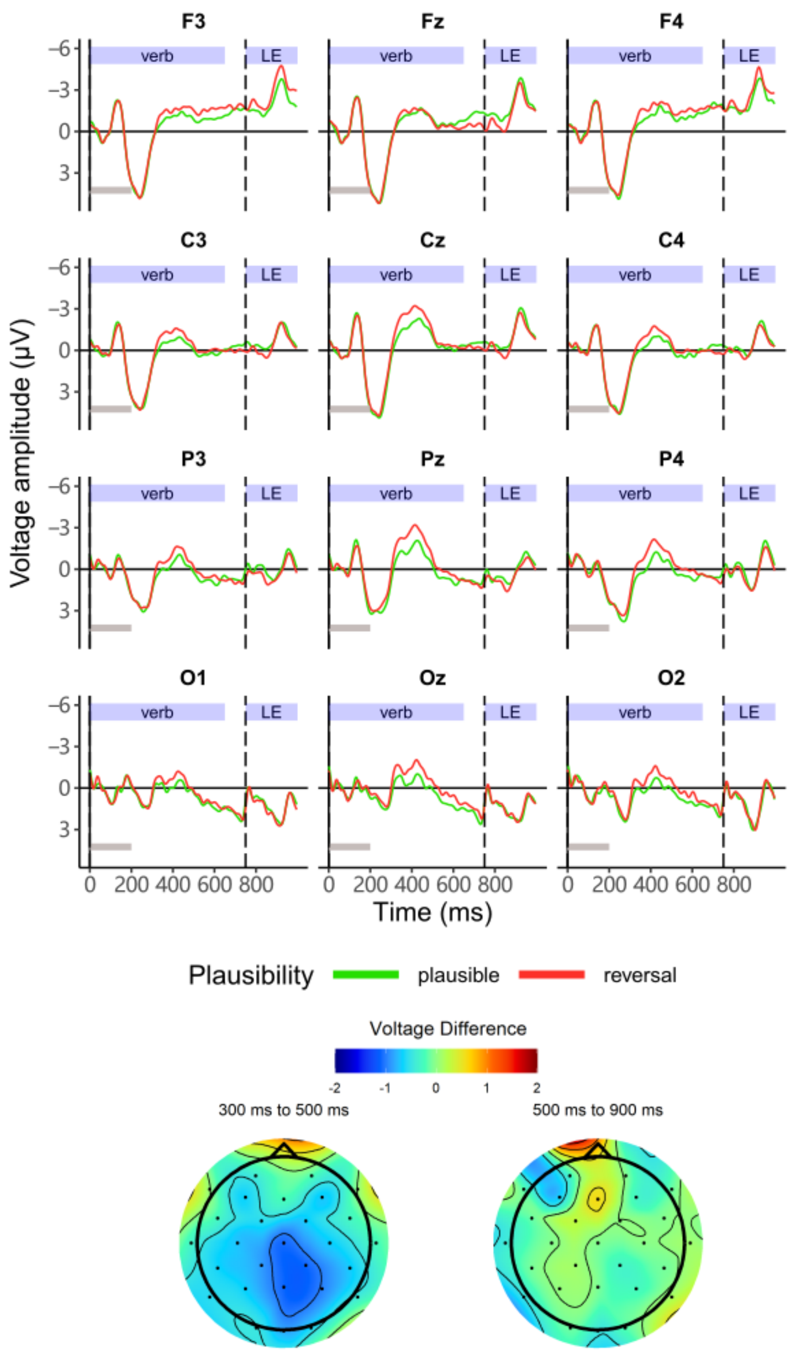

Figure 9. *ERPs for the effect of Plausibility on verb processing, averaged across Agent Animacy and Structure. Scalp maps shows reversal minus plausibility (averaged across other factors) for the N400 time window from 300 to 500 ms and the P600 time window from 700 to 900 ms. The 200-ms post-onset baseline interval is indicated with a gray rectangle.*

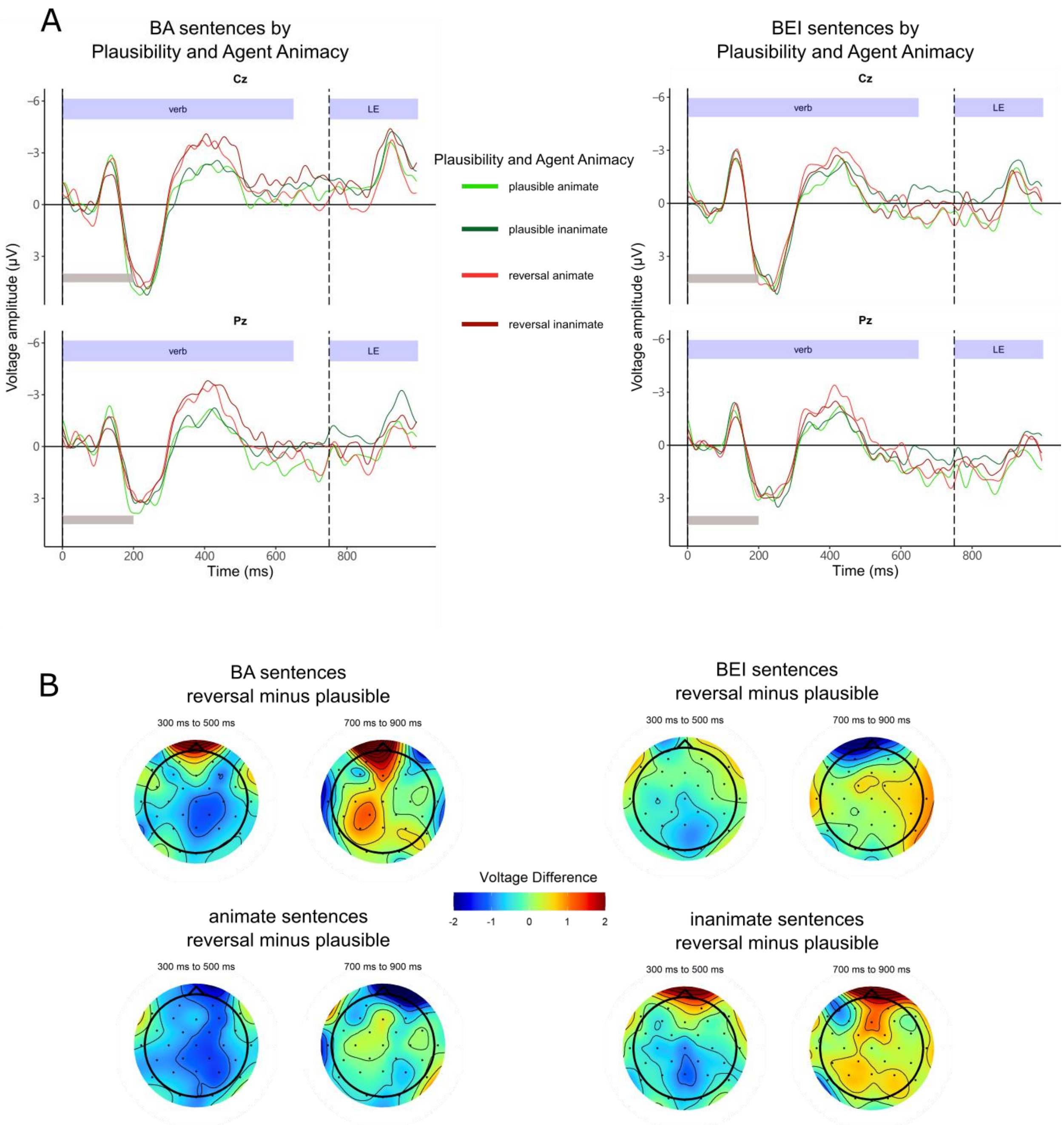

Figure 10. *A) ERPs for the effect of Plausibility on verb processing at select electrodes for BA (above) and BEI (below). The 200-ms post-onset baseline interval is indicated with a gray rectangle. B) Scalp maps show reversal minus plausibility for each of the labeled conditions (averaged across other factors) for the N400 time window from 300 to 500 ms and the P600 time window from 700 to 900 ms. We note that while the scalp maps for the subconditions of Structure and Agent Animacy suggest different patterns, not all effects are significant according to the model results, as detailed in the following sections.*

4.3c.i N400

As seen in Figure 9, role reversals elicited a greater N400 amplitude on posterior and central electrodes in comparison to plausible sentences. However, our initial optimized mixed effects model showed that this Plausibility effect was only marginally significant (β = -0.51 μV, SE = 0.32, *Z* = -1.61, *p* = 0.106). We compared this to a model at just the electrode Pz, which did show a significant effect of Plausibility (β = -0.92 μV, SE = 0.35, *Z* = -2.65, *p* = 0.008). Because Electrode and Plausibility were treatment coded with Pz and plausible as the reference levels, we expected these values to be identical; on closer investigation, we determined that the non-identical coefficient values were due to the larger model being limited to a maximum of three-way interactions, and running the full model on midline electrodes (i.e., including all interactions for Plausibility x Agent Animacy x Structure x Electrode) resulted in identical coefficients and a significant effect (β = -1.11 μV, SE = 0.34, *Z* = -3.23, *p* = 0.001). We further determined that the large random structure in the model for midline electrodes (sometimes including two- or three-way interactions and the factor Plausibility for both item and subject random structures) also contributed to the lack of significance. This example demonstrates some of the challenges of interpreting mixed effects models and the many degrees of freedom in building models.[10] We bring up this difficulty because such challenges are faced by many, if not all, users of mixed effects models and the field continues to develop standards for best practice (Meteyard & Davies, 2020). Ultimately, we report here in the text a model we believe most accurately represents our results, including all midline electrodes but a slightly reduced random structure removing the random slope for Plausibility from the item random structure. This adjustment yielded the model coefficients in Figure 11. We further apply the same steps to the selection of the model in the P600 time window. For transparency, all model outputs are reported in the supplementary materials.

Figure 10 shows differences and similarities between BA and BEI role reversal effects. In the N400 time window, both structures appeared to show greater negativity for role reversal sentences, with the negativity for BA reversals greater and broader, and the negativity for BEI reversals smaller and more localized. BA sentences also showed a frontal positivity, with a left and right anterior negativity. The lack of a significant Structure by Plausibility interaction

[10] We thank Dr. Meghan Clayards and Dr. Morgan Sonderegger for their input in discussion of these effects.

indicates that the model did not predict that these differences were important for capturing variability.

The final model in Figure 11 showed a main effect of Plausibility (β = -0.46 µV, SE = 0.18, Z = 2.55, p = 0.011), demonstrating that role reversals elicited greater negativity in the N400 time window at electrode Pz when averaged across Structure and Agent Animacy. The next significant effect was a three-way interaction among Structure, Agent Animacy, and Plausibility (β = 0.39 µV, SE = 0.18, Z = 2.16, p = 0.03), which is depicted in the predicted voltages in Figure 11. We calculated post-hoc pairwise comparisons both at the reference level Pz and averaged across electrodes for the three factors. All implausible role reversals were predicted to have a numerically larger N400 amplitude, but this difference was only significant for BEI sentences with plausible animate agents (for example, the reversal sentence *仆人被镜子擦亮了* "servant BEI mirror polished"; β = -1.39 µV, SE = 0.35, Z = 3.99, p < 0.001). Described more qualitatively as seen in the model prediction in Figure 11, implausible role reversals with BA sentences elicited a numerically greater N400 than plausible BA sentences regardless of Agent Animacy status (although not significant in pairwise comparison), while implausible role reversals with BEI sentences elicited a greater N400 for animate agents but not for inanimate agents.

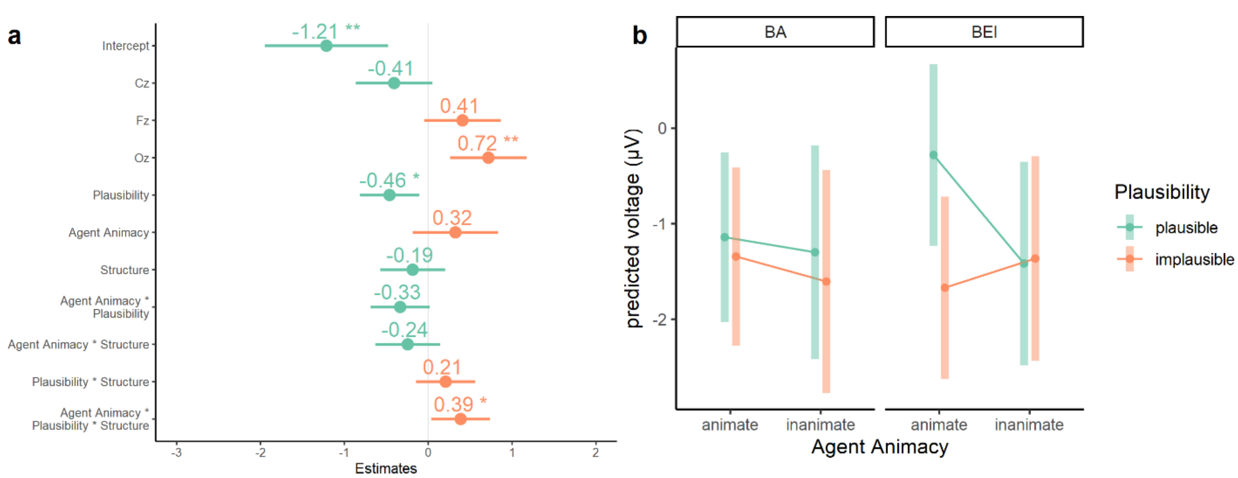

Figure 11. *(a) Model coefficients of linear mixed effects model for N400 amplitude at the verb. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001 (b) Model predictions for three-way interaction among Agent Animacy, Plausibility, and Structure. Error bars show 95% confidence intervals.*

4.3c.ii P600

As seen in Figure 9, in the 700 to 900 ms time window there was a small frontal positivity in the grand average for the effect of Plausibility. The model at midline electrodes, depicted in Figure 12, showed that the coefficient for Plausibility was not significant, which is consistent with Plausibility not modifying the P600 at Pz when averaged across other conditions (β = 0.02 μV, SE = 0.46, *Z* = 0.04, *p* = 0.969). Note that this reported model was obtained after removing the random slope for Plausibility in the random structure for item.

Considering the differences in the P600 time window between BA and BEI sentences shown in Figure 10, it appears that BA sentences showed a posterior and a frontal positivity, while BEI sentences showed an anterior negativity with a small, right-lateralized positivity. Despite these observed differences, the two-way interaction between Structure and Plausibility was not significant.

The model also showed a significant three-way interaction among Agent Animacy, Structure, and Plausibility (β = 0.62 μV, SE = 0.23, *Z* = 2.66, *p* = 0.008). From the predicted voltages for this interaction in Figure 12, it can be seen that the model predicted a numerically greater positivity for implausible sentences in each condition except for BEI reversals with plausible animate agents, where there was a significant negativity (β = 1.67 μV, SE = 0.46, *Z* = 3.63, *p* = 0.007).

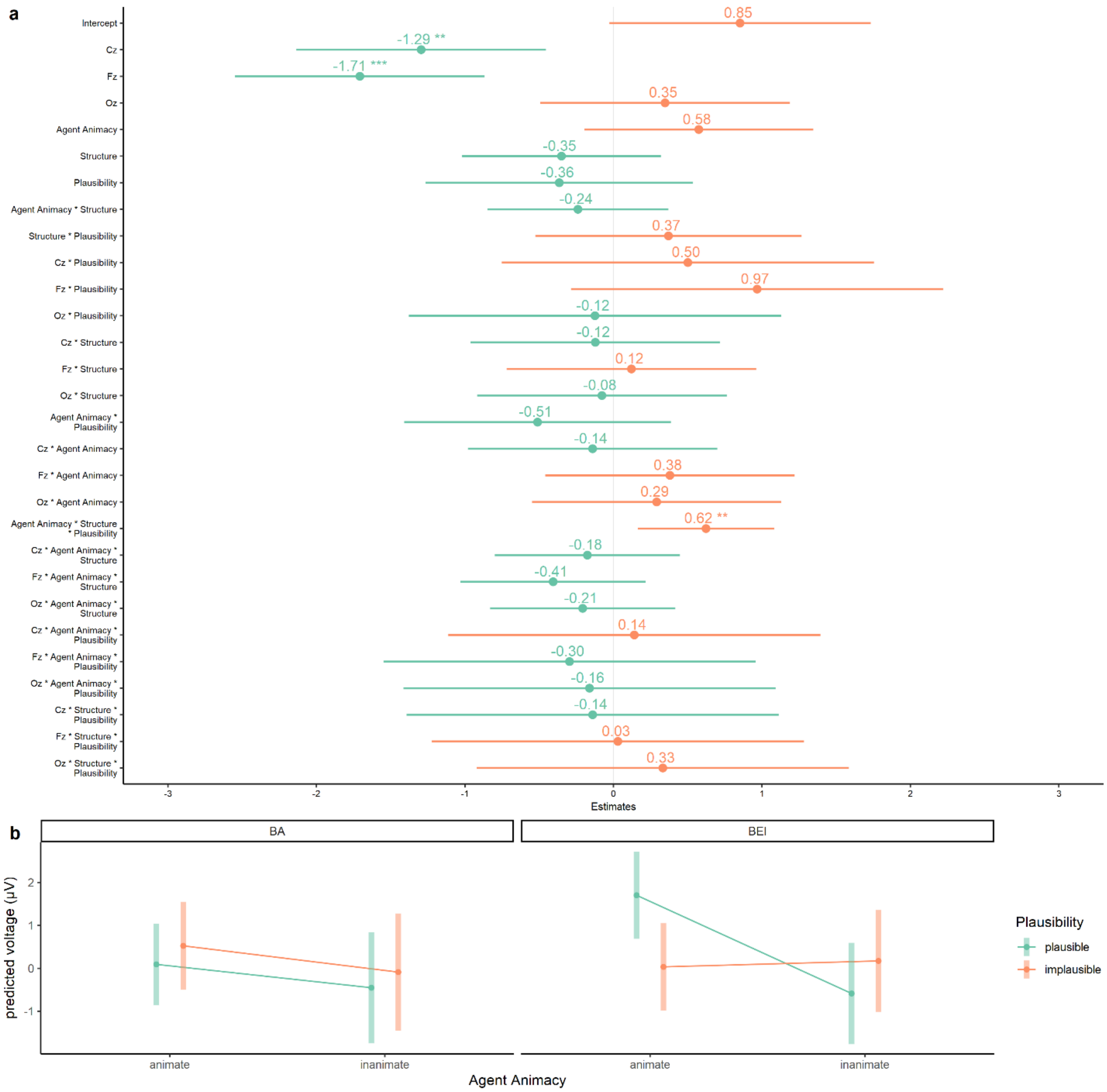

Figure 12*. (a) Model coefficients of linear mixed effects model for P600 amplitude at the verb. Error bars represent 95% confidence intervals. * indicates p < 0.05, ** indicates p < 0.01, *** indicates p < 0.001 (b) Model predictions for three-way interaction among Agent Animacy, Plausibility, and Structure. Error bars show 95% confidence intervals.*

5 Discussion

The present study investigated processing of argument structure in verb-final sentences in Mandarin, a language that has flexible word order and virtually no inflection. Our participants were native, monolingual Mandarin speakers who read verb-final sentences and completed an agent-assignment task with competing cues: Structure, Agent Animacy, Reversibility, and Order. We analyzed both behavioral and EEG data. Our behavioral results showed that 1) in the absence of other cues, word order was not used to assign argument structure; 2) the coverbs BA and BEI were the strongest cues for agent assignment but were differently impacted by Agent Animacy in the case of role reversals; and 3) participants showed individual differences in their reliance on Plausibility and Structure. Our EEG results showed that 1) sentence-initial noun animacy did not impact N400 amplitude; 2) BA elicited a reduced P200 amplitude relative to BEI and nouns; and 3) role reversal sentences, at the verb, elicited an N400 effect without a P600 effect.

To our knowledge, this is the first time this forced agent-assignment task has been used in an EEG experiment. A key advantage to this task is that, in at least some ways, it closely resembles the task of natural language comprehension, where comprehenders must understand who did what to whom. Additionally, all experimental sentences were, in principle, grammatical structures. In combination with the behavioral results, our ERP findings give further insight into online, incremental parsing decisions throughout the process of sentence comprehension. As we summarize below, there are key points of convergence among the agent assignment task, reaction times, and ERPs that provide convincing evidence for our characterization of Mandarin verb-final sentence processing.

We first overview our behavioral and ERP results and consider potential differences between the coverbs BA and BEI. We then relate our results to three sentence processing models that have considered data from Mandarin: the Competition Model, the eADM, and the Bag of Arguments account. We end with discussion of limitations and future directions.

5.1 Behavioral Results

5.1a Agent Assignment Task

In the agent assignment task, participants interpreted the competing cues of Structure, Agent Animacy, Reversibility, and Order to comprehend each sentence. Surprisingly, their responses de-emphasized the effects of word order in the absence of other cues. More predictably, BA and BEI were the strongest cues for agent assignment, and in their absence, participants strongly preferred the plausible agent in irreversible sentences. In contrast to prior reports (Li et al., 1992; Su, 2001), however, Agent Animacy was not a primary cue, only showing interactions with other factors.

Interactions further demonstrated a nuanced interplay among cues for agent assignment, beyond a simple linear hierarchy. The coverbs, BA and BEI, were each stronger when indicating a plausible animate agent than when indicating a plausible inanimate agent in reversible sentences. However, the coverbs showed differences in the case of role reversals: while BA was more successful at eliciting role reversal interpretations when the implausible agent was animate (e.g., *喜鹊把鸟笼困住了* “magpie BA birdcage trapped”), interpretations of BEI role reversals were not impacted by Agent Animacy. We discuss this further in Section 5.3.

Reversibility showed additional effects in NNV sentences where no cover was present. For irreversible sentences, participants strongly favored the plausible agent. For reversible sentences, however, participants showed no preference between subject-object-verb and object-subject-verb word order. We interpret this flexibility as evidence that Mandarin parsers do not normally rely on pre-verbal position of arguments as a cue for agent status; instead, the post-verbal position may be the only reliable word order cue, where nouns following verbs are preferentially interpreted as patients (as seen in Mandarin-English bilinguals in Li et al., 1992). We acknowledge that this finding conflicts with prior reports of Mandarin speakers preference object-subject-verb order (Li et al., 1992; Liu et al., 1992; Wang et al., 2012) and is more in line with accounts that suggest verb-final reversible sentences may be uninterpretable (Yu & Tamaoka, 2018). Some of the differences in findings for word order preferences in Mandarin may also be due to experimental design. In the present study, upcoming nouns had equal probability of being animate or inanimate and plausible or implausible agents, which may have led to participants disregarding the cue of Order.

Alongside these patterns, we also identified qualitative individual differences in cue-weighting strategies among participants, as shown in Figure 3. Comparison among individuals revealed three types of participant strategies: coverb-driven (N = 8), where agent assignment was primarily driven by BA and BEI; plausibility-driven (N = 6), where plausibility was used even when in conflict with a coverb; and balanced (N = 20), where coverbs were preferentially used when available but plausibility was employed for NNV structures without BA or BEI. Alongside visualization of these discrete strategies, we also computed a continuous Difference Score measure that we showed to impact reaction times. It may be tempting to attribute the variation in Difference Score as due solely to participants’ motivation and understanding of the experiment instructions; perhaps coverb-driven participants found an easy strategy of relying solely on the coverb in the two-thirds of sentences where one was available, while guessing randomly for NNV sentences. Such an approach could be limited to the experimental setting and not extend to real-world sentence processing. However, we suggest that reliance on different cues in different language contexts reflects strategies that are important for typical language use, such as using semantic cues to recognize an ironic statement or focusing on sentence structure to analyze the precise meaning of contract law. Additionally, our models for reaction times demonstrate that even coverb-drive participants were impacted by experimental conditions other than just Structure.

5.1b Reaction Times

Participants' reaction times corroborated many of the effects seen in agent assignment choices. BA and BEI were again shown to be the strongest cues, with the presence of a coverb dramatically reducing the time needed to select the agent. Reversible sentences also elicited longer reaction times than irreversible sentences, which is consistent with reversible sentences showing an equal probability of eliciting first or second noun agent selection. These effects generally correspond to the level of ambiguity in sentences, with more ambiguous sentences requiring longer times to interpret. Accordingly, reversible NNV sentences elicited the longest reaction times. Because neither agent is inherently more plausible in these sentences, their processing may in some ways be compared to an extreme type of garden path sentences (e.g., Ferreira & Henderson, 1991). The fact that these sentences took more time may also indicate that participants pursued parallel parses of both possible interpretations (e.g., Hickok, 1993).

Reaction times additionally showed several effects beyond those seen in choices in the agent assignment task and beyond what may be simply attributed to ambiguity resolution. First, inanimate agents elicited slightly longer reaction times than animate agents, even though there was no simple effect of Agent Animacy on participants' binary agent assignment choices. Second, BEI role reversals were found to elicit longer reaction times than BEI congruent sentences, while a similar effect was not found for BA sentences. This interaction difference in reaction times for BA and BEI sentences gives additional evidence of nuanced differences between the two coverbs, which we discuss in more detail in Section 5.3. Third, individual differences in participants' cue weighting also impacted reaction time, as demonstrated by the effect of Difference Score. Participants who preferentially relied on plausibility over coverbs took longer to process all sentences, and effect was strongest for the maximally ambiguous reversible NNV sentences. Being more plausibility-driven was also associated with longer reaction times for plausible inanimate agents than plausible animate agents, which suggests further differences in individual cue weighting strategies.

The present study is not a typical reaction time experiment, with over one second between verb onset and the response prompt to prevent motor artifacts in the ERPs. It is therefore unclear the extent to which our results capture online, automatic processing, as opposed to offline, reflective processes. Notwithstanding this limitation, our model results show that experimental conditions impacted reaction times in line with behavioral and ERP results.

5.2 ERPs

We analyzed average ERP amplitudes time-locked to three different sentence positions: the first noun, the second word in the sentence, and the verb. By considering multiple time windows, we captured multiple snapshots of participants' parsing during real-time construction of sentence structure, as well as gaining insight into how competing cues impact argument structure processing even prior to the verb.

5.2a Pre-verb effects

At the first noun position, we did not observe an effect of first noun animacy on N400 amplitude, in contrast to previous reports (Bourguignon et al., 2012; Kyriaki et al., 2020). Additionally, N400 amplitude may have been impacted by the types of nouns in our stimuli, which included inanimate nouns that have conceivable underlying animate agents, such as *寺庙* "temple" in *高僧寺庙请来了* "monk temple invited" where there are presumably other monks in the temple who do the inviting. Inanimate nouns like these may have "perceived agency", which has been suggested to be a more important cue than animacy for sentence processing (Lowder & Gordon, 2013, 2015). Participants may also have realized that animacy was an unreliable cue in our stimuli, leading them to consider sentence-initial animate and inanimate nouns as equally plausible agents. We also note that this lack of an effect of animacy at noun one is consistent with our behavioral findings showing that plausibility is more important than animacy. Given that previous reports of an N400 effect on sentence-initial inanimate nouns were for English (Bourguignon et al., 2012; Kyriaki et al., 2020), this may also be evidence that crosslinguistic differences in parsing are present from the first word of a sentence. Given Mandarin's typologically uncommon linguistic features (Chappell et al., 2007), Mandarin comprehenders may not benefit from an agent-first or subject-first processing strategy (c.f. Wang et al., 2009).

We also report an unexpected finding at the second word in the sentence. We observed a striking difference between the P200 amplitude for BA and BEI. This effect was confirmed by our models, which further showed that P200 amplitude was the same for BEI as for noun two in NNV sentences. We initially interpreted this P200 effect as due to differences in the number of strokes. BA and BEI are each a single character while our nouns were all two characters, and visual differences have been associated with P200 amplitude (Potts, 2004; Potts & Tucker, 2001). However, this visual difference cannot explain why BA has a smaller P200 and BEI elicited a P200 amplitude similar to that for nouns. There is also a previous report of a similar P200 difference between BA and BEI in an auditory experiment, where visual complexity could not be a factor (Philipp et al., 2008). Other evidence has also suggested that the P200 component is sensitive to word class, with a greater frontal positivity for verbs than for nouns (K. D. Federmeier, Segal, Lombrozo, & Kutas, 2000; Preissl, Pulvermüller, Lutzenberger, & Birbaumer, 1995). This word class explanation also fails to account for the P200 amplitude difference between BA and BEI.

Instead, we interpret that there is an additional cognitive component underlying the P200 difference relating to participants' strategies for the agent assignment task. For instance, on encountering the coverb BEI or the second noun in an NNV sentence, participants needed to wait to fill or confirm the proto-agent role in their syntactic parse. When seeing BA, however, participants may have committed to an argument structure for an agent-initial sentence, at least initially. This explanation supposes that assigning agent status is more important than assigning patient status. If the primacy of agent over patient was a result of our explicit agent assignment task, then a future experiment using a patient assignment task should find the opposite P200 pattern for BA and BEI. Our interpretation is consistent with the P200 as an index of task demands on memory (Schendan & Kutas, 2007).

5.2b Role Reversals

Of the approximately 360 sentences each participant read in the experiment, about 60 sentences were role reversals where semantic and syntactic cues contradicted each other. For these sentences and their matched controls, we analyzed average amplitude time-locked to verb onset in the N400 and P600 time windows to examine the effects of Plausibility and its interactions with Structure and Agent Animacy. Our results showed that role reversals elicited a greater N400 amplitude and a small, local frontal positivity, without a posterior P600 effect. In contrast to accounts that predict a P600 (e.g., Brouwer et al., 2012; Chow & Phillips, 2013; Kolk & Chwilla, 2007; Kuperberg, 2007), our findings suggest that Mandarin role reversal anomalies were detected via relatively early, automatic semantic processing and meaning retrieval mechanisms. This role reversal N400 effect may be due to the particular features of Mandarin (Bornkessel-Schlesewsky et al., 2011) or reflect our participants' ability to predict the verb based on the preceding context (Chow et al., 2018, 2016).

Our statistical model results suggest that N400 amplitude for role reversals further depended on Structure and Agent Animacy. While BA reversals were predicted to show a greater negativity regardless of Agent Animacy, the model predicted that BEI reversals only elicit an N400 effect for plausible animate agents (e.g., *仆人被镜子擦亮了* "servant BEI mirror polished"). This pattern is visible in the grand average ERPs in Figure 10, where the N400 effect for BA is more consistent. We note that similar three-way interactions among Structure, Agent Animacy, and Plausibility/Order were also found for the agent assignment task and reaction times. We discuss these interactions and potential differences between BA and BEI in more detail in Section 5.3.

In the P600 time window, our statistical model also showed a three-way interaction involving Plausibility. Post-hoc pairwise comparisons did not confirm significant differences, and Plausibility did not show a significant simple effect or lower-level interaction, so we interpret this interaction only cautiously. The model predicted that only role reversals with plausible animate agents (e.g., *镜子把仆人擦亮了* "mirror BA servant polished") elicited an effect for BA sentences, but only role reversals with plausible inanimate agents (e.g., *鸟笼被喜鹊困住了* "birdcage BEI magpie trapped") elicited an effect for BEI sentences. This interaction is in addition to animate plausible agents eliciting greater positivities than inanimate plausible agents.

We followed up on this analysis more closely and determined that the prediction was not for a typical centro-parietal P600; instead, the positivity was frontal and maximal at Fz. In contrast to the P600, frontal positivities are associated with prediction error (DeLong, Quante, & Kutas, 2014; Kuperberg, Brothers, & Wlotko, 2020) and have been dissociated from semantic P600 effects (Chow et al., 2018). Given the anterior scalp distribution and the lack of a simple effect of Plausibility in the P600 time window, we conclude that our Mandarin role reversals elicited an N400 effect without a semantic P600. In the interest of being critical of our own analysis, we note that our choice of post-onset baseline correction did impact the magnitudes of the effects in the N400 and P600 time windows. ERP figures, scalp maps, and model results for the pre-onset baseline are in our supplementary materials.

At least some of our differences with prior findings may also be due to task. Generally, the P600 component is vulnerable to task effects while the N400 is considered more automatic and robust (Bornkessel-Schlesewsky & Schlesewsky, 2019; Schacht et al., 2014). However, we note that that some tasks have been shown to impact N400 amplitude (Kyriaki et al., 2020; Schacht et al., 2014), and our experimental task of explicitly assigning agent may have encouraged particular parsing strategies indexed by the N400. Nonetheless, it is difficult to explain our observed pattern as due solely to task when many accounts predict that overt experimental tasks elicit larger P600s (as discussed in Brouwer & Crocker, 2017).

5.3 Differences between BA and BEI

A common thread across our results is that Structure consistently showed interactions with Agent Animacy and Plausibility. In the agent assignment task, BA and BEI appeared equally strong in driving agent selection, but we see a difference between them in role reversal sentences. BA was consistently a stronger cue when indicating an animate agent, even when working against plausibility. BEI, in contrast, showed no interaction with Agent Animacy in role reversal sentences. Participants' reaction times further showed that role reversals with BEI were more costly to process than those with BA. Participants' ERPs showed additional differences. While BEI elicited a P200 equal in size to that of nouns, BA's P200 amplitude was much reduced. At the verb in role reversal sentences, our statistical model results suggested that the N400 was effect was consistent for sentences with BA regardless of Agent Animacy, but only appeared for BEI sentences with plausible animate agents.

We interpret these findings together to suggest that BA is a stronger cue for expressing implausible, role reversal sentences. In the agent assignment task, BA became a stronger cue when indicating an animate agent in role reversals (e.g., *喜鹊把鸟笼困住了* "magpie BA birdcage trapped"). For other role reversal sentences, BA and BEI were equally strong as cues. The reaction time results showed that BEI role reversals were more costly to process than BEI congruent sentences, indicating that participants experienced conflict between the cues of Plausibility and Structure. BA reversals were also predicted to elicit numerically longer reaction times, but this did not reach significance and was not more costly than processing BA reversible sentences. This suggests that at some level participants were able to disregard Plausibility when processing BA role reversal sentences, indicating there was greater conflict between Plausibility and Structure for BEI role reversals.

