Quantifier Scope in L2 Learners: Interpretation, Processing, and Acquisition

by

Shaohua Fang

Bachelor of Arts, Ningbo University, 2014

Master of Arts, Shanghai Jiao Tong University, 2017

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis or dissertation was presented

by

Shaohua Fang

It was defended on

June 20, 2023

and approved by

Melinda Fricke, Assistant Professor, Department of Linguistics, University of Pittsburgh Tessa

Warren, Professor, Department of Psychology, University of Pittsburgh

Yi Xu, Associate Professor, Department of East Asian Languages and Literatures, University of Pittsburgh

Dissertation Director: Alan Juffs, Professor, Department of Linguistics, University of Pittsburgh

Quantifier Scope in L2 Learners: Interpretation, Processing, and Acquisition

Shaohua Fang, PhD

University of Pittsburgh, 2023

Scope ambiguity arises when a sentence contains scope-bearing logical operators like quantifiers and negation (Fox, 2003). For instance, in English, the sentence ‘Every horse didn’t jump over the fence’ can be interpreted as either ‘None of the horses jumped over the fence’ (surface scope) or ‘Not all of the horses jumped over the fence’ (inverse scope). Chinese exhibits a scope-rigid nature, arguably permitting only surface scope readings (Aoun & Li, 1989; Huang, 1982).

Furthermore, English-speaking learners of Chinese would presumably encounter the well-known poverty-of-the-stimulus (POS) problem (Pearl, 2021; White, 2022) due to the lack of positive evidence in the L2 input. This study utilized a bidirectional design to investigate the offline interpretation and online processing of quantifier scope by L2 learners of English and Chinese. Three experiments involving doubly quantified and negatively quantified sentences were conducted, marking the first investigation of its kind.

In Experiment 1 with sentence-interpretation matching tasks, both English and Chinese speakers predominantly favored surface scope interpretations without supportive discourse contexts. Notably, English speakers leaned more towards accepting inverse scope interpretations, except for NU sentences. In Experiment 2 involving L2 learners of English and Chinese, results from truth-value judgment tasks demonstrated their target-like acquisition of quantifier scope interpretations across most structures. In Experiment 3, employing the visual-world eye-tracking paradigm with the same participants, the findings closely paralleled those of Experiment 2,

especially regarding the observed patterns in eye-fixation data. This cross-linguistic study advances empirical research on L2 learners' quantifier scope interpretation, revealing their capacity to largely surmount the POS problem. Acquisition patterns were shaped by factors including L2 (triggering) input, grammatical constraints, L1 transfer, L2 proficiency, statistical inference, and experimental setups.

Another novel discovery is that Chinese exhibits the ability to obtain inverse scope interpretations for NU sentences across experiments, challenging the long-standing scope-rigidity claim associated with this language. This compelling evidence indicates that scope rigidity in Chinese may not be universally applicable, and certain interpretations that were previously assumed to be unavailable can be facilitated by factors like discourse contexts or experimental setups. Therefore, NU sentences are not a POS problem for learners of L2 Chinese.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I can vividly envision how, five years ago, I could only dream of reaching this moment of completing my dissertation. Now, against all odds, I have made it! Completing a dissertation project truly embodies a collaborative effort, and it would have been impossible without the unwavering support and contributions of all those who have aided in its success. I am immensely grateful for the assistance and encouragement that have propelled me forward on this challenging yet rewarding journey.

To Dr. Alan Juffs, who has been my advisor and mentor during my five years as a graduate student at Pitt, I owe the greatest debt of gratitude. His unparalleled influence on my intellectual development spans every facet of my academic journey, rendering words of appreciation inadequate. Alan's unwavering support and expertise have been instrumental in shaping my trajectory, as he generously devoted his time to numerous weekly meetings, discussing diverse topics in linguistics, language acquisition, sentence processing, teaching, dissertation project, and collaborative endeavors. Beyond his role as an exceptional advisor, Alan has been an inspiring source, a compassionate mentor, a supportive collaborator, an outstanding professor, and a true friend. My graduate experience would be unimaginable without his guidance, and I am forever grateful for the indelible mark he has left on my life. I wholeheartedly aspire to emulate his exemplary model should I have the privilege of advising graduate students in the future.

I extend my heartfelt gratitude to the other members of my committee: Dr. Melinda Fricke, for patiently listening to the setup of the eye-tracking experiment and providing valuable guidance on its design; Dr. Yi Xu, whose Chinese as a Second Language course sparked my interest in Chinese linguistics and second language acquisition of Chinese. This led to a project on the acquisition and processing of temporality by L2 learners of Chinese; Dr. Tessa Warren, whose psycholinguistics course (twice!) proved immensely helpful in developing a deeper understanding of motivations, methodologies, and other critical aspects of psycholinguistics.

I would like to express my sincere gratitude to faculty members at Shanghai Jiao Tong University, with a special acknowledgment to Dr. Hui Chang, who played a pivotal role in shaping my academic journey during my master's program in China. Under Dr. Chang's supervision, I delved into the fascinating realms of formal linguistics and second language acquisition, which ignited a passion for linguistic research that continues to drive me today. His mentorship and belief in my potential inspired me to pursue further studies in the United States. I am also grateful to the faculty members at Stony Brook University for their support and encouragement during my first year in the US, especially Dr. Mark Aronoff, Dr. John Bailyn, Dr. Ellen Broselow, Dr. Francisco Ordóñez, Dr. Lori Repetti, and Dr. Jiwon Yun.

I am eager to express my heartfelt gratitude to the numerous individuals who have offered their invaluable help and support throughout my journey in Pittsburgh and beyond, whether in person or virtually. My sincere thanks go to Sean Nonnenmacher, Juan Berrios, Joseph Patrick, Miroo Lee, Angela Krak, Ben Naismith, Enas Albasiri, Zhiyi Wu, Fei Yuan, Hui Zhang, Pengliang He, Han Li, Hongchen Wu, Keshu Xiang, Hanzhong Sun, Jia Wang, and Kaiyan Song for their assistance and encouragement. Furthermore, I extend my appreciation to Dr. Jon Sprouse for engaging discussions on acceptability judgment tasks, to Dr. Foong Ha Yap for enlightening discussions about temporality in Chinese and various other languages, to Dr. Brian MacWhinney for sharing interesting perspectives on quantifier scope, and to Dr. Lin Fan for providing warm encouragement during my time as an undergraduate first exploring the fascinating world of linguistics.

I am deeply grateful to all those who contributed to the success of my dissertation experiments. First and foremost, my heartfelt thanks go to the (hundreds of) research participants who generously gave their time and effort, both on-site and remotely, making this study possible. Emily Martin and Qing Ma's invaluable help in constructing experimental materials, along with Pengliang He and Lei Zhang's artistic talent in creating eye-catching pictures for the eye-tracking experiment, added depth to the study. The English stimuli recording by Mary Dryer and the Chinese stimuli recording by Qing Ma further enriched the research process. Jiani He, Hanzhong Sun, Foong Ha Yap, and Lisze Siaw provided crucial support in recruiting and testing participants in China, while Yi Xu, Gang Liu, Yunwen Su, and many others collaborated in recruiting participants learning Chinese as a foreign language in the US. I am also hugely appreciative of Pengliang He's invaluable assistance in compiling and processing eye-tracking data using computational tools through the development of Python scripts. Special thanks to Tingting Wang for insightful discussions on the eye-tracking experiment setup with Gorilla, as well as to Dr. Scott Fraundorf and Junyi Bai for their invaluable assistance with statistics.

Thanks also go to the financial support that made my dissertation project possible: the Andrew Mellon Predoctoral Fellowship and the Language Learning Dissertation Grant.

Finally, I reserve my most heartfelt thanks for the ones who have been my unwavering pillars of support throughout this journey. To my parents, grandparents, and sister, I am forever grateful for their boundless love and unwavering encouragement in pursuing my passion. This dissertation is dedicated to them, as their constant belief in me has been the driving force behind my achievements.

1.0 Introduction

The nature and emergence of linguistic competence by human speakers has long been a topic of inquiry for language researchers and learning scientists. Chomsky (2005) identified three essential factors that may be responsible for the characterization and emergence of human language.

1: Genetic endowment (Universal Grammar)

2: Experience

3: Principles not specific to the faculty of language

Universal Grammar (UG) refers to cognitive structures that limit the hypothesis space for language learning. The role of UG in first language (L1) acquisition has received substantial empirical support (e.g., Crain, 2008; Crain et al., 2021; Pagliarini et al., 2021). Experience broadly refers to the extent to which a language user has exposure to the language in question. According to some language acquisition models such as the ‘usage-based’ model (Tomasello, 2009), language experience has been argued to play a paramount role in shaping linguistic development. The third factor concerns resources afforded by domain general cognition (e.g., working memory) for language processing typically associated with performance factors.

These factors and their interaction have not only informed the inquiry into first language (L1) acquisition but also into second language (L2) acquisition (SLA), for which it is crucial to examine the roles of UG, L1 knowledge, and the input in shaping the representation, acquisition, and processing of languages other than one’s mother tongue acquired in adulthood (White, 2022). While this line of research has been extensively conducted in the domain of syntax and syntax-semantics among L2 learners, including definiteness (e.g., Cho & Slabakova, 2014; Ionin et al., 2008), argument structure (e.g., Inagaki, 1997; Juffs, 1996), grammatical aspect (e.g., Gabriele, 2009; Montrul, 2010), domains pertaining to logical operations on different logical operators such as quantifiers and negation have received relatively little attention in L2 interpretation and processing. One goal of this dissertation thus is to extend such an inquiry to this domain by investigating the interpretation and processing of quantifier scope in a bidirectional study with Chinese-speaking learners of English (henceforth L2 English learners) and English-speaking learners of Chinese (henceforth L2 Chinese learners).

While interpretation and processing may be related in the sense that online processing involves assigning syntactic structure(s) to language input for arriving at the correct interpretation, these will be investigated in their own right in the present study. In SLA research, it has been acknowledged that the field needs both a property theory for *what* is being acquired (White, 1989, 2003) and a transition theory for *how* acquisition is achieved[1] (Gregg, 1996; Juffs, 2004). Discussing one without alluding to the other would not allow us to gain a complete picture of SLA. The present study uses Truth Value Judgment Tasks (Crain & Thornton, 2000) to probe the nature of learners' grammatical representations by assessing how learners bring scope interpretation to language input strings. The processing of quantifier scope, as in (1) and (2), will be examined

[1] Processing and acquisition are intricately linked (Phillips & Ehrenhofer, 2015). One way to comprehend this relationship is that processing plays a role in acquiring novel linguistic representations, in that breakdowns in processing may lead learners to reconstruct and acquire such representations. Therefore, understanding processing in L2 learners is vital for any theory of second language acquisition. In this dissertation, exploring the processing of quantifier scope interpretation enriches our understanding of how L2 learners acquire scope interpretations.

through a visual-world eye-tracking experiment, which aims to uncover the online computational aspects of quantifier scope computation in adult L2 learners.

(1) Every child climbed a tree.

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

Whereas some studies regarding the online processes underlying quantifier scope computation have been done among children (e.g., Brooks & Sekerina, 2006; Lohiniva & Panizza, 2016; Minai et al., 2012; Sekerina & Sauermann, 2017; Sekerina et al., 2018) and adult L1 speakers (e.g., Anderson, 2004; Brooks & Sekerina, 2006; Dwivedi, 2013; Filik et al., 2004; Paterson et al., 2008; Zhou & Gao, 2009), much less is known about how L2 learners compute quantifier scope in real-time. Among the very few studies on the L2 processing of quantifier scope, most of them used the self-paced reading (SPR) task (Chung & Shin, 2022; Lee, 2009), and only one, to the best of my knowledge, so far have used the eye-tracking technique with heritage bilinguals (Sekerina & Sauermann, 2015). Aside from the potential impact of the choice of the online methods (self-paced reading vs. eye-tracking) on L2 results (Juffs & Fang, 2023), these studies yielded mixed results on the possibility of arriving at native-like processing among L2 learners, presumably due to the fact that these studies differ in scope phenomena (double-quantifier vs. quantifier-negation), target populations, and languages tested.

Regardless of camps of SLA, be they generative/nativist or usage-based/emergentist approaches, when it comes to acquiring a second language, the learner must be able to discover the mapping between form (in the sense of linguistic structure) and meaning (in the sense of semantic interpretation) at different levels of representation, i.e., lexical and sentence level, for the acquisition task to succeed (DeKeyser, 2005; Slabakova, 2013). Within a language, such mappings are not always systematic and consistent in cases where the syntactic form does not map onto one unambiguous meaning. The kind of quantifier scope under investigation in this dissertation can be represented by a sentence containing a quantifier (i.e., universal quantifier ‘every’ or existential quantifier ‘a) and/or negation and often has the potential for ambiguity (Fox, 2003). For example, although one interpretation may be strongly preferred, sentences in English, as in (1), in fact have two possible interpretations: a. For each child, there was a different tree that (s)he climbed, and b. there was a single tree that was climbed by each child. Similarly, the sentence as in (2), has two possible readings: a. None of the horses jumped over the fence, b. It was not the case that all the horses jumped over the fence (some did, some did not).

These cases have clearly shown that the meaning that the interaction between different logical operators brings to the sentence does not have a one-to-one correspondence with syntax. This complex form-meaning mapping presumably would present challenges for learners. Moreover, such form-meaning mappings with respect to quantifier scope exhibit cross-linguistic variation, for example, between English and Mandarin Chinese (henceforth Chinese). In the theoretical literature (e.g., Aoun & Li, 1989; Huang, 1982), Chinese, relative to English, has been argued to be much more rigid in its ability to permit both readings for sentences with different logical operators. Specifically, Chinese only allows one reading for (1), interpreted as ‘For each child, there was a different tree that (s)he climbed’, and for (2), interpreted as ‘None of the horses jumped over the fence’. As such, these cross-linguistic differences offer an opportunity for SLA research to explore the role of first language (L1) transfer in L2 acquisition and processing of quantifier scope. For example, if L2 Chinese learners are unable to acquire the absence of the ‘some’ reading for sentence (2) due to the influence from English, they might interpret that (2) in Chinese permits the ‘some’ reading.

Moreover, such cross-linguistic variation also points toward the fact that English and Chinese are in a superset-subset relationship with respect to quantifier scope interpretation where English constitutes the superset and Chinese the subset (Crain, 2012). In the realm of semantic acquisition among children, on the assumption that more than one interpretation was logically made available in the Universal Grammar and these interpretations create subset-superset scenarios (Crain, 2012; Crain & Thornton, 2000; Shimada & Goro, 2021), the learner must decide the interpretation that his/her grammar permits. A built-in constraint called the Semantic Subset Principle has been argued to operate in navigating through different interpretive possibilities, by which leaners are guided to generalize conservatively and set up their initial hypothesis (Crain, Ni, & Conway, 1994; Musolino, 2006; though see Musolino et al., 2019). Following the Semantic Subset Principle, it would be the case that the ‘none’ reading for sentence (2) constitutes the initial hypothesis for English-speaking children (See Musolino, 2006, for the argument). In the course of language development, the initial hypothesis could be revised based on the *positive* evidence[2] such that English eventually would allow both readings.

In a similar vein, the issue of *positive* evidence pertaining to the Semantic Subset Principle would also figure prominently in L2 acquisition of quantifier scope across languages. For Chinese learners of English, their initial hypothesis about English quantifier scope would stem from their

---

[2] *Positive* evidence refers to the evidence from language input for learners to figure out that certain structures or interpretations are available in one’s grammar. In addition, the existence of the Semantic Subset Principle has also been grounded in the assumption that *negative* evidence, information about ungrammaticality, is not available during language development (Musolino, 2006).

knowledge about Chinese such that they would initially allow the interpretation permitted only in Chinese. However, they should be able to acquire the presence of the other interpretation that is absent in their L1 grammar but available in the L2 input. In detail, *positive* evidence exemplified in the L2 input provides clues for L2 English learners that the ‘some’ reading, for example, for (2) is possible.

Nevertheless, language input on SLA seems inadequate for its role in achieving full competence in the target language among L2 learners when its quantity and quality are taken into serious consideration (Carroll, 2017). L2 input regarding quantifier scope interpretation is limited because semantic phenomena including, but not restricted to, quantification, disjunction and pragmatic entailment are seldom taught in instructional settings. Even worse is that explicit instruction seems to not work in affecting L2 learners’ ability to generalize the availability of quantificational interpretation (Wu & Ionin, 2021). In addition, L1 acquisition research has provided compelling evidence that the extent to which children can exhibit their underlying linguistic competence is affected by pragmatic/contextual factors (Musolino & Lidz, 2006; Viau et al., 2010). Therefore, the input used to set possible quantifier scope interpretation appears insufficient (one example is not enough) and the knowledge to be accessed for such interpretation is highly complex and subtle. This being said, the amount of L2 input for Chinese-speaking learners of English is expected to be significantly greater than that for English-speaking learners of Chinese. This is due to the possibility of the former encountering contexts that favor inverse scope interpretations, which are possible in English, whereas the latter would not have access to evidence from the L2 input to show that inverse scope interpretations are not allowed in Chinese. Therefore, English-speaking learners of Chinese are likely to encounter the learnability problem, also known as the *poverty-of-the-stimulus* problem (Berwick et al., 2011; White, 2003). This problem pertains to how learners could successfully acquire a language for which there is degenerate or just no evidence.

However, it should be acknowledged that the extent to which Chinese disallows inverse scope interpretations remains an empirical question and may depend on sentence structure. To establish whether Chinese truly qualifies as a subset language, warranting a learnability problem for English-speaking learners of Chinese, empirical evidence is required as a baseline. As a foreshadowing of the findings, NU sentences were discovered to permit inverse scope interpretations, challenging the notion that Chinese is strictly rigid in scope interpretation, previously supported mainly by observations in doubly quantified sentences. In the context of L2 acquisition, the acquisition of Chinese NU sentences for inverse scope interpretations by English speakers would not constitute a learnability problem, as the L2 input provides opportunities to encounter positive evidence that inverse scope interpretations are possible for NU sentences in Chinese. Furthermore, in other quantifier scope structures, certain interpretations not readily available in UG may be facilitated by factors like discourse contexts or experimental setups, as emphasized by approaches that consider the contribution of various factors in quantifier scope interpretation.

Although learners may face such a problem, albeit to varying degrees, it presents different challenges for learners with different language backgrounds. Chinese learners of English may eventually be able to acquire the presence of both readings for sentences as in (1) and (2) based on the *positive* evidence, limited and subtle though it may be, in the L2 input. In contrast, English learners of Chinese might be much less successful in this regard. Consider again the example sentences in (2) (e.g., Every horse didn't jump over the fence). While English allows both the 'none' reading and 'some' reading, Chinese allows the 'none' reading only. If English learners of

Chinese transfer both interpretations from their native language into their L2 Chinese, there will be no direct *positive* evidence from the L2 input to inform them that the 'some' interpretation is disallowed. To borrow an argument from statistics, 'absence of evidence is not evidence of absence' (Altman & Bland, 1995). It is then likely that English learners of Chinese would be unable or unlikely to unlearn, or preempt (Ambridge et al. 2018; Rutherford, 1989) the 'some' reading and hence overgeneralize the English patterns in the L2. Moreover, *negative* evidence is either absent or unhelpful, even if it is available in some form or other for L2 learners (Marsden, 2009; Trahey & White, 1993), which may also lead to difficulties associated with the acquisition of the absence of the 'some' interpretation for English learners of Chinese. As such, if learners were able to acquire the quantifier scope interpretation in the absence of sufficient evidence either from their native language or L2 input, then UG, the cognitive structure shared across languages, must be playing a role in bridging the gap between the degenerated language input and the target linguistic competence developed by L2 learners.

In second language acquisition, research that provides evidence in support of UG has primarily been in the domain of formal morpho-syntax (e.g., Heil & López, 2020; Montrul & Slabakova, 2003; Rothman & Iverson, 2008; Song & Schwartz, 2009). With only a few exceptions (e.g., Dekydtspotter et al., 2001; Dekydtspotter & Sprouse, 2001; Grüter et al., 2010; Marsden, 2009), relatively little research has been done in the domain of formal semantics and its interaction with other levels of linguistic representation. Given the potential influence of these factors including L1 transfer, UG, and language input, investigating the L2 acquisition and interpretation of quantifier scope in a bidirectional study offers an opportunity to explore how UG constraints interact with L1 transfer[3] and language input to allow L2 learners to overcome the learnability problem they face especially among English learners of Chinese in these processes.

Quantifier scope interpretation may also present challenges for L2 learners, theoretically because its interpretation lies at an interface between different linguistic modules and brings together at least three levels of representation: syntax (i.e., relative c-command relation between scope bearing expressions), semantics (i.e., logical operations for scope relation), and pragmatics/discourse context (e.g., conversational inference/contextual support for semantic interpretation). According to the Interface Hypothesis (IH) posited by Sorace (2011) and Sorace and Filiaci (2006), interface phenomena are challenging due to interplay between language domains. Such interplay requires extra processing resources for information integration, and thus exhibits more difficulties for learners (e.g., null subject interpretation conditioned by discourse context) compared to phenomena whose interpretation is made within a single module (e.g., syntactic licensing of null subject). Furthermore, interface properties can be categorized into two groups: internal interfaces, which involve properties (e.g., syntax, semantics) within the grammar itself, and external interfaces, which involve the interaction of formal linguistic properties with cognitive domains beyond the grammar, such as pragmatics and discourse. The IH posits that difficulties for learners mostly are associated with external interfaces.

Research in child language acquisition has revealed that discourse context enhances children's likelihood of accessing the less preferred interpretation in a scopally ambiguous

---

[3] Some models, such as Schwartz and Sprouse's (1994, 1996) full transfer/full access model of L2 acquisition, propose that the initial stage of the interlanguage is based on L1 grammar. According to this model, Universal Grammar (UG) constrains the restructuring of this L1-based interlanguage if L2 input fails to induce this change.

sentence (Musolino & Lidz, 2006; Viau et al., 2010). This finding is contrary to what the IH would otherwise predict for L2 acquisition. Nevertheless, pragmatics, if being distinguished from discourse context in a fine-grained fashion (Rothman & Slabakova, 2011), would likely yield differential effects on learners with different language backgrounds (Özçelik, 2018). For example, as discussed previously, for Chinese-speaking learners of English to acquire the native-like quantifier scope interpretation, they would have to navigate through two possible readings in a controlled story context by applying the syntactic and semantic knowledge. Additionally, they should correctly apply the Principle of Charity (Davidson, 1984; Grice, 1975; Gualmini, 2008), a pragmatic principle on which language users tend to access the interpretation that makes the sentence under consideration true. As such, despite that the 'none' reading, for example, for (2) (e.g., Every horse didn't jump over the fence.) is available in English, L2 learners of English should endorse (2) to a larger extent in a context more supportive for a 'some' reading than in a context supportive for a 'none' reading, because the 'some' reading is the preferred one on (2). By contrast, English-speaking learners of Chinese would not be bothered by this pragmatic principle, because Chinese is claimed to be unambiguous in scope interpretation. As such, IH would predict L2 English learners to have greater difficulty in scope interpretation than L2 Chinese learners, in the direction opposite to the prediction for the results due to the interaction between L1 transfer and language input.

This bidirectional study thus provides a unique window into the investigation of the roles of L1 transfer, language input, interfaces of linguistic modules, and formal accounts of quantifier scope interpretation among L2 learners. At the same time, it addresses the intersection of language modules on L2 interpretation, processing, and acquisition across different learner groups. This study includes an offline sentence-interpretation matching study (Experiment 1), which will test the interpretation of sentences with logical operators interacting with each other in the absence of context by native speakers of English and Chinese to determine whether there is cross-linguistic variation in quantifier scope interpretation. Experiment 2 investigates the offline interpretation of quantifier scope by Chinese-speaking learners of English on English sentences and English-speaking learners of Chinese on Chinese sentences. Experiment 3 explores the online processing of quantifier scope by L2 learners of English and Chinese.

The remainder of this dissertation is organized as follows. Chapter 2 discusses the main difference in how universal quantifiers and existential quantifier/negation interact for scope interpretation in English and Chinese by presenting theoretical accounts and empirical evidence pertaining to the phenomena in question. Chapter 3 reviews relevant previous literature on the interpretation, processing, and acquisition of quantifier scope by adult L2 learners. Based on reviewed literature and identified research gaps, this chapter concludes with the research questions and predictions for the experiments to be reported in the following chapters. Chapter 4 reports on two sentence-interpretation tasks with native speakers of English (henceforth L1 English speakers) and native speakers of Chinese (henceforth L1 Chinese speakers) with the aim of establishing the baseline of scope interpretation in the absence of context among native speakers. Chapter 5 reports on a written Truth Value Judgment Task (TVJT) with Chinese-speaking learners of English and English-speaking learners of Chinese. Chapter 6 reports on a visual-world eye-tracking experiment with L2 English learners and L2 Chinese learners. Chapter 7 concludes the dissertation by summarizing and discussing the research findings and highlighting limitations and recommendations for future research.

2.0 Quantifier scope across languages

This chapter aims to present an overview of how English and Chinese differ in quantifier scope interpretation. In this overview, theoretical approaches to quantifier scope interpretation across languages will be discussed and the experimental literature that contribute to the theoretical claims will also be reviewed. Section 2.1 details scope phenomena in the case of sentences containing scope-bearing logical operators such as quantifier (i.e., *every*, *a*) and/or negation (i.e., *not*) in English. Section 2.2 provides an overall picture of the same scope phenomena in Chinese. This chapter concludes with a comparison of English and Chinese with respect to quantifier scope interpretation.

2.1 English

2.1.1 Doubly quantified sentences

In English, doubly quantified sentences contain a universal quantifier and an existential quantifier as in (1) (repeated in (3) below) in which the universal quantifier ($\forall$) linearly precedes and hierarchically c-commands the existential quantifier ($\exists$) at the surface syntactic structure. Such doubly quantified sentences are ambiguous depending on the relative scope of 'every' and 'a' at the level of Logical Form (LF).

(3) Every child climbed a tree.

a. Surface scope: ($\forall > \exists$)

i. For each child, there was a (different) tree that (s)he climbed.

ii. $\forall x$ [child (x) $\rightarrow \exists y$ [tree (y) $\wedge$ climb (x, y)]]

b. Inverse scope: ($\exists > \forall$)

i. There was a single tree that was climbed by each child.

ii. $\exists y$ [tree (y) $\wedge \forall x$ [child (x) $\rightarrow$ climb (x, y)]]

In (3), two readings can be derived due to the interaction between the universal quantifier and existential quantifier. On one reading, (3) can be paraphrased as (3a(i)) and represented following the logical notation as in (3a(ii)). This reading is derived from the operation that the universally quantified subject (i.e., *every child*) takes scope over the object quantifier phrase (i.e., *a tree*). It is called the 'surface scope' reading in which the syntactic structure and the scope relation between these two operators are isomorphic. On another reading, (3) can be interpreted as (3bi) whose logical form is represented (3bii). The fact that the object quantifier phrase (i.e., *a tree*) takes scope over the quantifier phrase (i.e., *every child*) in the subject position gives rise to this reading. It is referred to as the 'inverse scope' reading in which the surface syntactic c-command domain and the semantic representation are not isomorphic. As such, despite the same structural configuration in the surface syntax, either scope interpretation is possible at the level of semantic representation in English.

Scope ambiguity still holds when the universal quantifier appears in the object position and existential quantifier occurs in the subject position. Consider the example sentence (4). There are two possible interpretations: the surface scope reading (4a(i)) for which the subject quantifier phrase (i.e., *a child*) takes a wider scope and the inverse scope reading (4b(i)) for which the object quantifier phrase (i.e., *every tree*) takes a wider scope. The logical forms associated with each interpretation are represented in (4a(ii)) and (4b(ii)), respectively.

(4) A child climbed every tree.

a. Surface scope: ($\exists > \forall$)

i. There was a single child who climbed multiple trees.

ii. $\exists x$ [child (x) $\wedge$ $\forall y$ [tree (y) $\rightarrow$ climb (x, y)]

b. Inverse scope: ($\forall > \exists$)

i. For each tree, there was a (different) child who climbed it.

ii. $\forall x$ [child (x) $\rightarrow$ $\exists y$ [tree (y) $\wedge$ climb (x, y)]]

It should be pointed out that the two interpretations for each doubly quantified sentence are not logically independent (Reinhart 1976; 1997). For (3), its inverse scope reading entails the surface scope reading[4]: If a specific tree was climbed by every child, it follows that every child climbed a tree. In other words, in any scenario where the inverse scope reading for (3), i.e., (3b(i)), is true, the surface scope reading is true, i.e., (3a(i)). This entailment problem may lead one to

[4] The dissertation does not primarily investigate the entailment relations between different interpretations for learners as a focus of inquiry. This is partly due to the highly abstract nature of these semantic phenomena, which will be addressed in future research.

question whether sentences such as (3) are really ambiguous (Szendroi, 2022). Moreover, it might be problematic to rely on data from doubly quantified sentences such as (3) to determine the availability of the inverse scope reading in a given language, because such sentences would also hold true on the surface scope interpretation even if they are judged to be true in a context where the inverse scope interpretation is intended to be elicited.

For (4) in which the existential quantifier precedes the universal quantifier in the surface syntax, the surface scope reading entails the inverse scope reading: If there was a single child who climbed a different tree, it then follows that each tree was climbed by a child (i.e., the children who climbed the trees happen to be a single individual). In contrast, the inverse scope reading does not entail the surface scope reading in this case, because for people to accept the sentences such as (4) in a context where a different child has a different tree to climb, they must have assigned this sentence with an inverse scope reading. Therefore, sentences with an existential quantifier in the subject position and a universal quantifier in the object position provide a better testing ground for determining whether a language allows the inverse scope interpretation.

Another case of this type of ambiguity arises in the interaction between universal quantifier and negation (Horn, 1989; Lasnik, 1972). Consider the sentence in (5).

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

a. Surface scope: ($\forall > \neg$)

i. None of the horses jumped over the fence.

ii. $\forall x$ [horse (x) $\rightarrow \neg$ jump over the fence (x)]

b. Inverse scop: ($\neg > \forall$)

i. Some of the horses jumped over the fence.

ii. $\neg\forall x$ [horse (x) $\rightarrow$ jump over the fence (x)]

In (5), the universally quantified phrase *every horse* precedes and c-commands the negative clitic *n't* in surface syntax. Its surface scope reading occurs when the universally quantified phrase takes scope over negation. On the surface scope reading, this sentence is interpreted as (5a(i)) and logically represented as (5a(ii)). Its inverse scope reading occurs when negation take scope over the universally quantified phrase. On the inverse scope reading, this sentence is interpreted as (5b(i)) and logically represented as (5b(ii)). Like doubly quantified sentences, negatively quantified sentences as in (5) allow either scope relation between universal quantifier and negation at the level of semantic representation even though the surface syntax stays the same for both the surface scope and invers scope readings.