These behavioral findings may be the underlying cause of the ERP differences for role reversals seen for BA and BEI. Participants may have assigned structural roles more quickly for BA sentences and thus were better able to predict the verb in a limited time frame (Chow et al., 2018). This interpretation is also consistent with our finding that BEI role reversals elicited a greater frontal positivity than BA role reversals. The mechanism by which BA is faster than BEI for assigning structural roles could further be due to fundamental differences between (proto-)agent and (proto-)patient roles. Given that not all verbs can or always take objects (patients), the agent role might be more salient or prioritized for argument structure assignment.

As we already touched on, the nature of our agent assignment task may have resulted in different parsing strategies when comprehending BA and BEI. Participants were focused on identifying the agent in each sentence, and BA indicated the agent was the noun they just saw while BEI indicated the agent was still upcoming. Beyond the experimental task, however, we cannot neglect the potential impact of inherent differences between the two coverbs. As mentioned in the introduction, Mandarin is a pro-drop language, and BA and BEI can both appear in pro-drop structures (Li & Thompson, 1989). BEI can occur with a null subject, null object, or null subject and object. For example, the sentences *鸭子被吃了* “duck BEI ate”, *被鸭子吃了* “BEI duck ate”, and *被吃了* “BEI ate” respectively correspond to the meanings “(something) ate the duck”, “the duck ate (something)”, and “(something) ate (something).” BA, however, must be followed by a noun phrase that will take the patient role. These differences could conceivably impact parsing decisions in the building of incremental structures.

There may also be frequency differences in the use of BA and BEI structures. Unfortunately, we do not have strong data as to the extent to which such differences exist and how they interact with oral and written registers. The same constraints hold for determining the relative frequencies of subject-object-verb and object-subject-verb sentences without coverbs. Some linguists have emphasized the similarities in grammatical functions to propose an overarching grammar that can account for both coverbs (Hsien-Yi, 1998). Despite these potential similarities, our findings showcase potentially important differences.

5.4 Implications for sentence processing models

To inform predictions for our experiment, we relied on three sentence processing models that have considered data from Mandarin: the Competition Model, the eADM, and the Bag of Arguments. These models have different scopes and motivations, but each makes testable claims about argument structure processing in Mandarin. As will be discussed below, the data are broadly consistent with the Competition Model and the eADM, but difficult to reconcile with predictions from the Bag of Arguments model. Importantly, each model falls short of fully explaining our results, and in what follows we make suggestions for additional factors that must be considered to understand sentence processing mechanisms.

5.4a Competition Model

The Competition Model is a cue-based account of language processing that has examined abundant data from behavioral experiments with orthogonal cue competition. According to the model, parsers develop cue-weighting strategies based on experience with the specific strength, availability, and reliability of cues in their language. Previous data using a Competition Model paradigm suggested that in the absence of competing cues, Mandarin parsers use word order to assign agent status, and object-subject-verb is the preferred interpretation for verb-final sentences; additionally, in the absence of coverbs, animate nouns should be reliably interpreted as agents over inanimate nouns (Li et al., 1992; Miao, 1981; Su, 2001). One study further reported that BEI and animacy are both stronger cues than BA (Li et al., 1992).

Our results reveal important contrasts with these two prior findings. Firstly, our data show that participants did not use word order to assign agent status when there were no other available cues. As described above in the discussion of our behavioral results, this may indicate that word order alone is not a reliable cue in Mandarin, even in the absence of competing cues; instead, post-verb position of a noun argument may instead be the only word order cue that impacts Mandarin parsing. Because all our sentence materials were verb-final, word order was not found to play a role in the absence of other cues. Secondly, we found that in the absence of coverbs, participants selected plausible nouns as agents whether these nouns were animate or inanimate. This finding demonstrates that animacy was overinterpreted as a cue in previous Competition Model reports, which did not employ plausible inanimate agents in their sentence materials (Li et al., 1992; Liu et al., 1992).

In the present study, we not only included plausible inanimate agents, but also dramatically increased the number of sentence tokens per condition and further manipulated the cue of

Reversibility. Our results reveal interactions of Agent Animacy with Structure and Order that likely better reflect the real impact of animacy in Mandarin parsing. As seen in Figure for reversible sentences, both BA and BEI were strongest as cues when agents were animate. We see this pattern for BA reversals as well, where BA was more likely to elicit a reversal interpretation if the implausible agent was animate (i.e., when the plausible inanimate agent was in position two, as in *喜鹊把鸟笼困住了* "magpie BA birdcage trapped"). Participants' judgments for BEI sentences, however, were unaffected by animacy. Additionally for irreversible sentences, plausible inanimate agents were overall less likely to be chosen than plausible animate agents, as seen for all NNV sentences and BA reversals. These effects demonstrate that while animacy is undoubtedly a cue for argument structure assignment in Mandarin parsers, its effects are more subtle when plausibility is not confounded with animacy, and the coverb BEI is apparently more acceptable with inanimate agents than the coverb BA (see also Fang & Juffs, 2020). The contrasts with previous findings indicate that cue strength is not best captured by a linear ranking. Nonetheless, if we apply the same approach to rank cues by relative strength, our results show that for verb-final sentences, BA = BEI > plausibility > animacy, while word order does not impact agent assignment.

Nonetheless, it is unclear how the model can account for the individual differences we found in the agent assignment task. Although the Competition Model has grown in scope in its more recent forms (MacWhinney, 2005), experimental work within the frame of the model has focused on individual languages' intrinsic properties of cue weighting, with limited consideration for differences among individual comprehenders. For instance, results for condition means have been plotted without error bars or confidence intervals, belying the variability among participants (e.g., Li et al., 1992; MacWhinney, Bates, & Kliegl, 1984; Su, 2001). The Competition Model task paradigm has undeniably revealed important variation across languages, but it is unclear how the model accounts for differences in sentence interpretation among individual parsers of a given language. We note that there has been thorough consideration of the effects of bilingual language experience and the potentials for forward and backward transfer (Liu et al., 1992; MacWhinney, 2017), and this perspective may be extended to encompass individual variability as well. For instance, while Liu and colleagues (1992) group their bilingual speakers in discrete categories, it may be more revealing to find continuous measures, such as age of acquisition or proficiency, to compare individuals.

Competition Model researchers have also noted that real-time processing or processing under time pressure may rely exclusively on the strongest cue in their language, leading them to make different decisions from offline judgments (MacWhinney, 2022a). In the case of our experiment, although their judgments were not made with time pressure, participants may have adopted strategies or processing heuristics to handle the monotony of interpreting more than 300 less-than-natural sentences. In this case, the Competition Model acknowledges that parsers may exclusively rely on a single cue. Our data, however, do not support this prediction. We found that coverbs and plausibility were both strong cues and that most participants relied on both of these cues to make their interpretations. Even if participants were only able to consider a single cue at a time, the existence of the three agent assignment strategies demonstrates that there were differences in individuals' reliance on the available cues. These findings leave an open question as to whether our participants differed merely in their preferred strategy for the task or if they have truly diverging cue weighting strategies in real-time sentence processing.

5.4b eADM

Like the Competition Model, the eADM is centered around crosslinguistic variation in sentence processing. The eADM is also an account of real-time processing where parsers weight cues word-by-word and do not wait for the verb before constructing an initial argument structure of the sentence. Accordingly, we consider how the eADM relates to our incremental ERP results from sentence onset to the verb.

At the sentence-initial noun, our results showed no effect of animacy on N400 amplitude, in contrast to prior reports in English (Bourguignon et al., 2012; Kyriaki et al., 2020). The eADM is compatible with a subject-first (or agent-first) preference for sentence comprehension (Krebs et al., 2018; L. Wang et al., 2009), where sentence-initial inanimate nouns elicit N400 effects because they are less ideal agents than animate nouns (Bourguignon et al., 2012; Kyriaki et al., 2020). Our own results suggest that Mandarin is either different from English or our subjects adapted to animacy being an unreliable cue in our experiment. At the position of the coverb, it is difficult to determine what the eADM might predict. BA and BEI assumedly have high prominence for assigning argument structure, and their particular features as coverbs may conceivably lead to their being processed as predicates in the eADM architecture, with a linking computation performed that could engender an N400 effect. We found no evidence of an N400 effect on a coverb, but the P200 difference between BA and BEI may be relevant to the linking computation. There is also the possibility that differences between the coverb explain the eADM's previous report of an N400 for BEI reversals but not for BA reversals (Bornkessel-Schlesewsky et al., 2011).

The eADM also makes explicit predictions for differences in role reversal processing across languages. Because Mandarin is a sequence-independent language, comprehenders must rely on other cues during online sentence processing, and the eADM predicts that the relative prominence of these other cues (e.g., plausibility, animacy) means that Mandarin role reversals should elicit N400 and not P600 effects (Bornkessel-Schlesewsky et al., 2011). In the vein of this prediction, a previous eADM study on Mandarin reported an N400 effect for role reversals with the coverb BEI but not with BA, with the latter showing no amplitude difference between reversal and plausible sentences (Bornkessel-Schlesewsky et al., 2011). In line with the eADM predictions, our own ERP results showed a main effect of Plausibility on N400 amplitude without a P600. However, we also found a frontal positivity that was not predicted by the model. Assuming that the eADM authors also consider this type of frontal positivity as functionally distinct from the P600, it is unclear if this effect can also be attributed to the model's well-formedness check or reanalysis and repair.

The eADM may provide some explanation of the interactions with Agent Animacy shown in our ERP results. Bornkessel-Schlesewsky and colleagues (2011) suggested the possibility of animacy differences for the difference in N400 effect between BA and BEI. Their BA role reversals always started with an animate noun (e.g., *侦探把子弹击中了* "detective BA bullet hit") while BEI role reversals always started with an inanimate noun (e.g., *子弹被侦探击中了* "the bullet BEI the detective hit"), thus resulting in equivalent cue conflicts in both sentences, with plausibility suggesting an inanimate agent while the proverb suggested an animate agent. The possibility of an animacy effect is interesting in light of the present study's finding of a three-way interaction among Plausibility, Agent Animacy, and Structure for the N400 role reversal effect (see Figure 12). This interaction showed that BEI role reversals elicited a significant N400 effect when the plausible agent was animate (e.g., *仆人被镜子擦亮了* "servant BEI mirror polished"), but no difference when the plausible agent was inanimate (e.g., *鸟笼被喜鹊困住了*"birdcage BEI magpie trapped"). BA role reversals, on the other hand, were predicted to have a numerically more negative N400 whether the plausible agent was animate or inanimate, although the pairwise comparisons for the Agent Animacy sub-conditions were not significant. This suggests that the prominence of Agent Animacy may be affected by the particular coverb involved, or vice versa.

5.4c Bag of Arguments

The Bag of Arguments account predicts that verb-final role reversals should always elicit a P600 effect because the sentence is implausible and there is a failure in comprehension; an N400 effect, however, can be generated if there is both sufficient time to generate a prediction (at least 800 ms) and a sufficient degree of relatedness between the verb and its arguments (Chow et al., 2018). To our understanding, the 800-ms limit proposed by the Bag of Arguments model is an experimental measurement, not a theoretical cognitive limit; thus, the present study's SOA of 750 ms may have been enough time for participants to compute a prediction of the upcoming verb. Assuming that role reversal verbs were sufficiently predictable and that 750 ms is enough time for prediction, then an N400 effect would be consistent with the Bag of Arguments predictions.

But while the Bag of Arguments may ultimately be able to account for our N400 effect, it is difficult to reconcile this model with the lack of a P600. Note that while our model results with a post-onset baseline did not show a clear effect of Plausibility in the P600 time window, there was a more prominent late, sustained positivity observed for the pre-onset baseline. It may initially appear that our baseline selection inflated the N400 effect and decreased the P600 effect, but in fact the N400 effect was still significant for the pre-onset baseline and the greater positivity was only seen at frontal electrodes. Multiple accounts have qualitatively distinguished between the classic posterior P600 effects and frontal positivities seen in the same time window, including the Bag of Arguments (Chow et al., 2018). According to their account, frontal positivities are associated with prediction error, while P600 effects are limited to integration and detection of implausibility. According to our interpretation of the Bag of Arguments account, the crux of their model is predicting the presence or absence of an N400 effect; therefore, the absence of a P600 effect may not necessarily be at odds with the predictions. Nonetheless, the Bag of Arguments may be aided by establishing a clearer functional role for the P600.

We note that there are also differences between our data and the evidence on which the Bag of Arguments account is based. While the present study included both BA and BEI coverbs, the Bag of Arguments has to date only considered processing of BA sentences. Given that BA and

BEI encompass different syntactic and semantic usage, it is possible that verb predictions may be different between the two coverbs. Indeed, the predictions from the models for P600 amplitude indicate that BA reversal sentences had an overall positivity, which is consistent with the Bag of Arguments' predictions, while the effect for BEI reversals was inconsistent between animate and inanimate plausible agents. Note that this positivity was still frontally distributed, as seen in Figure 9. Additionally, the Bag of Arguments has mostly considered role reversals with two animate arguments, while role reversals in the present experiment always had an animacy contrast between the two nouns and the plausible agent could be either animate or inanimate. Not only is this difference likely to have impacted verb prediction, shared or contrasting animacy status may also have influenced semantic integration. While the meaning of role reversals with two animate arguments in Bag of Arguments experiments were extremely unlikely (e.g., *侦探把子弹击中了* "detective BA bullet hit"), many role reversals in the present experiment were impossible (e.g., *仆人被镜子擦亮了* "servant BEI mirror polished").

5.5 Takeaways and limitations

Of the three sentence processing models that have considered Mandarin data, none can explain all of our findings. We identify three areas where our results suggest at least one of the models can be expanded. First, none of the three models considers individual differences in their accounts. By explicitly identifying areas of their proposed architecture where individual variability is expected to play a role and identifying limits or parameters in how that feature may vary across individuals, the models could provide testable predictions. Second, only the eADM makes predictions for parsing prior to the verb. While verbs are often at the core of argument structure, there is incremental sentence parsing prior to verb comprehension, especially for verb-final structures (Angela D. Friederici & Frisch, 2000; Levy & Keller, 2013). Third, languages vary in how they express cues for assigning argument structure, and each of the models can expand in their predictions or explanations for crosslinguistic differences. The Competition Model's conception of cue strength is limited to linear ranking, which is in contrast with our data that shows interactions among cues. The eADM has clear mechanisms for crosslinguistic differences in cue weighting, but we note that there may be the need to account for further crosslinguistic variability in additional stages of their model, such as in step one for syntactic category identification. For instance, Mandarin has flexible syntactic category and the coverbs BA and BEI may carry out some of the roles that the eADM assigns to verbs. The Bag of

Arguments model does not explicitly consider crosslinguistic variability, but the focus on prediction of upcoming verbs suggests a possible extension to compare languages with and without verb-final structures to evaluate if the timing restraints vary as a function of language experience.

We note that differences in experimental design and analysis complicate direct comparison of our results with other studies. Firstly, behavioral tasks for studying argument structure and role reversals have taken multiple forms, from acceptability or plausibility judgments (Chow & Phillips, 2013; Chow et al., 2018), comprehension questions (Bornkessel-Schlesewsky et al., 2011), and forced agent assignment (Li et al., 1992). We hold that the behavioral results of the agent assignment task provide insight beyond other tasks and combine with ERP data to give a more complete picture of processing. However, task is known to affect ERP results (Schacht et al., 2014), including for processing of role reversals (Kyriaki et al., 2020). Negative ERP components like the N400 are sometimes described as automatic and resistant to task manipulation (Bornkessel-Schlesewsky & Schlesewsky, 2019), but an explicit semantics-based task like in the present experiment may promote N400 effects (Schacht et al., 2014), while tasks like acceptability judgments may increase P600 amplitude (Brouwer & Crocker, 2017). These differences further extend to verb type and modality (Bourguignon et al., 2012; Kyriaki et al., 2020).

Experimental findings are also impacted by how data are analyzed. For instance, some effects may be present only when using ANOVAs or only when using linear models (Kyriaki et al., 2020). Mixed effects models like those in the present study have become a new standard in the field of language sciences (Baayen, Davidson, & Bates, 2008), but there is considerable variation in their implementation. Mixed effects models are championed for their capacity to factor in both item- and participant-level variability for predictions, but experts in the field hold disparate views (e.g., Barr et al., 2013 vs Matuschek et al., 2017). Critical for the interpretation of the present study's results, most prior behavioral results for Mandarin role reversals have not been analyzed with mixed effects models (except for Chow et al., 2016, who did not include random slopes), and to our knowledge no Mandarin ERP findings for role reversals have been analyzed with mixed effects models. In the broader field of psychological sciences, we are often interested in relatively small effects that, although of theoretical interest, may prove elusive to measure and demonstrate experimentally. Given these challenges, it is essential that we interpret findings in the greater context of other studies.

6 Conclusion

Mandarin is a language with flexible word order, and the results of this study showed that Mandarin parsers relied on multiple cues for interpreting verb-final sentences. Comprehenders preferentially relied on the coverbs BA and BEI if available, and otherwise chose the more plausible interpretation. Even when no other cues were available and neither argument was more plausible, word order was not used to inform agent assignment. Role reversals elicited an N400 effect with no subsequent P600, consistent with the eADM but difficult to reconcile with the Bag of Arguments account. This finding suggests that argument structure processing varies qualitatively across languages. However, none of the models that have considered Mandarin data can fully explain our findings. We suggest that sentence processing accounts incorporate not only basic parsing mechanisms with crosslinguistic variation, but also individual differences, incremental parsing through the entirety of a sentence, and the diversity of sentence structures both within and between languages.

7 References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, E, & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). Cambridge: Cambridge University Press.

Bates, Elizabeth, McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, *11*(3), 245–299. https://doi.org/10.1016/0010-0277(82)90017-8

Bender, E. (2000). The Syntax of Mandarin Ba: Reconsidering the Verbal Analysis. *Journal of East Asian Linguistics*, *9*(2), 105–145.

Bickel, B., & Nichols, J. (2013). Inflectional Synthesis of the Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/22

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. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016). The Argument Dependency Model. In *Neurobiology of Language* (pp. 357–369). https://doi.org/10.1016/B978-0-12-407794-2.00030-4

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(FEB), 1–17. https://doi.org/10.3389/fpsyg.2019.00298

Bornkessel, I., & Schlesewsky, M. (2006a). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

Bornkessel, I., & Schlesewsky, M. (2006b). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821.

https://doi.org/10.1037/0033-295X.113.4.787

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. https://doi.org/10.1016/j.bandl.2012.05.001

Brouwer, H., & Crocker, M. W. (2017). On the Proper Treatment of the N400 and P600 in Language Comprehension. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01327

Brouwer, H., Delogu, F., Venhuizen, N. J., & Crocker, M. W. (2021). Neurobehavioral Correlates of Surprisal in Language Comprehension: A Neurocomputational Model. *Frontiers in Psychology*, *12*(February), 1–16. https://doi.org/10.3389/fpsyg.2021.615538

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. https://doi.org/10.1016/j.brainres.2012.01.055

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0010729

Chan, I. L., & Chang, C. B. (2018). LEXTALE _ CH: A quick , character-based proficiency test for Mandarin Chinese. *Proceedings of the Annual Boston University Conference on Language Development*, *42*(1), 114–130. Retrieved from https://hdl.handle.net/2144/27833

Chappell, H., Ming, L., & Peyraube, A. (2007). Chinese linguistics and typology: The state of the art. *Linguistic Typology*, *11*(1), 187–211. https://doi.org/10.1515/LINGTY.2007.014

Chomsky, N. (1957a). 4. Phrase Structure. In *Syntactic Structures* (pp. 26–33). https://doi.org/10.1515/9783112316009-005

Chomsky, N. (1957b). Syntactic Structures. In *Matériaux pour l'histoire de notre temps* (Vol. 27). https://doi.org/10.1515/9783112316009

Chomsky, N. (1993). Chapter 2. Subsystems of core grammar. 2.6. LF-representation and θ-theory (2). In *Lectures on Government and Binding* (pp. 101–117). https://doi.org/10.1515/9783110884166.101

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, *1506*, 76–93. https://doi.org/10.1016/j.brainres.2013.02.016

Chow, W., Smith, C., Lau, E., & Phillips, C. (2016). A " bag-of-arguments " mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/23273798.2015.1066832

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), 803–828. https://doi.org/10.1080/23273798.2018.1427878

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. https://doi.org/10.1080/23273798.2016.1160135

Christiansen, M. H., & Chater, N. (2015). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X1500031X

Comrie, B., Dryer, M. S., Gil, D., & Haspelmath, M. (2013). Introduction. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/s1

Croft, W. (2011). Parts of speech as language universals and as language-particular categories. *Approaches to the Typology of Word Classes*, (21), 65–102. https://doi.org/10.1515/9783110806120.65

Cunnings, I., & Fujita, H. (2021). Quantifying individual differences in native and nonnative sentence processing. *Applied Psycholinguistics*, *42*(3), 579–599. https://doi.org/10.1017/S0142716420000648

DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, *61*, 150–162. https://doi.org/10.1016/j.neuropsychologia.2014.06.016

Deng, X., Mai, Z., & Yip, V. (2018). An aspectual account of ba and bei constructions in child Mandarin. *First Language*, *38*(3), 243–262. https://doi.org/10.1177/0142723717743363

Dowty, D. (1991). Thematic Proto-Roles and Argument Selection. *Language*, *67*(3), 547. https://doi.org/10.2307/415037

Dryer, M. S. (2013). Order of Subject, Object and Verb. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/81

Edwards, J. (2006). Foundations of Bilingualism. In *The Handbook of Bilingualism* (pp. 7–31). https://doi.org/10.1002/9780470756997.ch1

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fang, S., & Juffs, A. (2020). Offline processing of Ba- and Bei- constructions in Mandarin Chinese. *Proceedings of the 23rd North American Conference on Chinese Linguistics (NACCL-32)*, 422–437. Storrs, CT.

Federmeier, K. D., Segal, J., Lombrozo, T., & Kutas, M. (2000). Brain responses to nouns, verbs and class-ambiguous words in context. *Brain, 123*(12), 2552–2566. https://doi.org/10.1093/brain/123.12.2552

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of*

Memory and Language, *30*(6), 725–745. https://doi.org/10.1016/0749-596X(91)90034-H

Ferreira, F., & Patson, N. D. (2007). The 'Good Enough' Approach to Language Comprehension. *Language and Linguistics Compass*, *1*(1–2), 71–83. https://doi.org/10.1111/j.1749-818x.2007.00007.x

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

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. https://doi.org/10.1006/jmla.2000.2709

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. https://doi.org/10.1017/CBO9780511790942

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. https://doi.org/10.1080/01690969308407585

Hahn, M., Jurafsky, D., & Futrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(5), 2347–2353. https://doi.org/10.1073/pnas.1910923117

Herbay, A. (2022). *ERPscope*. Retrieved from https://github.com/aherbay/erpscope

Hickok, G. (1993). Parallel parsing: Evidence from reactivation in garden-path sentences. *Journal of Psycholinguistic Research*, *22*(2), 239–250. https://doi.org/10.1007/BF01067832

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. https://doi.org/10.1016/j.cogbrainres.2003.10.022

Hsien-Yi, G. W. (1998). *THE SYNTAX-SEMANTICS INTERFACE : THE BA AND BEI CONSTRUCTIONS IN MANDARIN*. National University of Singapore.

Huang, C.-R., Chen, C.-J., & Shen, C. C. C. (2002). The Nature of Categorical Ambiguity and Its Implications for Language Porceesing: A Corpus-based Study of Mandarin Chinese. *Sentence Processing in East Asian Languages*, 53–83.

Huang, C.-T. J., Li, Y.-H. A., & Li, Y. (2009). *The Syntax of Chinese*. https://doi.org/10.1017/CBO9781139166935

Iggesen, O. A. (2013). Number of Cases. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/49

Inkscape Project. (2020). *Inkscape*. Retrieved from https://inkscape.org

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, *1*(6), 571–591. https://doi.org/10.1111/j.1749-818x.2007.00037.x

Kasparian, K., & Steinhauer, K. (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. *Frontiers in Psychology*, *8*(MAR), 1–22. https://doi.org/10.3389/fpsyg.2017.00389

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

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

Kolk, H., & Chwilla, D. (2007). Late positivities in unusual situations. *Brain and Language*, *100*(3), 257–261. https://doi.org/10.1016/j.bandl.2006.07.006

Kolk, H. H. J., Chwilla, D. J., Van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language*, *85*(1), 1–36. https://doi.org/10.1016/S0093-934X(02)00548-5

Kos, M., Vosse, T., van den Brink, D., & Hagoort, P. (2010). About edible restaurants: Conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology*, *1*(DEC), 1–11. https://doi.org/10.3389/fpsyg.2010.00222

Krebs, J., Malaia, E., Wilbur, R. B., & Roehm, D. (2018). Subject preference emerges as cross-modal strategy for linguistic processing. *Brain Research*, *1691*, 105–117. https://doi.org/10.1016/j.brainres.2018.03.029

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49. https://doi.org/10.1016/j.brainres.2006.12.063

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A Tale of Two Positivities and the N400: Distinct Neural Signatures Are Evoked by Confirmed and Violated Predictions at Different Levels of Representation. *Journal of Cognitive Neuroscience*, *32*(1), 12–35. https://doi.org/10.1162/jocn_a_01465

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Kyriaki, L., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Semantic reversal anomalies under the microscope: Task and modality influences on language-associated event-related potentials. *European Journal of Neuroscience*, *52*(7), 3803–3827. https://doi.org/10.1111/ejn.14862

LaPolla, R. (1993). *Arguments Against ' Subject ' and ' Direct Object ' as Viable Concepts in Chinese Arguments Against ' Subject ' and ' Direct Object ' · as Viable c · o _ ncepts 1n Chinese* *. (January 1993).

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Levy, R. P., & Keller, F. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language*, *68*(2), 199–222. https://doi.org/10.1016/j.jml.2012.02.005

Li, C. N., & Thompson, S. A. (1976). On the issue of word order in a synchronic grammar: A case against 'movement transformations.' *Lingua*, *39*(3), 169–181. https://doi.org/10.1016/0024-3841(76)90001-2

Li, C., & Thompson, S. (1974). Co-verbs in Mandarin Chinese: verbs of prepositions? *Journal of Chinese Linguistics*, *2*(3), 257–278. Retrieved from https://www.jstor.org/stable/23749820

Li, P. (1998). Crosslinguistic Variation And Sentence Processing: The Case Of Chinese. In D. Hillert (Ed.), *Sentence Processing: A Crosslinguistic Perspective* (pp. 33–53). https://doi.org/10.1163/9780585492230_004

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1992). Cues as Functional Constraints on Sentence Processing in Chinese. In H. C. Chen & O. Tzeng (Eds.), *Language Processing in Chinese* (pp. 207–234). https://doi.org/10.1016/S0166-4115(08)61893-2

Li, P., & MacWhinney, B. (2012). Competition Model. In *The Encyclopedia of Applied Linguistics* (pp. 1–5). https://doi.org/10.1002/9781405198431.wbeal0168

Li, & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Liao, C. H., & Lau, E. (2020). Enough time to get results? An ERP investigation of prediction with complex events. *Language, Cognition and Neuroscience*, *35*(9), 1162–1182. https://doi.org/10.1080/23273798.2020.1733626

Liu, C. (2014). *A modular theory of radical pro drop*. Harvard University.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451–484. https://doi.org/10.1017/S0142716400005762

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01171

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis - The assumption that measurement error always reduces effect sizes is fals. *Science*, *355*(6325), 582–584.

Lowder, M. W., & Gordon, P. C. (2013). It's hard to offend the college: Effects of sentence structure on figurative-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 993–1011. https://doi.org/10.1037/a0031671

Lowder, M. W., & Gordon, P. C. (2015). Natural forces as agents: Reconceptualizing the animate–inanimate distinction. *Cognition*, *136*, 85–90. https://doi.org/10.1016/j.cognition.2014.11.021

Lu, C.-C., Bates, E., Li, P., Tzeng, O., Hung, D., Tsai, C.-H., … Chung, Y.-M. (2000). Judgements of grammaticality in aphasia: The special case of Chinese. *Aphasiology*, *14*(10), 1021–1054. https://doi.org/10.1080/02687030050156593

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. Retrieved from https://cran.r-project.org/package=sjPlot

Macwhinney, B. (2017). Are these approaches incompatible ? *Linguistic Approaches to Bilingualism*, *6*, 730–733. https://doi.org/10.1075/lab.00014.mac

MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

MacWhinney, B. (2018). *Chapter 14. A unified model of first and second language learning*. https://doi.org/10.1075/tilar.22.15mac

MacWhinney, B. (2022). The Competition Model: Past and Future. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A Life in Cognition* (pp. 3–16). https://doi.org/10.1007/978-3-030-66175-5_1

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Martin, A. E. (2016). Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. *Frontiers in Psychology*, *7*(February), 1–17. https://doi.org/10.3389/fpsyg.2016.00120

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McFadden, T. (2003). On Morphological Case and Word-Order Freedom. *Annual Meeting of the Berkeley Linguistics Society*, *29*(1), 295. https://doi.org/10.3765/bls.v29i1.997

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Miao, X. (1981). Word order and semantic strategies in Chinese sentence comprehension. *International Journal of Psycholinguistics*, *8*(3), 109–122.

Morís Fernández, L., & Vadillo, M. A. (2020). Flexibility in reaction time analysis: many roads to a false positive? *Royal Society Open Science*, *7*(2), 190831. https://doi.org/10.1098/rsos.190831

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. https://doi.org/10.1162/jocn.1991.3.2.151

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z

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. https://doi.org/10.1016/j.bandl.2007.09.005

Philipp, M., Graf, T., Kretzschmar, F., & Primus, B. (2017). Beyond verb meaning: Experimental evidence for incremental processing of semantic roles and event structure. *Frontiers in Psychology*, *8*(OCT), 1–16. https://doi.org/10.3389/fpsyg.2017.01806

Potts, G. F. (2004). An ERP index of task relevance evaluation of visual stimuli. *Brain and Cognition*, *56*(1), 5–13. https://doi.org/10.1016/j.bandc.2004.03.006

Potts, G. F., & Tucker, D. M. (2001). Frontal evaluation and posterior representation in target detection. *Cognitive Brain Research*, *11*(1), 147–156. https://doi.org/10.1016/S0926-6410(00)00075-6

Preissl, H., Pulvermüller, F., Lutzenberger, W., & Birbaumer, N. (1995). Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters*, *197*(1), 81–83. https://doi.org/10.1016/0304-3940(95)11892-Z

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510–532. https://doi.org/10.1037/0033-2909.114.3.510

Russell, L. (2019). *emmeans: estimated marginal means, aka least-squares means. R package version 1.4. 2*.

Sakamoto, T. (2015). 15. Processing of syntactic and semantic information in the human brain: Evidence from ERP studies in Japanese. In *Handbook of Japanese Psycholinguistics* (pp. 457–510). https://doi.org/10.1515/9781614511212-022

Schacht, A., Sommer, W., Shmuilovich, O., Martínez, P. C., & Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE*, *9*(3), 1–7. https://doi.org/10.1371/journal.pone.0091226

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for transfer appropriate processing of memory: Processing versus feature similarity. *Psychonomic Bulletin & Review*, *14*(4), 612–619. https://doi.org/10.3758/BF03196810

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162. https://doi.org/10.1016/j.bandl.2011.07.001

Steinhauer, K., Pancheva, R., Newman, A. J., Gennari, S., & Ullman, M. T. (2001). How the mass counts: An electrophysiological approach to the processing of lexical features. *NeuroReport*, *12*(5). Retrieved from https://journals.lww.com/neuroreport/Fulltext/2001/04170/How_the_mass_counts__An_electrophysiological.27.aspx

Su, I. R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112. https://doi.org/10.1017/S0142716401001059

Sun, C.-F., & Givon, T. (1985). On the So-Called Sov Word Order in Mandarin Chinese: A Quantified Text Study and Its Implications. *Language*, *61*(2), 329. https://doi.org/10.2307/414148

Ting, J. (1998). Deriving the bei-construction in Mandarin Chinese. *Journal of East Asian Linguistics*, *7*(4), 319–354. https://doi.org/10.1023/A:1008340108602

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

van de Meerendonk, N., Kolk, H. H. J., Chwilla, D. J., & Vissers, C. T. W. M. (2009). Monitoring in language perception. *Linguistics and Language Compass*, *3*(5), 1211–1224. https://doi.org/10.1111/j.1749-818X.2009.00163.x

Van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research*, *22*(2), 241–255. https://doi.org/10.1016/j.cogbrainres.2004.09.002

Voeten, C. C. (2021). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. Retrieved from https://cran.r-project.org/package=buildmer

Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the 'subject'-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language and Cognitive Processes*, *24*(7–8), 1180–1226. https://doi.org/10.1080/01690960802159937

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The Role of Animacy in Online Argument Interpretation in Mandarin Chinese. In *Case, Word Order and Prominence* (Vol. 40, pp. 91–119). https://doi.org/10.1007/978-94-007-1463-2_5

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Yan, J., & Huizhong, Y. (2006). The English Proficiency of College and University Students in China: As Reflected in the CET. *Language, Culture and Curriculum*, *19*(1), 21–36. https://doi.org/10.1080/07908310608668752

Yang, J. (2006). Learners and users of English in China. *English Today*, *22*(2), 3–10. https://doi.org/10.1017/S0266078406002021

Yang, N., Waddington, G., Adams, R., & Han, J. (2018). Translation, cultural adaption, and test–retest reliability of Chinese versions of the Edinburgh Handedness Inventory and Waterloo Footedness Questionnaire. *Laterality: Asymmetries of Body, Brain and Cognition*, *23*(3), 255–273. https://doi.org/10.1080/1357650X.2017.1357728

Ye, Z., & Zhou, X. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103–115. https://doi.org/10.1016/j.brainres.2008.01.090

Yu, S., & Tamaoka, K. (2018). Age-related differences in the acceptability of non-canonical word orders in Mandarin Chinese. *Lingua Sinica*, *4*(1). https://doi.org/10.1186/s40655-018-0035-x

Ziegler, J., Snedeker, J., & Wittenberg, E. (2018). Event Structures Drive Semantic Structural Priming, Not Thematic Roles: Evidence From Idioms and Light Verbs. *Cognitive Science*, *42*(8), 2918–2949. https://doi.org/10.1111/cogs.12687

Link between Manuscripts I and II

In Manuscript I, we considered transitive, verb-final Mandarin sentences and how the comprehension of argument structure is managed amidst competing cues. Our results demonstrated that this comprehension is based on more than a simple linear ranking of cue strength; instead, cues interact to boost or counteract each other in intricate ways. For instance, in role reversal sentences where sentence structure indicates an implausible agent, BA is a stronger cue when indicating an animate noun, while agent selection choices for BEI reversals are unaffected by animacy. Our data further showed that word order is not used as a cue for argument structure assignment when there is no indication from either structure or plausibility. This finding for word order was in contrast with some accounts of a preference for object-subject-verb order (P. Li et al., 1992; L. Wang et al., 2012). Our results for word order suggest that the cue of pre-verb nouns being agents may not be as strong as the cue of post-verb nouns being patients, at least in the case of verb-final sentences.

In Manuscript II, we turned to the case of Mandarin adjective-noun placement and comparing native and non-native processing. Unlike argument structure interpretation in verb-final sentences, word order is important for interpretation in adjective-noun phrases. Attributive adjectives in Mandarin typically appear in prenominal position—as in English—but may also follow their nouns as a predicate without an intervening copula (Paul, 2010). We constructed sentences where the preceding sentence context licensed only prenominal adjective placement to test native and non-native processing adjective order violations. In contrast to argument structure in Mandarin verb-final sentences, word order is important for Mandarin adjective placement, and the underlying rules and variation in word order for adjective structures differ in important but relatively subtle ways from English. Comparing the ERP profiles of the processing of adjective placement between native and non-native Mandarin and English, along with individual differences, allowed for an appreciation of typological differences between the two languages and their potential for interference in bilinguals.

Together, the two manuscripts overview processing of Mandarin sentence structures in a crosslinguistic context. Manuscript I considered a complicated structure with four competing cues for agent assignment. Manuscript II considered a relatively simpler structure, but with a direct comparison between English and Mandarin native processing and native and non-native Mandarin processing. Mandarin and English have been suggested as two “remarkably interpenetrable" languages for bilingual processing (H. Liu et al., 1992), and these two manuscripts lay the foundation for future work comparing Mandarin second language processing and first language attrition.

1 References

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1992). Cues as Functional Constraints on Sentence Processing in Chinese. In H. C. Chen & O. Tzeng (Eds.), *Language Processing in Chinese* (pp. 207–234). https://doi.org/10.1016/S0166-4115(08)61893-2

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451–484. https://doi.org/10.1017/S0142716400005762

Paul, W. (2010). *Adjectives in Mandarin Chinese*. (1994), 115–152. https://doi.org/10.1075/la.153.04pau

Wang, L., Schlesewsky, M., Philipp, M., & Bornkessel-Schlesewsky, I. (2012). The Role of Animacy in Online Argument Interpretation in Mandarin Chinese. In *Case, Word Order and Prominence* (Vol. 40, pp. 91–119). https://doi.org/10.1007/978-94-007-1463-2_5

Manuscript II

Native and non-native parsing of adjective placement – an ERP study of Mandarin and English sentence processing

Max Wolpert[1,2], Hui Zhang[3], Shari Baum[2,4], Karsten Steinhauer[2,4]

Affiliations:

[1]Integrated Program in Neuroscience, McGill University, Montreal, QC, Canada

[2]Centre for Research on Brain, Language and Music, Montreal, QC, Canada

[3]School of Foreign Languages and Cultures, Nanjing Normal University, Nanjing, Jiangsu, China

[4]School of Communication Sciences and Disorders, McGill University, Montreal, QC, Canada

1 Abstract

The structural rules governing adjective placement vary across languages. In English and Mandarin, adjectives are typically prenominal, appearing before the nouns they modify. Mandarin adjectives, however, may also appear in a postnominal position, where they serve as a sentence predicate without an intervening copula verb, a structure that is generally not permitted in English. To understand how these crosslinguistic differences affect native and non-native processing, we used event-related potentials (ERPs) to study processing of adjective placement in Mandarin monolinguals (n = 32) and English-Mandarin bilinguals (n = 20, first language English, second language Mandarin). We further compared individual datasets to understand the role of individual differences on ERP patterns, per recommendations from Tanner & Van Hell (2014) and Fromont, Steinhauer, & Royle (2020). For the English experiment, results confirmed the previous report of a biphasic N400-P600 response on the second word (Steinhauer, 2014), with quantitative differences in effect onset and duration that likely reflect task strategy. For the Mandarin experiment, native and non-native speakers alike showed a greater N400 response to the second word for the ungrammatical noun-adjective order than the grammatical adjective-noun order; however, only the native speaker group showed a subsequent P600 effect. At the group-level, these results mirror those from the English experiment, but suggest that even at high proficiency levels, non-native speakers may not have applied rule-based reanalysis to process Mandarin adjective-noun pairs, as evidenced by the lack of a P600 effect for non-native speakers. Individual difference results showed that N400 and P600 effect magnitudes were correlated, but this correlation disappeared in most conditions when applying stringent measures to minimize artifacts from comparing adjacent time windows on identical electrodes. Nonetheless, even when this correlation was not present, participants varied continuously from P600-dominant to N400-dominant response patterns. Mandarin native speakers additionally showed an N400 effect at the first word position, but with a greater N400 amplitude for the grammatical adjective-noun condition. We interpret this effect as relating to the lexical properties of monomorphemic adjective-noun pairs without the particle DE, and propose future experiments to test this hypothesis.

KEYWORDS: Mandarin, adjectives, non-native processing, ERPs, individual differences

2 Introduction

Most of the worlds' languages have a distinct word class to modify nouns: the adjective (R. M. Dixon, 1982). An adjective expresses a certain quality of the noun (R. M. Dixon, 2004), and the morphology and structural rules governing adjectives vary across languages. Ukrainian, for instance, inflects adjectives for both grammatical case, gender, and number. Alongside inflection for gender and number, French has some adjectives that typically precede or the nouns they modify, like *petit chat* "little cat", and other adjectives that typically follow the nouns they modify, like *chat mignon* "cat cute". Other languages have relatively fixed adjective-noun order with no inflection, as is the case for Mandarin and English. In both Mandarin and English, word order is an important feature of adjective-noun phrases, and adjectives typically precede the nouns they modify (Sproat & Shih, 1987).

For native speakers, these structural rules are acquired early in life (Nicoladis & Rhemtulla, 2012). When adults learn a new language, however, they must both learn a new set of rules for adjective phrase formation and inhibit conflicting rules from their native language. At high proficiency, non-native speakers are capable of achieving this task and can show native-like processing of adjective phrases. For example, Steinhauer (2014) used electroencephalography (EEG) to study the processing of adjective-noun placement violations in English, in sentences like "Tom puts the vase *tall on the table." Results showed that both native and high-proficiency non-native speakers could show the same neural response pattern. However, the pattern for even high-proficiency non-native speakers to a certain extent still depended on their language background: if participants' first language was French, their ERPs showed a different response to adjectives that were pre-nominal compared to post-nominal in their native French; if their first language was Mandarin, where adjectives broadly match English word order, they showed an identical pattern to the English native speakers. These findings demonstrate that even late bilinguals can process word order rules using native-like mechanisms, but overlap or conflict between native and non-native structural rules still influences this processing.

Although Steinhauer (2014) found the same neural response pattern in English and Mandarin native speakers for English adjective-placement, the underlying grammar of adjective phrases is not the same in the two languages. Although Mandarin adjectives and nouns often have the same linear order as in English, Mandarin adjectives have structural properties not present in their English counterparts. In the present study, we considered processing of Mandarin adjective-noun structures by native monolinguals and by English native speakers with advanced Mandarin proficiency. Below, we summarize some of the linguistic features of Mandarin adjectives, and then introduce predictions about non-native processing and individual differences.

2.1 Mandarin Adjectives

Like English[11], Mandarin adjectives typically precede the nouns they modify and undergo minimal morphological change (Li & Thompson, 1989; Sproat & Shih, 1987). Unlike English, however, Mandarin adjectives may sometimes occur in predicate position without a copula. These intransitive predicate adjectives are distinguished from adjectives that complement a copula (R. M. Dixon, 2004). In Mandarin, predicate adjectives follow the nouns they modify, although usually with an intervening degree adverb, as in (1). Theoretical linguistics accounts typically consider Mandarin adjectives and verbs as distinct syntactic categories (e.g., Arcodia, 2014; although some describe Mandarin predicate adjectives as 'adjectival verbs', as in Li & Thompson, 1989), but it is unclear how variability in adjective-noun order may impact processing.

(1) *张三真聪明。*

Zhangsan zhen congming

Zhangsan really smart

"Zhangsan is really smart."

Additionally, when Mandarin adjectives precede their nouns, it is most common to have the intervening particle DE, as in (2) (Z. Xu, 2018). DE is a particle that serves many grammatical functions, but typically connects a noun and its modifier—an adjective, a relative clause, or a possessor (L. L. S. Cheng, 1986; L. L. S. Cheng & Sybesma, 2009). There is extensive discussion of the proper categorization of DE in Mandarin grammar (e.g., Paul, 2012) and what its function is in adjective phrases (Paul, 2005, 2010). To facilitate comparison with English and to limit the scope of our investigation, we limited this study to adjective-noun phrases that do not contain the particle DE, as in (3).

[11] Note adjective-noun order exceptions in English, such as "heir apparent" or "president elect", or predicate adjectives appearing without a copula, such as "I like my coffee sweet".

(2) *这是很好吃的甜点。*

zhe shi hen haochi de tiandian

this is very tasty DE dessert

"This is a very tasty dessert."

(3) *我昨天买了新书。*

wo zuotian mai LE xin shu

I yesterday buy LE_{PERF}[12] *new book*

"I bought new books yesterday."

Although our focus on adjective-noun phrases without DE encompasses a minority of adjective-noun pairs, there is still considerable discussion by linguists about this structure. From a descriptive level, these adjective-noun pairs are most common with monosyllabic adjectives, as in (3). Xu (2018) analyzes this pattern as a "monomorphemic constraint," where only adjectives containing a single morpheme are permitted to occur without DE. Accordingly, polysyllabic adjectives in combinations such as *漂亮蝴蝶 piaoliang hudie* "pretty butterfly" or *聪明人 congming ren* "smart person" are still considered to comprise a single morpheme in modern Mandarin, although each syllable may have historically been a separate morpheme. Xu (2018) further proposes that these adjective-noun combinations have either been lexicalized or are used in a lexicalized manner, and therefore represent single words (for further evidence to this account, see also 陈刚, 2012). Xu elaborates that these combinations can span from novel adjective-noun pairings for describing objects, such as *红袜子 hong wazi* "red sock(s)", to fully lexicalized compound words, like *美女 meinü* "beautiful woman". Others contend that with or without DE, adjective-noun structures do not pass critical tests for "wordhood" and must therefore be phrases (Paul, 2005, 2010). One linguist has claimed that "no decision is possible" as to whether these structures are words or phrases (Schäfer, 2009).

As thorough as these theoretical analyses are, it is unclear how—or even if—these distinctions could impact sentence processing (Phillips & Wagers, 2012). Sproat & Shih (1987, 1991) made explicit attempts to relate crosslinguistic comparison of adjective-noun ordering to cognitive limitations, including Mandarin structures without DE. Although they describe the

[12] PERF = perfective aspect particle

limits of variability across languages, they note explicitly that they fail to explain "the relationship between the cognitive basis of [adjective-noun] ordering, and the effect it has upon the grammar" (Sproat & Shih, 1987).

Because the present study was based on the previously mentioned study of English adjectives (Steinhauer, 2014), we chose to consider only the adjective-noun structure without DE. While this is not the most typical form of Mandarin adjective-noun structure, it mirrors the structure in English, and word order violations may therefore be processed by a similar mechanism. In the discussion, we interpret our results to make predictions for processing of adjective-noun structures with DE.

2.2 Non-native Processing

Non-native speakers vary considerably in their language performance (Birdsong, 2021), and models of second language acquisition often include different stages of learning (Gitsaki, 1998; Hulstijn, Alderson, & Schoonen, 2010). While beginner learners rely on memorization and translation from their native language, advanced learners can achieve automatic, native-like performance (Steinhauer et al., 2009). This progression from beginner to advanced follows a continuum, and individuals will show distinct outcomes at different points in their learning process (Meisel, Clahsen, & Pienemann, 1981; Smith & Truscott, 2005).

To study real-time, automatic processes like sentence comprehension, ERPs are a fruitful research methodology. Specific ERP patterns are elicited by violations of a language's grammatical rules, including phrase structure (e.g., Osterhout & Holcomb, 1992), agreement (e.g., Molinaro, Barber, & Carreiras, 2011), and grammatical gender and case (e.g., Barber & Carreiras, 2003). In this context, there are three ERP components commonly associated with these grammatical structures. The first of these three is the N400, a negative ERP wave occurring 300-500 milliseconds post-stimulus onset linked to lexical and semantic processing (Kutas & Federmeier, 2011; E. F. Lau et al., 2008; Molinaro et al., 2010). N400 amplitude is classically associated with semantic anomalies (Kutas & Hillyard, 1980), but has also been shown in response to morphosyntactic manipulations (e.g., Bornkessel, McElree, Schlesewsky, & Friederici, 2004; Guajardo & Wicha, 2014). The second and third components are the LAN (left-anterior negativity, first reported by Osterhout & Mobley, 1995) and the P600 (first reported by Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992); these latter two components have often been reported as a paired, biphasic effect in response to morphosyntactic violations (as described by Steinhauer et al., 2009; Tanner & Van Hell, 2014). Note that studies have also reported biphasic N400-P600 responses (Mueller, 2006; Steinhauer, 2014) and single N400 or P600 responses without a biphasic pattern (Nieuwland, Martin, & Carreiras, 2013; Severens, Jansma, & Hartsuiker, 2008) as well, in contrast to the LAN-P600 response. There is continuing discussion concerning the interpretation of these effects (Caffarra, Mendoza, & Davidson, 2019), but each of these three components has been observed in non-native processing (for reviews, see Caffarra, Molinaro, Davidson, & Carreiras, 2015; Kotz, 2009).

Indeed, the pattern of ERP responses to certain grammatical structures has been proposed as an index of progress in second language learning. While native speakers have often shown a LAN-P600 or N400-P600 response to grammatical violations (Caffarra et al., 2019), non-native speakers may show only N400 effects (Steinhauer et al., 2009). To explain these differences, Steinhauer, White, & Drury (2009) outlined discrete stages of non-native learning as indexed by ERP response patterns. In their account, learners' earliest knowledge of second language grammar is limited to declarative memory, with morphosyntactic violations processed as lexical anomalies and therefore only eliciting N400 effects (Morgan-Short, Sanz, Steinhauer, & Ullman, 2010; Morgan-Short, Steinhauer, Sanz, & Ullman, 2012; Osterhout, McLaughlin, Pitkänen, Frenck-Mestre, & Molinaro, 2006; Osterhout et al., 2008; Ullman, 2001). As learners build a second language grammar and begin to follow procedural rules, they then show more native-like patterns (e.g., Bowden, Steinhauer, Sanz, & Ullman, 2013; Dowens, Vergara, Barber, & Carreiras, 2010). Steinhauer and colleagues (2009) suggested that the progression of these patterns, in order of increasing learner proficiency, begins with a small P600 effect, then a native-like P600 effect, and then finally a LAN-P600 pattern that fully resembles the response of native speakers (for review of non-native ERP patterns, see Caffarra, Molinaro, Davidson, & Carreiras, 2015). In cases where grammatical structures instead elicit only P600 or biphasic N400-P600 effects for native speakers, the non-native pattern would be expected to begin with an N400 effect and then transition to a P600 effect (or a biphasic N400-P600), in line with increasing second language proficiency, as seen for processing of adjective placement (Steinhauer, 2014).

While Steinhauer and colleagues' (2009) proposal compares group-level proficiency differences, other studies have highlighted the importance of variability between individuals. By comparing native and non-native processing of German subject-verb agreement, Tanner and colleagues (Tanner et al., 2013) showed that although increasing German proficiency resulted in a native-like P600 effect (as per Steinhauer et al., 2009), individual participants who exhibited low proficiency varied in showing either an N400 or a P600 effect. Additionally, P600 amplitude was positively correlated with accuracy on the experimental task, suggesting that more proficient participants showed greater P600 effects. Tanner and colleagues (2013) interpreted these results as showing that although learners progress through proficiency stages as indexed by their similarity to the native processing ERP profile, individual learners have varying trajectories and may adopt different strategies to comprehend sentences.

2.3 Individual Differences

Tanner and colleagues expanded consideration of individual differences to critique the tradition of studying ERPs using grand averages, where averaged data across participants are considered to faithfully represent an effect at the group level (for discussion see Gaspar, Rousselet, & Pernet, 2011). By comparing grand averaged ERP responses with those of individual subjects, Tanner & Van Hell (2014) showed that a biphasic pattern at the grand average level can be a product of participants who individually show solely either an N400 or a P600, with only a minority of participants showing a biphasic pattern. This finding emerged for participants processing their native language, demonstrating that variability is not limited to non-native processing. Tanner & Van Hell (2014) suggested that participants who showed an N400 effect were forming predictions of verb agreement based on words, while those who showed a P600 effect were instead following combinatorial rules. The authors also presented evidence that participants' N400 and P600 magnitudes were inversely correlated, such that the larger the N400 effect individuals showed, the smaller their P600 effect would be, and vice versa.

To evaluate and compare these individual differences, these above-mentioned analyses were further paired with two quantitative measures (Tanner et al., 2014). The first measure was the Response Magnitude Index (RMI), which is the square root of the sum of the squared differences of the N400 and P600 effects. A larger RMI value indicates a participant has shown a larger ERP response to an experimental condition, indicating increased sensitivity to that manipulation. To quantify individuals' ERP patterns as being N400-dominant or P600-dominant, there is the second measure, the Response Dominance Index (RDI). A negative RDI value indicates an N400-dominant response, while a positive RDI value indicates a P600-dominant response; an RDI value close to zero indicates the N400 and P600 effect sizes are approximately equal. For non-native processing, Tanner and colleagues (2014) further reported relationships between participants' biographic information and their RDI and RMI scores. RMI score was predicted by proficiency measures, showing that more proficient participants had stronger ERP responses to grammatical violations. In contrast, RDI score was predicted by age of acquisition and motivation to sound native-like, meaning that participants who started acquiring the second language earlier and who wanted to appear more native-like showed a more P600-dominant pattern.

Because study of individual differences has focused on processing of morphosyntax with manipulations involving morphological inflection, it is not immediately clear if the same individual differences will emerge for adjective-noun placement in Mandarin and English. Neither language inflects adjectives, so the morpho- part of morphosyntax is not at play. However, one study of adjective placement did show individual differences in ERP patterns (Kemmerer, Weber-Fox, Price, Zdanczyk, & Way, 2007). In this study, Kemmerer and colleagues manipulated adjective pairs so they were congruent ("huge, gray elephant") or in violation of rules for linear order ("*gray, huge elephant"). At the position of the second adjective, grand average results showed a counter-intuitive reduced (less negative) N400 component for the violation conditions, followed by a P600 effect. However, behavioral results showed that approximately half of the participants were accepting of the linear order violations ("*gray, huge elephant"). By dividing the participants into two groups based on their acceptability ratings, the researchers showed that only participants who did not accept linear order violations showed the effect at the second adjective. Participants who accepted linear order violations did not show an amplitude difference in the N400 or P600 time windows but did show a longer latency in N400 peak amplitude for the control condition.

Considering these studies as a whole, we have identified two principal predictions about individual differences in ERP patterns for the structures included in the current investigation. First, the amplitudes of the N400 and P600 effects should be correlated within individuals, such that participants who show a larger N400 response to a given condition should show a smaller P600 response to the same condition, and vice versa. Second, when calculating the ratio of N400 to P600 amplitude, there should be a range of participant responses, from P600-dominant to N400-dominant, as well as individuals who show a biphasic N400-P600 pattern with relatively equal magnitudes of component amplitude. Tanner and colleagues further suggest that this variety of individual responses reflects fundamental differences in parsing strategies (Tanner & Van Hell, 2014). Participants may also differ in their acceptability ratings of adjective-noun word order violations, in which case some participants may show no ERP effect (Kemmerer et al., 2007). Lastly, individual differences may be especially prevalent for second-language learning, where individuals show substantial variation in production and comprehension throughout the learning process (Birdsong, 2021).

Although Tanner and colleagues' explanation offers a clear account for individual differences in ERP patterns, there are two potential weaknesses in the comparison of N400 and P600 amplitudes, as described by Fromont and colleagues (2020). The first is a problem of timing. Because ERP components vary in their onsets, peaks, and offsets, comparing adjacent time windows runs the risk of component overlap. If the N400 and P600 components overlap, a significant correlation between time windows may be driven by a single components' amplitude, giving the false impression that the components' amplitudes are related. The second weakness is a problem of space. Autocorrelation of a signal can be a valid analysis approach (e.g., Brazier & Casby, 1952), but comparison of identical electrodes may reflect an intrinsic property of the electrode or its precise positioning on the scalp, where there is a tendency for certain patterns of electrical activity, or where components overlap in space.

Fromont and colleagues (2020) detailed a solution to correct for these two potential concerns. By selecting non-adjacent time windows for the N400 and P600 components and averaging over distinct electrodes for each component (namely, where each of them is most prominent), it is possible to minimize the problems of both time and space. In Fromont et al.'s (2020) analyses, applying these corrections resulted in no correlation between N400 and P600 amplitudes in most cases. Results further demonstrated that most individuals indeed showed a biphasic N400-P600 pattern, not solely an N400 or P600 as seen by Tanner & Van Hell (2014).

2.4 Present Study

We had three aims for the present study. First, we sought to replicate the findings from Steinhauer (2014) for processing of adjective placement in English by native speakers. We used the same materials as the original study, as summarized in Table . We expected to observe the same pattern as originally reported, namely a biphasic N400-P600 response to the ungrammatical noun-adjective order. Second, given the grammatical properties of Mandarin adjectives, we tested native processing of Mandarin adjective placement to determine how the ERP pattern compares to that associated with English adjective placement. Even though adjectives in both Mandarin and English typically precede the nouns they modify, there are non-trivial differences in the grammar that could potentially influence processing. Nonetheless, we predicted that we would observe an N400-P600 response because we used an adjective-noun structure similar to that used in the English experiment. Third, we extended consideration to non-native processing by testing advanced learners of Mandarin. Steinhauer (2014) showed that high-proficiency second language learners of English can exhibit a native-like ERP pattern for adjective-noun placement violations, even if their native language has a different word order. Because English-Mandarin bilinguals can ostensibly use either language's structural rules to successfully comprehend adjective-noun placement, we predicted that high-proficiency Mandarin learners should show a native-like ERP pattern.

Alongside these three aims, we further considered the role of variability among individuals in ERP patterns as described by Tanner & Van Hell (2014). Per two predictions in this account, N400 and P600 effects should correlate, and most participants should be either N400-dominant or P600-dominant, without a biphasic pattern. As per Fromont and colleagues (2020), however, N400-P600 effect correlations should be absent when comparing non-adjacent time windows and non-identical groups of electrodes, and individual participants should tend to show a biphasic N400-P600 effect. Adjective-noun order in English and Mandarin is distinct from the morphosyntactic manipulations explored by Tanner & Van Hell (2014) and the syntactic category manipulation by Fromont and colleagues (2020), so we were not certain that we would observe the same individual ERP patterns. Nonetheless, to our understanding, the notion of individuals separately recruiting either lexical (i.e., N400) or combinatorial (i.e., P600) processing streams (Tanner & Van Hell, 2014) is not intrinsically limited to certain types of linguistic manipulations.

Table 1. ***Example sentence materials for English experiment.*** ** indicates points of ungrammaticality. All sentence materials ended with a three-word prepositional phrase.*

Word Order	Example sentence
adjective-noun (grammatical)	Tom puts the **tall vase** on the table.
noun-adjective (ungrammatical)	Tom puts the **vase *tall** on the table.

Table 2. ***Example sentence materials for Mandarin experiment.*** ** indicates point of ungrammaticality. Although 新 "new" could conceivably occur after 书 "book" (e.g., 这些书新，那些书旧 "these books are new, those books are old"), the*

preceding sentence context makes this word order unacceptable. All sentence materials ended with the perfective aspect particle LE.

Word Order	Example sentence
adjective-noun (grammatical)	父亲 最近 想看 新 书 了。 Father recently wants to read new book LE Father recently wants to read new books.
noun-adjective (ungrammatical)	父亲 最近 想看 书 新 了。 Father recently wants to read book new LE Father recently wants to read books *new.

3 Methods

3.1 Participants

In the Mandarin experiment, a total of 54 participants participated, 32 who were monolingual, native speakers and 22 who were English native speakers with high Mandarin proficiency. The 22 English native speakers also participated in the English experiment. Two native-English-speaking participants' recordings had complications, so only 20 participants' data were analyzed for this group. For native Mandarin speakers, recruitment was limited to participants who primarily communicated in Mandarin and had limited knowledge of English and other languages, including Chinese languages and dialects. All Mandarin native speakers were right-handed as based on the Edinburgh Handedness Inventory (average score = 83, SD = 19, Mandarin version from Yang, Waddington, Adams, & Han, 2018). For native English speakers, we adjusted recruitment standards in response to the difficulty in finding high-proficiency second language learners of Mandarin. Of the 22 English native speakers, two participants were left-handed based on the Edinburgh Handedness Inventory[13] (average score = 62, SD = 47, Oldfield, 1971). Participants further completed a LEXTALE in Mandarin (I. L. Chan & Chang, 2018) and English (Lemhöfer & Broersma, 2012), as well as a detailed language background questionnaire summarized in Table 3 and Table 4.

All participants had normal or corrected-to-normal vision, did not have any history of neurological disorders, and were recruited via online advertisement and word of mouth in

[13] In the discussion, we briefly consider the impact of handedness on our results.

Nanjing and tested at Nanjing Normal University. Participants gave written informed consent and were compensated either 150 RMB (native-Mandarin speakers) or 600 RMB (native-English, non-native-Mandarin speakers) for their time. The difference in compensation amount was due to the bilinguals' experiment session being about twice as long and the greater difficulty in their recruitment.

Table 3. Biographic details for the native-Mandarin participants.

Variable	Mean	Standard deviation	Range
Self-reported Mandarin Proficiency (scale of 1 to 10)	8.6	0.9	6.7 - 10
Self-reported English Proficiency (scale of 1 to 10)	3.4	2.0	1 - 7.3
Self-reported Dialect Proficiency (scale of 1 to 10)	6.8	2.6	0 - 10
English Age of Acquisition (self-reported)	8.8	2.2	5 - 13
Mandarin LexTALE score (out of 50)	45 (90%)	3.9	32 - 50
English LexTALE score (out of 100)	54 (54%)	9.4	40 - 73
Mandarin exposure percent	47.4	29.5	2.7 – 98.7
English exposure percent	5.4	5.1	0 – 17.5
Dialect exposure percent	47.2	30.6	0 – 97
Mandarin usage percent	92.5	10.5	60.3 - 100
English usage percent	7.2	9.6	0 - 35.8

Dialect and other languages usage percent	0.33	1.7	0 - 9.6

Table 4. Biographic details for the native-English, non-native Mandarin participants.

Variable	Mean	Standard deviation	Range
Self-reported Mandarin Proficiency (scale of 1 to 10)	6.4	1.6	1.7 - 9.0
Self-reported English Proficiency (scale of 1 to 10)	9.8	0.3	8.7 - 10
Mandarin Age of Acquisition (self-reported)	16	2.7	12 - 21
Mandarin LexTALE score (out of 50)	19.5 (39%)	8.4	8 - 35
English LexTALE score (out of 100)	93.5 (93.5%)	5.3	81 - 100
Mandarin exposure percent	5.2	4.6	0.4 - 20.2
English exposure percent	94.8	4.6	79.8 - 99.6
Mandarin usage percent	34.7	14.3	5.0 - 68.8
English usage percent	65.3	14.3	31.3 - 95.0

3.2 Materials

For the English experiment, we used the same materials as Steinhauer (2014), which comprised 72 adjective-noun pairs with the form name or pronoun, past-tense verb, “the”, adjective-noun pair, and three-word prepositional phrase, as in “Tom puts the tall vase on the table.” For the Mandarin experiment, we created 60 sentences each with a unique adjective-noun pair. All sentences took the form noun (either a person or an animal), adverb, verb, adjective-noun pair, and the particle LE, as in *父亲最近想看新书了* “Father recently wants to read new books”. All nouns, adverbs, and verbs were two characters. Adverbs were selected such that they licensed the use of the sentence-final, perfective aspect particle LE. This sentence structure further restricted word order such that noun-adjective order would be ungrammatical. Adjectives and nouns were each one character and were chosen such that they formed a plausible pair while attempting to avoid highly lexicalized compounds. Adjectives and nouns were further controlled for frequency (mean adjectives log-frequency = 3.25, SD = 0.78; mean nouns log-frequency = 3.25, SD = 0.74; based on subtitle frequencies from Cai & Brysbaert, 2010).

The 60 Mandarin sentences were arranged into three lists, where in each list half of the adjective-noun pairs were in the grammatical order adjective-noun and the other half in the ungrammatical order noun-adjective. To pseudorandomize our stimuli, we used the program Mix (Van Casteren & Davis, 2006), constraining the randomization such that a given Word Order (adjective-noun, noun-adjective) could repeat a maximum of two times consecutively. Each list was pseudorandomized and arranged in forward and reverse orders. The 72 English sentences were arranged into two lists, and then underwent the same procedure for arrangement as the Mandarin sentences.

3.3 Procedure

All parts of the experiment were approved by the McGill Faculty of Medicine Institutional Review Board following the guidelines of the Canadian Tri-Council Policy Statement and by the School of Foreign Languages and Cultures at Nanjing Normal University (南京师范大学外国语学院). After reviewing and signing the consent form, participants sat in a sound-attenuated booth. All stimuli were presented with Presentation® software (Version 17.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) using the Windows XP operating system.

After reviewing and signing the consent form, participants sat in a sound-attenuated room. For the Mandarin experiment, sentences were presented visually word-by-word, for 650 ms per word and each word followed by a 100-ms blank screen (i.e., stimulus onset asynchrony (SOA) = 750 ms); for the English experiment, the presentation followed the same format, but duration was 300 ms with a 200-ms blank screen (SOA = 500 ms). Each trial began with a cue for the participants to blink ("(--)") for 2000 ms, then a fixation cross ("+") for 500 ms. Each trial ended with a prompt for the acceptability judgment (English experiment: "Acceptable?"; Mandarin experiment: *可否接受句子?* "Can you accept the sentence?"). Participants pressed the "A" key to indicate acceptable or the "L" key to indicate unacceptable. The Mandarin and English experiments each lasted 10 to 20 minutes, but participants also completed other experiments in the same experimental session. Including setup, cleanup, and the length of the other experiments, the total experiment time for the Mandarin native speakers was two to three hours and the total experiment time for the Mandarin non-native speakers was four to five hours.

3.4 EEG Recording and Preprocessing

Participants' EEG was recorded from 32 Ag/AgCl electrodes mounted on an elastic cap according to the international 10-20 system (EASYCAP, Herrsching, Germany). To monitor vertical and horizontal eye movement, one electrode was positioned below the right eye and another to the left of the left eye. Electrode impedance was kept below 10 kΩ. The recordings were amplified online with a bandpass filter of 0.05-100 Hz, referenced online to electrode FCz, and digitized at a sampling rate of 500 Hz.

The signal was then segmented into epochs containing the first and second word of the adjective-noun pair, with 1000 ms past the onset of the second word and a 200-ms baseline prior to the onset of the first word. All epochs were segmented in relation to the trigger time-locked to the presentation of the second word in the adjective-noun pair. For the English experiment, this meant a time window of -700 to 1000 ms around the onset of the second word. For the Mandarin experiment, this time window was -950 to 1000 ms around the onset of the second word.

Artifact rejection was performed across epochs with a moving window threshold of 80 µV (window size = 500 ms), and participants had an average of XX% of epochs rejected (range from YY% to ZZ%). During review of the artifact rejection process, we determined that electrodes Fp1 and Fp2 were exceptionally noisy across participants and excluded them from the final analysis. For select subjects whose automatic rejection resulted in greater than 15% of trials being rejected, epoched data were manually inspected to include additional trials and the overall quality of individual datasets. This review resulted in all 20 English-Mandarin bilingual and 32 Mandarin monolingual subjects being included for final analysis. Single trial data were exported from MATLAB.

3.5 Data Analysis

All data analysis was done using *R* version 4.02 (R Core Team, 2017). To account for variability within items and within participants, we computed mixed effects models using the *glmer* function from package *lme4* version 1.1-23 (Bates, Mächler, Bolker, & Walker, 2015), including the optimizer = 'bobyqa' parameter. Model coefficients were calculated by maximum likelihood estimates using the Laplace approximation. For random effects structures, all factors with possible variability within trials or participants were included in the maximal possible structure.

All *p*-values were calculated with the Satterthwaite approximation calculated based on Wald *Z*-scores in *lmerTest* package version 3.1-2 (Kuznetsova et al., 2017). To construct the maximum possible model and random structure, we used the *buildmer* package version 1.8 (Voeten, 2021) using the direction = 'order' parameter, which adds effects to the model in order of their contribution to log-likelihood. We then again used *buildmer* to do stepwise removal of model variables with the direction = 'backward' parameter to maximize log-likelihood score. These optimized models are reported in the text to maximize power and minimize overfitting (per Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), but the maximal models are reported for reference in the supplementary materials (per Barr, Levy, Scheepers, & Tily, 2013).

Significant interactions were followed up with post-hoc tests for pairwise comparisons using the *emmeans* package version 1.4.8 (Russell, 2019) and Tukey method for adjustment of *p*-values to correct for multiple comparisons. Interactions were visualized with the emmip function from the *emmeans* package. Note that while *p*-values are reported for model results, our inferences and interpretations were not limited to significance testing; instead, we further considered our hypotheses and predictions, effect sizes, and the limitations of data quantity and quality (Loken & Gelman, 2017).