In cases where negation precedes the/a universal quantifier in a sentence, the situation becomes somewhat different. Consider the sentence in (6).

(6) The horse didn't jump over every fence.

a. Surface scope ($\neg > \forall$)

i. The horse jumped over some of the fences, but not all.

ii. $\neg\forall x$ [fence (x) $\rightarrow$ jump over (horse, x)]

b. Inverse scope ($\forall > \neg$)

i. The horse jumped over none of the fences.

ii. $\forall x$ [fence (x) $\rightarrow \neg$ jump over (horse, x)]

Logically speaking, both readings are available for such sentences. On the surface scope reading as paraphrased in (6a(i)), the universally quantified phrase *every horse* is interpreted within the scope of negation, as illustrated in (6a(ii)). By contrast, for the inverse scope reading as paraphrased in (6b(i)) to be made available, the universally quantified phrase *every horse* is interpreted outside the scope of negation. In the literature on the interpretation of negatively quantified sentences (Musolino et al., 2000; Musolino & Lidz, 2006), English-speaking children and adults found it extremely difficult to access the inverse scope reading. While the grammar of English makes available both readings for sentences with negation preceding the universal quantifier, pragmatic factors have been argued to override syntactic and semantic factors, leading to a strong bias toward the surface scope reading (Musolino & Lidz, 2006; Zhou & Crain, 2009).

The interpretation of such sentences involves scalar implicature, a class of conversational implicature (Horn, 1989). It is a pragmatic mechanism by which communication proceeds smoothly between interlocutors. Scalar implicature usually operates on sentences involving quantificational elements which are ordered on a scale in terms of information strength (Horn, 1989; Levinson, 2000). According to Levinson (2000, p. 254-255), typical positive scalar items include *some*, *all* with *some* being the weak item and *all* the strong item. The given positive scale has a corresponding negative scale whose items include *not-some* (i.e., *none*) (strong item) and *not-all* (i.e., *some*) (weak item). Strong scalar items entail the weak scalar items, and only the weak items give rise to scalar implicature. Widely discussed in the literature are examples involving positive scalar items, as in (7).

(7) John ate some of the cookies.

(8) John didn't eat all the cookies.

(9) John ate all the cookies.

The hearer would infer that (8) is what the speaker would intend for the utterance in (7) given that the interlocutors follow the Cooperative Principle during communication (Grice, 1989). Under this principle, the Maxim of Quantity is most relevant for scalar implicature to be computed with respect to (9). The Quantity Maxim (Grice, 1989, p. 26) specifically holds:

1. Make your contribution as informative as is required.
2. Do not make your contribution more informative than is required.

Constrained by this maxim, the speaker is saying as much as he/she can to optimize information exchange. Thus, the hearer has a good reason to infer that the stronger term does not hold in this context since the speaker otherwise would have uttered (9) to respect the Quantity Maxim. As such, the most plausible scalar implicature for (7) to be induced is (8) rather than (9).

This line of reasoning similarly works for (6). In (6), the inverse scope reading entails the surface scope reading. In other words, in any scenario where the inverse scope is true, the surface scope would also be true: If none of the fences are such that the horse jumped over them, then it follows that not all the fences are such that the horse jumped over, but not the vice versa. Therefore, the weak item 'not every' represented by the surface scope reading induces the scalar implicature such that the hearer who accesses the surface scope reading would infer that the speaker is not ready to use the strong item 'none', the inverse scope reading in this case. This explains why (6) almost only permits the surface scope reading – (6) is pragmatically compatible only with the surface scope reading although it is semantically compatible with both the surface scope and inverse scope readings.

The presence of the inverse scope reading for (5) is not incompatible with the pragmatic account. In (5), its surface scope reading entails the inverse scope reading: If it is true that none of the horses jumped over the fence, then it is also true that only some of the horses jumped over the fence. The inverse scope reading thereby is the weak item. Thus, the fact that the speaker chooses the weak item 'not all' leads the hearer to infer that the strong item 'none' is not readily accessible. Zhou and Crain (2009) also argued that the access of the inverse scope reading can also be explained by the Principle of Charity (Davidson, 1984; Grice, 1975). On the Principle of Charity, the hearer would accept (5) on the inverse scope reading because this sentence is deemed true on this reading. In this case, the inverse scope reading is accessible since the Principle of Charity is consistent with the Gricean Cooperative Principle. By contrast, even if the inverse scope reading for (6) can be obtained given the Principle of Charity, the Cooperative Principle would eventually override such a reading.

In general, English grammar permits scope ambiguity resulting from the interaction of various logical operators, despite pragmatic contexts favoring certain interpretations. There has been a substantial body of literature devoted to the discussion on the development of different theoretical proposals on scope ambiguity (see Kiss & Pafel, 2017, for a detailed review). The generative approach, among others, stands as a widely adopted framework in this regard (May, 1977; 1985; Fox, 2000). In the framework of generative grammar (Chomsky, 1995), scope ambiguities result from covert displacement (Lidz & Musolino, 2002), by which certain elements in a sentence are not interpreted in the position where they appear in the surface syntactic position, albeit the fact that these elements are not overtly moved during syntactic derivation. As a contrast, overt movement appears to be more intuitive, by which a linguistic constituent is syntactically moved out of its base-generated position and gets interpreted in its argument position. An example is the interpretation of *wh*-questions. Consider (10), for example.

(10) What did John buy ___?

Although the *wh*-phrase occurs in the grammatical subject position as a result of overt syntactic displacement for [NOM] and [+WH] features to be checked (Carnie, 2021, p.362), it is interpreted in the object position where *what* is assigned the theta role by the matrix verb *buy*.

Scope ambiguity is syntactically derived at the level of Logical Form (LF), a syntactic representation which interfaces with the semantic component of the grammar and where logical operators can optionally undergo covert movement. Under the T-model of grammar (Chomsky, 1986, p. 68), as illustrated in Figure 2.1, LF operates on its own without being influenced by phonological form (PF). Thus, the covert displacement operation at LF has no phonological consequence at PF in the surface syntax.

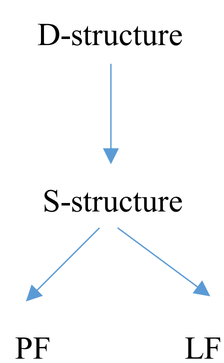

Figure 2.1 T-model of grammar

The rules for deriving quantifier scope are Quantifier Raising (QR) or Quantifier Lowering (QL) (Fox, 2000, Heim & Kratzer, 1988; May, 1977; 1985) and these rules arguably are applied through the mapping from SS to LF.

Under a QR or QL approach, let us consider how surface scope and inverse scope readings are derived for doubly quantified and negatively quantified sentences in English. First, as for the surface scope reading of (4),[5] it can be derived through QR by which The DP *every tree* moves from the object position and is adjoined to the left of the VP boundary. This movement is driven by the assumption that the quantified DP in the object position can be interpreted only when it is covertly moved at LF to create an interpretable structure containing correct semantic types (see Heim & Kratzer, 1988, p. 185-186 for a detailed discussion). The subject (i.e., *a child*) moves out of the VP-internal position[6] and adjoins to TP in the derivation from SS to LF, as schematized in (11) LF. As a result, *a child* takes scope over *every tree*.

(11) [$_{TP}$ a child$_i$ …[7] every tree$_j$ [8]… [$_{VP}$ t$_i$ climbed t$_j$]]

---

[5] (4) is discussed prior to (3) for two reasons: (a) because universal quantifiers in the object position have received more attention than that in the subject position in the semantics literature, (b) the former poses more issues when it comes to the determination of the nature of quantifiers (Portner, 2005, p. 127). However, a unified framework, if adopted, can be easily applied to explain scope ambiguity across different types of quantifier scope.

[6] According to VP-Internal Subject Hypothesis (Kitagawa, 1994; Koopman & Sportiche, 1991; Kuroda, 1988), the grammatical subject of a clause is base-generated in the Specifier position of the predicate.

[7] "…" indicates that constituents are not immediately adjacent to each other and in fact these are positions assumed to host variable binders (Heim & Kratzer, 1988).

[8] Quantified expressions including the ones base-generated in the object position, like Wh-operators, are not referring expressions, must be moved from the θ-position at LF (Kuno et al., 1999), because θ-criterion would otherwise require

The inverse scope reading of (4) can be derived through either QR or QL. An additional QR applied to (11) would yield the inverse scope reading, as represented in (12a). The inverse scope reading could also be obtained through QL, as can be seen in (12b). On either approach, the resultant structure is interpreted in a way that *every tree* takes a wider scope.

(12) a. [$_{TP}$ every tree$_j$ [a child$_i$ …~~every tree$_j$~~ … [$_{VP}$ t_i climbed t_j]]]

b. [$_{TP}$ ~~a child~~ … every tree$_j$ … [$_{VP}$ a child climbed t_j]]

Similarly, as for (3), its surface scope and inverse scope reading are derived through QR, as schematized in (13) and (14), respectively.

(13) [$_{TP}$ every child$_i$ …a tree$_j$ … [$_{VP}$ t_i climbed t_j]]

(14) [$_{TP}$ a tree$_j$ … every child$_i$ … [$_{VP}$ t_i climbed t_j]]

QR and QL can be similarly applied to the derivation of surface and inverse scope readings for negatively quantified sentences. As for (5), the surface scope reading comes about when *every*

---

these expressions be assigned with a θ-role, contrary to the fact that nonreferential expressions do not qualify as a legitimate argument.

horse moves from its VP-internal subject position and adjoins to the grammatical subject position in the surface syntax, as schematized in (15). The inverse scope reading for (5) can be derived through QL in that *every horse* lowers from the grammatical subject position back to the VP-internal subject position, as schematized in (16).

(15)[$_{TP}$ Every horse$_i$ [$_{TP}$ didn't [$_{VP}$ t$_i$ jump over the fence]]]

(16) [$_{TP}$ ~~Every horse$_i$~~ [$_{TP}$ didn't [$_{VP}$ every horse jump over the fence]]]

Under the structure-based approach (Aoun & Li, 1993), (6) would allow only the surface scope reading where negation takes scope over the universally quantified phrase in the object position, *every fence*. In a similar vein, the base position for *the horse* is VP-internal and it moves to surface as the grammatical subject in the surface syntax. The inverse scope reading would otherwise be obtained if *every fence* is raised to a position higher than negation in order to take wide scope. According to this account, this raising has been blocked by the intervening negation. Before detailing the reasoning for this blocking effect due to negation, I present some widely accepted assumptions in generative syntax to account for the interaction between quantified phrases and negation especially when such phrases are in the object position. First, Pollock (1989) and Chomsky (1992) assume that negation heads its own phrase, i.e., Negation Phrase (NegP). Second, it is also assumed that a negation operator (Neg-Op) takes up the Specifier of NegP

(Ouhalla, 1990). The structural representation of (6) for its surface scope reading is illustrated in Figure[9] (2.2).

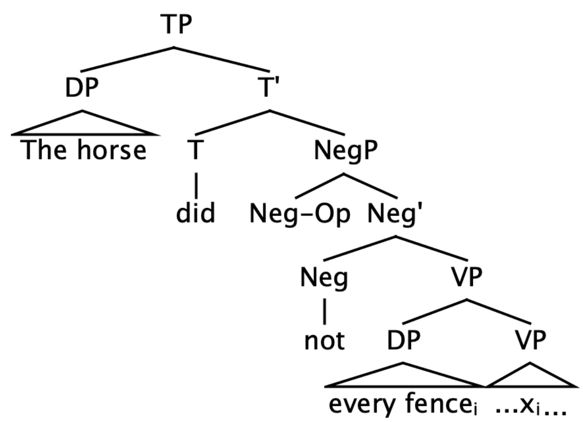

Figure 2.2 LF representation for the surface scope reading of (6)

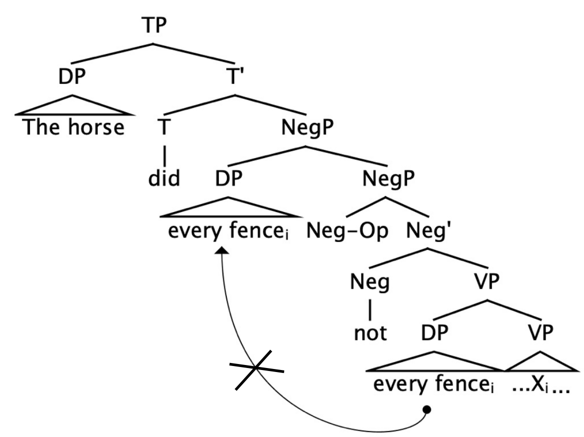

Figure 2.3 LF representation for the potential inverse scope reading of (6)

[9] All syntactic trees in this dissertation were drawn with TreeForm (Derrick & Archambault, 2010). In addition, for the sake of simplicity, CP (Complementizer Phrase) projection was not presented in the tree diagram.

The inverse scope reading would be possible only if *every fence* is raised to a position higher than negation, as represented in Figure (2.3). However, this movement, crossing negation, would violate the Locality Requirement, as stated in (17).

> (17) A variable, if it is subject to the Locality Requirement, must be bound by an Ā-binder α within the minimal maximal category containing α and the variable (Aoun & Li, 1993, p. 172).

As shown in Figure (2.2), the minimal maximal category containing the variable x_i and Negation operator (Neg-Op) as the Ā-binder α is NegP. It is clearly shown that the moved object variable is not bound by the Neg-Op. The fact that *every fence* cannot move crossing negation due to the Locality Requirement explains why only the surface scope reading for (6) is allowed.

The structure-based analysis aligns with the earlier presented analysis, which was based on scalar implicature calculations for sentences with negation preceding universally quantified phrases. Both the structure-based analysis and pragmatics-based analysis support the conclusion that inverse scope interpretations for (6) are not possible in English. Nevertheless, it is worth noting that scalar implicature has been reported to be cancellable (Politzer-Ahles et al., 2013). As such, English speakers would be expected to accept that the utterance: *The horse didn't jump over every fence. In fact, none of the fences were jumped over* can be interpreted as *The horse jumped over none of the fences*. As such, these two approaches differ in that the structure-based approach suggests that quantifier scope and its interpretation are driven by the c-commanding relationship between logical operators only (e.g., May, 1977, 1985; Aoun & Li, 1989, 1993; Tang, 2001), while approaches that have taken into account other factors such as pragmatics would argue for the interaction of different sources of information: syntax, semantics, and pragmatics, among others, in quantifier scope computation (e.g., Kuno et al., 1999; Paterson et al., 2008; Ionin & Luchkina, 2018).

Before proceeding to the discussion of quantifier scope in Chinese, I present an overview of the psycholinguistic research on the processing of quantifier scope with adult speakers of English. This systematic review has two main purposes. First, the empirical basis for scope ambiguity can be revealed by looking into psycholinguistic investigations. Second, performance on quantifier scope interpretation by adult English native speakers provides the baseline to which L2 performance can be compared.