For ease of interpretability, model results are reported with graphical depictions of coefficients and confidence intervals generated by the plot_model function from the *sjPlot* package version 2.8.9 (Lüdecke, 2021); full model output, including random effects, are reported in the supplementary materials in tables generated from the tab_model function from the *sjPlot* package. For simplicity, model results reported in the text are limited to significant effects or effects that were related to initial predictions. Additionally, simple effects of the factors Electrode, Anteriority, and Laterality, or interactions involving only these factors, are not reported or discussed in the text because they are not related to the experimental manipulations. Nonetheless, all model outputs can be consulted in the supplementary materials.

3.5a ERPs

For the statistical models for both the English and Mandarin experiments, the maximal models at midline electrodes included the factors Word Order and Electrode. For the models at lateral electrodes, the factors Anteriority (frontal, central, posterior) and Laterality (left, right) replaced the factor Electrode. For the models for the Mandarin experiment, the additional factor of Group (native, non-native) was added. The factors Word Order, Electrode, and Anteriority were all treatment coded, such that the respective reference levels were adjective-noun order, Pz, and posterior. All other factors were sum coded. All mixed effects models and all ERP plots were based on single trial data.

ERPs were analyzed in the N400 time window for the first word (300-500 ms post-onset) and in the N400 and P600 time windows (300-500 ms and 500-1000 ms, respectively) for the second word. The time windows for ERPs time-locked to the second word were based on a priori predictions from Steinhauer (2014). Analysis of the N400 time window for the first word was motivated by visual observation of a strong effect of Word Order and similar observations by Steinhauer (2014). Note that for the English results only, the reported time window for analysis of the P600 was 500-700 ms because the grand average ERPs only showed an effect in this interval. The model results for additional time windows are reported in the supplementary results.

Because the components of interest in the present study (N400, P600) are typically maximal near the midline (Kaan, 2007), we primarily report in the text only results from models at the midline; results from lateral electrodes are reported in the text if they show effects beyond the models at midline electrodes. Note that models for lateral electrodes excluding the midline are reported in the supplementary materials.

All final analyses were performed on single trial average amplitudes, and plotted figures represent grand averages calculated from subject averages of single trials. All figures showing ERP voltage against time and scalp maps reflect these average ERPs and were plotted using the R package *ERPscope* (Herbay, 2022).

3.5b Individual Differences

Analysis of individual differences in ERP amplitudes was carried out based on 1) the original descriptions of the RDI and RMI measures (Tanner et al., 2014) and 2) modifications suggested by Fromont and colleagues (Fromont et al., 2020). We conducted this analysis for both the Mandarin and the English results. This analysis included 1) correlations of N400 and P600 amplitudes, 2) calculation of RDI and RMI measures, the correlation between them, and, for the non-native speakers only, the correlations of these measures with experiment accuracy and Mandarin proficiency.

For the Mandarin experiment, we used the adjacent time windows of 300-500 ms for the N400 effect and 500-1000 ms for the P600 effect. To consider non-adjacent time windows (Fromont et al., 2020) where the effects were still maximal, we looked at average amplitude from 300-400 ms for the N400 effect and 750-1000 ms for the P600 effect. For the English experiment, we observed that the P600 effect at the grand-average level was limited to approximately 500-700 ms. With this consideration, we used the adjacent time windows of 300-500 ms for the N400 effect and 500-700 ms for the P600 effect. For comparison of non-adjacent time windows, we used 300-400 ms for the N400 effect but kept the same 500-700 ms window for the P600 effect.

For the Mandarin results, we calculated these measures for native and non-native Mandarin participants both separately and together as a larger group. Because predictions for non-native processing are different from those for native speakers (Tanner et al., 2014, 2013), we chose to report the analyses separately for native and non-native participants. The additional analyses not reported in the text are available in the supplementary materials.

4 Results

4.1 Behavioral Results

Native English and native Mandarin speakers showed high accuracy for their native language in the acceptability judgment task, as shown in Table 5 and Table 6. One native Mandarin participant showed low accuracy, but their d-prime score showed their performance was still above chance (d-prime = 0.1). We compared analysis results including and excluding their results and determined that there was not a major impact, so they were included in the final analysis.

For non-native Mandarin speakers, accuracy results showed more variability, as seen in Table 6. To further evaluate non-native performance, we calculated d-prime scores using the dprime function in the R package *psycho* (Makowski, 2018). The d-prime score results showed that all non-native participants performed above chance on the acceptability judgment task in Mandarin (mean = 0.505, sd = 0.22, range = 0.1 – 0.93).

Table 5. Accuracy by condition for English experiment. Values represent data from 17 participants.

Condition	Mean % accuracy (standard deviation)	Range of accuracy

Adjective Noun	98 (3)	92 – 100
Noun Adjective	97 (4)	86 – 100

Table 6. Accuracy by condition for Mandarin experiment. Values represent data from 17 participants.

Group	Condition	Mean % accuracy (standard deviation)	Range of accuracy
Native	Adjective Noun	94 (4)	77 – 100
	Noun Adjective	92 (5)	33 – 100
Non-native	Adjective Noun	83 (8)	65 – 97
	Noun Adjective	69 (10)	40 – 97

4.2 ERP Results

4.2a English

ERPs for the English experiment showed a similar pattern to that reported by Steinhauer (2014), with a posterior N400 followed by a P600 for the ungrammatical condition time-locked to the second word, as seen in Figure . Closer examination showed a small negativity distributed across the scalp prior to the second word onset, but this negativity was smaller in magnitude than in the N400 time window for the second word. In contrast to Steinhauer (2014), the P600 effect was smaller and spread over a smaller time interval, showing maximal amplitude from approximately 500 to 700 ms after word two onset.

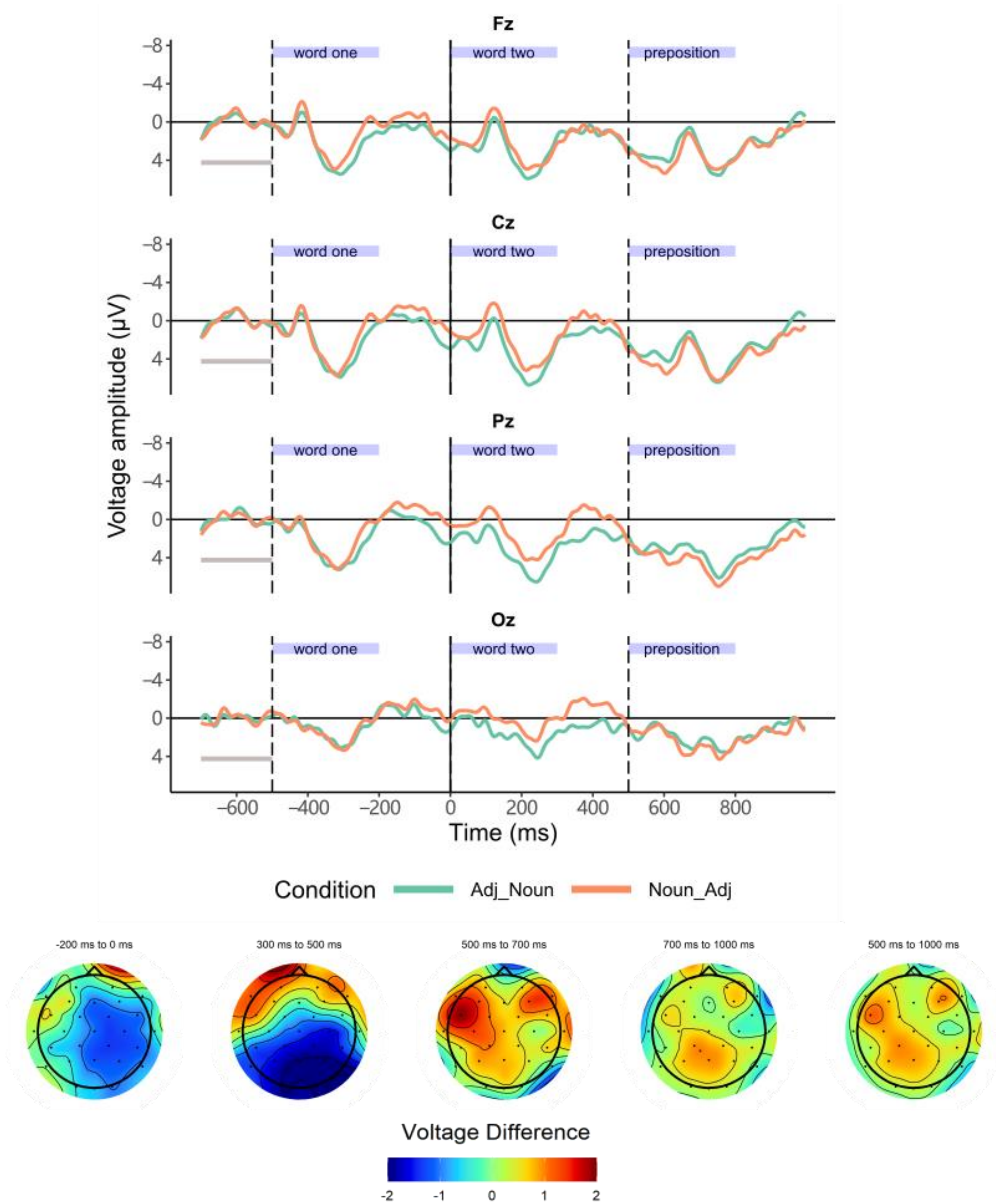

Figure 1. *ERPs for English experiment. Scalp maps show ungrammatical minus grammatical for the indicated time windows. The 200-ms pre-onset baseline interval is indicated with a gray rectangle.*

4.2a.i Word One N400

The statistical model for the N400 amplitude at the first word (corresponding to -200 to 0 ms in Figure) showed no significant effect of Word Order in posterior electrodes, which suggests that voltage was not impacted by our experimental conditions in this time window. The model at lateral electrodes showed a two-way interaction between Word Order and Anteriority, such that the difference between adjective-noun and noun-adjective orders was smaller at frontal electrodes than at posterior electrodes. However, post-hoc tests showed that none of the pairwise comparisons was significant, so we refrain from interpreting these results further.

4.2a.ii Word Two N400

The statistical model at midline electrodes for the N400 amplitude at the second word (corresponding to 300 to 500 ms in Figure), yielded the coefficients summarized in Figure 2. The intercept value ($\beta = 1.89$ µV, SE = 0.65, $Z = 2.89$, $p = 0.004$) reflects the predicted voltage for adjective-noun order at electrode Pz. There was then a simple effect of Word Order, indicating that noun-adjective order sentences had a larger N400 amplitude than adjective-noun order sentences at Pz ($\beta = -1.89$ µV, SE = 0.87, $Z = -2.17$, $p = 0.030$). This result is in line with our predictions for a greater N400 response to the ungrammatical Word Order condition.

There was additionally a two-way interaction between Word Order and Electrode at Fz. Post-hoc pairwise comparisons showed that while the electrode Pz showed a significant effect of Word Order on voltage, as indicated by the coefficient for the simple effect of Word Order, the electrode Fz did not show this effect ($\beta = -0.06$ µV, SE = 0.87, $Z = -0.067$, $p = 0.943$). This interaction indicates the N400 effect was greater at posterior electrodes.

The model at lateral electrodes showed the same pattern of effects, but this time with two-way interactions between Word Order and Electrode for both anterior and central electrodes. Post hoc pairwise comparisons showed that the N400 effect was significant only at posterior electrodes ($\beta = -1.49$ µV, SE = 0.62, $Z = -2.41$, $p = 0.02$) and not at central or frontal (central: $\beta = -0.081$ µV, SE = 0.62, $Z = 0.13$, $p = 0.90$; frontal: $\beta = 0.95$ µV, SE = 0.62, $Z = -1.53$, $p = 0.13$). In fact, at frontal electrodes, the voltage pattern was in the opposite direction, consistent with a frontal positivity.

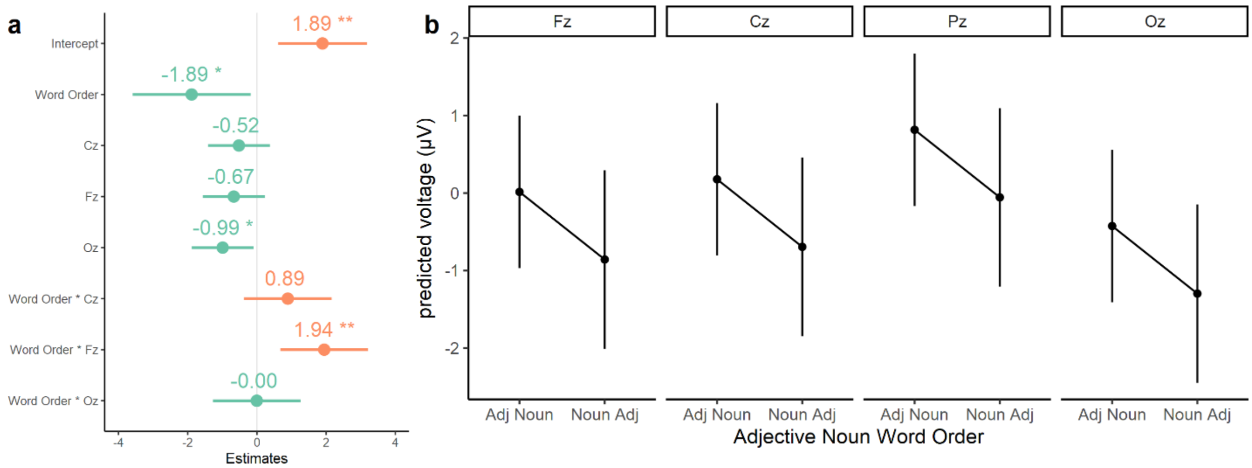

__Figure 2__. a) Model coefficients and b) predictions for N400 effect at Word Two in English experiment.

4.2a.iii Word Two P600

As visible in Figure 1, P600 amplitude was maximal in the 500 to 700 ms time window. Here we report the model results for average amplitude in this time range, with results for additional time windows reported in the supplementary results. At midline electrodes, the P600 amplitude was not significantly affected by Word Order, despite a predicted numerical difference ($\beta = 0.82$ µV, SE = 0.74, $Z = 1.11$, $p = 0.27$). The statistical model for lateral electrodes, however, showed a significant interaction between Word Order and Anteriority at frontal electrodes ($\beta = 0.81$ µV, SE = 0.38, $Z = 2.16$, $p = 0.03$). Post hoc pairwise comparisons showed that at frontal electrodes, noun-adjective order elicited greater positivity in this time window than adjective-noun order ($\beta = 1.27$ µV, SE = 0.59, $Z = 2.16$, $p = 0.03$). These model predictions are reported in Figure 3. Interpreted together with the scalp distribution of the ERPs in Figure , these model results confirm a frontal positivity effect.

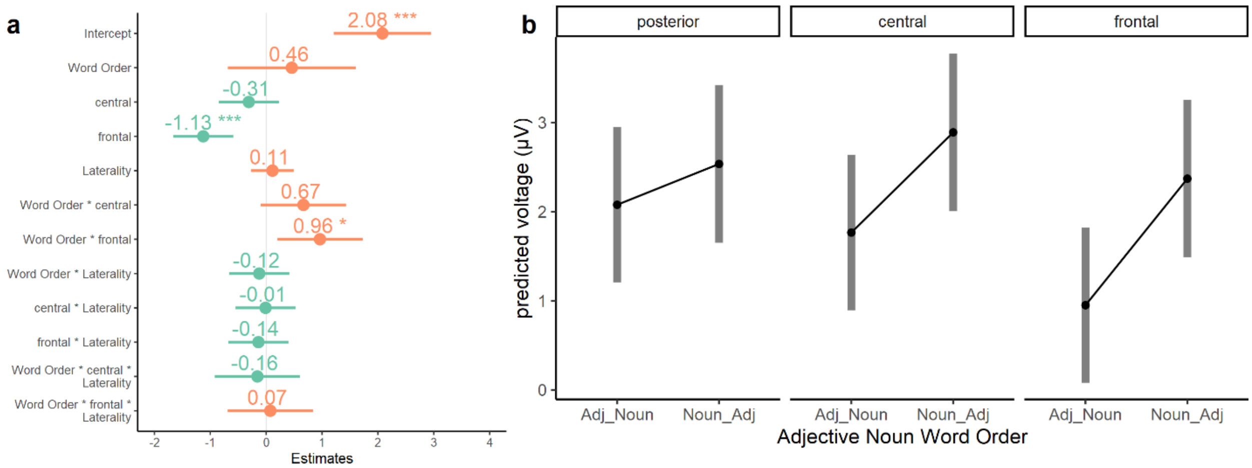

Figure 3. *a) Model coefficients and b) predictions for P600 effect at Word Two in English experiment.*

4.2b Mandarin

ERP results for the Mandarin experiment were analyzed with native and non-native data combined, with the addition of Group as a predicting factor. To appreciate the differences between the native and non-native results, Figure 4 depicts the separate ERP results by Group. For the native speakers, the Mandarin ERP results at word two are similar to the those reported by Steinhauer (2014). However, we observed an unexpected result at word one, where the grammatical condition showed a greater N400 amplitude than the ungrammatical condition. Steinhauer (2014) reported a similar finding for French native speakers processing English adjectives whose French translations typically follow the nouns they modify. This condition does not apply to Mandarin and English, so this effect at word one cannot be explained by the same reasoning. In the discussion, we elaborate on the possible lexical status of Mandarin adjective-noun pairs as a potential explanation. For the non-native speakers, the grand average also showed an N400 effect at the second word, but no subsequent P600.

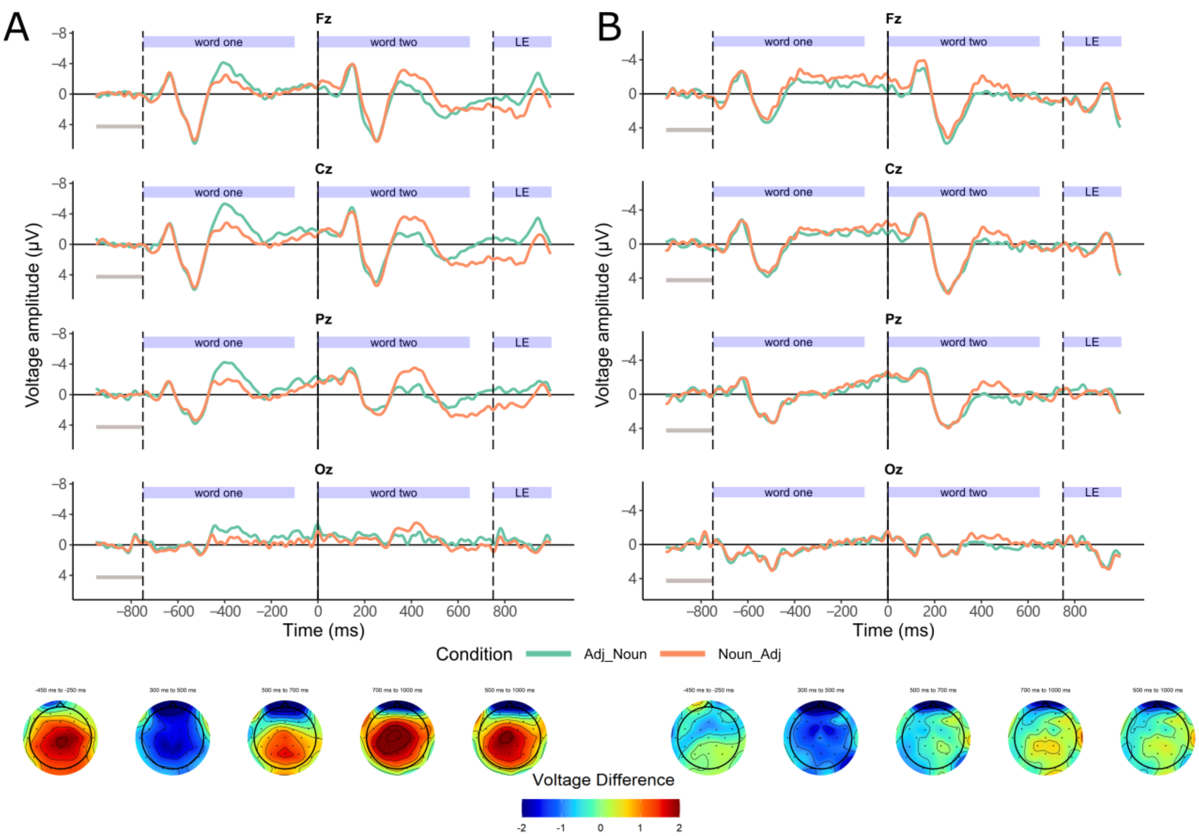

Figure 4. ERPs and topoplots for the Mandarin experiment, for native (a) and non-native (b) participants.

4.2b.i Word One N400

Coefficients and predictions from the statistical model are shown in Figure 5. The model intercept value represents the predicted voltage for adjective-noun word order at electrode Pz, averaged across native and non-native participants (β = -1.15 µV, SE = 0.29, *Z* = -3.92, *p* < 0.001). There were also simple effects of Group and Word Order. The coefficient for Group shows that at Pz, non-native participants had an overall more positive N400 amplitude than native participants (β = -1.50 µV, SE = 0.31, *Z* = -4.91, *p* < 0.001). This likely reflects that non-native processing often shows a reduced, broadly distributed N400 amplitude with longer latency, often linked to delays in lexical retrieval (Hahne, 2001; Mueller, 2005; Weber-Fox & Neville, 1996). The coefficient for Word Order shows that the adjective in grammatical adjective-noun order elicited a greater N400 effect than the noun in ungrammatical noun-adjective order, at electrode Pz averaged across Group (β = 0.58 µV, SE = 0.22, *Z* = 2.68, *p* = 0.007).

The interaction terms clarify the effect of Word Order. Group and Word Order had a significant interaction (β = 0.89 µV, SE = 0.22, *Z* = 4.12, *p* < 0.001), and post hoc pairwise comparisons showed that the effect of Word Order was significant for the native speakers, but not for the non-native speakers (native: β = -1.47 μV, SE = 0.27, Z = -5.50, $p < 0.001$; non-native: β = 0.312 μV, SE = 0.338, Z = 0.923, p = 0.8). This interaction demonstrates that only the native speakers exhibited an N400 difference in this time window.

Group showed an additional interaction with Electrode (β = 0.93 μV, SE = 0.20, Z = 4.59, $p <$ 0.001). Post hoc comparisons showed that Fz was more negative than Pz for non-native speakers, but this difference between electrodes was not significant for native speakers (native: β = -0.10, μV, SE = 0.25, Z = -0.40, p = 1; non-native: β = 1.76 μV, SE = 0.32, Z = 5.53, $p < 0.0001$). This difference again reflects that in comparison to their native counterparts, non-native participants showed an overall reduced N400 amplitude for the first word across both levels of Word Order. We interpret this result as stemming from variability in N400 onset and peak latency in non-native speakers.

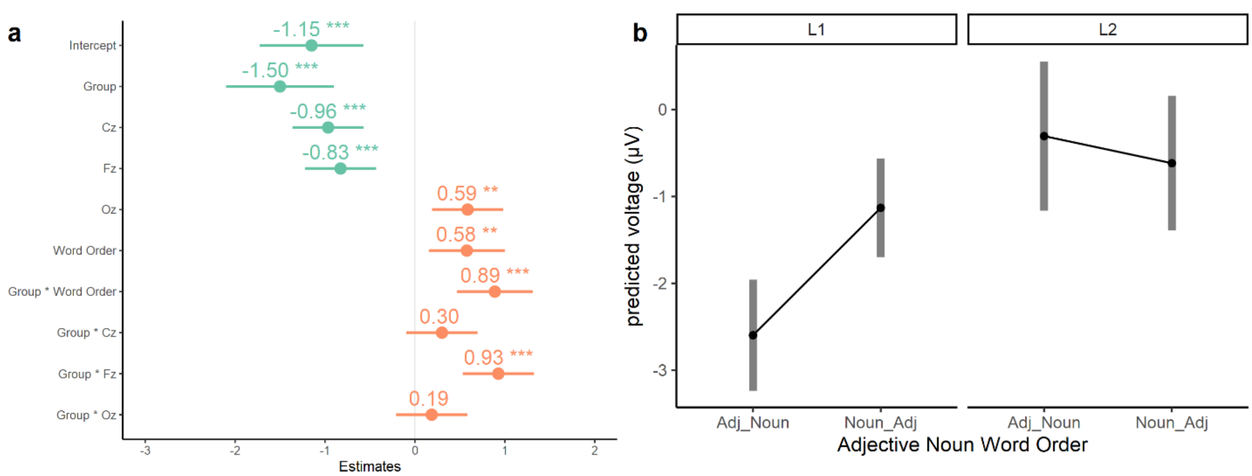

Figure 5. *a) Model coefficients and b) predictions for N400 effect at Word One in Mandarin experiment.*

4.2b.ii Word Two N400

The optimized model of N400 amplitude at word two resulted in only two simple effects, for Word Order and Group, as shown in Figure 6. The coefficient for Word Order shows that the ungrammatical noun-adjective word order elicited a greater magnitude N400 amplitude compared to the grammatical adjective-noun word order (β = -1.33 μV, SE = 0.29, Z = -4.59, $p <$ 0.001). This coefficient represents the effect of Word Order across the average of Group. Given that interactions did not improve log-likelihood score and were not included in the optimized model, the Word Order coefficient demonstrates that both native and non-native participants' N400 amplitude was modified by Word Order. Just as for the model for N400 amplitude at word one, the second simple effect of Group shows that non-native speakers had overall reduced N400 amplitude in comparison to native speakers, consistent with the N400 results at word one (β = -0.65 µV, SE = 0.29, *Z* = -2.22, *p* = 0.03).

\

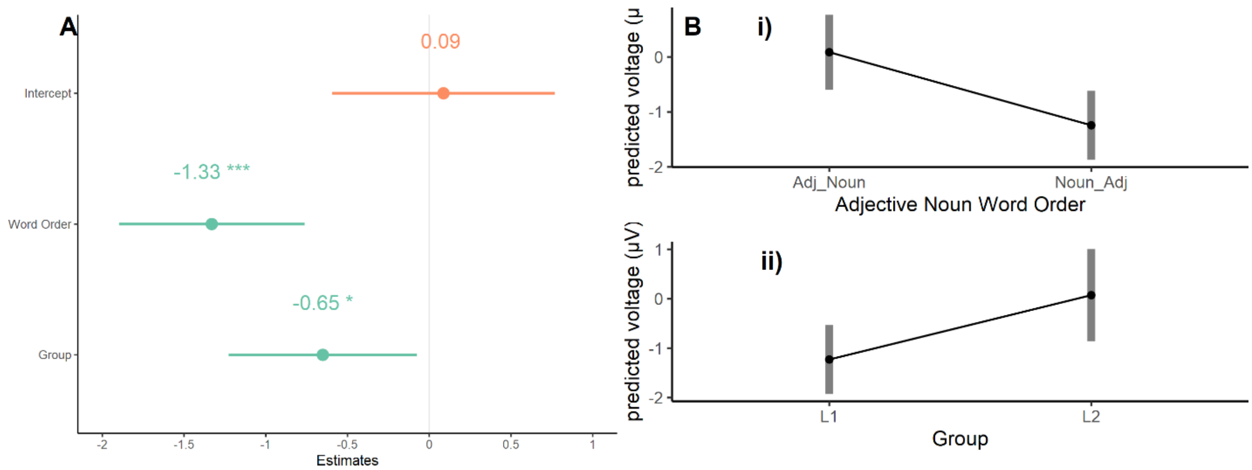

Figure 6. *A) Model coefficients and B) predictions for N400 effect at Word Two in Mandarin experiment.*

4.2b.iii Word Two P600

Model coefficients and predictions for the P600 amplitude at word two are shown in Figure 7. Similar to the model for N400 amplitude, there was a main effect of Word Order, where (ungrammatical) noun-adjective order elicited a greater positivity than (grammatical) adjective-noun order (β = 0.89 µV, SE = 0.34, *Z* = 2.61, *p* = 0.009). However, this was accompanied by an interaction between Group and Word Order (β = 0.62 µV, SE = 0.26, *Z* = 2.40, *p* = 0.02). Post hoc comparisons showed that the effect of Word Order on P600 amplitude was only significant for the native participants, with the non-native participants not showing any effect of Word Order on average amplitude (native: β = 1.23 µV, SE = 0.32, *Z* = 3.84, *p* = 0.0001; non-native: β = 0.01 µV, SE = 0.32, *Z* = 0.02, *p* = 0.98).

The model at lateral electrodes showed three additional interactions. The first was between Anteriority and Word Order (β = -0.58 µV, SE = 0.23, *Z* = -2.49, *p* = 0.01), and the second was between Anteriority and Group (β = 0.50 µV, SE = 0.16, *Z* = 3.01, *p* = 0.003). Post hoc comparisons did not show significant pairwise differences as part of this interaction. However, the third interaction, among Group, Word Order, and Anteriority, did show significant differences on follow up with post hoc pairwise comparisons (β = -0.49 µV, SE = 0.23, *Z* = -2.08, *p* = 0.04). These comparisons confirmed that the significant P600 effect for the L1 group was limited to central and posterior electrodes (posterior: β = 0.93 μV, SE = 0.41, $Z = 2.25$, $p = 0.02$; central: β = 0.91 μV, SE = 0.41, $Z = 2.21$, $p = 0.03$), and that the effect was not significant at any region for the L2 group.

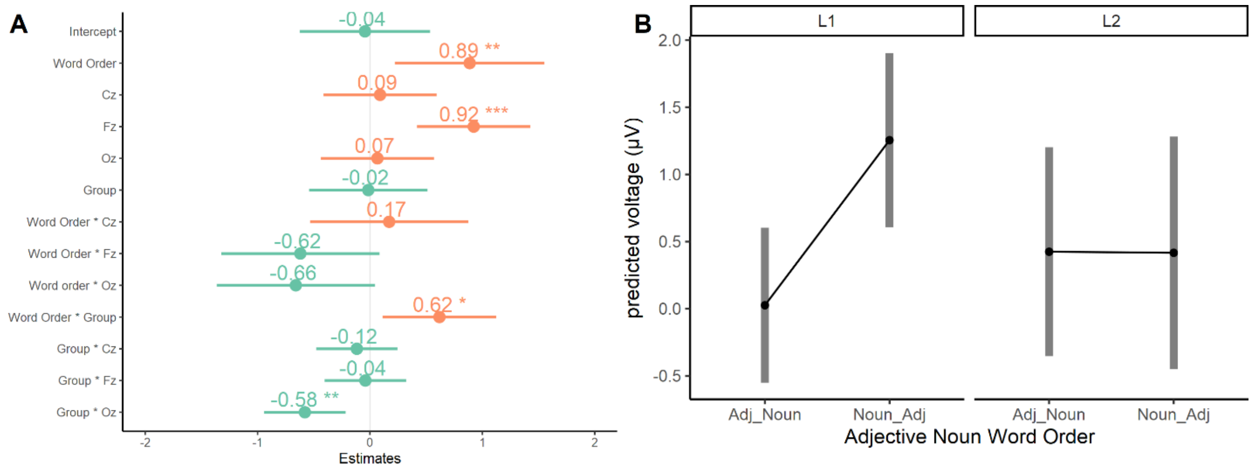

Figure 7. *Model coefficients and predictions for P600 effect at Word Two in Mandarin experiment.*

4.3 Individual Differences

4.3a Correlation between N400 and P600

Per Tanner & Van Hell (2014), we correlated N400 and P600 amplitudes at the same electrodes and with adjacent time windows for both the English and the Mandarin experiments. Per Fromont and colleagues (2020), we next applied more stringent measures by correlating data at different electrodes and non-adjacent time windows. To appreciate whether the correlation was more affected by comparing adjacent time windows or by comparing data on the same electrode, we also applied each of these two measures on their own. As detailed below, these more stringent measures resulted in several of the N400-P600 effect correlations disappearing or becoming reduced in significance and effect size.

The results for English are shown in Figure 8. Figure 8a shows that the correlation with adjacent time windows and on the single electrode Pz was highly significant and explained more than half of the data variability. The correlation remained significant even for non-adjacent time windows (300-400 ms for the N400 difference and 500-750 ms for the P600 difference) and different electrodes, as seen in Figure 8b. The follow-up correlations for non-adjacent time windows (Figure 8c) and non-identical electrodes (Figure 8d) suggest that the relationship in N400 and P600 amplitudes is more driven by overlap in time than overlap in space.