Research that has examined sentences containing two quantifiers versus one quantifier and one negation will be reviewed separately. In the literature, language users have been found to display a strong preference for surface scope interpretations, despite with some differences in the measurement adopted, target population tested, and type of quantifier scope investigated across studies (Szendroi, 2022). The 'Processing Scope Economy Principle' has been proposed to account for such an interpretive preference (Anderson, 2004). This principle has tied the relative ease with which quantifier scope is computed to abstract linguistic structures configured ready for scope calculation, as formulated in (18). That is, inverse scope interpretations tend to incur processing costs because of the extra syntactic displacement at LF, and thus are much less preferred at the level of interpretation even though they are accessible especially in supportive contexts.

(18) The human sentence processing mechanism prefers to compute a scope configuration with the simplest syntactic representation (or derivation). Computing

a more complex configuration is possible but incurs a processing cost (Anderson, 2004: 46).

For the inquiry into the interpretation of quantifier scope among English speakers, the focus has been primarily on doubly quantified sentences (e.g., Anderson, 2004; Dotlačil & Brasoveanu, 2015; Dwivedi, 2013; Filik et al., 2004; Fodor, 1982; Frazier et al., 1999; Kurtzman & MacDonald, 1993; Paterson et al., 2008; Patson & Warren, 2010; Wu & Ionin, 2022; Scontras et al., 2017). Under the constraint-based lexicalist framework for sentence processing (MacDonald et al., 1994), Kurtzman and MacDonald (1993) investigated interactive constraints from multiple principles on the resolution of scope ambiguity with doubly quantified sentences. In this study, they manipulated factors including linear order of quantifiers (every-a vs. a-every), clause voice (active vs. passive), and verb type (action verb vs. perception verb). English native speakers were presented with a quantified sentence followed by a reasonable continuation of that sentence under one of its interpretations and this presentation was self-paced. The participants were instructed to judge whether the continuation sentence made sense and was a natural continuation of the first sentence. For example, the singular continuation, *the tree was in the park*, would be reasonable for *Every kid climbed a tree* on its inverse scope interpretation (Every kid climbed the same tree), and the plural continuation, *the trees were in the park*, would be reasonable for the same sentence on its surface scope interpretation (Every kid climbed a different tree). For the active sentences, the results showed that the surface scope reading was strongly preferred regardless of whether the universally quantified phrase was in the subject or object position.

Moreover, such a preference was stronger for *a-every* sentences than for *every-a* sentences, demonstrating the effect of the order of quantifiers. This finding led the authors to propose the

Single Reference principle, according to which *a*-phrase is more likely to be interpreted as a single entity when it occurs in the grammatical subject position compared to when it is in the object position, because such an interpretation is simpler when the processor first encounters this element during sentence processing. In addition, the verb type effect was more pronounced in *every-a* sentences than in *a-every* sentences, in that the preference for the surface scope reading was stronger for *every-a* sentences containing action verbs than those containing perception verbs, which was explained by thematic hierarchy principles. As such, the surface scope reading was favored because different principles converged to support this reading in the case of active sentences. By contrast, judgements about passive sentences were much more variable because not all principles converged on one single interpretation in this case. The overall results emerging from this study is that inverse scope readings relative to surface scope readings are harder to access for English doubly quantified sentences and the extent to which a particular interpretation is accessible is constrained by multiple factors.

Evidence from online methods such as self-paced reading also suggests that inverse scope readings incur processing difficulty and are thus harder to access even if they are available for English native speakers. Anderson (2004) systematically investigated the processing of doubly quantified sentences among English native speakers. She adopted a similar paradigm as in Kurtzman and MacDonald (1993), but measured speakers' scope interpretation with word-by-word self-paced reading tasks. Basically, each quantified sentence was followed by a continuation sentence (intended for disambiguating the preceding quantified sentence). For sentences with *a* preceding *every*, the residual RTs calculated over the entire continuation sentence revealed a significant difference in RTs between the singular and the plural continuation sentences, with RTs higher for plural continuation than for singular continuation. No difference was observed for any individual region except for the final region (wrap-up for semantic integration) with a marginal difference in RTs between the singular and the plural continuation sentence on the item analysis. As for sentences with *every* preceding *a*, the residual RTs calculated over the entire continuation sentence and each region revealed no evidence for a difference in RTs between the singular and the plural continuation. A similar pattern of results was also observed in a self-paced reading study by Dwivedi (2013) particularly when participants were not forced to answer follow-up questions targeting quantifier scope interpretation. Results from Anderson (2004) seem to suggest that assigning an inverse scope interpretation yields a measurable processing cost especially in the case of *a-every* sentences. However, this effect was not pronounced in *every-a* sentences. As Anderson herself argued and Szendroi (2022) further added, one possible reason, due to the nature of this task, is that slower RTs at the plural continuation sentence preceded by a *a-every* sentence may simply be an indication that the processor was surprised to encounter an unexpected plural if they had entertained a surface scope interpretation of the preceding quantified sentence. In other words, the cost associated with the processing of the plural continuation may not necessarily indicate that the inverse scope assignment itself is costly. Rather, the information of the continuation sentence in its early region, which counters the interpretation the processor has generated in the preceding sentence, has led to this processing cost.

To minimize the complication caused by the disambiguating sentence, Anderson (2004) conducted another experiment in which the continuation did not disambiguate the quantified sentence (e.g., *A helpful member tested every recipe. The club's president did, too*). The online comprehension of quantified sentences was examined using the self-paced reading technique. Participants' interpretation of the test sentences was assessed based on their performance on the comprehension questions that followed each test sentence. The results showed that English native speakers read the entire quantified sentence more slowly when a *a-every* sentence was assigned an inverse scope reading than when the sentence was assigned a surface scope reading. It is interesting to note that this pattern numerically held for each region but RTs for each region were not significantly different between the surface scope and inverse scope interpretations, suggesting that costs in the processing of theses quantified sentences could only be measured globally. This is not surprising because these quantified sentences are globally ambiguous but not locally ambiguous and ambiguity for the whole sentence would not be resolved by the parser until the sentence wrap-up. These findings led Anderson to propose the Processing Scope Economy principle, positing that scope processing was largely dependent on the grammatical operation with which each reading (surface vs. inverse) was engaged. According to this account, the surface scope reading was preferred over the inverse scope reading, because the latter, which involves a more complex linguistic representation (Fox, 2000; May, 1977), was harder to process.

Different from Kurtzman and MacDonald (1993), in which multiple sources of information were shown to jointly determine the preferred interpretation of a doubly quantified sentence, Anderson (2004) represented an attempt to attribute costs associated with the processing of each scope interpretation only to how complex the syntactic representation of that interpretation is. Although Anderson (2004) has provided evidence in support of this account, a few methodological limitations should be noted in this study. First, verb type (with a mixture of agentive and psych verbs, e.g., hit, greeted) and clause tense (with a mixture of simple past tense and past perfect tense, e.g., asked, had used) were not properly controlled for, making them potential confounds. Second, RTs obtained from self-paced reading provide information regarding scope interpretation preference only. Therefore, this measurement does not allow for disentangling preference from access, as a strong dispreference for a particular interpretation does not necessarily indicate that this interpretation cannot be accessed by the reader. I return to this issue in the discussion on the choice of methods for Experiment 3.

Previous work so far from acceptability judgement and self-paced reading paradigms suggests that readers have a strong preference for the surface scope reading over the inverse scope reading, albeit on different accounts. Moreover, this surface scope interpretation preference was more evident in *a-every* sentences than in *every-a* sentences, which seems to align with the Single Reference principle proposed by Kurtzman and MacDonald (1993). However, a few studies that used eye-tracking reading paradigms have yielded a somewhat different picture (Filik et al., 2004; Paterson et al., 2008). For example, in Filik et al. (2004), English native speakers were presented with sentences such as *The celebrity gave an in depth interview to every reporter from the newspaper* (a(n)…every) or *The celebrity gave every in depth interview to a reporter from the newspaper* (every…a(an)), followed by continuation sentences, such as *but the reporter(s) was/were not very interested*, to disambiguate towards either a surface scope reading or an inverse scope reading. It was found that RTs were longer for the NP anaphora in the continuation sentence when it was plural compared to when it was singular regardless of the linear order of quantifiers in the preceding sentences. This further suggests that the surface scope reading at least for *every-a* sentences has not been accessed. The authors offered one possible explanation for the processing difficulty: readers might have underspecified quantifier scope due to a morphological mismatch between the plural anaphora and its singular antecedent.

The fact that the effect of the number feature of the first region in the continuation sentence was observed in the eye-tracking experiment but not in the self-paced reading experiment might simply be because eye-tracking was more appropriate for capturing fine-grained cognitive processes with respect to quantifier scope interpretation. If the experimental results are compared across tasks, it is clear that the processing of relative scope is sensitive to factors such as linear order of quantifiers in the acceptability judgment task by Kurtzman and MacDonald (1993) but not the eye-tracking task by Filik et al. (2004) and Paterson et al. (2008). The observed task effect may have arisen because tasks involving acceptability judgments allowed participants to evaluate possible semantic interpretations at the same time to a greater extent through metalinguistic reasoning compared to eye-tracking methods. Moreover, the fact that RTs were sensitive to factors manipulated in the context sentence during self-paced reading was observed for the continuation sentence as a whole, but not its individual regions (except for the final region in the case of *a-every* sentences). This finding suggests that quantifier scope is globally ambiguous, and its interpretation preference tends to manifest whenever semantic integration is evident.

Given the mixed results obtained in the timed context potentially due to the methodological limitations of sentence continuation paradigms, Patson and Warren (2010) used a number judgment task to investigate whether singular indefinite nouns in the distributive condition (e.g., *Each of the men carried a box*) vs. the collective condition (e.g., *Together the men carried a box*) were interpreted as conceptually plural during online reading by having participants judge whether one or two words appeared on the screen in a self-paced reading task on sentences presented one word or two words at a time. The advantage of this task is that features of the conceptual representations of quantification in question can be probed at the critical word. ‘One’ decisions took longer at the critical noun phrase in the distributive condition than in the collective condition, suggesting that singular noun phrases in the distributive condition can be interpreted plurally and thus its surface scope reading is accessed. Because this study was mainly set up to examine whether comprehenders could conceptually build plural referents for singular indefinite noun phrases, relative preference for surface scope reading vs. inverse scope reading for sentences in the distributive condition was not directly addressed in their study.

Other studies that have probed the interpretation of quantified sentences have used truth-value judgment tasks (TVJTs) in an offline setup (Crain & McKee 1985; Crain & Thornton 1998; Crain et al., 1996; Scontras et al., 2017; Wu & Ionin, 2022). This method has its origin in child language acquisition research (Crain & Thornton 1998; Pagliarini et al., 2021), and has been adapted for testing other populations including adult participants (Scontras et al., 2017; Spychalska et al., 2016). In its simple version, the task asks participants to decide whether the test sentence is an appropriate description of a given context, be it written or pictorial. From a semantic perspective, knowing the meaning of a sentence is to know the condition for that sentence to be true. As reasoned by Scontras and Pearl (2021, p. 2), one way of diagnosing sentence meaning is to identify the contexts that a sentence can (contexts in which a sentence is true) or cannot (contexts in which a sentence is false) describe. Thus, evaluating the degree to which a sentence aligns with its context can provide valuable insights into the meaning of that sentence. As discussed previously, one assumption to follow in the use of this task is that a sentence would be judged true, if one reading under consideration is made true, as postulated in the Principle of Charity (Grice, 1975). In addition, Condition of Plausible Dissent should be satisfied for this task to be appropriately set up (Crain & Thornton, 1998, p. 225). It requires that the discourse context for the sentence to be judged be pragmatically felicitous, because any unnatural pragmatics that participants might bring to the experimental setup may inadvertently contaminate research findings. In other words, Condition of Plausible Dissent requires decisions in TVJTs be made based on one's grammar rather the pragmatic elements of the context provided. This being the case, one's linguistic competence could be revealed through TVJTs (Thornton, 2017).

For example, Scontras et al. (2017) used TVJTs to examine how heritage speakers of Chinese with English as their dominant language interpreted doubly quantified sentences with universal quantifiers in either the subject (e.g., *Every shark attacked a/one pirate*) or object position (e.g., *A/one shark attacked every pirate*). As a note, in addition to the linear order of quantifiers, the contrast of indefinite article *a* and numeral *one* was also manipulated. Data from English native speakers served as the baseline for the target population performance. In a picture-based TVJT task, participants judged whether the sentence heard appropriately described the picture on a 7-point Likert Scale (1 = 'completely inappropriate', 7 = 'completely appropriate'). For sentences containing the indefinite article *a* and the universal quantifier *every*, pictures corresponding to the inverse scope reading were found to be less acceptable than those corresponding to the surface scope reading, consistent with previous findings. For sentences containing the numeral *one* and the universal quantifier *every*, they also found pictures matching inverse scope reading were less acceptable than those matching surface scope reading. As such, surface interpretation preference was consistently observed among English native speakers across experiments.

It is worth noting that although English native speakers were found to differ in their ratings between surface scope readings and inverse scope readings, average ratings on inverse scope readings were 4.46 (out of 7), due to which the author claimed that inverse scope readings were available to English speakers. Such characterization has remained largely descriptive and no 'one-size-fits-all' threshold exists for deciding whether one reading is available or not. In fact, to substantiate whether a certain reading is (relatively) available in a given language, it is in fact more critical to compare ratings on that reading between languages (English vs. Chinese). Interestingly, they additionally found an interaction between linear order of quantifiers and interpretation in the case of sentences containing *one*, driven by the fact that the inverse scope reading was more acceptable in sentences with *a* than in sentences with *one*. There are two implications for this finding: (a) not surprisingly, the inverse scope reading was more challenging for interpretation than the surface scope reading, due to the former being syntactically more complex than the latter, and (b) lexical information (*a* vs. *one*) influenced the extent to which surface scope reading was favored in that preference for the surface scope reading was more robust in *one* sentences than in *a* sentences. A similar pattern of results was also obtained in Ionin and Luchkina (2018). Citing the Single Reference principle from Kurtzman and MacDonald (1993), Scontras et al. (2017) suggest that differences in lexical items play a role affecting quantifier scope interpretation. As they argued, *one* is phonologically more salient than *a*[10], it is thus more likely to build a single-reference parse with *one* occurring initially and this tendency carries over through the course of parsing, leading to a strong preference for surface scope readings.

In the baseline group with English native speakers, Wu and Ionin (2022) also found a preference for surface scope readings on 'a/one…every' sentences even though both readings were possible. However, the previous finding that inverse scope readings were more pronounced in sentences with *a* than in sentences with *one* was not replicated in their study. Reasons for such a discrepancy were not quite clear. The fact that Wu and Ionin (2022) used a combination of written context and picture and included psych verbs (e.g., *scare*) may have complicated their results. Nevertheless, it has been consistently shown that both the surface scope and the inverse scope

[10] The use of 'one' as a numeral is more likely to prompt the interpretation of a noun phrase as a single entity compared to using 'a' as an indefinite determiner.

readings are available in English speakers and inverse scope readings are harder to access and process than surface scope readings on doubly quantified sentences.

2.1.2 Negatively quantified sentences

Compared to doubly quantified sentences, English negatively quantified sentences have received relatively less attention in psycholinguistic studies. Most studies that have looked at interpretation and processing of negatively quantified sentences by adult speakers of English focused on child language acquisition and data from adult speakers simply served as the baseline (Musolino et al., 2000; Musolino & Lidz, 2006). For example, Musolino and Lidz (2006) examined how children interpreted negatively quantified sentences with the universal quantifier *every* in the subject as in (5) (i.e., Every horse didn't jump over the fence) or object position as in (6) (i.e., Tthe horse didn't jump over every fence). Data from adult native speakers of English served as the baseline. They used the TVJT. In the experiment, one experimenter acted out short stories in front of the participants using toys and props. The other experimenter played the role of a puppet who at the end made a statement about what s/he thought happened in the story. The task for the participants was to decide whether the puppet's statement was right or wrong. For instance, if adult speakers of English were tested on their interpretation of sentences such as *Every horse didn't jump over the fence*, a story context in which two out of three horses jumped over the fence but the third horse did not, was provided for the experimenter to act out. If participants accepted this test sentence under this context, this would be taken as evidence that they interpreted *every* within the scope of negation and the inverse scope reading was accessed. If they rejected the test sentence, it indicated that they interpreted *every* outside the scope of negation and leant towards the surface scope reading only. Different from children, adult speakers were found to accept the inverse scope reading 92.5% of the time. In another condition, where the three horses first succeed in a log-jumping event before they tried the fence-jumping event and the test sentence was preceded by an affirmative statement (e.g., *Every horse jumped over the log, but every horse didn't jump over the fence*), both child and adult participants were found to increase their acceptance of the inverse scope reading. Therefore, subtle changes in context have led to differences in the extent to which inverse scope readings are available. Following a similar procedure, participants were tested on their interpretation of sentences such as *The strong guy didn't/won't put every elephant on the table*, adult speakers of English have been shown to accept such sentences on their inverse scope readings to a very small extent (20% of the time). This limited acceptance is due to the strong influence of pragmatic implicatures, favoring the surface reading as the preferred interpretation.

In a more recent study, Attali et al. (2021) found that English native speakers showed a strong preference for the inverse scope reading (74 % of the time) even though the surface scope reading was still available albeit to a much less degree (26 % of the time) in a paraphrase-validation task. However, the evidence regarding the interpretation of *every-negation* sentences by English native speakers is far from conclusive especially given that such sentences on their inverse scope readings are predicted to be less preferred by the Processing Scope Economy principle, as has been found in previous studies on doubly quantified sentences. The inverse scope reading for *every-negation* sentences was found to be much less preferred compared to the surface scope reading and calculating the inverse scope reading incurred a higher processing cost (Chung & Shin, 2022; Lee, 2009, 2010). In their study, Lee (2010) conducted a self-paced reading study in which English native speakers read an *every-negation* sentence in a non-cumulative moving window fashion and this sentence was preceded by a context that supported either a surface scope reading or an inverse scope reading. After reading the sentence, they were additionally instructed to decide whether the sentence was a truthful description of the context. The results showed that the test sentence was judged to be true on the surface scope reading 71 % of the time as opposed to 37 % of the time on the inverse scope reading. Compared to the surface scope reading, the inverse scope reading was also found to induce slower reading times and response times in this case. Lee (2009) further observed that both the surface scope and invers scope readings for *every-negation* sentences were almost equally available in the absence of contextual support.

The divergence in English speakers' performance on *every-negation* sentences thus far seems to be due to a task effect. Studies with offline tasks tend to demonstrate an inverse scope reading preference. A preference for the surface scope reading has been found primarily in studies using online methods (Chung & Shin, 2022; Lee, 2009, 2010). An alternative account that has been embraced to elucidate this disparity is the efficiency-based processing approach proposed by O'Grady (2005, 2009). According to this approach, grammatical rules do not play any role in shaping the core properties of language and the ease at which language is processed is driven by how efficient the parser could be for its operation when drawing on working memory resources. There are two assumptions for this approach to hold (O'Grady et al., 2009, p.7).

(i) As the processor works its way through a sentence, it immediately assigns each NP an interpretation, based on available clues such as position, determiner type, case marker, context, and so forth.

(ii) The revision of a previously assigned interpretation is costly since it disrupts the normal linear operation of the processor, which forms and interprets sentences in real time under conditions that value quickness.

On this account, it is not surprising that the inverse scope reading for *every-negation* sentences would be less preferred than the surface scope reading, because the former is more difficult to access especially in timed experimental settings. Specially, the surface scope reading would be first available during sentence processes on assumption (i), and the initial interpretation would have to be revised for the inverse scope reading – creating an additional burden on working memory – if assumption (ii) is respected. It should be noted that this processing-based account is completely different from Anderson's Processing Scope Economy principle in the sense that the former attributes processing difficulty solely to processor-driven efficiency complicated by working memory resources while the latter links processing cost to structural complexity.

Some studies using offline methods have also found English native speakers to prefer the surface scope reading on *every-negation* sentences (Chung, 2009; Chung & Shin, 2022, Wu & Ionin, 2019), which is in fact inconsistent with what the processing-based account would predict. For example, in a context-based acceptability judgment task in which participants rated a negatively quantified sentence with a universal quantifier in the subject position on a scale from 1 (totally unacceptable) to 4 (totally acceptable) in the context of a short story, English native speakers in Chung (2009) were found to show a greater acceptance of the test sentence following a context in support of the surface scope reading compared to that following a context in favor of the inverse scope reading, suggesting that English native speakers preferred the test sentence on its surface scope reading. With a similar design except that the context was accompanied by a picture, Wu and Ionin (2019) also found that English native speakers preferred *every-negation* sentences to be interpreted in a way that *every* took a wider scope over negation. Although ratings on the test sentences following a context favoring the inverse scope reading were much lower than the one favoring the surface scope reading in both studies, the participants had a tendency to not reject sentences in the inverse scope context (mean ratings over 2 out of 4), which appears to point toward the fact that the inverse scope reading for these sentences was not impossible.

The above literature review shows mixed results regarding the preferred interpretation for *every-negation* sentences among English native speakers. Some studies have found the inverse scope interpretation to be preferred (Attali et al., 2021; Musolino et al., 2000; Musolino & Lidz, 2006) whereas others have found the surface scope reading to prevail (Chung, 2009; Chung & Shin, 2022; Lee, 2009; Lee, 2010; Wu & Ionin, 2019). Whether experimental materials are tested in the timed context or not does not seem to explain the mixed evidence at least in the case of native speakers. Moreover, the way test sentences are presented in relation to their preceding contexts has been found to modulate scope interpretation (Chung & Shin, 2022; Özçelik, 2016). A close scrutiny into how the context was presented in those studies allowed us to speculate that studies involving the primary use of act-outs and/or pictures in the context (Attali et al., 2021; Musolino et al., 2000; Musolino & Lidz, 2006) tend to manifest a preference for the inverse scope reading, and studies involving the use of written context in which numbers and/or quantifiers were explicitly mentioned in the description of story context tend to demonstrate a bias toward surface scope interpretation among participants. As argued by Thornton (2017), act-outs provide an ideal scenario in which it is easy to build in a justification for the test sentence being false so that participants can easily deny the sentence under consideration. As such, particularly in the case of evaluating *every-negation* sentences against the surface scope context, participants would be shown to access the inverse scope if they were able to plausibly deny such sentences in the act-out task.

Nevertheless, given the mixed results, the interpretation of *every-negation* sentences remains a topic of further research. To the best of my knowledge, no research has been done to examine the interpretation of *every-negation* sentences in one single study that uses a combination of different methods.

Compared to *every-negation* sentences, *negation-every* sentences generated results that were rather consistent regardless of the method used (Chung, 2009; Chung, 2013; Musolino & Lidz, 2006). The surface scope reading usually prevails over the inverse scope reading for sentences such as *Scott didn't eat every meal*. In Chung (2013), English native speakers were presented with sentences such as 'Scott didn't eat every meal'[11] and asked to rate their acceptability in contexts supporting either the surface scope reading (e.g., Scott missed some meals but not all) or the inverse scope reading (e.g., Scott missed every single meal) on a 1-4 scale. The results indicated that the surface scope context received significantly higher acceptability ratings (3.49 out of 4) compared to the inverse scope context (2.36 out of 4). This suggests a strong preference for the surface scope reading, although the inverse scope reading was not entirely rejected by English native speakers.

However, the pragmatic bias caused by scalar implicatures makes the inverse scope reading pragmatically infelicitous. The preference for the surface scope reading in *negation-every*

[11] Additionally, aspects of world knowledge, such as plausibility, may influence one's interpretation of such sentences. In this case, it is more plausible that someone is less likely to skip some meals rather than skipping every meal. Further investigation is needed to explore the role of plausibility in quantifier scope interpretation.

sentences is also consistent with Anderson's Processing Scope Economy principle, as the surface scope reading can be derived without covert displacement at the structural level. Furthermore, the preference for the surface scope reading can be explained by O'Grady's efficiency-based processing theory. In sentences where negation occurs before the verb phrase (VP), both surface scope and inverse scope readings are equally available when encountering the universal quantifier phrase in the object position. The higher frequency of encountering the surface scope reading in language use leads to its greater activation, thus resulting in its preference. However, O'Grady's processing-based emergentist account has faced criticism for its failure to account for various other Chinese scope phenomena, including cases where the same word order yields different scope preferences (Crain, 2013).

2.2 Chinese

Like English, Chinese also allows the presence of logical operators (i.e., quantifier, negation) in different grammatical positions (i.e., subject or object), and a particular interpretation arises as a function of the way these operators are arranged in the surface syntax. Unlike English, Chinese has been claimed to be much more rigid in its ability to allow the inverse scope reading for a quantified sentence (Aoun & Li, 1989, 1993; Huang, 1981; Huang, 1982). According to this claim, Chinese, in contrast to English, lacks scope ambiguity. However, this claim was not fully supported in empirical studies with mixed evidence for the absence of the inverse scope reading for a sentence containing quantifier and/or negation. This section introduces basic linguistic facts concerning quantifier scope interpretation in Chinese and their theoretical analyses. Empirical studies regarding the interpretation of quantified sentences by Chinese native speakers are also reviewed. This section is comprised of two sub-sections with section 2.2.1 for doubly quantified sentences and section 2.2.2 for negatively quantified sentences.

2.2.1. Doubly quantified sentences

The Chinese counterparts of (3) and (4) are illustrated in (19) and (20) respectively.

(19) Mei-yi-ge haizi dou pa-le yi-ke shu.

Every-one-CL child DOU climb-PERF[12] one-CL tree

'Every child climbed a tree.'

(20) You yi-ge haizi pa-le mei-yi-ke shu.

Have/exist one-CL child climb-PERF every-one-CL tree

'A/one child climbed every tree.'

Before analyzing the interpretation of (19) and (20), it is important to address some specific linguistic properties found in Chinese quantified sentences. These properties apply to sentences with double quantifiers and those with a combination of a universal quantifier and negation. First,

---

[12] Abbreviations used in this dissertation are as follows: DOU = distributive universal quantifier (roughly corresponding to 'all' in English); PERF = perfective marker (roughly corresponding to '-*ed*/-*en*' in English); CL: classifier (absent in English); DE = genitive marker (roughly corresponding to *'s* in English); LOC = localizer.

sentence (20) is not literally translated from English whose Chinese counterpart would otherwise be *Yi-ge xiaohai pa-le mei-yi-ke shu*[13] if translated in a word-by-word fashion. The predicate 'you' is added right before the numeral phrase patterned as [numeral + classifier + NP], which is because numeral phrases in the subject position[14], if not preceded by an existential predicate *you*, would be considered as indefinite and therefore interpreted as non-specific and non-referential (Lee, 1986). Given the topic-prominent nature of Chinese (Li & Thompson, 1989), NPs in the subject position tend to be interpreted specifically and referentially (Yang & Wu, 2020). The addition of the existential predicate *you* makes it the case that the pattern of '*you* + numeral phrase' receives a specific and referential interpretation. As such, (20) could be interpreted as: there must be (at least) one individual assumed by the interlocutor in the discourse who has climbed every tree. Such semantic properties may potentially influence how quantified sentences with numeral phrases in the subject position are interpreted.

Second, unlike English, Chinese lacks an article system. Although *yi* 'one' is numeric in nature, it is not only doing the job of counting the referents associated with the nouns following it

---

[13] As an exploration, I conducted a survey of 12 adult Chinese speakers on their acceptability of (20) and the one without *you*. Specifically, they were asked to indicate whether (20) and the one without *you* were acceptable or not and further indicate which one (in comparison) was more acceptable. It was found that 11 out of 12 (92%) accepted both sentences and all of them indicated that the one with *you* was more acceptable. There could be two possible reasons for the finding that sentences without *you* are acceptable. First, albeit structurally being indefinite phrases, patterns in *yi* + classifier + NP are in any case interpreted as specific when they appear in the subject position. Second, the extent to which a sentence with a numeral subject is acceptable may depend on whether its object is quantified or not. Compared to sentences containing a quantified object and a numeral subject without *you*, sentences such as *yige-nanhai chi-le wo de pingguo* (A boy ate my apple) are much less acceptable.

[14] Sentences such as (19) would be completely acceptable when they contain a numeral phrase in the object position.

but also used to mark indefiniteness for the NP with which it co-occurs. When *yi* ‘one’ is unstressed, it arguably functions like an English indefinite article (Li & Thompson, 1989). Therefore, it is not entirely clear whether *yi* ‘one’ behaves merely as a numeral or to some extent as an indefinite article especially when compared to its English counterpart.

Third, while the particle DOU in Chinese has *all* in English as its rough counterpart, its functionality and the grammatical environments in which DOU can occur appear to be much wider in range in comparison to *all* in English, due to which many of the structures containing DOU in Chinese cannot find an equivalent in English. Because of its multi-faceted nature, DOU can be treated as a universal quantifier distributer and free choice item licensor, among others (Cheng et al., 2013; Xiang, 2016; Zhou & Crain, 2011). On the assumption that quantifier scope interpretation in Chinese may relate to the linguistic knowledge one has about DOU, the use of DOU through those different mechanisms are briefly reviewed. First, DOU is used as a universal quantifier distributer that universally quantifies and distributes over expressions preceding it (Cheng, 1995; Lee, 1986; Pan, 2006). One grammatical context in which DOU is used as such consists of a quantifier phrase in the subject position and DOU in a preverbal position, as in the case of (19). This sentence means that each of the relevant children climbed their own tree. It would otherwise be false if one of the children failed to climb a tree. Therefore, since DOU largely renders sentences as (19) a distributive reading, there is a good reason to think that DOU may also contribute to the surface scope reading for (19) in addition to interaction between the universal quantifier and the existential quantifier. Also, in the context where DOU is used as a universal quantifier distributer, the quantified elements include not only subjects but also topics as in (21) and place adverbials as in (22) (Li, 2012).

(21) Suoyou de pingguo wo dou chi-le.

All DE apple I DOU eat-PERF

'I have eaten all apples.'

(22) Fangjian-li daochu dou hen luan.

House-LOC everywhere DOU very messy

'It is very messy everywhere in the house.'

DOU can also be used to license a free choice item, such as preverbal *wh*-words, i.e., *shenme* 'what' and *shui* 'who', as in (23).

(23) Wo shenme rou dou bu chi[15].

I what meat DOU not eat

'I don't eat any meat.'

In this case, *wh*-words are not interpreted as interrogatives (e.g., *Zhangsan mai-le shenme*? 'What did Zhangsan buy?'), but as universally quantified NPs. For the *wh*-word to be interpreted as universal (Ladusaw,1979; Saebo, 2001), DOU must be present serving as a binder (Cheng, 1994).