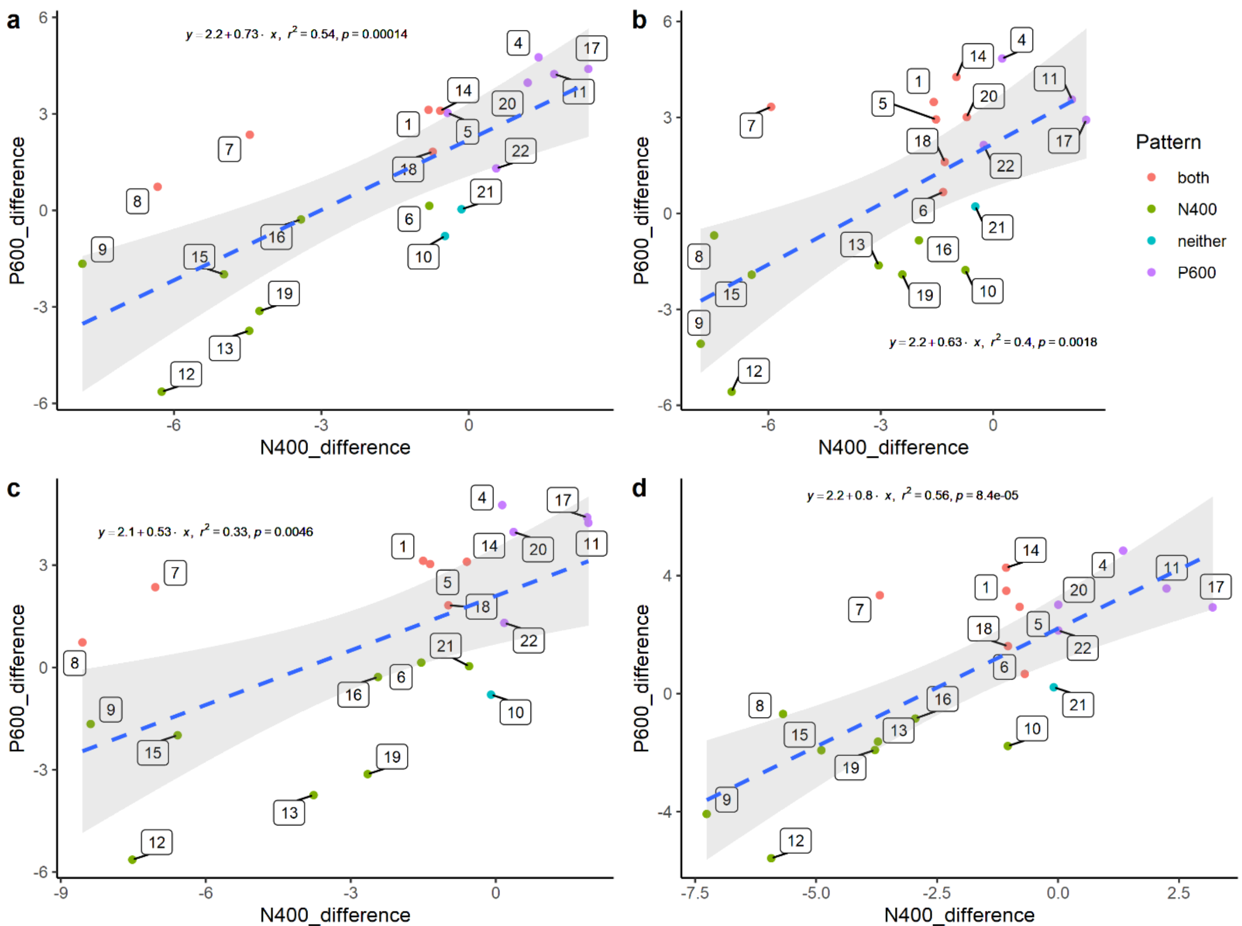

__Figure 8.__ Correlations between P600 and N400 effects for English. Clockwise from top left: a) electrode Pz with adjacent N400-P600 time windows; b) split electrode groups (C3, Cz, C4 for P600 and P3, Pz, P4 for N400) and non-adjacent time windows; c) electrode Pz with non-adjacent time windows; d) split electrode groups (C3, Cz, C4 for P600 and P3, Pz, P4 for N400) and adjacent time windows. R^2 values represent adjusted- R^2.

The results for Mandarin are shown in Figure 9, where there was a significant correlation with adjacent time windows on the single electrode Pz. However, when comparing the data for non-adjacent time windows and non-identical electrodes, the correlation was no longer significant. With these restrictions on the data, the line of best fit captured almost none of the variability. Figure 9c shows the correlation at Pz for non-adjacent time windows (300-400 ms for the N400 difference and 750-1000 ms for the P600 difference). While this correlation was significant, the R^2 value was less than a third of the value for the original comparison at Pz with adjacent time windows. Figure 9d shows the correlation at non-identical electrodes (C3, Cz, C4 for the P600 difference and P3, Pz, P4 for the N400 difference). This correlation was significant and with a moderate value for R^2. From these results, it seems that a significant correlation between the N400 and P600 effect magnitudes depends on either comparing across identical electrodes or comparing adjacent time windows. While both of these follow-up correlations were significant, it appears that comparing adjacent time windows was the more important factor for achieving a significant correlation. Non-native participants' Mandarin results broadly showed the same pattern as for native participants (see supplementary results).

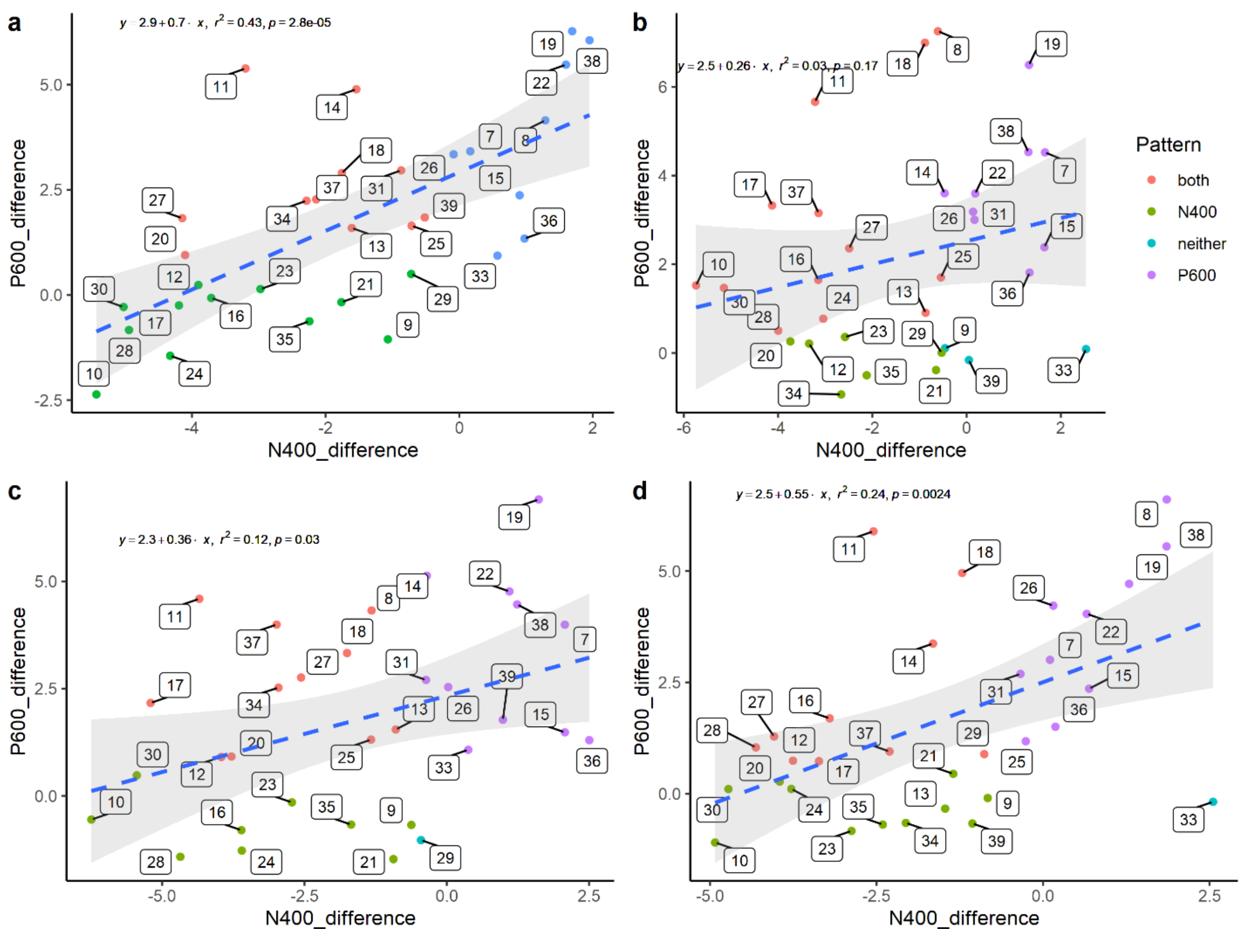

Figure 9. *Correlations between P600 and N400 effects for native Mandarin participants. Clockwise from top left: a) electrode Pz with adjacent N400-P600 time windows; b) split electrode groups (C3, Cz, C4 for P600 and P3, Pz, P4 for N400) and non-adjacent time windows; c) electrode Pz with non-adjacent time windows; d) split electrode groups (C3, Cz, C4 for P600 and P3, Pz, P4 for N400) and adjacent time windows.* R^2 *values represent adjusted-* R^2.

4.3b RDI and RMI calculations

RDI and RMI values were calculated as per Tanner et al. (2014) and as per Fromont et al.'s (2020) recommendations. Although correlations between N400 and P600 amplitudes depended on our method for selecting the data, we found that even the most stringent approach of using non-adjacent time windows and distinct electrodes still resulted in a range of participant response patterns, from N400-dominant to balanced to P600-dominant.

Results for the English experiment are shown in Figure 10. Note that the P600 time window used for the RDI and RMI calculations was 500 to 700 ms because this was the only interval where the P600 effect was visible in the grand average. While the N400 and P600 effects were inversely correlated in this time window, the RDI values showed that nine participants were P600-dominant, eight participants were N400-dominant, and three participants showed approximately equal N400 and P600 effect sizes. As revealed by the RMI values, the three balanced participants' ERPs showed the smallest effect of Word Order. Thus, while there appears to be a continuum from P600- to N400-dominant response profiles according to the RDI values, participants with RDI values closer to zero tended to show smaller effects overall.

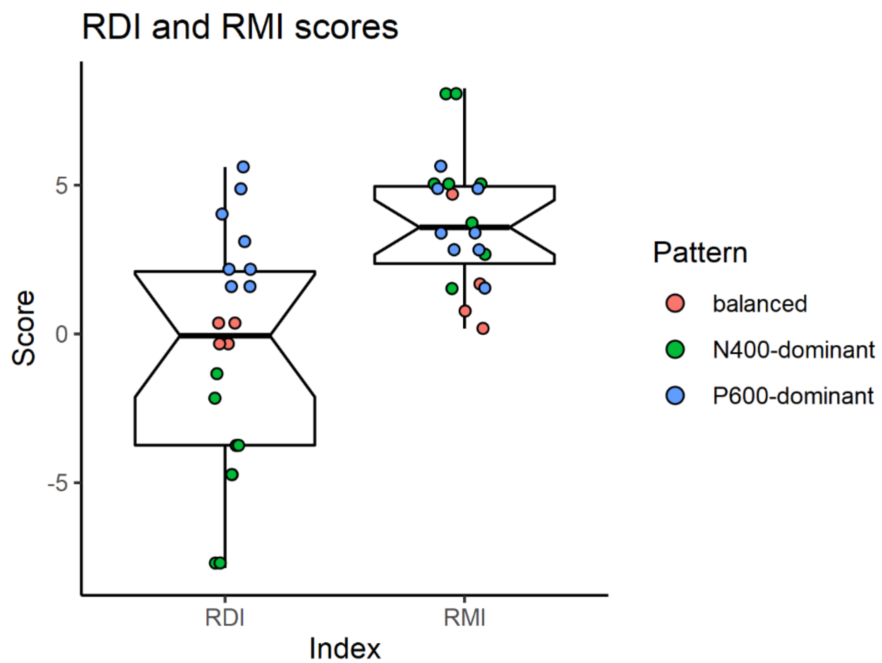

Figure 10. RDI and RMI values for the English experiment. Color represents dominance pattern according to RDI, with -1 to 1 considered to be 'balanced'.

Results for the Mandarin experiment are shown in Figure 11. For the native speakers, the RDI and RMI patterns were similar to those for the English experiment. RDI profiles show a continuum of response patterns, with thirteen P600-dominant participants, ten N400-dominant, and nine balanced. While the balanced participants tended to show the lowest RMI values, there were two balanced participants who showed a response magnitude closer to the group average.

For the non-native speakers, the pattern of RDI and RMI values was again similar, with seven 600-dominant participants, 5 N400-dominant participants, and 8 balanced participants. The RMI values again showed that while most balanced participants had the smallest magnitude effects, there were 4 balanced participants whose response magnitudes were closer to the average value. In comparison to the native participants, non-native RMI values tended to be lower, indicating an overall decreased response sensitivity to the Word Order manipulation. Per Tanner et al. (2014), we further computed correlations between the individuals' RDI and RMI values and their language biographic information, including self-reported proficiency, months of Mandarin immersion, months of Mandarin study, performance on a Mandarin LEXTALE, and their accuracy on the experimental task. None of these correlations with RDI or RMI was significant, and no correlations between biographic information and ERP amplitudes were significant after correction for multiple comparisons.

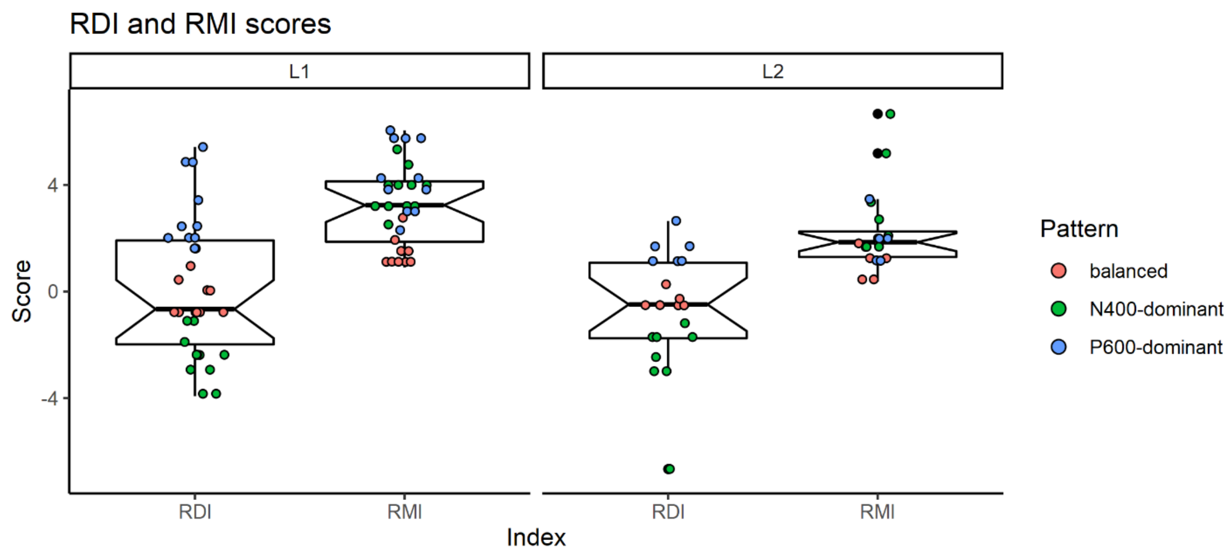

***Figure 11**. RDI and RMI values for the Mandarin experiment, showing values for native participants (L1) and values for non-native participants (L2). Color represents dominance pattern according to RDI, with -1 to 1 considered to be 'balanced'.*

5 Discussion

In this study, we have used ERPs to better understand the processing time course of adjective-noun placement violations in Mandarin for both native and non-native speakers. We also replicated a similar ERP study in English (Steinhauer, 2014) to directly compare ERPs from corresponding constructions in both languages within the same participants. For the Mandarin experiment, we found that native speakers showed a biphasic N400-P600 response to adjective-noun order violations at the second word, alongside a greater N400 at the first word for the correct condition. Non-native speakers showed only an N400 effect in the predicted time window, without a subsequent P600. We further showed that these group-level effects were the result of qualitative individual differences in ERP responses, even when accounting for artifacts from adjacent time windows or identical electrodes.

At the position of the second word, our results show a similar biphasic N400-P600 pattern between Mandarin and English. This shared pattern suggests that adjective-noun placement violations are detected via similar mechanisms in both languages. However, the presence of an N400 effect at word one in the Mandarin experiment diverges from the English results, which indicates that Mandarin adjectives are in some ways different from their English counterparts. Additionally, although non-native Mandarin speakers exhibited an N400 effect to adjective misplacement, they did not show the native-like P600 pattern, suggesting they did not engage in a rule-based reanalysis of the adjective-noun pairs. This apparent failure of high-proficiency non-native speakers to show a native-like processing pattern may be related to the grammatical differences between Mandarin and English adjectives.

While the above interpretations stem from the group-level results, individual participants' data show qualitatively diverse patterns. Across all participants, there was a range of responses from N400-dominant to P600-dominant, as proposed by Tanner & Van Hell (2014). Despite this variability, at least a third of our native participants showed a biphasic N400-P600 response. When we followed Fromont and colleagues' (2020) suggestions for comparing non-adjacent time windows and non-identical electrodes, the N400 and P600 effects were not always correlated, which goes against Tanner & Van Hell's (2014) findings. Additionally, the RDI and RMI scores were not correlated with the non-native speakers' proficiency or language experience, in contrast to previous reports (Tanner et al., 2013). Below, we expound on the group-level and individual-level interpretations in more detail.

5.1 English experiment

In line with Steinhauer (2014), our results for the English experiment showed a biphasic N400-P600 response at the word order violation. Notably, the ERP visualization in Figure suggests that the N400 negativity may have begun prior to the presentation of the second word and that the P600 effect was limited to an early time window of approximately 500-700 ms. Although the negativity was not significant in the time window of the N400 for the first word, the ERP plots suggest that this early effect is a sustained negativity, different from the typical N400 effect which occurs in a shorter time window (Kaan, 2007).

These results may indicate an effect of participant strategy. Participants completed the English experiment right after the Mandarin experiment and conceivably relied on strategies they developed from their first exposure to the adjective-noun placement manipulation. Difference waves (supplementary materials) show that the magnitude of the negativity for violations increases after the onset of the second word and is most prominent in the second word N400 interval (300-500 ms). The preceding sustained negativity may be related to participants' adopting an explicit strategy of identifying the violation condition at the first word, in line with suggested links between sustained posterior negativities and task effects (Jacobsen, Klein, & Löw, 2018). Such a strategy may also have reduced the reanalysis or syntactic repair computations that participants carried out, which could explain the short duration of the P600 effect.

5.2 Mandarin experiment

Statistical model results show that both native and non-native participants showed a greater N400 amplitude in response to the second word of the ungrammatical noun-adjective condition. Native speakers additionally showed a subsequent P600 effect and an N400 effect on the first word for the correct adjective-noun condition.

5.2a Native speakers

The native speakers' biphasic N400-P600 pattern was consistent with that previously reported for English (Steinhauer, 2014). Alongside this predicted effect, participants showed an unexpected N400 effect at the first word. A first interpretation could be that lexical differences between adjectives and nouns, like frequency, could be driving the effect. After all, the data show that the adjective elicited a greater N400 than the noun in both the grammatical and ungrammatical Word Order conditions. However, frequency data (as detailed in the Methods section) show that frequency could not be the cause of this effect; if anything, nouns were numerically (although not significantly) less frequent than adjectives. It is also unlikely that adjectives or nouns would inherently elicit different N400 amplitudes, as both word classes require meaning retrieval. Adjectives may generally be more abstract than nouns, but abstract words have been shown to elicit smaller N400 amplitudes than concrete words (P. J. Holcomb, Kounios, Anderson, & West, 1999), which goes in the opposite direction of the present findings. Alternatively, adjectives may represent a less frequent or less predictable continuation of our sentences, which could be in line with cloze probability driving the N400 effect (W. Chow et al., 2016). However, there was no such N400 effect for adjectives in the correct condition seen in the English experiment, either in Steinhauer (2014) or our replication.

A second possibility is that there is an inherent difference between the adjective-noun structures used between the English and Mandarin experiment. As detailed in the introduction, adjective-noun pairs without the particle DE represent a minority of cases in typical language use. Some accounts further suggest that these adjective-noun structures without DE are actually single words (Z. Xu, 2018). With this consideration in mind, we must interpret differently participants' incremental parsing. Because adjective-noun pairs were presented one character at a time (i.e., a separate screen for the first and second word with an SOA of 750 ms), participants first saw one of the characters and had time to evaluate their sentence parse up to that moment. For the ungrammatical noun-adjective order, participants first read a noun that could plausibly complete a grammatical sentence, such as *父亲最近想看书*... "Father recently wants to read books…". Only on encountering the adjective in the second word position would the sentence become ungrammatical. For the grammatical adjective-noun order, however, participants first read an adjective that would be an ungrammatical completion for the sentence, such as *父亲最近想看*新*... "Father recently wants to read *new…". At this point, the sentence would be ungrammatical until encountering the noun in the second word position.

This parsing sequence is ostensibly the case for both Mandarin and English, but the difference in ERP patterns may be due to structural differences between the two languages. We suggest that unlike English, where adjective-noun pairs are phrases composed of two separate words, Mandarin adjective-noun pairs without the particle DE show features of a lexicalized single word. On encountering a single-morpheme (per the monomorphemic constraint, as in Xu, 2018) adjective without a noun, participants would have lexical retrieval difficulty, as evidenced by an increased N400 amplitude. There may be additional insight from data on pseudoword processing. Pseudowords, such as 'plab' in English, have been shown to elicit greater N400 amplitudes (e.g., McLaughlin, Osterhout, & Kim, 2004). This effect has also been shown for Chinese words when characters are combined in nonexistent pairs (Gao et al., 2022; Q. Wang & Yuan, 2008). Accordingly, when monomorphemic adjectives are presented without DE, their processing response may be akin to that of pseudowords.

5.2b Non-native speakers

Like their native counterparts, Mandarin non-native participants showed an N400 effect time-locked to the presentation of the second word in the adjective-noun pair violation. The statistical model for the N400 effect, as shown in Figure 6, showed that native and non-native speakers had a similar magnitude effect, but non-native speakers had overall decreased N400 amplitude.

Considering the ERP plots in Figure 4, the N400 components showed differences between native and non-native participants for both the first and second words. Non-native speakers' first and second word N400 component amplitude was reduced and did not show a clear peak, suggesting that the N400 had variable latency among participants. A similarly reduced N400 component has been reported for non-native processing of semantic violations (Newman, Tremblay, Nichols, Neville, & Ullman, 2012). In the ERP figures, it appears that the N400 effect on the second word actually began prior to the analyzed N400 time window, at least on frontal and central electrodes. The scalp maps in Figure 4 show there was a relatively small difference in the earlier time window, and the statistical model results show that this negativity was not significant in the time window for the N400 of the first word.

At the group level, the lack of a P600 effect for the non-native participants suggests that they represent a lower Mandarin proficiency level, according to the predicted pattern outline in the introduction (Steinhauer et al., 2009). The individual differences results, however, reveal that several participants indeed showed a P600 effect, as can be seen in Figure 11, but there were no significant correlations between ERP effect amplitude and proficiency measures. This was also the case for the RDI and RMI scores.

However, the Mandarin adjective-noun structure, in particular, may not follow a typical pattern for second language acquisition. In the early stages of classroom instruction, Mandarin students learn that adjectives can show properties of stative verbs without an intervening copula, such as *我很开心* "I am happy". When learning about the DE structure with adjectives, learners may see structures such as *开心的朋友* "happy DE friend". At more advanced language levels, however, learners will encounter both prenominal and postnominal adjective use, such as *我喜欢开心的朋友* "I like happy friends" and *我喜欢朋友开心* "I like (my) friends (to be) happy". The alternative adjective position reflects subtle differences in meaning. At lower proficiency, learners may process postnominal adjectives of this type as violations. At advanced proficiency levels, learners may know that both pre- and postnominal adjective placements are possibilities but be unclear as to the appropriate circumstances in which to use them. Only at much higher proficiencies would learners process such phrases in a native-like way. As such, our relatively high-proficiency, non-native Mandarin participants may be in this intermediate stage, which would be consistent with the lack of correlations between proficiency measures and the ERP effects.

A final consideration for the non-native speakers is that of handedness. Because high proficiency Mandarin second language speakers are challenging to recruit, we accepted two participants who were left-handed, despite the common approach to recruit only right-handed participants for language and psychology experiments (Bailey, McMillan, & Newman, 2020). Handedness has been shown to relate to lateralization of the primary language networks in the brain, with left-handers more likely to show non-lateralization or right-lateralization (Szaflarski et al., 2002). These structural differences would likely impact ERP patterns and may be responsible for some of the reported handedness effects in the ERP literature (Grey, Tanner, & van Hell, 2017; Lee & Federmeier, 2015). However, it should still be noted that a majority of left-handers show right lateralization of language networks like their right-handed counterparts (Szaflarski et al., 2002). It is also important to not exclude left-handers from research, and left-handed processing cannot be described as intrinsically atypical (Bailey et al., 2020). Nonetheless, we acknowledge the additional variability that left-handed participants potentially brought to our sample. Tanner & Van Hell (2014) reported that right-handed participants with a family history of left-handedness showed reduced P600 effects to morphosyntactic violations. The present study manipulated word order, not morphosyntax, so it is unclear whether this finding could extend to the present results.

5.3 Individual differences

We tested two claims about individual differences in N400 and P600 effects. The first claim is that the effects should be negatively correlated, where greater N400 effects correspond to smaller P600 effects. To evaluate this claim, we correlated effects at a single electrode with adjacent time windows; we also followed suggestions from Fromont and colleagues (2020) to compare non-adjacent time windows at non-identical electrodes. For the English experiment, we found that the N400 and P600 effects showed a modest correlation in all of these conditions. For the Mandarin experiment, we found that the correlation between the effects was more dramatically affected by the stringent measures suggested by Fromont and colleagues. While some of the correlations seen in Figure 8 and Figure 9 indicate that participants varied in a linear way between N400 and P600 effect magnitudes, it is also apparent that comparing adjacent time windows contributed substantially to the correlations. This effect of adjacent time window indicates that component overlap may at least partially have driven correlations (Delogu, Brouwer, & Crocker, 2021).

The second claim is that individuals should show ERP response patterns ranging from a pure N400, to a balanced, biphasic N400-P600, to a pure P600. Our RDI score results broadly confirmed this claim, even when applying the stricter criteria suggested by Fromont and colleagues (2020). The Mandarin data for the non-native speakers also showed that five participants showed a P600-dominant response, even though there was no visible P600 effect at the group level. Importantly, this full range of patterns was present even when N400 and P600 effects did not correlate for data from non-identical electrodes and non-adjacent time windows. This finding demonstrates that a correlation between N400 and P600 effects is not necessary for there to be a range of individual response patterns. While the RDI score distributions confirm the second claim, the story is not as clear for individuals with a more balanced pattern of effects. As can be seen in Figure 10 and Figure 11, individuals with a balanced RDI score tended to have a minimal RMI score, suggesting that rather than showing a balanced N400-P600 effect, these so-called balanced participants simply showed no effect. It is also important to note that a non-balanced RDI score does not mean that an individual did not show a biphasic N400-P600 response.

While the present study showed effects of individual differences, it must be noted that prior studies considered different grammatical structures. Some of the first studies to report these individual differences (Tanner et al., 2014, 2013; Tanner & Van Hell, 2014) centered around processing of morphosyntax, notably verb tense and subject-verb agreement in German and English. This type of agreement is realized by morphological inflection on the verb, and processing of these forms may draw on both lexicalized representations and grammatical rules. The authors suggest that individuals' N400 or P600 patterns may reflect different processing strategies, and in the case of non-native processing, represent their progress in lexical-driven or syntax-driven (and native-like) parsing. For Fromont and colleagues (2020), their experiment was about recognition of syntactic category when the French words *le* and *la* are temporarily ambiguous between interpretation as clitics or determiners, and must then be followed by verbs or nouns, respectively. These constructions did not involve morphosyntactic operations, and Fromont and colleagues (2020) results showed that the biphasic N400-P600 pattern was seen across individuals. Unlike these experiments, the present study considered phrase structure violations with no morphosyntactic operations and without a syntactic category ambiguity. As such, it would be presumptuous to assume that each of the varied experimental paradigms targeted identical processing mechanisms, even though the results all showed N400-P600 effects. Accordingly, the range and type of individual differences in processing strategies may also change between types of grammatical violations.

The signal-to-noise ratio is low for ERPs from a single individual, which can make it challenging to meaningfully conclude that a single participant shows a specific response pattern. It is sometimes only when considering the grand average of many participants that effects can be reliably measured, even with a large number of trials per participant (Boudewyn, Luck, Farrens, & Kappenman, 2018). In addition to variability at the participant level, variability at the item or trial level can also impact results in ways that are invisible when only considering grand averages. One solution can be to analyze ERP results with mixed effects models, as done in the present study and as recommended as a field standard (Baayen et al., 2008; D. Bates et al., 2015). The random structure of mixed effects models allows fitting of slopes and intercepts for individual participants, items, trials, or other factors that represent random samples from a larger population. By considering the elements involved in the random structure, individual differences can be considered alongside analyzing the effects of experimental factors (Schepens, van der Slik, & van Hout, 2018; Speelman, Heylen, & Geeraerts, 2018)

5.4 Limitations and future directions

Although the Mandarin and English experiments considered a highly similar adjective-noun structure, there were differences in the experimental design that may have impacted results. Firstly, the SOA for the English materials was 500 ms (as used in Steinhauer, 2014), while the SOA for the Mandarin materials was 750 ms (based on SOAs used by Bornkessel-Schlesewsky et al., 2011). For the first word N400 time window of 300-500 ms post-onset, participants in either experiment would still have only seen the first word of the adjective-noun pair. After 500 ms, however, participants' English ERPs may have been impacted by the appearance of the second word. Secondly, although the chosen Mandarin adjective structure is largely similar to its English counterpart, adjective-noun pairs without DE are not the most typical form of Mandarin adjectives. To understand typical Mandarin adjective comprehension, additional structures must be studied.

To this end, the current results present a logical continuation for a future experiment. Courteau, Misirliyan, Royle, & Steinhauer (2020) conducted a relevant study in French, comparing processing of pre- and post-nominal adjectives, their word order inversions, and their appearance in noun-drop structures. A corollary ERP experiment in Mandarin could embed adjectives in a syntactic structure that allows them to either follow or precede their nouns, such as *汽车正在等绿灯/灯绿了* "the car is waiting for the green light / the light (to turn) green". Including the particle DE allows us to generate additional structures to test experimentally. Adjective inversions would follow the format *张三是聪明的孩子/孩子的*聪明* "Zhangsan is a smart DE kid / kid DE *smart". Importantly, because of DE's role as a possessive marker, there can be a plausible, grammatical continuation to the sentence with the noun + DE structure, such as *张三是孩子的朋友* "Zhangsan is the kid's DE friend". Of note in comparison to French noun-drop, Mandarin permits noun drop with adjectives in these structures, such as *这件衣服不行，我要穿漂亮的* "this piece of clothing isn't good, I want to wear a pretty (piece) DE".

In a reading experiment, DE could be presented as a character on its own (as in 新/的/书 new/DE/book) or occur together with its preceding noun or adjectives (as in 新的/书 new DE/book). Although reading presentations like these are unnatural in relation to language use in the real world, this decomposition allows for a mechanistic separation of discrete processing steps. Splitting the phrase into three separate pieces in the visual modality allows the parser time to anticipate upcoming words. It is precisely this anticipation step that may reveal processing differences between languages and between individuals. Courteau and colleagues (2020) have further shown that effects in such an experiment can be measured from auditory ERP data, so an extension to auditory comprehension could also be feasible.

6 Conclusion

Although both languages' adjectives typically precede their nouns, English and Mandarin have important grammatical differences that can impact processing. In spite of these differences, our results showed that native speakers of both languages show a biphasic N400-P600 ERP pattern to adjective placement violations, indicating a shared comprehension mechanism.

However, Mandarin native speakers also showed an unexpected N400 effect on the first word for the correct condition. We interpret this effect as resulting from the semi-lexicalized nature of Mandarin adjective-noun pairs without the particle DE. This hypothesis must be tested with future experiments on different Mandarin adjective structures. In contrast to their native counterparts, non-native Mandarin participants showed only an N400 effect at the group level, in line with some predictions for lower-proficiency second language learners who may process grammatical violations by comparison to lexicalized forms instead of using rule-based procedural memory. The lack of a non-native P600 effect may be due in part to the flexible word order of Mandarin adjectives, whose position in relation to their nouns can communicate subtle differences, with structural restrictions that may be challenging for second language learners to comprehend natively. Crucially, these group-level results must be considered in the context of individual differences. By comparing individual participant results, we showed that correlations between N400 and P600 effect magnitudes can disappear under stringent comparison of non-adjacent time windows and non-identical electrodes. This shows that significant correlations between N400 and P600 effects may in part be due to component overlap or comparison of non-independent sources. However, even when effect magnitudes did not correlate, participants showed the full range of N400-dominant to balanced to P600-dominant. We suggest that these individual differences reflect divergent task and comprehension strategies among participants, and these strategies likely depend on the type of structures used in language experiments.

5 References

Arcodia, G. F. (2014). *The Chinese adjective as a word class*. https://doi.org/10.1075/cilt.332.06arc

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bailey, L. M., McMillan, L. E., & Newman, A. J. (2020). A sinister subject: Quantifying handedness-based recruitment biases in current neuroimaging research. *European Journal of Neuroscience*, *51*(7), 1642–1656. https://doi.org/10.1111/ejn.14542

Barber, H., & Carreiras, M. (2003). Integrating gender and number information in Spanish word pairs: An ERP study. *Cortex*, *39*(3), 465–482. https://doi.org/10.1016/S0010-9452(08)70259-4

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Birdsong, D. (2021). Analyzing variability in L2 ultimate attainment. *Language, Interaction and Acquisition*, *12*(1), 133–156. https://doi.org/10.1075/lia.21001.bir

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. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel, I., McElree, B., Schlesewsky, M., & Friederici, A. D. (2004). Multi-dimensional contributions to garden path strength: Dissociating phrase structure from case marking. *Journal of Memory and Language*, *51*(4), 495–522. https://doi.org/10.1016/j.jml.2004.06.011

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Bowden, H. W., Steinhauer, K., Sanz, C., & Ullman, M. T. (2013). Native-like brain processing of syntax can be attained by university foreign language learners. *Neuropsychologia*, *51*(13), 2492–2511. https://doi.org/10.1016/j.neuropsychologia.2013.09.004

Brazier, M. A. B., & Casby, J. U. (1952). Crosscorrelation and autocorrelation studies of electroencephalographic potentials. *Electroencephalography and Clinical Neurophysiology*, *4*(2), 201–211. https://doi.org/10.1016/0013-4694(52)90010-2

Caffarra, S., Mendoza, M., & Davidson, D. (2019). Is the LAN effect in morphosyntactic processing an ERP artifact? *Brain and Language*, *191*, 9–16. https://doi.org/10.1016/j.bandl.2019.01.003

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews*, *51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, *5*(6). https://doi.org/10.1371/journal.pone.0010729

Chan, I. L., & Chang, C. B. (2018). LEXTALE _ CH: A quick , character-based proficiency test for Mandarin Chinese. *Proceedings of the Annual Boston University Conference on Language Development*, *42*(1), 114–130. Retrieved from https://hdl.handle.net/2144/27833

Chen, G. (2012). *A Contrastive Study of Adjectival Class in Chinese and English (汉英形容词对比研究)*. Zhejiang University (浙江大学).