[15] The sentence without DOU, if not ungrammatical, would be interpreted as an interrogative, meaning 'what kind of meat do I not eat?'.

Given that DOU is mandatory through all these cases and complicates the interpretation of quantifier phrases, it is reasonable to assume that a learner's knowledge about this morpheme may somehow interact with the interpretation of the quantified sentences in question. In this case, evidence appears to be available from the input and may pre-empt L1 transfer in the absence of explicit classroom instruction. (Note that learner's ability to use explicit input is called into question by some theorists, e.g., Truscott & Sharwood Smith, 2011 and discussion in Juffs & Fang, 2022).

As has been discussed in 2.1.1, doubly quantified sentences are scopally ambiguous in English, permitting two possible reading through QR rules. According to the theoretical literature, Chinese has been argued to exhibit scope rigidity in the sense that it permits only the surface scope reading but not the inverse scope reading. Accordingly, the only interpretation available for (19) is that every child climbed a different tree and for (20) is that the same child climbed every tree.

To account for the cross-linguistic difference between Chinese and English in exhibiting scope ambiguity, Huang (1982, p.220) proposed the *General Condition on Scope Interpretation*, which was later dubbed as the *Isomorphic Principle* (IP)[16] as in (24) by Aoun and Li (1989, p142; 1993, p15).

---

[16] It should be noted that this principle has its predictive power restricted to simple active sentences such as doubly quantified sentences for their distribution of scope interpretation. Complex sentences such as passives and double object constructions for their scope interpretation cannot be predicted by this principle, because these sentences in Chinese would otherwise be unambiguous under this principle but in fact, they are ambiguous. Empirical evidence for the observation that complex structures are scopally ambiguous has not been available until very recently (Gan & Tsai, 2020; Larson & Wu, 2018). A unified formal analysis for scope interpretation across languages was put forward by Aoun and Li (1989, 1993), which I introduce later in this section.

(24) Suppose A and B are QPs. Then if A-commands B at S-Structure, A c-commands B at LF.

The essence of this principle is that the c-commanding relationship between quantifiers at S-Structure that should be preserved at LF determines scope interpretation. On this principle, for (19) and (20), because the subject quantifier phrase c-commands the object quantifier phrase at S-Structure, the subject c-commands and hence takes scope over the object at LF, thus only giving rising to the surface scope reading.

Despite the Isomorphic Principle, English may permit a restructuring process, due to which English can be scopally ambiguous. According to Huang (1982), English and Chinese parametrically differ in head-directionality. Simply put, whereas English is a head-initial language, Chinese is a head-final language (cf. Cheng & Sybesma, 1999). The main argument is that the inverse scope reading in English for sentence such as (19) and (20) can be derived through restructuring by which the object NP in (19), for example, can be analyzed as a phrase adjoined to the head of TP (T) for the inverse scope reading to be generated. By contrast, restructuring is prohibited in Chinese because cases where the object NP adjoins to T would otherwise violate the head-final constraint. Thus, Chinese only allows the surface scope reading which is made possible in cases where the object NP is analyzed as the sister of V and the head-final constraint is thus respected[17]. To accommodate a wider range of data

[17] It is important to note that the view presented here represents the generative linguistic perspective among others. According to this view, scope interpretation is primarily a syntactic operation driven by the c-command relationship

including simple active clauses, Aoun and Li (1989, 1993) argued that the contrast between English and Chinese in scope interpretation results from structural differences between the two languages.

According to Aoun and Li (1989, 1993), English and Chinese fundamentally differ in constituent structure at the clause level – namely that the subject in English is base-generated in the Spec of VP and raises to the Spec of TP during syntactic derivation while the subject in Chinese is base-generated in the Spec of VP and stays *in situ* without through subject raising at the surface structure. The lack of subject raising in Chinese arguably is due to the nature of Infl(ection) in this language. Under the framework of Chomsky (1986), subject raising is made possible only when the process of V-raising to Infl could take place. The reasoning then is that if Infl in Chinese is claimed to be 'degenerate'[18], V-raising should be prohibited, and consequently subject raising would not take place. The alternative account would be that the NP trace left after subject raising would not be lexically governed (Stowell, 1985), on the assumption that Chinese Infl is not lexical due to its degenerate nature. As a

---

between various quantifier phrases. The cross-linguistic variation in scope interpretation preferences, as discussed in this study, is argued to arise from structural differences between English and Chinese. On the other hand, alternative approaches attribute scope interpretations to the interplay of diverse sources of linguistic information, positing that different linguistic levels interact to yield specific scope interpretations. Furthermore, the scope-rigidity argument concerning Chinese quantifier scope has been primarily based on doubly quantified sentences. The evidence presented in this dissertation, particularly from Chinese negatively quantified sentences, introduces fresh perspectives for this argument.

[18] This may also be attributed to the absence of overt tense and agreement markers in Chinese, although they are not explicitly morphologically marked.

result, English and Chinese would be mapped to (25a) and (25b) respectively – which are taken to be two possible primitive LF configurations for any structure containing two quantifiers in any language.

(25) a. Q_i [… t_i … Q_j [… t_j …]

b. Q_i [… t_i … Q_j [… t_i … t_j …]

In (25b) as for the case of English, either Q_i or Q_j could take scope over the other, because 1) Q_i c-commands Q_j c-commands Q_j ; 2) Q_j c-commands the NP trace t_i left by Quantifier Raising, which is a member of the chain containing $Q_{i,}$ following the Scope Principle (SP) proposed by Aoun and Li (1989, p.141) - namely that A quantifier A has scope over a quantifier B if A c-commands a member of the chain containing B (for a detailed review on these processes, see Kiss & Pafel, 2017, p. 14-15). As a result, on the surface, Q_i symmetrically c-commands $Q_{j,}$ deriving both readings – hence the ambiguity of doubly quantified sentences in English; by contrast, as represented in (25a), Q_i asymmetrically c-commands Q_j in Chinese, giving rise to the surface scope reading only. With such analyses in mind, one may wonder why Chinese does not allow covert quantifier displacement (via quantifier raising or lowering) for the inverse scope reading to be derived, as in the case of English discussed in 2.1.1.

Although QR presumably is operative across languages, Aoun and Li's theory (1989, 1993), if adopted, would complicate the answer to this question. Take (3) and its Chinese counterpart as in (19) as an example. Assuming that QR were applicable in both languages, this sentence could be ambiguous between the surface scope reading (26a) and the inverse scope reading (26b), depending on when QR applies to the quantified expressions in the sentence.

(26) a. [Every $child_j$ [a $tree_i$ [t_j climbed t_i]]]

b. [[A $tree_i$ every $child_j$ [t_j climbed t_i]]]

Aoun and Li (1989, 1993) also proposed the Minimal Binding Requirement (MBR) – namely that variables must be bound by the most local potential antecedent (Ā-binder), according to which the LF in (26a) and (26b) should be ruled out. In (26a), *a tree* is the most local Ā-binder for t_j and t_i. However, *every child* rather than *a tree* is the antecedent of t_j, hence violating MBR. MBR is also violated in the case of (26b), because *every child* is the most local Ā-binder for t_j and t_i, but t_i in fact has *a tree* as its real antecedent. The LF for Chinese quantified sentences has been argued to be represented as in (25a) where QR is present by which quantified expressions (Q_i, Q_j) are moved and MBR is not violated at the same time. (25b) is the abstract LF representation for English where SP is operative and MBR is also not violated particularly in the case of deriving the inverse scope reading, because the NP trace (not a variable) of Q_i (the t_i between Q_j and t_j in (22b)) c-commanded by Q_j is not subject to MBR.

In summary, there are at least two takeaways from the theoretical frameworks discussed above: (1) the presence of scope ambiguity across languages does not depend on whether a language allows covert quantifier displacement as a way to generate the inverse scope reading but rather depends on the way LF is represented in each language and language-specific constituent structures affect the form of LF; (2) the approaches to Chinese scope interpretation represented by Huang (1982) and Aoun and Li (1989, 1993), albeit different in their predictive power and eventually the range of data to which these approaches apply, claim that Chinese lacks the inverse scope interpretation for doubly quantified sentences in the context of simple actives.

As for the empirical work, much less has been done on doubly quantified sentences in Chinese than in English. Except for some L1 and L2 acquisition studies where the data on the interpretation of Chinese doubly quantified sentence by native speakers serve as the baseline (e.g., Chu et al., 2014; Wu & Ionin, 2022), there are only a few studies that have so far directly addressed the extent to which the inverse scope reading was allowed in Chinese doubly quantified sentences (Scontras et al., 2017; Tsai et al., 2014; Zhou & Gao, 2009). As one of the earliest studies, Zhou and Gao (2009) examined the interpretation of Chinese doubly quantified sentences in which the universal quantifier precedes the existential quantifier. In a truth-value judgement task where *Mei-ge ren dou qu-le yi-jia gongchang* (Everyone went to a factory) was presented under context in favor of either a surface scope reading (e.g., for each person in the context, she/he went to a different factory) or inverse scope reading (e.g., for each person in the context, they went to the same factory), Chinese native speakers were asked to rate on a 5-point scale whether each sentence matches the meaning of the corresponding context. The finding that mean ratings of test sentences under the inverse scope context were greater than 3 (out of 5) was interpreted as evidence that inverse scope readings are available in Chinese under appropriate contexts even though the surface scope reading was more acceptable than the inverse scope reading. In addition, this study also found that the availability of inverse scope readings was affected by verb type in that such readings were less readily available in quantified sentences containing action verbs compared to when test sentences contained locative or psych verbs, again confirming the role of lexical knowledge in affecting scope interpretation as has been found in the case of English doubly quantified sentences (Kurtzman & MacDonald, 1993).

A set of other studies found that inverse scope readings for doubly quantified sentences were not available in Chinese (Scontras et al., 2014; Scontras et al., 2017; Tsai et al., 2014). In support of this finding, those studies mainly examined the interpretation of doubly quantified sentences in which the existential quantifier precedes the universal quantifier as sentences in (20). The use of such sentences would avoid the potential *entailment* issue. It remains unclear whether sentences like (19) are truly scopally ambiguous because its inverse scope reading entails its surface scope reading: If there was a specific tree that was climbed by every child, it follows that every child climbed a tree. As a result, it would be more appropriate to showcase whether Chinese doubly quantified sentences are scopally ambiguous by examining doubly quantified sentences with an existential quantifier phrase in the subject position.

In this spirit, Scontras et al. (2017) investigated interpretation of doubly quantified sentences by Chinese native speakers using a sentence-picture matching task. In this task, a single trial consisted of an aurally presented sentence and a picture. Participants were asked to rate whether the heard sentence appropriately matched the picture on a 7-point Likert Scale. Both types of doubly quantified sentences were tested. For *every-a* sentences in Chinese, although the surface scope reading on average was much higher than the inverse scope reading, the inverse scope reading-whose mean ratings were above 3.5 (out of 7)-was not impossible. The critical results were those from *a-every* sentences: Its inverse scope reading almost reached a floor level with the mean rating of 1.56 (out of 7). Tsai et al. (2014) used the same method and replicated Scontras et al. (2017). Although the patterns of results in Scontras et al. (2014) resembled those in Scontras et al. (2017), Chinese participants of Scontras et al. (2014), which used a binary truth-value judgment task, had judged the *a-every* sentence against the picture for an inverse scope reading 0% of the time. These studies seem to provide strong evidence that Chinese lacks inverse scope readings for doubly quantified sentences. Nevertheless, given that only a quite limited number of studies have been conducted, more research in this regard is of course needed.

One important methodological consideration to highlight is the choice between scalar judgment and binary judgment tasks. In the field of experimental semantics, studies have traditionally employed binary truth-value judgment tasks that require participants to make a binary choice (e.g., True-False) regarding the truth-value of a statement. For instance, studies conducted by Scontras et al. (2014) and Montrul and Ionin (2010) utilized such binary judgment tasks. However, it has been suggested that scalar judgments have a particular advantage over binary judgments in detecting semantic ambiguities: Scalar judgments could reveal an interpretation that would otherwise be kept hidden with binary judgments, because binary judgments tend to be driven by a reading that is much preferred and more readily accessible (Marty et al., 2015, 2020). While it remains an empirical question as to whether different types of scale would give rise to different experimental effects, it is generally accepted that the use of scalar judgment allows for better capturing the gradience and nuanced contrast of grammar.

2.2.2 Negatively quantified sentences

It is also important to understand how scope interactions between negation and quantifier work in Chinese. Depending on the relative word order of negation and quantifier, two types of negation-quantifier scope interactions can be identified. The surface forms of these two types of interactions for the Chinese counterparts of (5) and (6) are shown in (27) and (28).

(27) Mei-yi-pi ma dou meiyou tiao guo liba.

Every-one-CL horse DOU not jump over fence

‘Every horse didn’t jump over the fence.’

(28) Zhe-pi ma meiyou tiao guo mei-shan liba.

This-CL horse not jump over every-CL fence

‘The horse didn’t jump over every fence.

In the literature, Chinese has been reported to allow only surface scope readings for both types of negation-quantifier interactions (Fan, 2017; Zhou & Crian, 2009). This appears to suggest that the generalization for scope rigidity in Chinese can be extended from doubly quantified sentences to negatively quantified sentences. However, to my knowledge, no theoretical explanation has been advanced specifically for the lack of inverse scope readings in negatively quantified sentences. Recall in section 2.2.1, the exclusive permission of surface scope readings for doubly quantified sentences suggests that scope in Chinese seems to be exclusively determined by the surface c-commanding relation between logical operators, following the Isomorphic Principle as formulated in (24). In the same spirit, the lack of inverse scope readings for negatively quantified sentences can also be explained with this principle. Furthermore, differences in the range of scope interpretations allowed by Chinese and English negatively quantified sentences should also be due to the cross-linguistic variation in constituent structures at the clause level, the reasoning that has been applied for the case of doubly quantified sentences.

There has been no systematic research thus far testing the availability of surface scope and inverse scope readings in Chinese negatively quantified sentences specifically among adult native speakers. Data about adult speakers’ interpretation of negatively quantified sentences primarily come from child language acquisition as the baseline (Fan, 2017; Zhou & Crain, 2009). In a study set up for uncovering how Chinese-speaking children understood the scope relation between the universal quantifier and negation, Fan (2017) investigated the interpretation preference for sentences structurally similar to (27) and (28) using a truth-value judgment task. Slightly different from (28), they tested sentences in which negation precedes a universal quantifier and appears in the subject position, as exemplified in (29).

(29) Bushi mei-pi ma dou tiaoguo-le liba.
Not-be every-CL horse all jump over-PERF fence
'Not every horse jumped over the fence.'

The results showed that adult speakers of Chinese accepted (27) as descriptions of contexts in support of surface scope readings 100% of the time, but utterly rejected such sentences as descriptions of contexts towards inverse scope readings. As for (29), its surface readings were completely accepted by adult speakers and its inverse scope readings were much less preferred with the participants accepting it only 26.67 % of the time.

As the baseline data, adult speakers were also tested for their interpretation of interactions between universal quantifier and negation in Zhou and Crain (2009). Different from Fan (2017), the sentence stimuli in Zhou and Crain (2009) were bi-clausal sentences with an affirmative sentence preceding the critical test sentence. The lead sentence arguably could satisfy the felicity condition on the use of negation, thus making the access of inverse scope readings more readily (Viau et al., 2010). It turned out that the adult participants accepted both (27) and (29) for their surface scope readings only (although children were much more tolerant of inverse scope readings across the board). The discourse context due to the lead sentence does not seem to facilitate the access of the inverse scope reading in this case. That said, these two studies differ in the context for the interpretation to be evaluated against, making their results hard to compare. The fact that (29) on its inverse scope reading was not completely impossible in Fan (2017) might be due to the combined use of both video story and pictures to present test stimuli. Unlike Fan (2017), Zhou and Crain (2009) used act-out stories. As a further note, it remains unknown how interpretation preference may differ between (28) and (29) because of the differences in the linear distance between negation and quantifier and the kinds of negators (*mei* vs. *bu*) used in the two sentences. Nevertheless, it provides empirical evidence for the generalization that Chinese speakers consistently prefer the surface scope reading for negatively quantified sentences. Moreover, since both studies used binary truth-value judgment tasks, studies in the use of scalar judgments are needed to avoid the aforementioned issues with such tasks.

2.3 A comparison between English and Chinese

In comparison, English and Chinese seem to differ in their ability to allow surface scope and inverse scope interpretations for sentences with interactions between logical operators such as universal quantifier, existential quantifier, and negation. Table 2.1 summarizes scope interpretation possibilities in English and Chinese based on the L1 theoretical and empirical literature. While inconclusive, the theoretical and empirical literature reviewed thus far seem to point towards a tendency that English permits the availability of inverse scope interpretations to a much larger extent than Chinese. Therefore, English sentences with two quantifiers or one quantifier and a negation are scopally ambiguous, allowing for two possible interpretations when a clear enough context is available. In contrast, Chinese is largely unambiguous with the surface construal prevailing, even in the face of a felicitous context. In English, inverse scope readings for quantified sentences are generated through quantifier movement either by raising or lowering quantifiers at LF. In Chinese, the lack of inverse scope readings is explained by the Isomorphic Principle, according to which the c-commanding relationship between logical operators in the surface syntax is preserved at LF. As a postulation, fundamental differences in the constituent structure between English and Chinese give rise to the observed contrast in the extent to which English and Chinese permit inverse scope interpretations.

The grammar of English syntax and semantics predicts the availability of both scope interpretations; yet these two interpretations are not equally accessible with one being more preferred than the other. As reviewed, one's preferred scope interpretation is not only determined by grammatical factors but also influenced by a range of other factors (external to the grammar) such as pragmatics, task demands, parsing, and discourse context. Overall, surface scope readings are preferred over inverse scope readings for English quantified sentences, which can be well accounted for by the Processing Scope Economy principle on which inverse scope readings derived through covert movement would yield a higher processing cost than surface scope readings. Preference for surface scope readings has been quite consistently observed for doubly quantified sentences. In the literature, the evidence for the preferred interpretation for *every-negation* sentences is mixed, arguably due to task effect (offline tasks for inverse scope preference vs. online tasks for surface scope preference). It would then be helpful to examine whether the same groups of participants perform differently across tasks in one single study. As for *negation-every* sentences, albeit not impossible, the inverse scope readings would be pragmatically incompatible with such sentences, hence much less preferred compared to the surface scope readings.

In the case of Chinese, Chinese speakers seem to consistently accept surface scope readings only for doubly quantified sentences and for *every-negation* sentences as evidenced in very few empirical studies. In a study by (Fan, 2017), inverse scope readings for *negation-every* sentences were not entirely impossible, presumably because of the rich contextual support from the combined presentation of picture and video story. Nevertheless, there are relatively few L1 studies on the interpretation of Chinese quantified sentences, and these studies have primarily relied on offline tasks, making it challenging to compare empirical evidence between English and Chinese. Therefore, the L1 data collected in this study from Chinese speakers serves as both the baseline for L2 performance and provides additional empirical evidence for the preferred scope interpretation of quantified sentences in native speakers and the factors influencing their interpretation.

Table 2.1 Summary of scope interpretation possibilities across languages based on L1 literature

Scope relation	Word order	Example sentences	Language	Interpretation	
				Surface	Inverse
Doubly quantified	$\forall > \exists$	Every child climbed a tree.	English	√√	√
			Chinese	√	×
	$\exists > \forall$	A child climbed every tree.	English	√√	√
			Chinese	√	×
Negatively quantified	$\forall > \neg$	Every horse didn't jump over the fence.	English	√	√
			Chinese	√	×
	$\neg > \forall$	The horse didn't jump over every fence.	English	√√	?√
			Chinese	√	?×

In Table 2.1, I use '√√' to represent a reading that is strongly preferred over the other reading which is nevertheless accessible as represented by the symbol '√'. The symbol '×' represents a reading that is inaccessible and unacceptable to native speakers of a language. Cases where a reading is not stipulated by UG but is not utterly impossible due to pragmatic support are represented by '?×'. The symbol is '?√' is used to stand for cases where a reading is made possible through UG but is marginally (un)acceptable due to pragmatically being infelicitous.

3.0 Second language research on scope phenomena

There have roughly been two areas of investigations on quantification in L2. One area of concern has been the distribution of existential quantifiers (e.g., *any, wh-existentials*[19]) (form of the grammar) in L2 learners, as evidenced by grammaticality judgment data (Gil et al., 2019; Gil & Marsden, 2010; Yuan, 2010). For example, sentences like ‘I have already drunk anything’ would be considered ungrammatical because the use of ‘any’ as an existential quantifier typically occurs in negative or interrogative contexts (downward entailing environments). The other area has focused on knowledge of L2 grammar and learners’ knowledge of the potential to accept/recognize the range of interpretations. The most extensive work in this area has been on the interaction between quantifiers and/or negation (Chu et al., 2014; Chung, 2009, 2013; Chung & Shin, 2022; Kim, 2010; Kwak, 2010; Lee, 2009, 2010; Marsden, 2009; Özçelik, 2018; Scontras, et al., 2017; Shen & Chen, 2022; Wu & Ionin, 2019, 2022). The target group of learners tested so far has been disproportionately limited to Korean speakers, especially for the interpretation of quantifier-negation interaction. Moreover, conclusions drawn from the few studies that exist have mainly relied on offline measures.

This section is devoted to a systematic review of prior studies on the acquisition and processing of quantifier scope interpretation by adult L2 learners. I also summarize and critique relevant theoretical explanations to the findings of existing studies. Section 3.1 provides a review of studies on quantifier scope in L2 acquisition at length. These studies mainly deployed offline

[19] *Any* in English has its counterpart realized as *wh*-existentials in some East Asian languages including Chinese, Korean, and Japanese (Cheng, 1994; Nishigauchi, 1990, inter alia).

methods to tap into possible meaning(s) of quantified sentences. Section 3.2 reviews studies on quantifier scope in L2 processing whose focus of inquiry was on how such meanings are derived and accessed by learners. Despite this distinction, it is nevertheless hard to discuss one without referring to the other in the sense that learners' acquisition and processing are linked developmentally (Carroll, 2001; Phillips & Ehrenhofer, 2015).

3.1 Quantifier scope in L2 acquisition

Investigation of the L2 acquisition of quantifier scope varies in many aspects such as testing learners with different L1s, the use of different methodologies, and the stimuli being different in one way or another (Chu et al., 2014; Chung, 2009, 2012; Kim, 2010; Kwak, 2010; Lee, 2009; Marsden, 2009; Scontras, et al., 2017; Shen & Chen, 2022; Özçelik, 2018; Wu & Ionin, 2019, 2022). Due to such variation, the evidence for whether L2 learners can achieve native-like competence in scope interpretation is mixed. In the case of doubly quantified sentences with offline methods (e.g., 'Someone dropped every plate'), Chu et al. (2014) and Wu and Ionin (2022) are the only two studies that investigated the interpretation of English sentences with a universal quantifier and an existential quantifier by Chinese-speaking learners of English. Chu et al. (2014) investigated the interpretation of such sentences as *Someone dropped every plate* by Chinese-speaking learners of English with a sentence-picture matching task. They found that learners accepted such sentences in contexts that supported surface scope readings (only one person) and yet (even advanced learners) rejected them in contexts in support of inverse scope readings (one person for every plate dropped), unlike English native speakers who accepted both readings although inverse scope readings were less preferred than surface scope readings. They attributed learners' pattern of results to Anderson's (2004) Processing Scope Economy principle on which surface scope readings were derived with less processing costs than inverse scope readings. One problem exists for this explanation: If this principle was at play, learners and native speakers should behave similarly in inverse scope readings; however, native speakers accepted such readings to a larger extent than learners.

Similar patterns of results were also observed in Wu and Ionin (2022) who examined the interpretation of sentences like 'A dog scared every man' in a context-based acceptability judgment task. Native speakers and L2 learners were found to accept surface scope readings for test sentences. While less acceptable than surface scope readings, inverse scope readings overall were accessible to English native speakers. However, given that Wu and Ionin (2022) included a mixture of action verbs and psych verbs in their test sentences, their results need to be interpreted with caution. Moreover, neither study examined the influence of the relative word order of quantifiers. According to Kurtzman and MacDonald's (1993) Single Reference Principle, all things being equal, it should be the case that the surface scope reading for sentences with an existential quantifier in the subject position is more salient than that for sentences with an existential quantifier in the object position. It remains to be seen whether L2 learners of English would demonstrate sensitivity to this principle. This can be determined by comparing their interpretation of sentences with an existential quantifier in subject vs. object positions for their surface scope readings.

L2 acquisition of quantifier scope has also been studied in languages other than English. For example, Marsden (2009) tested learners of Japanese speaking English and Korean as their L1s on the acquisition of the interpretation of doubly quantified sentences in Japanese. Japanese and Korean patterned similarly in terms of doubly quantified sentences such as 'Someone read every book'. This sentence in its canonical word order is not ambiguous and both languages allow only the surface scope reading. However, its scrambled counterpart permits both surface scope and inverse scope readings, patterns like those in English. One major finding is that intermediate English-speaking learners mistakenly accepted the canonical doubly quantified sentences for their inverse scope readings whereas Korean-speaking learners rejected such readings. This was taken to suggest a role for L1 transfer in affecting scope interpretation. The other critical finding is that some advanced English-speaking learners correctly rejected the canonical sentences on their inverse scope readings, indicating that they were able to acquire the absence of inverse scope readings in the target language. Scrambled sentences on both readings presented no difficulty for all learners. Advanced learners contrasted in their performance on inverse scope readings observed in Chu et al. (2014), Wu and Ionin (2022), and Marsden (2009), due to which it is reasonable to assume that the acquisition of scope interpretation in L2 learners benefits significantly from positive evidence in L2 input and the application of constraints proposed in formal grammar. However, these factors have distinct impacts on the target-like acquisition process.

There has been limited investigation on double quantifier scope in L2 Chinese, except for Scontras et al. (2017), who tested English-dominant Chinese heritage speakers[20] on doubly quantified sentences such as 'A shark attached every pirate/Every shark attacked every pirate'. During a picture-based acceptability judgment task, heritage speakers exhibited a tendency to reject inverse scope readings in Chinese. This was evident in their lower ratings compared to the English baseline for inverse scope readings (e.g., 2.79 vs. 4.46 on a 7-point scale, for *a-every*

[20] Due to presumed differences in the quality of input received and age of first exposure, the learning outcomes of L2 learners and heritage speakers may differ drastically. Therefore, it is ideal to examine them separately.

sentences). This was taken to suggest that heritage speakers of Chinese seemed to parallel with the Chinese baseline for lacking scope ambiguities. However, heritage speakers of Chinese rated inverse scope readings higher than the Chinese baseline (e.g., 2.79 vs. 1.56, for *a-every* sentences). The authors did not attribute the higher ratings to transfer from English because another group of heritage speakers of Chinese were found to reject inverse scope readings on English doubly quantified sentences. Instead, ratings being higher on inverse scope readings for heritage speakers than for native Chinese speakers were argued to result from yes-bias in experimental settings because heritage speakers presumably were less confident in making judgments in their weaker grammar. As a result, the finding that the heritage speakers lack scope ambiguities in their dominant language, English, and weak language, Chinese, was interpreted as evidence that heritage speakers (presumably much less influenced by language transfer compared to L2 learners) prefer a grammar that would yield scope interpretations with smaller processing costs due to structural complexity and this grammar of choice happens to be Chinese. However, the linguistic profile of the tested heritage speakers in their study remains unknown and variation in age of acquisition and language learning experience may lead to different leaning outcomes among individual heritage speakers or between heritage speakers and adult L2 learners (Romano & Guijarro-Fuentes, 2023).

With respect to the interaction between a universal quantifier and an existential quantifier as reviewed above, the interaction between quantifier and negation constitutes another type of quantifier scope. Doubly quantified sentences and negatively quantified sentences appear to share similarities in L2 acquisition, as fine-grained knowledge for interpreting scope relations is unlikely to be readily available in surface input. This observation suggests the possibility of a classical Poverty-of-the-Stimulus learning situation (Chomsky, 1986; White, 2003). These structures differ in at least two respects in their potential to inform L2 acquisition theories. One is that interpretation of *negation-every* sentences involves the computation of scalar implicatures, an investigation of which relative to *every-negation* sentences may allow for a more fine-grained inquiry into different types of interface conditions (syntax-pragmatics vs. syntax-discourse) in L2 acquisition. The other is that (in addition to DOU in both classes of quantifier scope) the existential quantifier (pertaining to the definiteness properties of the noun phrase represented by an existential quantifier, e.g., 'a child') present in double quantifier phenomena may serve as triggering input to inform learners of a linguistic property that would otherwise render much unlikely for acquisition (if without recourse to UG), since immediate evidence for it is absent in the input. Alternatively, the requirement of DOU in both classes of quantifier scope may be positive triggering input.

While some attention has been devoted to the interaction between quantifier and negation in L2 studies, much of the research has focused on L2 acquisition of English by native Korean speakers (Chung, 2009; Chung, 2013; Kim, 2010; Lee, 2009) and very few studies have been conducted with learners speaking other L1s including Chinese (Özçelik, 2018; Wu & Ionin, 2019). Given that Korean relative to Chinese behaves differently in the preferred reading for sentences containing a universal quantifier that follows negation (Lee, 2009), investigations on L2 learners with Chinese as their L1 would provide new insights into the role of cross-linguistic differences in scope interpretation.

Most of the prior studies that looked at the acquisition of the interaction between quantifier and negation in L2 English have been conducted with Korean-speaking learners of English, of which some have only examined *negation-quantifier* sentences (Chung, 2013; Kwak, 2010) and a few others have explored both types of such interactions manifested in sentences with quantifier in either the subject or the object position (Chung, 2009; Kim, 2010; Lee, 2009). While under different theoretical approaches, namely UG-based (e.g., White, 2003) vs. efficiency-based processing (e.g., O'Grady et al., 2009), the general patterns of results emerging from these studies are that learners of English, though not necessarily at low proficiency, tended to accept only the *every-not* reading no matter where the quantifier occurs, different from native English speakers who showed a preference for the surface scope reading with *negation-quantifier* sentences and for the inverse scope reading with *quantifier-negation* sentences. For example, using a truth-value judgment task, Chung (2009) tested Korean-speaking learners of English in their interpretation of both types of interactions in English. The results showed that learners of low proficiency accepted *every-not* readings but rejected *not-every* readings for both types of negatively quantified sentences. Although advanced learners like native speakers seemed to be ambivalent in judging most of the combinations, they consistently rejected the *every-not* reading for sentences with the quantifier in the object position.