Cheng, L. L. S. (1986). de in Mandarin. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, *31*(4), 313–326. https://doi.org/10.1017/S0008413100011865

Cheng, L. L. S., & Sybesma, R. (2009). De 的 as an underspecified classifier: first explorations. 语言学论丛, *39*, 1–19.

Chow, W., Smith, C., Lau, E., & Phillips, C. (2016). A “ bag-of-arguments ” mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/23273798.2015.1066832

Courteau, E., Misirliyan, C., Royle, P., & Steinhauer, K. (2020). Prenominal and postnominal adjectives in French: An ERP study. *Proceedings of the Society for the Neurobiology of Language (SNL2021)*, 5–8.

Delogu, F., Brouwer, H., & Crocker, M. W. (2021). When components collide: Spatiotemporal overlap of the N400 and P600 in language comprehension. *Brain Research*, *1766*, 147514. https://doi.org/10.1016/j.brainres.2021.147514

Dixon, R. M. (1982). Where have all the adjectives gone? In R. M. Dixon (Ed.), *Where have All the Adjectives Gone? And Other Essays in Semantics and Syntax* (pp. 1–62). https://doi.org/10.1515/9783110822939.1

Dixon, R. M. (2004). Adjective Classes in Typological Perspective. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Adjective Classes : A Cross-Linguistic Typology* (pp. 1–45). Retrieved from http://ebookcentral.proquest.com/lib/mcgill/detail.action?docID=422819

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic Processing in Late Second-Language Learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Fromont, L. A., Steinhauer, K., & Royle, P. (2020). Verbing nouns and nouning verbs: Using a balanced design provides ERP evidence against “syntax-first” approaches to sentence processing. *PLOS ONE*, *15*(3), e0229169. https://doi.org/10.1371/journal.pone.0229169

Gao, F., Wang, J., Wu, C., Wang, M. yun, Zhang, J., & Yuan, Z. (2022). The neural dynamics associated with lexicality effect in reading single Chinese words, pseudo-words and non-words. *Cognitive Neurodynamics*, *16*(2), 471–481. https://doi.org/10.1007/s11571-021-09720-y

Gaspar, C. M., Rousselet, G. A., & Pernet, C. R. (2011). Reliability of ERP and single-trial analyses. *NeuroImage*, *58*(2), 620–629. https://doi.org/10.1016/j.neuroimage.2011.06.052

Gitsaki, C. (1998). Second language acquisition theories: Overview and evaluation. *Journal of Communication and International Studies*, *4*(2), 89–98. Retrieved from http://eprint.uq.edu.au/archive/00002325/01/L2-theories.htm

Grey, S., Tanner, D., & van Hell, J. G. (2017). How right is left? Handedness modulates neural responses during morphosyntactic processing. *Brain Research*, *1669*, 27–43. https://doi.org/10.1016/j.brainres.2017.05.024

Guajardo, L. F., & Wicha, N. Y. Y. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *NeuroImage*, *91*, 262–272. https://doi.org/10.1016/j.neuroimage.2013.09.077

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. https://doi.org/10.1080/01690969308407585

Hahne, A. (2001). What's Different in Second-Language Processing? Evidence from Event-Related Brain Potentials. *Journal of Psycholinguistic Research*, *30*(3), 251–266. https://doi.org/10.1023/A:1010490917575

Herbay, A. (2022). *ERPscope*. Retrieved from https://github.com/aherbay/erpscope

Holcomb, P. J., Kounios, J., Anderson, J. E., & West, W. C. (1999). Dual-coding, context-availability, and concreteness effects in sentence comprehension: An electrophysiological investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(3), 721–742. https://doi.org/10.1037//0278-7393.25.3.721

Hulstijn, J. H., Alderson, J. C., & Schoonen, R. (2010). Developmental stages in second-language acquisition and levels of second-language proficiency: Are there links between them. In R. Alanen, A. Huhta, M. Martin, M. Tarnanen, K. Mäntylä, P. Kalaja, & A. Palviainen (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 5–10). https://doi.org/10.1108/S1572-6304(2012)0000025053

Jacobsen, T., Klein, S., & Löw, A. (2018). The Posterior Sustained Negativity Revisited—An SPN Reanalysis of Jacobsen and Höfel (2003). *Symmetry*, *10*(1), 27. https://doi.org/10.3390/sym10010027

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, *1*(6), 571–591. https://doi.org/10.1111/j.1749-818x.2007.00037.x

Kemmerer, D., Weber-Fox, C., Price, K., Zdanczyk, C., & Way, H. (2007). Big brown dog or brown big dog? An electrophysiological study of semantic constraints on prenominal adjective order. *Brain and Language*, *100*(3), 238–256. https://doi.org/10.1016/j.bandl.2005.12.002

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2–3), 68–74. https://doi.org/10.1016/j.bandl.2008.06.002

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. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lau, E. F., Phillips, C., & Poeppel, D. (2008, December). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, Vol. 9, pp. 920–933. https://doi.org/10.1038/nrn2532

Lee, C., & Federmeier, K. D. (2015). It's All in the Family. *Psychological Science*, *26*(7), 997–1005. https://doi.org/10.1177/0956797615575743

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Li, & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis - The assumption that measurement error always reduces effect sizes is fals. *Science*, *355*(6325), 582–584.

Lüdecke, D. (2021). *sjPlot: Data Visualization for Statistics in Social Science*. Retrieved from https://cran.r-project.org/package=sjPlot

Makowski, D. (2018). The Psycho Package: An Efficient and Publishing-Oriented Workflow for Psychological Science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*(7), 703–704. https://doi.org/10.1038/nn1264

Meisel, J. M., Clahsen, H., & Pienemann, M. (1981). On determining developmental stages in natural second language acquisition. *Studies in Second Language Acquisition*, *3*(2), 109–135. https://doi.org/10.1017/S0272263100004137

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, *47*(8), 908–930. https://doi.org/10.1016/j.cortex.2011.02.019

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, *1*(1), 1–7. https://doi.org/10.1080/17588920903373952

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second Language Acquisition of Gender Agreement in Explicit and Implicit Training Conditions: An Event-Related Potential Study. *Language Learning*, *60*(1), 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933–947. https://doi.org/10.1162/jocn_a_00119

Mueller, J. L. (2005). Electrophysiological correlates of second language processing. *Second Language Research*, *21*(2), 152–174. https://doi.org/10.1191/0267658305sr256oa

Mueller, J. L. (2006). L2 in a Nutshell: The Investigation of Second Language Processing in the Miniature Language Model. *Language Learning*, *56*, 235–270. https://doi.org/10.1111/j.1467-9922.2006.00363.x

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The Influence of Language Proficiency on Lexical Semantic Processing in Native and Late Learners of English. *Journal of Cognitive Neuroscience*, *24*(5), 1205–1223. https://doi.org/10.1162/jocn_a_00143

Nicoladis, E., & Rhemtulla, M. (2012). Children’s acquisition of word order depends on syntactic/semantic role: Evidence from adjective-noun order. *First Language*, *32*(4), 479–493. https://doi.org/10.1177/0142723711422629

Nieuwland, M. S., Martin, A. E., & Carreiras, M. (2013). Event-related brain potential evidence for animacy processing asymmetries during sentence comprehension. *Brain and Language*, *126*(2), 151–158. https://doi.org/10.1016/j.bandl.2013.04.005

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly.

Journal of Memory and Language, *31*(6), 785–806. https://doi.org/10.1016/0749-596X(92)90039-Z
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*(SUPPL. 1), 199–230. https://doi.org/10.1111/j.1467-9922.2006.00361.x
Osterhout, L., & Mobley, L. A. (1995). Event-Related Brain Potentials Elicited by Failure to Agree. *Journal of Memory and Language*, *34*(6), 739–773. https://doi.org/10.1006/jmla.1995.1033
Osterhout, L., Poliakov, A., Inoue, K., McLaughlin, J., Valentine, G., Pitkanen, I., … Hirschensohn, J. (2008). Second-language learning and changes in the brain. *Journal of Neurolinguistics*, *21*(6), 509–521. https://doi.org/10.1016/j.jneuroling.2008.01.001
Paul, W. (2005). Adjectival modification in Mandarin Chinese and related issues. *Linguistics*, *43*(4), 757–793. https://doi.org/10.1515/ling.2005.43.4.757
Paul, W. (2010). *Adjectives in Mandarin Chinese*. (1994), 115–152. https://doi.org/10.1075/la.153.04pau
Paul, W. (2012). Why Chinese de is not like French de : A critical analysis of the predicational approach to nominal modification. *Studies in Chinese Linguistics*, *33*(3), 183–210.
Phillips, C., & Wagers, M. (2012). Relating structure and time in linguistics and psycholinguistics. *The Oxford Handbook of Psycholinguistics*, (2), 739–756. https://doi.org/10.1093/oxfordhb/9780198568971.013.0045
R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Retrieved from https://www.r-project.org/
Russell, L. (2019). *emmeans: estimated marginal means, aka least-squares means. R package version 1.4. 2*.
Schäfer, M. (2009). A+N constructions in Mandarin and the ‘compound vs. phrase’ debate. *Word Structure*, *2*(2), 272–293. https://doi.org/10.3366/E1750124509000452
Schepens, J., van der Slik, F., & van Hout, R. (2018). *The L2 Impact on Learning L3 Dutch: The L2 Distance Effect*. https://doi.org/10.1007/978-3-319-69830-4_3
Severens, E., Jansma, B. M., & Hartsuiker, R. J. (2008). Morphophonological influences on the comprehension of subject–verb agreement: An ERP study. *Brain Research*, *1228*, 135–144. https://doi.org/10.1016/j.brainres.2008.05.092
Smith, M. S., & Truscott, J. (2005). Stages or continua in second language acquisition: A MOGUL solution. *Applied Linguistics*, *26*(2), 219–240. https://doi.org/10.1093/applin/amh049
Speelman, D., Heylen, K., & Geeraerts, D. (2018). *Introduction*. https://doi.org/10.1007/978-3-319-69830-4_1
Sproat, R., & Shih, C. (1987). North East Linguistics Society Prenominal Adjectival Ordering in English

and Mandarin. *East Linguistics Society*, *18*(3).

Sproat, R., & Shih, C. (1991). The Cross-Linguistic Distribution of Adjective Ordering Restrictions. In *Interdisciplinary Approaches to Language* (pp. 565–593). https://doi.org/10.1007/978-94-011-3818-5_30

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Szaflarski, J. P., Binder, J. R., Possing, E. T., McKiernan, K. A., Ward, B. D., & Hammeke, T. A. (2002). Language lateralization in left-handed and ambidextrous people. *Neurology, 59*(2), 238–244. https://doi.org/10.1212/WNL.59.2.238

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism, 17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*(1), 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

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. https://doi.org/10.1017/s1366728901000220

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Voeten, C. C. (2021). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. Retrieved from https://cran.r-project.org/package=buildmer

Wang, Q., & Yuan, J. (2008). N400 lexicality effect in highly blurred Chinese words: Evidence for automatic processing. *NeuroReport*, *19*(2), 173–178. https://doi.org/10.1097/WNR.0b013e3282f3e3f2

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. https://doi.org/10.1162/jocn.1996.8.3.231

Xu, Z. (2018). The word status of Chinese adjective-noun combinations. *Linguistics*, *56*(1), 207–256. https://doi.org/10.1515/ling-2017-0035

Yang, N., Waddington, G., Adams, R., & Han, J. (2018). Translation, cultural adaption, and test–retest reliability of Chinese versions of the Edinburgh Handedness Inventory and Waterloo Footedness Questionnaire. *Laterality: Asymmetries of Body, Brain and Cognition*, *23*(3), 255–273. https://doi.org/10.1080/1357650X.2017.1357728

General Discussion

Although Manuscript I and Manuscript II of the thesis have been prepared as stand-alone articles, their results and interpretation combine to present a wide view of the factors involved in Mandarin sentence processing. Manuscript I considered the role of competing cues for argument structure assignment in verb-final sentences, while Manuscript II compared processing of adjective-noun and noun-adjective word orders in native and non-native Mandarin speakers. Both manuscripts manipulated syntax and both considered individual differences in participants' results. For each manuscript, we first summarize the results before discussing the broader context.

We note explicitly that the original proposal for the experiments in this thesis was in the context of a larger project to study first language attrition of Mandarin. We selected Mandarin verb-final sentences because there are specific conflicts with English sentence structure (i.e., the primary cue for English argument structure interpretation is word order, while word order is the weakest cue for Mandarin) that we predicted to be appropriate targets for studying attrition effects. On the other hand, we selected adjective-noun placement because both English and Mandarin have prenominal adjectives, which led us to the prediction that processing of this structure would be resistant to attrition. Unfortunately, the COVID-19 pandemic halted our collection of attrition data and forced a reevaluation of the thesis aims. Nonetheless, the two articles come together cohesively to provide advances in our understanding of cross-linguistic sentence processing.

For Manuscript I, our behavioral and ERP results support the idea that Mandarin is different from other languages in how parsers process argument structure. In line with previous Competition Model results (Li, Bates, Liu, & MacWhinney, 1992; Liu, Bates, & Li, 1992), our agent assignment task showed that BA and BEI were the strongest cues for interpreting the agent role; in contrast to prior results, however, we found that word order did not play a role in the absence of other cues. The interpretation and reaction time results together presented a range of interactions that was more nuanced than the simple ranking of cue strength suggested by previous Competition Model studies. Task results also revealed individual differences, where participants varied in whether they relied more on plausibility or more on coverbs to interpret sentences. We also observed an unexpected P200 difference at the position of the coverb, where BA had a significantly smaller P200 amplitude than both BEI and nouns. We interpret this as related to the explicit agent assignment task, where BA confirmed that the previous noun should be interpreted as the agent, while BEI pointed to an agent that was not yet available. At the verb, role reversal sentences elicited an N400 response without a subsequent P600 effect. This is in line with predictions from the eADM, but not the Bag of Arguments (Bornkessel-Schlesewsky et al., 2011; Chow, Momma, Smith, Lau, & Phillips, 2016). However, none of the three models could fully account for our findings. We suggest that sentence processing models must combine consideration of individual differences, incremental parsing throughout sentences, and crosslinguistic variability in sentence structures.

Manuscript II included two experiments: a replication of a prior experiment with English adjective placement and a new, corollary experiment with Mandarin adjective placement. We tested two groups of participants, functionally monolingual Mandarin native speakers and high-proficiency Mandarin non-native speakers whose native language was English. For our English-Mandarin bilinguals, our English adjective-noun results replicated the previously reported N400-P600 response pattern (Steinhauer, 2014). This was also the case for native Mandarin speakers for the Mandarin adjective-noun pairs, but the Mandarin data also showed an unexpected N400 effect at the first word in the correct condition. We could rule out the most likely lexical explanations for this difference and proposed that this effect is due to the lexical status of Mandarin adjective-noun pairs with the particle DE. For the non-native Mandarin speakers, we found a group-level N400 effect, which is consistent with lower-proficiency language learners. We suggested that Mandarin adjective placement is more nuanced than in English and so is a challenging structure for Mandarin learners to acquire. Thus, despite their relatively high proficiency, our non-native Mandarin participants did not show the native-like biphasic N400-P600 response. Additionally, we investigated individual differences in ERP patterns. Our findings showed that even though N400 and P600 effects did not always correlate with each other (Tanner et al., 2014), our participants showed a range of individual responses from N400-dominant to P600-dominant. We suggest that individual response patterns reflect individual differences in processing strategies for the experimental task, and these patterns likely depend on the type of sentence structure studied.

Below, we discuss the findings from the two manuscripts in a broader context. Namely, we 1) defend the idea that core mechanisms of sentence processing show crosslinguistic variability; 2) suggest that non-native sentence processing has nuances of cue weighting (or parameters or features) that require consideration of typology; 3) detail considerations and predictions for future work on language attrition; 4) highlight the importance and meaning of individual differences; and 5) discuss past and future challenges associated with this work.

1 Crosslinguistic comparison of argument structure processing

Human language shows wide variability (Evans & Levinson, 2009) and this variability may have implications for core mechanisms of sentence processing (Bornkessel-Schlesewsky et al., 2011). The present thesis explored two Mandarin structures: verb-final transitive sentences and adjective-noun pairs. In the context of previous data from English and other languages, our behavioral and ERP results support the idea that Mandarin sentence processing exhibits differences from that of other languages. Below, we first consider the findings of each manuscript in the context of crosslinguistic differences, and then present a broader discussion of the interpretation of such differences.

1.1 Manuscripts I and II

For verb-final sentences, Mandarin permits both subject-object-verb and object-subject-verb word order, and our data suggest there is no inherent preference between these orders, so context and plausibility are more important than word order for comprehension. To make argument structure more explicit in verb-final sentences, the coverbs BA and BEI can be used to assign patient or agent status, respectively, to their subsequent nouns. Our results demonstrate interactions among these cues for the assignment of argument structure in Mandarin. While BA and BEI were the strongest cues when available, they were both slightly weaker when they indicated an inanimate agent. In the case of role reversals, however, BEI was not affected by agent animacy, while BA was weaker when the implausible agent was inanimate. In contrast to prior results from Competition Model studies (H. Liu et al., 1992; Su, 2001), word order was not used to assign argument structure in the absence of other cues, and our inclusion of plausible inanimate agents demonstrated that plausibility, not animacy, is the primary cue for agent assignment in the absence of a coverb.

The sentence interpretation results from Manuscript I demonstrate that argument structure cues do not follow an additive, linear pattern of ranking in strength, in contrast to what has often been suggested in Competition Model studies (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; Li et al., 1992). There are instead subtle interactions among cues for sentence comprehension, as demonstrated by the interactions among coverb, agent animacy, and word order, as shown in the model predictions in Figure below (reproduced from Manuscript I). Nonetheless, we expect that the pattern of cue ranking across languages such that each language (or language type) shows a different profile of cue weights.

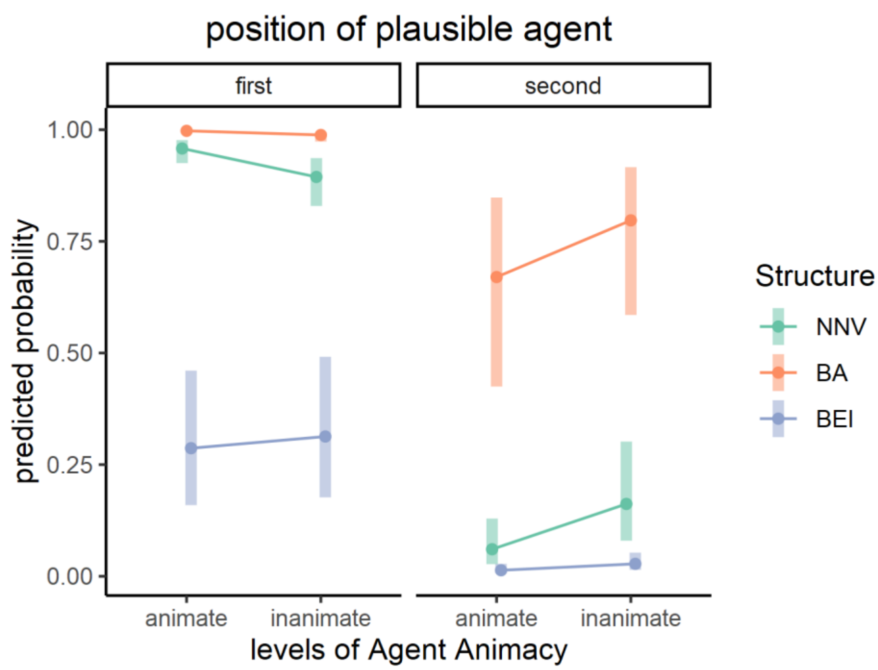

Figure 1*. Interaction among Order, Agent Animacy, and Structure for first agent noun selection in* ***irreversible sentences****. Error bars show 95% confidence intervals.*

The ERP results for role reversals give further evidence of crosslinguistic differences in argument structure processing. The two accounts that have previously examined Mandarin role reversals lead to different predictions: according to the eADM, Mandarin's flexible word order requires Mandarin parsers to consider additional cues for constructing templates for argument structure, which should result in N400 effects to role reversals (Bornkessel-Schlesewsky et al., 2011); the Bag of Arguments, in contrast, predicts there should always be P600 effects to role reversals, and an N400 effect may only be observed if there is both enough time to form a prediction and sufficient relatedness between a verb and its arguments (W. Chow et al., 2016; W. Y. Chow et al., 2018). Our results are more in line with predictions from the eADM, suggesting that role reversal anomalies may be detected by different mechanisms in different languages.

For adjective-noun order, we first consider the results from native speakers, and turn to the non-native speakers' data in Section 2. Behaviorally, all participants were able to detect that the adjective placement was a grammatical violation in their native language based on the sentence context, and their group-level ERP pattern showed a biphasic N400-P600 pattern that mirrored that previously reported for English (Steinhauer, 2014). This shared response pattern to English and Mandarin adjective placement violations is consistent with a shared mechanism for detecting the syntactic anomaly. At some level, English and Mandarin have the common feature of adjectives preceding nouns, and our ERP data suggest that this rule is applied in a similar way in online processing for native speakers of both languages.

However, the native Mandarin parsing data showed a qualitative difference from the English results: an N400 effect on the first word of the adjective-noun pair, where the correct word order elicited the larger N400. We excluded the simple lexical effect of frequency for this effect in our discussion in Manuscript II, instead suggesting that monomorphemic adjective-noun pairs in Mandarin represent lexicalized or partially lexicalized units. At one level, this explanation is orthogonal to adjective placement and closer to the study of compound words. However, if true, the lexical status of monomorphemic adjective-noun pairs demonstrates that Mandarin adjective argument structure is qualitatively different from English, and true adjective phrases (i.e., not lexicalized units) in Mandarin may require the connecting particle DE.

1.2 Variation in core mechanisms

As touched on in Manuscript I, crosslinguistic comparison can give insight in two directions. The first direction is identifying language and processing universals which apply no matter what language is under consideration. We note that this is an implicit assumption of the Bag of Arguments account (and most language processing models), which uses Mandarin sentence processing data to understand verb prediction across languages. There is clearly merit to this approach, and a variety of ERP components have been documented for similar phenomena across multiple languages, including the N400 and P600 (Kaan, 2007). Despite differences in culture and society, humans are a single species, and all languages must use the same cognitive architecture, and there is no reason individuals' innate cognitive abilities should vary across languages

However, although humans share broadly similar neuroanatomical structures and functions, it is well-established that the brain is shaped in profound ways by experience (Stiles, 2011). This impact of experience brings us to the second direction to which crosslinguistic comparison can contribute: revealing processing patterns that are quantitatively or qualitatively different across languages. Using a language encompasses hundreds of thousands of hours over the course of a lifetime. Within the first ten thousand of those hours, we know that the infant brain becomes perceptually tuned to the language environment (Polka, Rvachew, & Mattock, n.d.). At the level of grammatical rules, there is evidence that young children show different patterns of learning depending on the features of the language they are acquiring, such as acquiring morphological rules earlier for languages with rich morphology (Dressler, 2003). It then stands to reason that the experience of using particular language features may shape the core processing of sentence elements. For instance, might a speaker of a language with many grammatical cases be perceptually focused on the structural roles these cases communicate, or a speaker of an ergative language have a different perspective of transitivity than a speaker of an accusative language? Is this potential perceptual difference captured by the cue weighting or prominence filtering mechanisms proposed by the Competition Model and the eADM (Bornkessel & Schlesewsky, 2006a; MacWhinney, 2022b), or is there possibly a larger piece of the core cognitive architecture that is shaped differently?

Some may worry that this line of reasoning may lead to a strong version of the linguistic relativity hypothesis, where our language experience shapes the fundamental way we perceive the world (Kay & Kempton, 1984; Lupyan, 2012; Wolff & Holmes, 2011). This view has even gained traction in popular media, such as the movie *Arrival* where learning an alien language allows a linguist to see forward and backward in time (Villeneuve, 2016). Linguistic relativity is often staunchly denounced (e.g., Pinker, 1994) and even a weak version incites controversy (e.g., January & Kako, 2007). Nonetheless, it serves as a point of comparison for determining the potential of crosslinguistic differences to shape processing. Experimental effects have been demonstrated for the mapping of time and space (Y. Li, Casaponsa, Wu, & Thierry, 2019), categorical perception of color (Roberson, Davidoff, Davies, & Shapiro, 2005), and maybe even how eye witnesses conceive of the agents of events (Fausey & Boroditsky, 2011). These experimental findings suggest that language experience can, in at least limited ways, shape bottom-up perceptions and processing of the external environment.

This thesis makes no contribution to whether or not language shapes non-linguistic perception. Instead, we raise the example of linguistic relativity as a counterpoint. If there is some acknowledgement that at least small pieces of non-linguistic perception are impacted by language experience, and that we can base perceptual differences in neurobiology (Athanasopoulos & Casaponsa, 2020; Thierry, 2016), how can we dismiss the idea that the linguistic domain of sentence processing may have quantitative and qualitative crosslinguistic differences in basic mechanisms? If we take again the example of predicting an upcoming verb from its preceding arguments, users of verb-final languages and languages with frequent verb-final structures (including Mandarin) have more experience and a greater need for predicting verbs than users of non-verb final languages. This experience with greater demands on working memory could conceivably lead to core differences in sentence processing mechanisms.

The Competition Model at least implicitly considers that linguistic cues are not divided into separate domains, such as divisions between syntax and semantics, but could there be categories of cues that, through linguistic experience, become more important for specific languages? This idea of discrete categories of cues could be consistent with classification of Mandarina as a "semantics-based language" (Su, 2001). The Bag of Arguments proposes that parsing requires at least 800 ms to assign structural roles for prediction, but since this number is an empirical observation and not an intrinsic cognitive limit, there could be quantitative variation among languages. The eADM explicitly includes a prominence computation that weights cues according to their language-specific relevance, but users of all languages are assumed to create a template of structural roles from the earliest stages of sentence comprehension. The eADM limits the impact of crosslinguistic differences to steps two and three of the model, but even the proposed step one of identification of syntactic category may also conceivably show crosslinguistic variation, especially in a language with loose syntactic categories like Mandarin.

Nonetheless, the eADM goes a long way in proposing what pieces of the processing architecture are or are not vulnerable to language-specific influence, with testable predictions at each step. We recommend that other models integrate crosslinguistic variability, or at least make explicit where they predict no crosslinguistic variability. Identifying systemic patterns for syntactic rules across languages is challenging (e.g., Dryer, 1988), and we should not assume that the task is any simpler for identifying patterns for processing rules is any easier. Having sentence processing models that make crosslinguistic predictions is the first step, and the next step is to test these predictions systematically across languages. In this respect, the present thesis is not sufficient – for all its special features that make it interesting to study, Mandarin must be systematically compared to typologically distinct languages.

2 Native and non-native processing

2.1 Manuscripts I and II

Manuscript II directly compared native and non-native processing of adjective-noun placement. Because we recruited high-proficiency bilinguals whose first language (English) has the same preferred adjective-noun order as Mandarin, we predicted that non-native ERP results would resemble those of native speakers. Ultimately, this was not the case, with non-native participants showing an N400 effect compared to the native participants' biphasic N400-P600 effect. Neither did we find a relationship between ERP effects and Mandarin proficiency data. According to the model by Steinhauer, White, & Drury (2009), our findings are predicted for lower proficiency second language learners who are not yet applying combinatorial, rules-based processing.

To extend these considerations to the experiment in Manuscript I, we mention data from Mandarin-English bilinguals that are currently in analysis for Mandarin verb-final sentences. Preliminary findings suggest that role reversal sentences also showed an N400 effect for non-native Mandarin learners, albeit with a potential interaction with agent animacy. Preliminary behavioral results also suggest that non-native Mandarin speakers overall relied on similar cues to their native counterparts, without showing signs of their native preference for object-subject-verb interpretation. However, many participants struggled to assign agent status to plausible inanimate agents, suggesting that they used animacy as a cue instead of plausibility, which may have been due to their not having fully understood the event structures of the verbs.

The two Mandarin structures studied in this thesis have different relationships to their counterparts in English. For adjective placement in Manuscript II, both languages ostensibly have the same feature of adjective-noun order, such that transfer from L1 to L2 was expected. By contrast, for verb-final sentences in Manuscript I, English does not permit subject-object-verb word order; English also does not have coverbs, although there are at least nominal similarities between the coverb BEI and the English passive voice preposition "by," which can permit object-verb-subject word order ("he was bitten by the dog"). A priori, we expected that English-Mandarin bilinguals would have no trouble processing adjective placement rules in either language, while verb-final sentences in Mandarin could present difficulties and English speakers may default to their preferred object-subject-verb interpretations in the absence of other cues. However, our preliminary behavioral data for the verb-final processing experiment (as mentioned above) suggests that English-Mandarin bilinguals were able to apply both subject-object-verb and object-subject verb interpretations and could successfully use the cues of BA and BEI, in addition to appearing to show a similar ERP pattern to role reversal sentences. For adjective-placement, however, task accuracy was lower and ERP results were qualitatively different from native speakers. Together, these results suggest a more complex relationship between the grammatical structures in Mandarin and English than we had anticipated.

Adjective-noun placement is a good example to consider because it is relatively simple: attributive adjectives of a given language are placed postnominally or prenominally (or some combination thereof), and Mandarin and English both have prenominal adjective placement (Dryer, 2013a). Although Mandarin learners should have the benefit of positive transfer from their first language rules, our results showed that even high-proficiency learners showed only an N400 effect with no P600, suggesting they were not applying a proceduralized, rules-based reanalysis of the order violations (Steinhauer et al., 2009). The idea of a binary feature for adjective-noun order is not in keeping with our results, and this has forced us to consider the idea more deeply.

2.2 Complex Cues

The focus of the present thesis is on language processing, and our claims are ultimately about linguistic performance and not competence per se. However, we note theoretical linguistics accounts have extensively considered grammatical features, including in bilingualism and second language learning (e.g., Lardiere, 2009; Sorace, 2011). MacWhinney himself, one of the creators of the Competition Model, has written that nativist (i.e., based on ideas of Universal Grammar and rules and parameters) approaches are not incompatible with the emergentist and connectionist focus on processing (MacWhinney, 2017). We interpret MacWhinney's (2017) statement to express the idea that underlying the terminology differences, these perspectives share some common ideas as to the explanations for second language variability, and that despite their different foundations, nativist and emergentist accounts can inform each other.

For the non-native acquisition of Mandarin adjective-noun placement by English native speakers, we note that the shared feature of prenominal adjective placement (Dryer, 2013a) belies the complexity underlying adjective structures. In our interpretation of Manuscript II results, we suggest that lexical status itself interacts with monomorphemic adjective-noun pairs in Mandarin, while English adjectives are always their own words. Both English and Mandarin have cases where the adjective can directly follow the verb, but there are differences in these cases. When Mandarin adjectives follow their nouns, there is often a sense of "becoming the state of that adjective", such as *等灯绿了才可以走* /wait light green LE CAI can go/ "wait for the light to turn green before going". English learners of Mandarin likely regularly encounter these noun-adjective structures but may not have the precise knowledge of when this use is appropriate. In the behavioral data of Manuscript II, we also saw that accuracy for the incorrect noun-adjective order condition was relatively low for our non-native speakers, suggesting that instead of having completely reassembled their parameters (or features, or cues, or rules) for adjective placement, they instead simply accept noun-adjective order when in doubt.

The interplay among these linguistic features cannot be described in terms of a linear ranking of cue strength like Competition Model results typically do (Li et al., 1992; MacWhinney, Bates, & Kliegl, 1984). In the context of our two manuscripts, we recommend that cues should be considered to have complex interactions by default. We also note that first language knowledge and processing are not invulnerable to influence from the second language. As shown specifically for Mandarin-English bilinguals (H. Liu et al., 1992), there are both forward and backward transfer of processing strategies for argument structure assignment. Nonetheless, many studies do not consider the impact of or report their participants' knowledge of other languages, including studies by the Bag of Arguments and the eADM research groups (Bornkessel-Schlesewsky et al., 2011; Chow et al., 2016; Chow et al., 2018). We discuss the relation to first language attrition in more detail in the following section.

3 Language attrition

In the original proposal for the present projects, our primary aim was to use these experiments to study first language attrition. This effort was interrupted by the COVID-19 pandemic, but work is ongoing to make this extension. First language attrition is the non-pathological forgetting of the first language and has been documented at the levels of sound (e.g., Hopp & Schmid, 2013), word (e.g., Kasparian & Steinhauer, 2016), and sentence (Kasparian & Steinhauer, 2017). Attrition can range from dramatic loss of ability or motivation to use a first language, as documented in Holocaust survivors for their native German (Schmid, 2002), to adopting variations from a different dialect of the native language, as reported in Spaniards who relocated to Miami (Domínguez, 2013). Researchers have called for language attrition to be a part of models of bilingualism, with some have even suggesting that all bilinguals should be considered attriters (Schmid & Köpke, 2017).

Which domains of language are more vulnerable to attrition has also been a topic of debate. According to Schmid & Köpke (2017), native (morpho)syntactic rules are more resistant to attrition than other domains of language. Evidence for this includes lack of attrition effects for German morphosyntax (Bergmann, Meulman, Stowe, Sprenger, & Schmid, 2015). However, there is also ERP and behavioral evidence for attrition of both syntax and morphosyntax (Kasparian & Steinhauer, 2017; Kasparian, Vespignani, & Steinhauer, 2016). We note that some of this debate on syntactic processing does not distinguish between morphosyntax and phrase structure (per the example of Caffarra, Molinaro, Davidson, & Carreiras, 2015). Morphosyntactic changes, such as inflection for case marking or subject-verb agreement, can depend on long-distance dependencies, but can also likely rely in part on lexical memory (Krause, Bosch, & Clahsen, 2015). Phrase structure and word order, in contrast to morphosyntax, are inherently connected to the incremental building of syntactic hierarchies. In this way, explanations and predictions for syntactic attrition effects should distinguish between phrase structure (syntax) and morphosyntax.

Systematic comparison of languages with typological differences is of special value for the study of language attrition. We propose that comparing congruent, conflicting, and unshared features among languages can help to distinguish the roles of active interference and decay in language attrition. Interference refers to active interaction among units of memory (Underwood, 1957), while decay indicates independent weakening over time (Ebbinghaus, 1885). These two concepts have been important points of distinction in the science of memory (Hardt, Nader, & Nadel, 2013; Wixted, 2004). Direct evidence of the role of decay in language attrition can be seen in a study of Russian immigrants to Israel who showed changes to their native Russian despite having no second language knowledge (Baladzhaeva & Laufer, 2017). Because there was no second language knowledge, the changes in these immigrants' Russian were likely due to the change in environment and usage that resulted from their moving from one country to another. Thus, while decay is clearly a factor, it remains to be demonstrated that interference plays a role in language attrition.

3.1 Interference

Active interference in bilinguals further relates to the idea of coactivation. In this view, when bilinguals retrieve a word (or a rule, or a sound) from one language, they automatically activate the corresponding element in their other language. On the lexical front, convincing ERP evidence of this came from Thierry & Wu (2007), who showed that Mandarin-English bilinguals had reduced N400 effects when completing a lexical-relatedness task for unrelated English word pairs, but only when those pairs shared one character in Mandarin. For instance, the words ham and train are semantically unrelated in both English and Mandarin, but their Mandarin equivalents *火车* "train" and *火腿* "ham" share the same first syllable. Because the task was completed in English, the results are evidence that bilinguals automatically activate their native language translation equivalents. Similar effects have been reported for Welsh consonant mutation for Welsh-English bilinguals parsing English sentences (Vaughan-Evans, Kuipers, Thierry, & Jones, 2014). Importantly, this consonant mutation coactivation suggests that bilinguals not only simultaneously access their two lexicons, but also their two grammars.

Nonetheless, there has also been a proposal that coactivation is not necessary to explain these effects. Costa, Pannunzi, Deco, & Pickering (2016) proposed a model in which forward transfer during learning could result in the so-called coactivation results. In forward transfer, second language learners at beginning proficiency levels initially rely on the activation of their first language words and grammar, including the interconnections between these elements, such as the closer relationship between "train" and "ham" in Mandarin. At higher proficiency, learners may exclusively activate the new words and grammar of the second language, but the interconnections from their native language have been transferred intact to their second language knowledge representations. Thus, when they activate their second language, connections between elements of representations that transferred during learning could account for the effects reported by Thierry & Wu (2007), with no need for coactivation of both languages. One way to test this hypothesis proposed by Costa and colleagues (2016) would be to investigate the effect of the second language on the first; because their mechanism for forward transfer does not include a step for reverse transfer, any interference effects observed from the second language while processing the first would be due to coactivation.

This approach was just the strategy taken by a study of Spanish attriters residing in the US (Pu, Medina, J. Holcomb, & J. Midgley, 2019). ERP data showed that, compared to non-attriters, Spanish attriters showed a greater N400 effect for semantically unrelated English words that were lexically related in their native Spanish (e.g., *hueso* “bone” and *huevo* “egg”). If Costa and colleagues’ (2016) forward transfer during learning account were true, then there would be no explanation for the observed attrition effect. These data are thus strong evidence that bilinguals do indeed coactivate their lexical representations in both languages during word retrieval. Relating this back to interference vs decay in language attrition, there is evidence that interference effects in attrition are plausible. We hope to test this further by examining whether Mandarin attriters have active interference from English during sentence processing in their native language.

3.2 Ongoing and future efforts

Attriters have been and will continue to be recruited for testing, including with the experiments presented in this thesis. From the current results, we can make several predictions for effects in attriters. From Manuscript I, about argument assignment in verb-final sentences, we predict that extended English experience will change native Mandarin preference for sentence interpretation, as seen in prior Competition Model studies of Mandarin-English bilinguals (H. Liu et al., 1992; Su, 2001). The factor that is potentially most vulnerable to change is the cue strength of word order, which is not important in Mandarin but almost solely important in English. If Mandarin attriters adopt a more English-like strategy for word order interpretation, they should prefer object-subject-verb over subject-object-verb word order. From Manuscript II, about adjective-noun placement in Mandarin, we initially predicted that this structure would be resistant to attrition effects because both English and Mandarin have prenominal adjectives. In light of our findings, however, it is possible that Mandarin attriters could show differences in the N400 effect on the first word of the adjective-noun pair, which we interpreted as related to the lexical status of monomorphemic adjective-noun pairs.

4 Individual differences

In the present thesis, individual differences emerged in all experiments. In Manuscript I, participants’ agent selection showed that individuals varied in adopting a plausibility-driven, coverb-driven, or balanced approach. We calculated a continuous measure of this strategy and showed that individuals’ strategy influenced their reaction times, with more plausibility-driven participants taking more time, especially for sentences where word order was the only cue. In

Manuscript II, individuals varied in their ERP patterns, with some showing an N400-dominant response and others a P600-dominant response. Here, we first highlight some of the challenges of individual differences, in particular for ERP methodologies. We then discuss possibilities for interpretation beyond the experimental setting.

4.1 Challenges

Differences among individual participants often exceed the effect size for an experimental manipulation. For instance, in the reaction time data in Manuscript I, we report a group-level effect of structure such that the presence of a coverb results in an average of 400 ms faster decision making for the agent assignment task compared to when no coverb is present. However, individual participants and individual trials were much more variable. We should not interpret this as meaning that the group-level effect is not reliable (assuming sufficient power to detect our effect) but we must acknowledge that the majority of the variability in our sample is not explained by our experimental manipulation. If we stick to group-level hypotheses, this variability is merely noise obscuring the true effect. If we instead find meaningful measures to capture some of this variability, we may identify factors that are worth including in sentence processing models.

Aside from not explaining variability, neglecting individual differences may lead to spurious interpretations (Tanner et al., 2014; Tanner & Van Hell, 2014), who showed that group-level effects (e.g., biphasic N400-P600) may not reflect any individual's processing (e.g., only N400 or P600, no biphasic effects). In Manuscript II, we showed that our participants did exhibit a range of ERP responses from N400-dominant to P600-dominant, but at least one-third did show a biphasic N400-P600 pattern. While we discussed some issues with the notion of what a balanced N400-P600 response truly is in Manuscript II, we note that Tanner and colleagues' (2013) approach of using the Response Dominance Index (RDI) can capture some of the variability in ERP data. We further suggest that the RDI should be used in conjunction with the Response Magnitude Index (RMI) to confirm that RDI profiles correspond to a variety of RMIs.

The nature of ERPs themselves presents an additional challenge for appreciating individual differences. ERP waveforms represent the sum of multiple discrete neural sources that overlap in space and time. If we just consider the N400 and P600 components, individuals may vary in either or both of these signals, but we must interpret a waveform that is the sum of overlapping signals. This overlap may lead to incorrect interpretations. As depicted in Figure 2, Brouwer &

Crocker (2017) outline a scenario where two conditions show a true difference only in the amplitude of the N400 component. However, because of component overlap, we would also be inclined to wrongly interpret that the first condition shows a large P600 effect while the second condition shows a minimal P600 effect. In fact, the P600 component is the same in both cases. This example represents an issue for interpreting group-level effects, but also demonstrates how individual variability could potentially compound and lead us to incorrect interpretations of group-level effects. Suppose that a particular individual, for whatever reason, has a smaller N400 component to begin with but a strong P600 component. Given the nature of what EEG scalp measurements truly reflect from neural activity, this could be as trivial as a difference in orientation for the neural source of the N400 (Luck, 2014). In fact, that individual's neural response indexed by the N400 and P600 are equal in strength, but the EEG signal captures variability that makes it challenging to compare the magnitudes of the two components. This example is only hypothetical, but even at a smaller scale, such a scenario could extend across individuals to then give a misleading picture at the group-level. Nonetheless, we still note that RDI and RMI values could solve some of this issue. Supposing that the example in the figure were individual data, then the RDI for Contrast 1 would show a balanced response while the RDI for Contrast 2 would show an N400-dominant response. The RMI should then be the same for both contrasts.

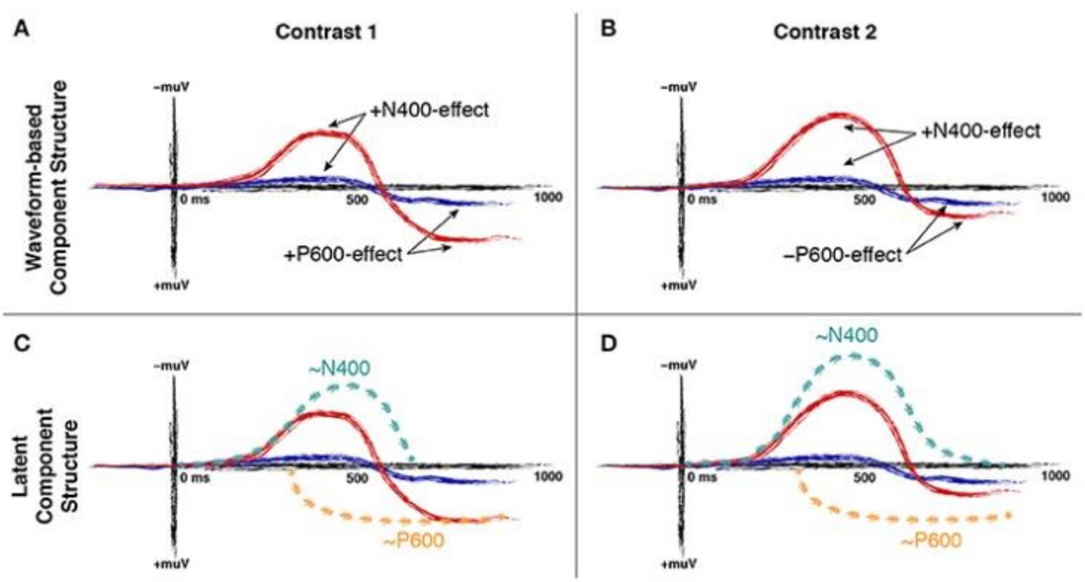

Figure 2. *Hypothetical wave forms for observed and underlying N400 and P600 effects. Reproduced from Brouwer & Crocker (2017) per Creative Commons License.*

In at least some ways, the consideration of individual differences as a worthy topic of study is relatively recent in the language sciences, and this new approach has the potential to make significant changes to our experimental design and analysis (Goodhew & Edwards, 2019). Previously, our statistical approaches have usually considered group-level effects, with individual variation around that effect an obstacle to overcome, and the ERP methodology is vulnerable to group-level effects misrepresenting individual data (Tanner & Van Hell, 2014). As demonstrated by multiple researchers, there is a major replication crisis in experimental psychology and cognitive neuroscience where reported effects are often Type I errors or may not extend beyond the precise experimental conditions of the laboratory setting (Ioannidis, 2005; Romero, 2019; Szucs & Ioannidis, 2017; c.f. Leek & Jager, 2017). Including evaluation of individual differences in experiments can be a tool for holding science to higher rigor (Rouder & Haaf, 2021).

4.2 Interpretation

Although individual differences are acknowledged as an important area for research (Kidd et al., 2018), it is at least initially difficult to appreciate whether or not our findings extend beyond the experimental tasks to real-world sentence processing. In Manuscript I, participants may have differed in their motivation or understanding of the task and thus adopted different approaches for either relying on the coverb or plausibility. Likewise in Manuscript II, individuals may have differed in their strategy for detecting the adjective-noun placement manipulation, either going by the first word in the adjective-noun pair or waiting for the full pair. This brings us to a difficult question: are individual differences in language sciences completely metalinguistically driven by task demands?

We consider two possibilities for how these differences could manifest in more naturalistic settings. First, our findings may reflect real, consistent variability among individuals. This would mean that our plausibility-driven participants in Manuscript I would indeed take longer to process sentences before interpreting meaning, while N400-dominant participants in Manuscript II would adopt more prediction-driven strategies to parse sentences in everyday life and P600-dominant participants instead reanalyze and apply combinatorial rules to their parses. While we cannot rule out this possibility, we acknowledge that our experimental findings cannot convincingly support a strong version of this interpretation. To justify an interpretation along these lines, we would need to demonstrate test-retest reliability for individuals.

Instead, a second possibility is that our findings demonstrate that individuals may at different times and in different tasks apply qualitatively and quantitatively different strategies for parsing. For instance, our plausibility-driven participants may have been more motivated to identify the "correct" agent in each sentence, so they took extra time to evaluate their parsing. Our N400-dominant participants may have used the first word of the adjective-noun pair to somehow anticipate the correct or incorrect adjective-noun order, while the P600-dominant participants instead waited to parse the full adjective phrase before applying a rules-based reanalysis or repair. This second possibility suggests that individuals can adopt different "modes" of sentence processing, which should indeed extend to real-life language use. For example, we have moments when we need to carefully listen to the precise meaning of an utterance and hang on to every word, predicting what might be said next, and other times when we just need to get the gist of what is said and only reanalyze a structure when we have difficulty parsing. Thus real-life, task-dependent parsing strategies may be similar to Grosjean's descriptions of language modes in bilinguals (Grosjean, 2012).

5 Past and future challenges

5.1 Meaningful statistics

Mixed effects models have recently been touted as a standard for the field of language sciences (Baayen et al., 2008), including for ERP research (Davidson, 2009; Nickels & Steinhauer, 2018; Tremblay, 2005). In comparison to the ANOVA, mixed effects models allow for including random structure (akin to the error term in ANOVA) for both item and participant in the same model, as well as being flexible for missing data points for one or more conditions (Baayen et al., 2008). The choice between ANOVA and mixed effects models can also lead to disparate analysis results (Kyriaki et al., 2020). Over the course of analysis for the present thesis, we have invested considerable time in determining our optimal method for model building and comparison. There is continued discussion in the literature as to the best ways of using mixed effects models (Meteyard & Davies, 2020), but we believe our approach (as outlined in Methods section of Manuscripts I and II) has advantages.

For analysis of any dataset, there must be a decision between hypothesis-driven and data-driven analysis (Shih & Chai, 2016). For the present thesis, we believe we found a compromise between hypothesis-driven and data-driven approaches. We have used the R package *buildmer*

(Voeten, 2021) to fit the maximally-converging model for both fixed effects and random structure (including both participant and item/trial as random effects) and then optimizing from the maximal model to the model with the highest log-likelihood score. Log-likelihood increases with explanatory power of the model but also receives a penalty for each additional variable included to minimize the risk of overfitting. Because interaction coefficients in regression models can be challenging to interpret (Hayes, Glynn, & Huge, 2012), we further limited the highest-order effects to three-way interactions. For the results sections, we then reported the optimized models and included the maximal models in the supplementary materials. We reviewed each of the models in the context of predictions and meaningful interpretations. In a single case, for the role reversal effects in the N400 time window, we opted to report a slightly different model in the text that we believe best reflects the actual variability in the data, accompanying this with a justification in the manuscript text. For transparency, we still report all models in supplementary materials so that our results can be critically evaluated.

One area we wish to expand on in future work is interpretation of random effects. Random effects are routinely not reported for mixed effects models (Meteyard & Davies, 2020), which we believe to be an oversight of a valuable opportunity to engage with individual and item variability. A primary strength of mixed effects models is the opportunity to include variation by participant and by item at the same time (Baayen et al., 2008), but this procedure should be an opportunity to consider this additional variability in a rigorous way. One way to report this is to show graphs as in Figure 3, which allow identification of factors on which participants vary in a meaningful way.

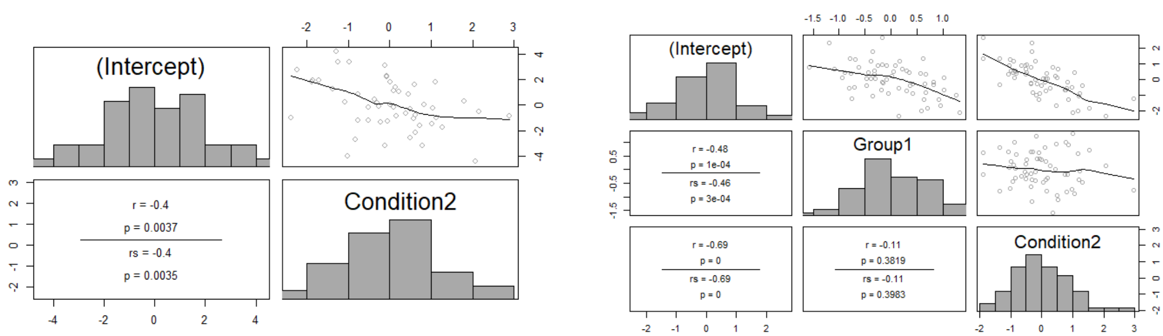

Figure 3. *Graphical representation of the random effects for the mixed effects model of N400 amplitude in the Mandarin adjective-noun placement experiment.*

We make a final comment on baseline correction in ERP analyses in relation to mixed effects models. As suggested by Alday (2019), an alternative to baseline correction is to include the baseline interval as a factor in the model. As the authors detail, baseline correction can reduce signal-to-noise ratio, so statistical power can be increased by including the baseline interval in the model. We note two difficulties for applying their recommendation in the context of the present thesis work. First, there are multiple considerations for selecting the appropriate time window for applying a baseline correction, and inappropriate selection runs the risk of introducing artifacts into the data (Steinhauer & Drury, 2012). For instance, in Manuscript I, we deliberated the selection of a pre-onset or post-onset baseline for capturing the effect of our experimental manipulation, ultimately deciding that a post-onset baseline minimized differences that occurred prior to the critical word (but still reporting the pre-onset baseline in supplementary materials). Alday's suggested alternative to baseline correction unfortunately does not address this prime challenge of time window selection for an appropriate baseline. Of course, we could still have applied Alday's approach to both our pre-onset and post-onset baseline intervals, but we note that this does not "solve" the issue of baseline correction because it is still essential to choose an appropriate time window to place the baseline. Second, we note that when an experimental design is more complex than the 2x1 design example given by Alday, including the additional factor of baseline and its interactions may make interpretation more challenging and detract from the experimental manipulations of interest. One response to this second criticism is that experimental design should be simple and that higher-order interactions may be underpowered (Brysbaert & Stevens, 2018; Card et al., 2020). While we acknowledge the importance of this concern, we also recognize that controlling for multiple factors in sentence processing experiments is necessary to capture the intrinsic interactions among linguistic factors in processing and allows for data-driven analysis of datasets that is useful beyond the scope of a single experiment (Sassenhagen, 2019).

5.2 Task and modality

In a recent report, Kyriaki, Schlesewsky, & Bornkessel-Schlesewsky (2020) compared the impacts of task and modality on ERP effects in response to role reversal sentences. While their results showed some overarching pattern of N400 and P600 effects from prior studies (Bourguignon et al., 2012), the results showed differences both by task and by modality. We briefly consider the influence of task and modality for the results of the present thesis, and finish with some thoughts on how these issues may be addressed in future work.

The experiments in the present thesis are all reading studies with rapid serial visual presentation (RSVP) of sentences word-by-word. While many ERP components, including the N400 and P600, have corollaries in both the auditory and visual domains, it is not trivial to directly compare visual and auditory ERP results. While RSVP experiments allow for precision in the onset and duration of each word, eliciting a clear pattern of N1-P2-N400 components for each unit (Kaan, 2007), auditory presentation does not have clear onsets or offset between words. In the context of role reversal experiments, we note that previous experiments of Mandarin sentence processing differed in auditory (Bornkessel-Schlesewsky et al., 2011) versus visual (Chow & Phillips, 2013; Chow et al., 2018) presentation. Kyriaki and colleagues' (2020) own results showed that N400 and P600 effects may appear and disappear depending on the modality of presentation. Given the ongoing debate about role reversals (Seyednozadi, Pishghadam, & Pishghadam, 2021), we note that in the context of the ERP methodology there may be advantages to first considering the relatively simpler case of reading studies where stimulus onset and duration are easier to control. Because the auditory domain presents special challenges for experimental design and analysis, it should be considered if the additional data gained are inherently more meaningful than reading data. In many cases, the answer will be yes (e.g., the impact of prosody, experiments of naturalistic sentence comprehension, the role of speaker identity), but other cases may be better suited to at least initial investigation in the visual domain. In that sense, there are still open questions concerning the impact of modality on the N400 and P600 components.

Task is another well-reported source of variability in ERP effects (e.g., Gajewski & Falkenstein, 2013). In Manuscript I, we used the classic task from Competition Model studies for the first time in an ERP setting, which is in contrast to the acceptability judgments (Chow & Phillips, 2013; Chow et al., 2018) and comprehension questions (Bornkessel-Schlesewsky et al., 2011) employed in previous experiments on Mandarin role reversals. We note that while our task encourages metalinguistic awareness of the experimental manipulation, it is also closer to the real-world comprehension goal of parsing agent and patient roles in sentences. Although some accounts have proposed that the N400 component (and other negative-going ERP components) represent automatic processing resistant to task effects (Bornkessel-Schlesewsky & Schlesewsky, 2019), we note that Kyriaki et al. (2020) showed some impact of task on both the N400 and P600 components. For our own data, our experimental tasks in Manuscripts I and II may have promoted more P600-driven processing (Schacht et al., 2014). On the other hand, the explicit agent assignment task in Manuscript I could conceivably resulted in greater awareness of semantic roles and perhaps encouraged an N400 response. Additionally, the Mandarin experiments in both manuscripts did not use filler sentences, and participants were thus likely aware of at least some of the experimental manipulations, which may have impacted ERP results (Brouwer & Crocker, 2017). We acknowledge this limitation in our work, but note that, as detailed above, our results cannot be dismissed as due solely to participant metalinguistic awareness. Instead, more work is needed to understand task effects on sentence comprehension.

With the goal of identifying the effects of task and modality, we commend the systematic approach of Kyriaki et al. (2020). For the effect of task specifically, we suggest that future work apply a “necessary and sufficient” approach (Brennan, 2022) like that often used in biological sciences (or instead the terms “indispensable and inducing”, as suggested by Yoshihara & Yoshihara, 2018). This necessary and sufficient approach to experiments is closely related to the idea of a “rescue” condition, wherein a previously absent effect is restored by the application of necessary and sufficient measures. We note that the approach used in the Bag of Arguments account for “rescuing” the N400 effect in role reversal sentences by allowing for longer time for prediction is in line with our proposal (Chow et al., 2018). We return also to the idea of “modes” that was discussed in Section 4.2. Is it possible to use a task to induce a certain processing mode in individuals and then use a different task to have participants employ a different processing mode? Such an experiment would likely need to be much simpler than the role reversal manipulations in the present thesis or in Kyriaki et al. (2020), but an approach along these lines has the potential to demonstrate meaningful effects of both task and individual differences that we can take into consideration for more complex experiments. We also note that tasks vary in how well they capture individual differences (Hedge, Powell, & Sumner, 2018), and it may be of interest in experimental design to include both tasks known to produce both high and low inter-individual variability.

5.3 Towards complete models of sentence processing

From the results of this thesis, we have several recommendations for what sentence processing models must include. First, we acknowledge what the three models considered in the present thesis have already captured. The Competition Model is a comprehensive account of language learning that can be applied to first and second language acquisition, native and non-native processing, and can account for some extent of crosslinguistic differences (MacWhinney, 2005, 2022a). The eADM focuses on core sentence processing mechanisms with predictions for crosslinguistic impacts on specific, delineated processing steps (Bornkessel-Schlesewsky et al., 2011; Bornkessel-Schlesewsky & Schlesewsky, 2008; Bornkessel & Schlesewsky, 2006a). The Bag of Arguments considers the processes underlying prediction with the expectation that the account extends to all languages (Chow et al., 2016; Chow et al., 2018, 2016). All three models present testable predictions for experimental work and have greatly contributed to the planning and interpretation of the work in the present thesis.

However, the models all fall short in accounting for both bilingualism and crosslinguistic differences and their interplay with individual differences. At first glance, capturing all three of these elements in a single model may seem too broad of a scope. However, we note that the Competition Model itself is already broad in applying to learning and processing (MacWhinney, 2022a). Additionally, we note that the eADM's predictions are confounded by the potential interaction with participants' bilingualism, which may obscure the crosslinguistic differences for monolinguals that the model aims to capture. The Bag of Arguments forgoes any crosslinguistic variability in its account, instead highlighting what should be universal across languages. Sentence processing clearly has common mechanisms across humans regardless of their specific language(s), but we should not suppose that these mechanisms are invulnerable to the variability in language structure.

During the course of our analysis and interpretations in the present thesis, we have found the structure of discrete processing stages as outlined in the eADM to be useful in making predictions and interpretations. We propose that in addition to the prominence, linking, and plausibility computations in step two of the eADM, steps one and three may also show crosslinguistic variability. In a recent account of negative ERP components, the authors behind the eADM substantially extend potential impacts on the N400 component (Bornkessel-Schlesewsky & Schlesewsky, 2019), but there is still no account for the roles of individual variability or bilingualism.

While a single psychological model cannot hope to explain all aspects of cognition, the field of sentence processing is clearly interested in the nuances of the impacts of multiple factors. One approach would be to apply a checklist of relevant factors that each model should consider. Several research groups have highlighted the need for checklists in ERP and neuroimaging research (Gau et al., 2022; Šoškić, Jovanović, Styles, Kappenman, & Ković, 2022). We propose that applying a checklist of relevant variables to processing models can help to hold model proposals to a high standard and encourage explicit, testable predictions. With a checklist of this nature, authors of new and old model proposals can make explicit predictions for factors they may not have otherwise considered, which can contribute to community efforts for open and reproducible science (Bosco et al., 2017). Below in Table 1, we have outlined such a hypothetical checklist for the eADM and Bag of Arguments models, filling in the answers based on our own understanding of the models' predictions.

*Table 1. **Example checklist for three factors applied to the Competition Model, eADM, and Bag of Arguments**.*

Factor to Consider	Competition Model	eADM	Bag of Arguments
Crosslinguistic variation	Languages vary in strength and validity of different cues; cue strength is hierarchical and in the case of cue conflict the strongest cue drives processing	Core mechanisms of processing distinguish predicate and non-predicate sentence elements; languages vary in their encoding of prominence cues after identification of an element as predicate or non-predicate	Prediction of upcoming verbs proceeds identically across languages
Bilingualism	Native and non-native language processing share similar cognitive mechanisms; bilingualism entails positive and negative, forward and reverse transfer that interacts with proficiency and age of acquisition	Bilingual experience may impact the same stages that are susceptible to crosslinguistic variation	Prediction depends on cloze probability, so bilingualism may impact in so far as language experience impacts event knowledge
Individual differences	Individuals should follow cue rankings according to their specific language experience	Individuals should follow cue rankings according to their specific language experience	Individuals' predictions may be influenced by factors such as working memory and event knowledge

6 Conclusion

Mandarin and English have a combination of shared and unshared linguistic features that are a valuable source of comparison for understanding basic mechanisms of sentence processing. In this thesis, we have used behavioral and ERP methods to study processing of two Mandarin sentence structures: verb-final sentences with varying cues for argument structure and adjective-noun pairs with varying word order. For verb-final sentences, we have shown that cues for argument structure assignment interact in ways beyond simple linear order and word order is not used in the absence of other cues. Our ERP results for verb-final sentences with role reversals are consistent with models predicting crosslinguistic differences in argument structure processing (Bornkessel-Schlesewsky et al., 2011). For adjective-noun placement, our results show that Mandarin and English native speakers process word order violations via a similar neural mechanism, but that certain kinds of Mandarin adjective-noun pairs may have undergone lexicalization processes. Additionally, non-native Mandarin speakers showed an ERP profile with qualitative differences from native speakers, suggesting that, despite relatively high proficiency, Mandarin and English adjective structures have key feature differences that are challenging for second language acquisition. The results of this thesis center around variability across languages, individuals, and native and non-native processing. Our results demonstrate that these three sources of variation must be considered together and neglecting any of them may confound research findings.

7 References

Alday, P. M. (2019). How much baseline correction do we need in ERP research? Extended GLM model can replace baseline correction while lifting its limits. *Psychophysiology*, *56*(12). https://doi.org/10.1111/psyp.13451

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baladzhaeva, L., & Laufer, B. (2017). Is first language attrition possible without second language knowledge? *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2016-0066

Bates, E., McNew, S., MacWhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, *11*(3), 245–299. https://doi.org/10.1016/0010-0277(82)90017-8

Bergmann, C., Meulman, N., Stowe, L. A., Sprenger, S. A., & Schmid, M. S. (2015). Prolonged L2 immersion engenders little change in morphosyntactic processing of bilingual natives. *NeuroReport*, *26*(17), 1065–1070. https://doi.org/10.1097/WNR.0000000000000469

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. https://doi.org/10.1016/j.bandl.2010.09.010

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on “semantic P600” effects in language comprehension. *Brain Research Reviews*, *59*(1), 55–73. https://doi.org/10.1016/j.brainresrev.2008.05.003

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2019). Toward a neurobiologically plausible model of language-related, negative event-related potentials. *Frontiers in Psychology*, *10*(FEB), 1–17. https://doi.org/10.3389/fpsyg.2019.00298

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review*, *113*(4), 787–821. https://doi.org/10.1037/0033-295X.113.4.787

Bosco, F., Carp, J. M., Field, J. G., IJzerman, H., Lewis, M., Munafo, M., … Giner-Sorolla, R. (2017). Maximizing the Reproducibility of Your Research. In *Psychological Science Under Scrutiny* (pp. 1–21). https://doi.org/10.1002/9781119095910.ch1

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. https://doi.org/10.1016/j.bandl.2012.05.001

Brennan, A. (2022). Necessary and Sufficient Conditions. In E. N. Zalta & U. Nodelman (Eds.), *The {Stanford} Encyclopedia of Philosophy* ({F}all 202). Metaphysics Research Lab, Stanford University.

Brouwer, H., & Crocker, M. W. (2017). On the Proper Treatment of the N400 and P600 in Language Comprehension. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01327

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.10

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience and Biobehavioral Reviews*, *51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Card, D., Henderson, P., Khandelwal, U., Jia, R., Mahowald, K., & Jurafsky, D. (2020). With Little Power Comes Great Responsibility. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 9263–9274. https://doi.org/10.18653/v1/2020.emnlp-main.745

Chow, W.-Y., & Phillips, C. (2013). No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese. *Brain Research*, *1506*, 76–93. https://doi.org/10.1016/j.brainres.2013.02.016

Chow, W., Smith, C., Lau, E., & Phillips, C. (2016). A " bag-of-arguments " mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596. https://doi.org/23273798.2015.1066832

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), 803–828. https://doi.org/10.1080/23273798.2018.1427878

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. https://doi.org/10.1080/23273798.2016.1160135

Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2016). Do Bilinguals Automatically Activate Their Native Language When They Are Not Using It? *Cognitive Science*, (November). https://doi.org/10.1111/cogs.12434

Davidson, D. J. (2009). Functional Mixed-Effect Models for Electrophysiological Responses. *Neurophysiology*, *41*(1), 71–79. https://doi.org/10.1007/s11062-009-9079-y

Domínguez, L. (2013). *Understanding Interfaces*. https://doi.org/10.1075/lald.55

Dressler, W. (2003). Morphology and Linguistic Typology. *On-Line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)*.

Dryer, M. S. (1988). Object-verb order and adjective-noun order: Dispelling a myth. *Lingua*, *74*(2–3), 185–217. https://doi.org/10.1016/0024-3841(88)90059-9

Dryer, M. S. (2013). Order of Adjective and Noun. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Retrieved from https://wals.info/chapter/87

Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker & Humblot.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–448. https://doi.org/10.1017/S0140525X0999094X

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*(1), 150–157. https://doi.org/10.3758/s13423-010-0021-5

Gajewski, P. D., & Falkenstein, M. (2013). Effects of task complexity on ERP components in Go/Nogo tasks. *International Journal of Psychophysiology*, *87*(3), 273–278. https://doi.org/10.1016/j.ijpsycho.2012.08.007

Gau, R., Gould van Praag, C. D., van Mourik, T., Wiebels, K., Adolfi, F. G., Scarpazza, C., … Moreau, D. (2022). *COBIDAS checklist*. Retrieved from https://doi.org/10.17605/OSF.IO/ANVQY

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition*, *69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Grosjean, F. (2012). Bilingual and Monolingual Language Modes. In *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0090

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences*, *17*(3), 111–120. https://doi.org/10.1016/j.tics.2013.01.001

Hayes, A. F., Glynn, C. J., & Huge, M. E. (2012). Cautions Regarding the Interpretation of Regression Coefficients and Hypothesis Tests in Linear Models with Interactions. *Communication Methods and Measures*, *6*(1), 1–11. https://doi.org/10.1080/19312458.2012.651415

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hopp, H., & Schmid, M. S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, *34*(2), 361–394. https://doi.org/10.1017/S0142716411000737

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky (2001). *Cognition*, *104*(2), 417–426. https://doi.org/10.1016/j.cognition.2006.07.008

Kaan, E. (2007). Event-Related Potentials and Language Processing: A Brief Overview. *Language and Linguistics Compass*, *1*(6), 571–591. https://doi.org/10.1111/j.1749-818x.2007.00037.x

Kasparian, K., & Steinhauer, K. (2016). Confusing similar words: ERP correlates of lexical-semantic processing in first language attrition and late second language acquisition. *Neuropsychologia*, *93*(June), 200–217. https://doi.org/10.1016/j.neuropsychologia.2016.10.007

Kasparian, K., & Steinhauer, K. (2017). When the second language takes the lead: Neurocognitive processing changes in the first language of adult attriters. *Frontiers in Psychology*, *8*(MAR), 1–22. https://doi.org/10.3389/fpsyg.2017.00389

Kasparian, K., Vespignani, F., & Steinhauer, K. (2016). First Language Attrition Induces Changes in Online Morphosyntactic Processing and Re-Analysis: An ERP Study of Number Agreement in Complex Italian Sentences. *Cognitive Science*, 1–44. https://doi.org/10.1111/cogs.12450

Kay, P., & Kempton, W. (1984). What Is the Sapir-Whorf Hypothesis? *American Anthropologist*, *86*(1), 65–79. https://doi.org/10.1525/aa.1984.86.1.02a00050

Kidd, E., Donnelly, S., & Christiansen, M. H. (2018). Individual Differences in Language Acquisition and Processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. https://doi.org/10.1016/j.tics.2017.11.006

Krause, H., Bosch, S., & Clahsen, H. (2015). MORPHOSYNTAX IN THE BILINGUAL MENTAL LEXICON. *Studies in Second Language Acquisition*, *37*(4), 597–621. https://doi.org/10.1017/S0272263114000564

Kyriaki, L., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2020). Semantic reversal anomalies under the microscope: Task and modality influences on language-associated event-related potentials. *European Journal of Neuroscience*, *52*(7), 3803–3827. https://doi.org/10.1111/ejn.14862

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, *25*(2), 173–227. https://doi.org/10.1177/0267658308100283

Leek, J. T., & Jager, L. R. (2017). Is Most Published Research Really False? *Annual Review of Statistics and Its Application*, *4*(1), 109–122. https://doi.org/10.1146/annurev-statistics-060116-054104

Li, P., Bates, E., Liu, H., & MacWhinney, B. (1992). Cues as Functional Constraints on Sentence Processing in Chinese. In H. C. Chen & O. Tzeng (Eds.), *Language Processing in Chinese* (pp. 207–

234). https://doi.org/10.1016/S0166-4115(08)61893-2

Li, Y., Casaponsa, A., Wu, Y. J., & Thierry, G. (2019). Back to the future? How Chinese-English bilinguals switch between front and back orientation for time. *NeuroImage*, *203*, 116180. https://doi.org/10.1016/j.neuroimage.2019.116180

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, *13*(4), 451–484. https://doi.org/10.1017/S0142716400005762

Luck, S. J. (Steven J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Macwhinney, B. (2017). Are these approaches incompatible ? *Linguistic Approaches to Bilingualism*, *6*, 730–733. https://doi.org/10.1075/lab.00014.mac

MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

MacWhinney, B. (2022a). The Competition Model: Past and Future. In B. MacWhinney (Ed.), *Mechanisms of Language Acquisition* (pp. 3–16). https://doi.org/10.1007/978-3-030-66175-5_1

MacWhinney, B. (2022b). The Competition Model: Past and Future. In J. Gervain, G. Csibra, & K. Kovács (Eds.), *A Life in Cognition* (pp. 3–16). https://doi.org/10.1007/978-3-030-66175-5_1

MacWhinney, B., Bates, E., & Kliegl, R. (1984). Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 127–150. https://doi.org/10.1016/S0022-5371(84)90093-8

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Nickels, S., & Steinhauer, K. (2018). Prosody–syntax integration in a second language: Contrasting event-related potentials from German and Chinese learners of English using linear mixed effect models. *Second Language Research*, *34*(1), 9–37. https://doi.org/10.1177/0267658316649998

Pinker, S. (1994). *The Language Instinct*. New York: Harper Perennial Modern Classics.

Polka, L., Rvachew, S., & Mattock, K. (n.d.). Experiential Influences on Speech Perception and Speech Production in Infancy. In *Blackwell Handbook of Language Development* (pp. 153–172). https://doi.org/10.1002/9780470757833.ch8

Pu, H., Medina, Y. E., J. Holcomb, P., & J. Midgley, K. (2019). Testing for Nonselective Bilingual Lexical Access Using L1 Attrited Bilinguals. *Brain Sciences*, *9*(6), 126. https://doi.org/10.3390/brainsci9060126

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, *14*(11). https://doi.org/10.1111/phc3.12633

Rouder, J. N., & Haaf, J. M. (2021). Are There Reliable Qualitative Individual Difference in Cognition? *Journal of Cognition*, *4*(1). https://doi.org/10.5334/joc.131

Sassenhagen, J. (2019). How to analyse electrophysiological responses to naturalistic language with time-resolved multiple regression. *Language, Cognition and Neuroscience*, *34*(4), 474–490. https://doi.org/10.1080/23273798.2018.1502458

Schacht, A., Sommer, W., Shmuilovich, O., Martínez, P. C., & Martín-Loeches, M. (2014). Differential task effects on N400 and P600 elicited by semantic and syntactic violations. *PLoS ONE*, *9*(3), 1–7. https://doi.org/10.1371/journal.pone.0091226

Schmid, M. S. (2002). *First Language Attrition, Use and Maintenance*. https://doi.org/10.1075/sibil.24

Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637–667. https://doi.org/10.1075/lab.17058.sch

Seyednozadi, Z., Pishghadam, R., & Pishghadam, M. (2021). Functional Role of the N400 and the P600 in Language-related ERP Studies with Respect to Semantic Processing: An Overview. *Archives of Neuropsychiatry*. https://doi.org/10.29399/npa.27422

Shih, W., & Chai, S. (2016). Data-Driven vs. Hypothesis-Driven Research: Making sense of big data. *Academy of Management Proceedings*, *2016*(1), 14843. https://doi.org/10.5465/ambpp.2016.14843abstract

Sorace, A. (2011). Pinning down the concept of “interface” in bilingualism. *Linguistic Approaches to Bilingualism*, *1*(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Šoškić, A., Jovanović, V., Styles, S. J., Kappenman, E. S., & Ković, V. (2022). How to do Better N400 Studies: Reproducibility, Consistency and Adherence to Research Standards in the Existing Literature. *Neuropsychology Review, 32*(3), 577–600. https://doi.org/10.1007/s11065-021-09513-4

Steinhauer, K. (2014). Event-related Potentials (ERPs) in Second Language Research: A Brief Introduction to the Technique, a Selected Review, and an Invitation to Reconsider Critical Periods in L2. *Applied Linguistics*, *35*(4), 393–417. https://doi.org/10.1093/applin/amu028

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162. https://doi.org/10.1016/j.bandl.2011.07.001

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Stiles, J. (2011). *Brain development and the nature versus nurture debate*. https://doi.org/10.1016/B978-0-444-53884-0.00015-4

Su, I. R. (2001). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied Psycholinguistics*, *22*(1), 83–112. https://doi.org/10.1017/S0142716401001059

Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, *15*(3), e2000797. https://doi.org/10.1371/journal.pbio.2000797

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism*, *17*(2), 277–293. https://doi.org/10.1017/S1366728913000370

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism*, *16*(2), 367–382. https://doi.org/10.1017/S1366728912000302

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*(1), 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Tremblay, A. (2005). Word Order in Mandarin Chinese and Grammatical Relations. *Proceedings of the 19th Pacific Asia Conference on Language, Information and Computation*, (1976), 333–340. Retrieved from http://aclweb.org/anthology/Y05-1033

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49–60. https://doi.org/10.1037/h0044616

Vaughan-Evans, A., Kuipers, J. R., Thierry, G., & Jones, M. W. (2014). Anomalous transfer of syntax between languages. *Journal of Neuroscience*, *34*(24), 8333–8335. https://doi.org/10.1523/JNEUROSCI.0665-14.2014

Villeneuve, D. (2016). *Arrival*. Paramount Pictures.

Voeten, C. C. (2021). *buildmer: Stepwise Elimination and Term Reordering for Mixed-Effects Regression*. Retrieved from https://cran.r-project.org/package=buildmer

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *WIREs Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Yoshihara, M., & Yoshihara, M. (2018). 'Necessary and sufficient' in biology is not necessarily necessary – confusions and erroneous conclusions resulting from misapplied logic in the field of biology, especially neuroscience. *Journal of Neurogenetics*, *32*(2), 53–64. https://doi.org/10.1080/01677063.2018.1468443

Stimuli for Manuscript I

Reversibility	Agent Animacy	Item Number	N1	N1_Eng	N2	N2_Eng	Verb	Verb_Eng
reversible	animate	101	表兄	cousin	姐妹	sisters	赶上	catch up
reversible	animate	102	乌龟	tortoise	兔子	rabbit	赶上	catch up
reversible	animate	103	小狗	puppy	公鸡	rooster	追上	catch up
reversible	animate	104	大象	elephant	猴子	monkey	追上	catch up
reversible	animate	105	店主	shopowner	员工	staff	夸奖	to praise
reversible	animate	106	股东	shareholder	总裁	chairman	夸奖	to praise
reversible	animate	107	新人	newcomer	领导	leader	拥抱	embrace
reversible	animate	108	教授	professor	学生	student	拥抱	embrace
reversible	animate	109	岳母	mother in law	前夫	ex-husband	欺骗	rag
reversible	animate	110	特务	spy	国王	king	欺骗	rag
reversible	animate	111	委员	council member	部长	department head	表扬	to praise
reversible	animate	112	同学	classmate	闺蜜	female best friend	表扬	to praise
reversible	animate	113	法官	judge	民众	people	抱怨	complain
reversible	animate	114	儿子	son	父母	parents	抱怨	complain
reversible	animate	115	猎人	hunter	狮子	lion	扑倒	throw down
reversible	animate	116	保安	security guard	疯子	madman	扑倒	throw down
reversible	animate	117	臭虫	bed bugs	蚂蚁	ant	攻击	attack
reversible	animate	118	鲨鱼	shark	虎鲸	killer whale	攻击	attack
reversible	animate	119	犀牛	rhinoceros	老虎	tiger	威胁	to threaten
reversible	animate	120	小偷	thief	侦探	detective	威胁	to threaten
reversible	animate	121	记者	reporter	编辑	editor	批评	criticize
reversible	animate	122	首长	senior official	教官	instructor	批评	criticize
reversible	animate	123	选手	player	队友	teammate	鼓励	encourage
reversible	animate	124	书记	secretary	党员	party member	鼓励	encourage
reversible	animate	125	仙女	fairy	巫师	wizard	催眠	hypnotize
reversible	animate	126	和尚	monk	怪物	monster	催眠	hypnotize
reversible	animate	127	外公	grandfather	女婿	son in law	原谅	forgive
reversible	animate	128	经理	manager	董事	board member	原谅	forgive
reversible	animate	129	乞丐	beggar	富人	rich person	戏弄	make fun of
reversible	animate	130	市长	mayor	选民	voter	戏弄	make fun of
reversible	animate	131	士卒	soldier	猎犬	hound	拖走	tow away
reversible	animate	132	母狼	rag	狐狸	fox	拖走	tow away

reversible	animate	133	狸猫	leopard cat	棕熊	rag	唬住	to scare
reversible	animate	134	战士	warrior	营长	commander	唬住	to scare
reversible	animate	135	劫匪	bandit	强盗	robber	绑架	kidnap
reversible	animate	136	恶鬼	evil spirit	怪兽	monster	绑架	kidnap
reversible	animate	137	院长	dean	医者	doctor	冤枉	to treat unjustly
reversible	animate	138	舞者	dancer	歌手	rag	冤枉	to treat unjustly
reversible	animate	139	证人	witness	被告	defendant	举报	to denounce
reversible	animate	140	老板	boss	技工	mechanic	举报	to denounce
reversible	animate	141	共犯	accomplice	杀手	rag	抛下	throw down
reversible	animate	142	猎狗	hound	主人	host	抛下	throw down
reversible	animate	143	木工	rag	石匠	mason	起诉	prosecute
reversible	animate	144	乘客	passenger	船员	sailor	起诉	prosecute
reversible	animate	145	律师	lawyer	专家	expert	拉黑	to blacklist
reversible	animate	146	科长	section chief	主任	director	拉黑	to blacklist
reversible	animate	147	丈夫	husband	妻子	wife	宠坏	to spoil (e.g. a child)
reversible	animate	148	儿媳	daughter-in-law	外婆	grandmother	宠坏	to spoil (e.g. a child)
reversible	animate	149	房东	landlord	访客	visitor	打断	interrupt
reversible	animate	150	班长	class monitor	学委	school committee member	打断	interrupt
reversible	animate	151	门卫	gatekeeper	居民	resident	监视	to monitor
reversible	animate	152	特工	special agent	敌人	enemy	监视	to monitor
reversible	animate	153	新娘	bride	表妹	cousin	说明	explain
reversible	animate	154	省长	governor	百姓	common people	说明	explain
reversible	animate	155	球迷	fan	教练	coach	赶走	drive away
reversible	animate	156	暴民	rag	议员	senator	赶走	drive away
reversible	animate	157	岳父	father in law	新郎	groom	杀害	rag
reversible	animate	158	外宾	foreign guest	大使	ambassador	杀害	rag
reversible	animate	159	会计	accountant	秘书	secretary	通知	notify
reversible	animate	160	家属	family member	法医	forensic investigator	通知	notify
reversible	animate	161	作家	writer	诗人	poet	错过	miss (e.g. opportunity)
reversible	animate	162	处长	section chief	黑客	hacker	错过	miss (e.g. opportunity)
irreversible	animate	201	皇子	prince	王位	title of king	继承	inherit
irreversible	animate	202	兄长	brother	手法	technique	继承	inherit
irreversible	animate	203	公主	princess	宫殿	palace	展示	to display
irreversible	animate	204	花匠	gardener	成果	result	展示	to display

irreversible	animate	205	兽医	veterinarian	感染	infection	治疗	to treat (illness)
irreversible	animate	206	军医	military doctor	疟疾	malaria	治疗	to treat (illness)
irreversible	animate	207	商人	businessman	戏台	stage	搭起	to put up
irreversible	animate	208	导游	tourist guide	帐篷	tent	搭起	to put up
irreversible	animate	209	牙医	dentist	肿块	tumor	切除	excise
irreversible	animate	210	护士	nurse	疤痕	scar	切除	excise
irreversible	animate	211	厨师	chef	羽毛	feather	拔掉	pluck
irreversible	animate	212	首领	leader	插头	plug	拔掉	pluck
irreversible	animate	213	姑娘	girl	音乐	music	欣赏	rag
irreversible	animate	214	游客	tourist	美景	scenery	欣赏	rag
irreversible	animate	215	干事	administrator	快递	delivery	分类	classify
irreversible	animate	216	馆长	building head	植物	rag	分类	classify
irreversible	animate	217	导演	director	剧情	plot	讲解	explain
irreversible	animate	218	大师	grandmaster	系统	system	讲解	explain
irreversible	animate	219	干部	official	条件	rag	提高	improve
irreversible	animate	220	原告	plaintiff	错误	error	提高	improve
irreversible	animate	221	技师	technician	皮卡	pickup truck	修理	repair
irreversible	animate	222	老公	husband	水管	water pipe	修理	repair
irreversible	animate	223	歹徒	rag	公物	public property	毁坏	rag
irreversible	animate	224	演员	actor	道具	prop (theater)	毁坏	rag
irreversible	animate	225	水手	sailor	岛屿	islands	轰炸	to bomb
irreversible	animate	226	海盗	pirate	港口	port	轰炸	to bomb
irreversible	animate	227	总统	president	真相	the actual facts	忽视	neglect
irreversible	animate	228	大爷	uncle	闹钟	alarm clock	忽视	neglect
irreversible	animate	229	男士	man	裤子	pants	剪短	to cut short (e.g. hair)
irreversible	animate	230	大妈	aunt	头发	hair	剪短	to cut short (e.g. hair)
irreversible	animate	231	观众	audience	感想	impression	交流	communicate with
irreversible	animate	232	空姐	flight attendant	航线	air route	交流	communicate with
irreversible	animate	233	官员	official	腐败	corruption	否认	deny
irreversible	animate	234	球员	ballplayer	传闻	rumor	否认	deny
irreversible	animate	235	表哥	cousin	念头	idea	放弃	give up
irreversible	animate	236	矿工	miner	梦想	rag	放弃	give up
irreversible	animate	237	助手	assistant	简历	resume	制作	manufacture
irreversible	animate	238	裁缝	tailor	西装	suit	制作	manufacture

irreversible	animate	239	勇士	warrior	奖品	prize	拿到	to get
irreversible	animate	240	助教	teaching assistant	职称	staff	拿到	to get
irreversible	animate	241	高管	executive	行情	market price	讨论	discuss
irreversible	animate	242	州长	governor	政治	politics	讨论	discuss
irreversible	animate	243	讲师	lecturer	数据	data	捏造	to make up (e.g. a story)
irreversible	animate	244	人们	people	谣言	rumor	捏造	to make up (e.g. a story)
irreversible	animate	245	司机	driver	汽油	gasoline	耗尽	to deplete
irreversible	animate	246	武士	rag	体力	physical strength	耗尽	to deplete
irreversible	animate	247	大叔	uncle	钥匙	key	丢失	to lose
irreversible	animate	248	旅客	traveler	护照	passport	丢失	to lose
irreversible	animate	249	哨兵	sentry	路牌	street sign	修改	to alter
irreversible	animate	250	职员	staff	谎言	lie	修改	to alter
irreversible	animate	251	祖母	grandmother	记忆	memory	保存	to keep
irreversible	animate	252	母亲	rag	照片	photo	保存	to keep
irreversible	animate	253	孙子	grandson	食物	food	浪费	to waste
irreversible	animate	254	少年	juvenile	青春	youth	浪费	to waste
irreversible	animate	255	乐师	musician	路费	toll	节省	to save
irreversible	animate	256	将军	general	精力	energy	节省	to save
irreversible	animate	257	大蛇	big snake	树枝	rag	环绕	to surround
irreversible	animate	258	昆虫	rag	大树	big tree	环绕	to surround
irreversible	animate	259	富豪	rich and powerful person	财产	property	捐赠	to contribute
irreversible	animate	260	农民	farmer	宝石	gem	捐赠	to contribute
irreversible	animate	261	大姐	big sister	皮鞋	leather shoes	擦亮	to polish
irreversible	animate	262	仆人	servant	镜子	mirror	擦亮	to polish
irreversible	inanimate	301	疗法	treatment	酋长	chief	治愈	cure
irreversible	inanimate	302	草药	herbal medicine	村民	rag	治愈	cure
irreversible	inanimate	303	疾病	disease	老人	old man	传染	infect
irreversible	inanimate	304	流感	influenza	客户	client	传染	infect
irreversible	inanimate	305	审判	trial	大臣	chancellor	革职	dismissal (from job)
irreversible	inanimate	306	条约	rag	首相	prime minister	革职	dismissal (from job)
irreversible	inanimate	307	子弹	bullet	孔雀	peacock	打中	hit
irreversible	inanimate	308	砖头	brick	恶棍	villain	打中	hit
irreversible	inanimate	309	声音	voice	小孩	child	吓到	scared
irreversible	inanimate	310	传说	legend	皇后	empress	吓到	scared

irreversible	inanimate	311	故障	malfunction	电工	electrician	难住	to stump
irreversible	inanimate	312	数学	mathematics	老师	teacher	难住	to stump
irreversible	inanimate	313	毯子	blanket	小鹿	fawn	裹住	to wrap
irreversible	inanimate	314	浴巾	bath towel	婴儿	infant	裹住	to wrap
irreversible	inanimate	315	针剂	injection	病人	patient	麻醉	anesthetize
irreversible	inanimate	316	药物	medication	患者	patient	麻醉	anesthetize
irreversible	inanimate	317	歌声	rag	导游	tourist guide	安慰	comfort
irreversible	inanimate	318	鲜花	fresh flowers	助理	assistant	安慰	comfort
irreversible	inanimate	319	法术	magic spell	骑士	knight	变丑	disfigure
irreversible	inanimate	320	巫术	witchcraft	女王	queen	变丑	disfigure
irreversible	inanimate	321	难题	problem	外行	layman	考住	test
irreversible	inanimate	322	围棋	Wei Qi (go)	老外	foreigner	考住	test
irreversible	inanimate	323	手铐	handcuffs	骗子	conman	锁住	to lock
irreversible	inanimate	324	铁链	chain	叛徒	traitor	锁住	to lock
irreversible	inanimate	325	箱子	box	老鼠	mouse	困住	to trap
irreversible	inanimate	326	鸟笼	birdcage	喜鹊	magpie	困住	to trap
irreversible	inanimate	327	寒流	cold stream	难民	refugee	冻死	freeze to death
irreversible	inanimate	328	冬天	winter	蝗虫	locust	冻死	freeze to death
irreversible	inanimate	329	麦子	wheat	家畜	livestock	喂饱	to give food
irreversible	inanimate	330	鸟食	bird food	鸽子	pigeon	喂饱	to give food
irreversible	inanimate	331	灯光	light	猫咪	cat	吓跑	scare away
irreversible	inanimate	332	火光	flame	野狼	wild wolf	吓跑	scare away
irreversible	inanimate	333	寺庙	temple	高僧	senior monk	请来	invite
irreversible	inanimate	334	大学	university	院士	scholar	请来	invite
irreversible	inanimate	335	国学	national studies	孩童	child	启发	enlighten
irreversible	inanimate	336	历史	history	子孙	offspring	启发	enlighten
irreversible	inanimate	337	电梯	elevator	伯母	aunt	送来	send
irreversible	inanimate	338	卡车	truck	山羊	goat	送来	send
irreversible	inanimate	339	咒语	curse	妖怪	monster	唤醒	to wake up (somebody)
irreversible	inanimate	340	号角	horn (bugle)	士兵	soldier	唤醒	to wake up (somebody)
irreversible	inanimate	341	阳光	sunlight	男孩	boy	温暖	to warm
irreversible	inanimate	342	政策	policy	军人	soldier	温暖	to warm
irreversible	inanimate	343	军刀	saber	对手	opponent	刺死	stab
irreversible	inanimate	344	匕首	dagger	间谍	spy	刺死	stab

irreversible	inanimate	345	雷达	radar	野兽	beast	捕捉	capture
irreversible	inanimate	346	陷阱	trap	刺猬	hedgehog	捕捉	capture
irreversible	inanimate	347	玩具	toy	盲人	blind person	绊倒	to trip
irreversible	inanimate	348	台阶	step	行人	pedestrian	绊倒	to trip
irreversible	inanimate	349	祭祀	sacrifice	鬼魂	ghost	招来	attract
irreversible	inanimate	350	馒头	steamed bread	苍蝇	fly	招来	attract
irreversible	inanimate	351	地图	map	船长	boat captain	误导	mislead
irreversible	inanimate	352	传统	tradition	后人	posterity	误导	mislead
irreversible	inanimate	353	热汤	hot soup	保姆	babysitter	烫伤	ignite
irreversible	inanimate	354	开水	boiling water	明星	rag	烫伤	ignite
irreversible	inanimate	355	文章	article	读者	reader	鼓舞	boost morale
irreversible	inanimate	356	昆曲	rag	听众	audience	鼓舞	boost morale
irreversible	inanimate	357	圣旨	imperial edict	罪犯	criminal	赦免	to pardon
irreversible	inanimate	358	法庭	court	囚犯	convict	赦免	to pardon
irreversible	inanimate	359	大风	gale	人群	crowd	吹倒	blow down
irreversible	inanimate	360	台风	typhoon	海鸥	seagull	吹倒	blow down
reversible	inanimate	401	石头	stone	树根	rag	包围	surround
reversible	inanimate	402	房屋	house	树林	rag	包围	surround
reversible	inanimate	403	盒子	box	礼品	gift	压碎	crush
reversible	inanimate	404	吉他	guitar	竖琴	harp	压碎	crush
reversible	inanimate	405	会议	meeting	活动	activity	改变	change
reversible	inanimate	406	企业	enterprise	社区	community	改变	change
reversible	inanimate	407	电脑	computer	程序	computer program	启动	start up (e.g. computer)
reversible	inanimate	408	暖气	rag	热水	hot water	启动	start up (e.g. computer)
reversible	inanimate	409	软件	software	网站	website	更新	rag
reversible	inanimate	410	调查	survey	项目	project	更新	rag
reversible	inanimate	411	矛盾	contradiction	感情	feeling	引发	trigger
reversible	inanimate	412	士气	morale	自信	confidence	引发	trigger
reversible	inanimate	413	刷子	brush	抹布	rag	清理	clean up
reversible	inanimate	414	喷头	nozzle	抹布	rag	清理	clean up
reversible	inanimate	415	游艇	yacht	轮船	steamship	指引	to guide
reversible	inanimate	416	贸易	trade	建设	construction	指引	to guide
reversible	inanimate	417	拖轮	tugboat	小船	boat	牵引	to pull
reversible	inanimate	418	板车	rag	摩托	motorbike	牵引	to pull

reversible	inanimate	419	材料	rag	液体	liquid	污染	pollute
reversible	inanimate	420	泥土	soil	沙子	sand	污染	pollute
reversible	inanimate	421	火柴	matches	壁炉	fireplace	点着	ignite
reversible	inanimate	422	雪茄	cigar	香烟	cigarette	点着	ignite
reversible	inanimate	423	火车	train	坦克	tank	碰撞	collide
reversible	inanimate	424	卫星	satellite	火箭	rocket	碰撞	collide
reversible	inanimate	425	婚礼	wedding	贷款	loan	推迟	put off
reversible	inanimate	426	蜜月	honeymoon	怀孕	pregnancy	推迟	put off
reversible	inanimate	427	合同	contract	协议	agreement	换新	upgrade
reversible	inanimate	428	时事	rag	广播	broadcast	换新	upgrade
reversible	inanimate	429	渔网	fishing net	海带	seaweed	缠住	tangle up
reversible	inanimate	430	围巾	scarf	衣袖	sleeve	缠住	tangle up
reversible	inanimate	431	经济	economy	消费	consumption	预测	forecast
reversible	inanimate	432	温度	temperature	气压	air pressure	预测	forecast
reversible	inanimate	433	婚姻	marriage	爱情	love	拯救	to rescue
reversible	inanimate	434	舆论	public opinion	案件	rag	拯救	to rescue
reversible	inanimate	435	交流	communicate with	和平	peace	促进	promote (a cause)
reversible	inanimate	436	画室	studio	投资	investment	促进	promote (a cause)
reversible	inanimate	437	工资	wages	奖金	bonus	包含	contain
reversible	inanimate	438	展览	exhibition	晚宴	rag	包含	contain
reversible	inanimate	439	动画	cartoon	商品	commodity	宣传	to give publicity to
reversible	inanimate	440	广告	ad	寺院	temple	宣传	to give publicity to
reversible	inanimate	441	海啸	tsunami	地震	earthquake	触发	to spark
reversible	inanimate	442	想法	idea	计划	plan	触发	to spark
reversible	inanimate	443	灵感	inspiration	直觉	intuition	引起	give rise to
reversible	inanimate	444	食欲	appetite	胃口	appetite	引起	give rise to
reversible	inanimate	445	环境	surroundings	气氛	atmosphere	破坏	to damage
reversible	inanimate	446	箭头	arrow	飞镖	dart	破坏	to damage
reversible	inanimate	447	桌子	rag	椅子	rag	支撑	to prop up
reversible	inanimate	448	理论	theory	实践	practice	支撑	to prop up
reversible	inanimate	449	食堂	canteen	卫生	health	改进	improve
reversible	inanimate	450	结果	result	论文	paper	改进	improve
reversible	inanimate	451	体重	body weight	饮食	diet	减少	reduce
reversible	inanimate	452	抗议	protest	捐款	donation	减少	reduce

reversible	inanimate	453	期刊	rag	博客	blog	抄袭	plagiarize
reversible	inanimate	454	杂志	rag	报纸	newspaper	抄袭	plagiarize
reversible	inanimate	455	文明	civilization	城市	city	创造	create
reversible	inanimate	456	法律	law	权利	rag	创造	create
reversible	inanimate	457	课程	course	实习	internship	包括	include
reversible	inanimate	458	游行	march	典礼	ceremony	包括	include
reversible	inanimate	459	合作	cooperation	关系	relationship	强化	strengthen
reversible	inanimate	460	纲领	program	制度	system	强化	strengthen

Stimuli for Manuscript II

Item number	Subject	Adverb	Verb	Adjective	Noun	Sentence final particle
1	小明	现在	住在	老	区	了。
	Xiaoming	now	lives in	old	district	LE
2	大卫	昨天	修缮	旧	房	了。
	Dawei	yesterday	repaired	old	house	LE
3	舅舅	终于	做完	累	活	了。
	Uncle	finally	finished	tiring	work	LE
4	树林	果然	出现	红	鸟	了。
	Forest	as expected	appeared	red	bird	LE
5	短信	里面	出现	别	字	了。
	Text message	inside	appeared	wrong	character	LE
6	小妹	已经	加上	辣	油	了。
	Little sister	already	added	spicy	oil	LE
7	表姐	竟然	买到	假	货	了。
	Cousin	unexpectedly	bought	fake	product	LE
8	奶奶	早上	去买	糙	米	了。
	Grandma	morning	goes to buy	coarse	rice	LE
9	小花	上午	去烫	直	发	了。
	Xiaohua	morning	goes to iron	straight	hair	LE
10	姐姐	上午	去过	深	山	了。
	Older sister	morning	went	deep	mountains	LE
11	母亲	去年	逛过	北	欧	了。
	Mother	last year	visited	north	Europe	LE
12	小红	突然	发现	暗	门	了。
	Xiaohong	suddenly	discovered	secret	door	LE
13	海岛	上面	发现	野	鸟	了。
	Island	on	discovered	wild	birds	LE
14	小偷	终于	遭受	重	刑	了。
	Thief	finally	suffered	serious	punishment	LE
15	楼下	社区	变成	闹	市	了。
	Downstairs	community	became	noisy	city	LE
16	某人	刚刚	吃掉	甜	桃	了。
	Certain people	just	ate up	sweet	peach	LE

17 她们	已经	吃过	晚	餐	了。
They	already	ate	late	meal	LE
18 青蛙	刚才	吞掉	活	虫	了。
Frog	just	swallowed up	live	insect	LE
19 玛丽	夏天	套上	短	袖	了。
Mary	summer	wore	short	sleeve	LE
20 婶婶	果然	嫁给	暖	男	了。
Aunt	as expected	married	warm	boy	LE
21 姑姑	最后	学做	素	汤	了。
Aunt	in the end	learned to make	plain	soup	LE
22 春天	已经	带来	嫩	叶	了。
Spring	already	brought	newly-grown	leaves	LE
23 西北	已经	出现	旱	地	了。
Northwest	already	appeared	dry	earth	LE
24 老人	居然	想喝	苦	茶	了。
Old people	unexpectedly	want to drink	bitter	tea	LE
25 顾客	最后	想点	凉	菜	了。
Customer	in the end	wants to order	cold	dish	LE
26 父亲	最近	想看	新	书	了。
Father	recently	wants to read	new	book	LE
27 旅客	周末	想看	高	楼	了。
Traveler	weekend	wants to look at	tall	buildings	LE
28 小龙	刚才	戳到	右	眼	了。
Xiaolong	just	poked	right	eye	LE
29 小胖	终于	找到	西	街	了。
Xiaopang	finally	found	west	street	LE
30 阿姨	终于	找到	白	狗	了。
Auntie	finally	found	white	dog	LE
31 警察	昨天	抓到	笨	贼	了。
Police	yesterday	arrested	stupid	crook	LE
32 市民	又在	抱怨	浓	雾	了。
Citizen	again	complains	thick	fog	LE
33 小刚	可以	接受	低	价	了。
Xiaogang	can	receive	low	price	LE

34	姑妈	已经	收养	黑	猫	了。
	Aunt	already	took in	black	cat	LE
35	老外	竟然	敢吃	肥	肠	了。
	Foreigner	unexpectedly	dared to eat	fatty	intestine	LE
36	世界	已经	没有	好	人	了。
	World	already	doesn't have	good	people	LE
37	森林	现在	没有	小	兔	了。
	Forest	now	doesn't have	little	rabbits	LE
38	小王	早上	泼掉	冷	水	了。
	Xiaowang	morning	spilled	cold	water	LE
39	厨师	竟然	烫伤	左	手	了。
	Chef	unexpectedly	burned	left	hand	LE
40	爷爷	现在	爱吃	瘦	肉	了。
	Grandpa	now	loves to eat	lean	meat	LE
41	人们	有幸	看见	紫	霞	了。
	People	luckily	saw	purple	clouds	LE
42	罪犯	刚才	破坏	公	物	了。
	Criminal	just	broke	public	property	LE
43	孩子	早上	穿上	长	裤	了。
	Child	morning	wore	long	sleeves	LE
44	汽车	现在	等待	绿	灯	了。
	Car	now	waiting	green	light	LE
45	白醋	当然	算是	弱	酸	了。
	White vinegar	of course	counts as	weak	acid	LE
46	老人	已经	能走	平	路	了。
	Old people	already	can walk	flat	road	LE
47	螃蟹	逐渐	蜕掉	硬	壳	了。
	Crabs	gradually	shed	hard	shell	LE
48	小张	已经	蠢成	笨	猪	了。
	Xiaozhang	already	stupidly became	dumb	pig	LE
49	同桌	现在	说出	真	话	了。
	Deskmate	now	speaks	true	words	LE
50	朋友	晚上	走进	窄	巷	了。
	Friends	evening	walk into	narrow	alley	LE

51	他们	突然	越过	南	墙	了。
	They	suddenly	crossed	south	wall	LE
52	老师	马上	追上	全	班	了。
	Teacher	at once	caught up to	full	class	LE
53	男生	已经	采到	香	花	了。
	Boy	already	stepped on	fragrant	flowers	LE
54	爸爸	路上	闻到	臭	味	了。
	Father	on the way	smelled	smelly	smell	LE
55	美玲	最后	看到	远	景	了。
	Meiling	in the end	saw	distant	scenery	LE
56	弟弟	终于	买到	薄	被	了。
	Brother	finally	bought	thin	blanket	LE
57	小芬	今晚	想吃	细	面	了。
	Xiaofen	tonight	wants to eat	thin	noodles	LE
58	小西	终于	放弃	软	床	了。
	Xiaoxi	finally	gave up	soft	bed	LE
59	银行	最后	清理	坏	账	了。
	Bank	in the end	cleaned	bad	account	LE
60	马丁	果然	走过	近	路	了。
	Martin	as expected	walked	nearby	road	LE