Taken together, these findings point towards the role of L1 transfer in scope interpretation. In Korean, due to the SOV basic word order, object always precedes negation, giving rise to the fact that Korean exhibits the same scope configuration, i.e., every > negation, no matter where the quantifier occurs. As such, particularly for negation-quantifier sentences, Korean and English would differ in which reading prevails for the surface scope reading. Hence, the role of L1 transfer should be more evident and better manifested in the case of negation-quantifier sentences for Korean-speaking learners of English. For instance, Kim (2010) found that both intermediate and advanced Korean learners, in contrast to native English speakers, accepted the *every-not* reading for sentences with quantifiers in the object position, suggesting that L1 transfer effects persist even in advanced learners. However, caution should be taken when making a direct comparison between these two studies, due to differences in the modality of presenting test sentences (written vs. oral) and the type of scale used (binary vs. scalar). The inconclusive findings nonetheless leave unresolved questions about the roles of L1 transfer and UG, among others, in the domain of quantifier scope interpretation in L2.

As can been seen from current studies, acquiring the target-like interpretation of scope relations can be very difficult for L2 learners. This difficult arises partly due to the negative influence of L1 properties, but also because scope phenomena involve the interplay of different levels of linguistic representation, as outlined in Chapter 1: syntax, semantics, and pragmatics/discourse. According to the Interface Hypothesis in its more recent version (e.g., Tsimpli & Sorace, 2006; Sorace, 2011), *external* interfaces that require one to bring together modules of grammar (syntax, semantics, morphology, and phonology; their interactions to yield *internal* interfaces) with modules external to grammar such as pragmatics and discourse to arrive at interpretation(s) permitted by the grammar are particularly difficulty for learners even at the advanced level, leading to interlanguage grammars that demonstrate optionality and indeterminacy. Among the current studies on scope interpretation in L2, Özçelik (2018) stands out for its explicit examination of the impact of interface conditions. This study was to investigate how L2 learners of English and Turkish respectively interpreted sentences containing negation and a cardinal number, as in (30). In English, this sentence is ambiguous between (30a) as the surface scope reading and (30b) as the inverse scope reading. In Turkish, it is unambiguous, only with (30a) permitted. Özçelik argued that whereas both groups of participants must acquire quantifier scope for its operation at the syntax-semantics interface, learners of English (unlike learners of Turkish) may additionally be implicated by the Principle of Charity, a pragmatic principle to be employed for access of inverse scope readings (also see a discussion on the operation of this principle in Zhou and Crain (2009)). As such, L2 learners of English were predicted to experience greater difficulty than L2 learners of Turkish in behaving similarly to native speakers for scope interpretation. Of different proficiency levels, a group of 26 Turkish-speaking learners of English and a group of English-speaking learners of Turkish were tested in a truth-value judgment task.

(30) Donald didn't find two guys.

a. It is not the case that Donald found two guys. (Surface: not > two)

b. There are two guys that Donald didn't find. (Inverse: two > not)

Two main findings arose from this study: a. Turkish-speaking learners of English (particularly those at the intermediate and advanced levels) behaved similarly to native English speakers, accepting the inverse scope reading for sentences like (30); b. English-speaking learners of Turkish did not converge with the native control even at the advanced level, manifested in the contrast that such sentences on their inverse scope readings were always accepted by the learners but consistently rejected by native speakers of Turkish. As a result, the prediction that learners of English would experience a greater difficulty was not born out. The claim based on this finding is that external interfaces are not necessarily problematic and internal interfaces are not necessarily unproblematic, at least in the domain of scope interpretation.

Several points can be made with respect to this study. First, test sentences of Özçelik (2018) exemplified in (30) are the ones with negation appearing before the cardinal quantifier. A similar configuration that has been reviewed in other L2 studies, and one that the current dissertation investigates, involves a universal quantifier, as in (6) (i.e., The horse didn't jump over every fence).

Gricean implicature would arise for (6) only, but not for (30)[21], leading to the difference in the extent to which they allow inverse scope readings (Gennari & MacDonald, 2006; Lidz & Musolino, 2002). Second, it might be simply the word order cue (i.e., two > not; Donald iki (two) çocuk bul-ma (negation) -dı.) that Turkish-speaking learners of English relied on to determine the interpretation of (30) in its Turkish counterpart, leading them to always accept the inverse scope reading in Turkish. Therefore, the influence of interface conditions may somehow be complicated by the mismatch between the surface word order of logical operators and their c-commanding relationship. Third, the sample sizes of the groups in this study are relatively small, which may result in the interpretation of their results being inconclusive.

As far as the quantifier-negation interaction is concerned, this phenomenon has rarely been studied in the context of L2 Chinese. A quite recent study that reports on the interpretation of Chinese negatively quantified sentences has concerned English-dominant heritage speakers of Chinese (Shen & Chen, 2022). In Shen and Chen (2022), a group of 24 English-dominant heritage Chinese speakers were tested for their offline interpretation of sentences with universal quantifier preceding negation in both the English and Chinese versions (e.g., All teachers didn't use Donald's car). The authors found that the extent to which inverse scope readings in Chinese were allowed varied among heritage speakers, of whom only a few accepted such readings (9 out of 24). This is contrary to the claim that heritage speakers tend to simplify their grammars posited by Scontras et al. (2017) and Polinsky and Scontras (2020). However, this study suffers from several notable methodological issues that undermine its generalizability. The first issue is that input has not been

[21] As argued by Lidz and Musolino (2002), it is unclear how to precisely define the relative informative strength of the two scope readings for (30) and consequently how a scalar implicature could emerge.

considered for its potential interaction with the acceptance of inverse scope readings between participants, given that no information was provided regarding whether these heritage speakers of Chinese received any additional instruction in Chinese, either in or outside of the classroom. One additional issue is that test sentences were presented in the written format for the English version, but in both the written and aural format for the Chinese version, which could complicate patterns of results observed across languages. The further issue is that the experiment was not conducted with separate groups of participants for each version. Because the English version of their study was implemented before the Chinese version, participants who accepted the inverse scope readings for Chinese sentences may have been primed by similar readings that were activated during the English task.

3.2 Quantifier scope in L2 processing

Relative to the L2 acquisition of quantifier scope informed by offline measures, much less is known about how L2 learners process quantifier scope for its interpretation to be derived in real-time. To date, research on L2 processing of quantified sentences has been limited in both the scope of phenomena investigated and the diversity of target populations tested. Specifically, the focus of inquiry has predominantly been on the processing of sentences containing quantifier and negation by L2 learners with Korean as L1 (Chung, 2022; Lee, 2009, 2010, 2018). For performance on *negation-quantifier* sentences by Korean-speaking learners of English, evidence suggests that L1 transfer is operative and language proficiency overall does not modulate its influence (Lee, 2009, 2010). For example, in the study conducted by Lee (2010), a group of 42 Korean-speaking learners and a group of 24 native English speakers completed a self-paced reading experiment in which participants first read a short story in favor of either a surface scope interpretation or an inverse scope interpretation. A test sentence (e.g., The boy didn't eat every cookie.) was then presented segment-by-segment in a non-cumulative manner. The task for the participants was to respond whether the test sentence was a true or false statement for the story. Consistent with the literature, native speakers of English showed a strong preference for the surface scope reading (not > every) for sentences with a quantifier in the object position, manifested in higher acceptance for test sentences in contexts favoring surface scope readings and shorter reading times (RTs) regarding the True or False response of the surface scope reading. Learners diverged from the native control: Learners of low proficiency preferred the inverse scope reading (every > not) although learners of high proficiency seemed somewhat ambivalent without clear preference in terms of acceptance rates and response times for judging test sentences. L1 influence was in evidence here because quantifier always occurs prior to negation in Korean quantified sentences.

As for *quantifier-negation* sentences (e.g., Every horse didn't jump over the fence), Korean-speaking learners of English were found to consistently prefer surface scope readings (every > not) also in a self-paced reading task (Lee, 2009, 2018; Chung, 2022), which could also be explained in terms of L1 transfer. In fact, results of these studies of negatively quantified sentences were interpreted for their findings with reference to O'Grady et al.'s (2009) efficiency-based processing account for both L1 speakers and L2 learners, according to which *quantifier-negation* sentences, for instance, on the surface scope readings should be easier to access because such readings would be taxing working memory to a lesser extent compared to inverse scope readings.

This account has some problems for several reasons. First, studies with Korean-speaking learners of English that used offline measures also demonstrated a preference for ‘every > not’ interpretations regardless of the relative positions of quantifier and negation, presumably due to L1 transfer (Chung, 2009; Kim, 2010). The processing-based account would predict this to happen only when interpretation was evaluated under time pressure. Second, L1 speakers in Chung (2022) have been noted to be sensitive to the context under manipulation: Unless the context strongly biased towards surface scope readings, ‘not > every’ readings stood out as the preferred reading for *every-negation* sentences, contrary to what O’Grady et al.’s (2009) account would anticipate. In this sense, instead of simply applying certain parsing heuristics, comprehenders consulted different sources of linguistic information during the processing of quantified sentences. Third, according to O’Grady (2022), what has been transferred to L2 is the processing operation rather than the linguistic representation from L1 when it comes to the interpretation of sentences containing quantifier and negation. It is therefore unclear about the source of L1 transfer observed in offline performance.

3.3 Summary

In sum, while a few studies have been conducted to investigate whether L2 learners can interpret and process quantifier scope of the L2 as native speakers of that language, the results have been mixed. Relevant to the present investigation, it should be noted that prior offline studies with Chinese-speaking learners of English have predominantly focused on doubly quantified sentences featuring an existential quantifier preceding a universal quantifier (e.g., A child climbed every tree) and on negatively quantified sentences featuring a universal quantifier preceding negation (e.g., Every horse didn't jump over the fence). Except for the very few studies with English-dominant heritage speakers of Chinese on quantifier scope interpretation, this phenomenon thus far has not been systematically studied in English-speaking learners of Chinese. To date, although there have been several studies examining Korean-speaking learners of English for their online processing of quantifier scope in relation to negation, no such studies have yet been conducted with Chinese learners of English or English learners of Chinese.

In the studies reviewed, Chinese-speaking learners of English even at the advanced level were found to accept surface scope readings only but reject inverse scope readings for 'A/someone…every' sentences (Chu et al., 2014; Wu & Ionin, 2022) and for 'Every…not' sentences (Wu & Ionin, 2019). One explanation is that learners like native speakers may simply adhere to the principle of Processing Scope Economy and thus prefer the reading that involves a relatively simpler syntactic derivation; however, this explanation remains somewhat problematic because if this principle holds true for both L2 learners and native speakers, both groups would not differ in the extent to which they prefer inverse scope readings. What native English speakers differ from L2 learners is that they almost never rejected inverse scope readings (which are in fact more prevalent for *every-negation* sentences observed in some studies (e.g., Chung, 2009; Lee, 2010). It is also reasonable to assume that negative transfer from L1 may have contributed to Chinese speakers' difficulty in accepting inverse scope readings, as it does for Korean-speaking learners of English especially on their performance in negation-quantifier sentences (e.g., Chung, 2009; Kim, 2010).

Negative L1 transfer presumably could be overcome in the learning scenario where the learner's task is to add a new reading to their existing grammar and such a learning task in principle can be achieved based on the positive evidence from language input. As such, the inconsistent findings obtained in different studies regarding L2 learners' ability to arrive at the target-like interpretation for quantified sentences (cf. Chu et al., 2014; Chung, 2009; Kim, 2010; Marsden, 2009; Özçelik, 2018; Wu & Ionin, 2022) suggest that positive evidence for its effect on quantifier scope interpretation may not always be reliable or used. Therefore, it is important to scrutinize its influence on an individual, case-by-case basis. The mixed results could also be attributed to differences in experimental setups across studies, as well as confounding factors related to stimuli that were not adequately controlled for. Additionally, pragmatic and processing factors that may contribute to non-native behavior have not yet been considered in scope interpretation by Chinese learners of English and English-speaking learners of Chinese, despite some recent attention to these factors in the L2 acquisition of quantifier scope in other languages (e.g., Chung, 2022; Lee, 2009; Özçelik, 2018). Another important issue that previous L2 studies on quantifier scope have not addressed is that access and preference are not properly disentangled.

In the following section, I discuss in detail some theoretical issues in relation to L2 quantifier scope including the factors believed to affect adult L2 knowledge of quantifier scope interpretation/processing and potential sources of the inconsistent findings in the literature. The aim is to contextualize for the present investigation for research questions and predictions to be introduced.

3.4 Theoretical issues in relation to L2 quantifier scope

3.4.1 The role of language input

Input has been a central concern in any models of language acquisition, whether they are generative or usage-based/emergentist theories (for discussions see Lidz & Gagliardi, 2015; Rankin & Unsworth, 2016; Slabakova et al., 2014; Zyzik, 2009). The extent to which target language input is available, noticed, processed, and used is crucial for language acquisition and development. At issue among different theoretical approaches is whether input alone suffices for successful acquisition. While input has been argued to play an important role in triggering linguistic representations, according to the generative approach to language acquisition, input on its own is not sufficient for learners to develop subtle and complex knowledge of human languages (Schwartz & Sprouse, 2013; White, 2003). This approach claims that in situations where learners attain knowledge for which scarce or even no evidence is available, they must have turned to an innate linguistic system, namely UG, for the gap to be bridged between what is available from the input and what is instantiated in the target grammar. Such learning situations constitute the classical Poverty of the Stimulus (POS) problem both in L1 and L2 acquisition (Hornstein & Lightfoot, 1985; Rothman & Slabakova, 2018; Schwartz & Sprouse, 2000; White, 2003). Investigations into POS problems in L2 are theoretically informative for at least two reasons. First, they shed light on the role of UG and its interaction with input during second language acquisition. Second, they address one of the central questions in SLA as to whether adult L2 acquisition is fundamentally different from child language acquisition (Bley-Vroman, 1989). The reasoning is that if L2 acquisition occurs despite POS problems, it suggests that UG guides this process, as it does in child language acquisition, indicating that L1 and L2 acquisition are not fundamentally different (Schwartz & Sprouse, 2000).

As such, to determine the severity of the POS problem faced by L2 learners, it is crucial to carefully identify and evaluate learning scenarios to ascertain the availability of certain language input. In the case of L2 acquisition, evidence could come from different sources: L2 input, L1 knowledge (via transfer), and classroom instruction, in the form of *positive* or *negative* evidence. If L2 learners potentially demonstrate linguistic knowledge beyond the absence of evidence from these three sources, this could offer some support for the domain-specificity of L2 representations. Scope interpretation is one of the best-known examples that provides compelling evidence for the relative contributions of input and UG (Juffs, 2022; Slabakova, 2016, p.35), hence attesting to the POS problem. The scope phenomena in question pose varying levels of POS-related difficulties for L2 learners, depending on their learning scenarios complicated by the type of evidence of language input involved. I evaluate each type of evidence for its potential influence in the context of the present bidirectional study in which English-speaking learners of Chinese and Chinese-speaking learners of English are examined for the interpretation of quantifier scope in their respective target language.

Positive evidence, be it from naturalistic input or explicit classroom instruction, is considered the primary linguistic data obligatory in second language acquisition. The learning task for learners should be easy if it is simply to add new interpretations. For sentences such as (5) (repeated in (31) below), its inverse scope reading (31b) should be eventually acquirable for English-speaking learners of Chinese, for which there is no evidence from L1 but positive evidence available and exemplified in the L2 input. Its surface scope reading (31a) should be fully acquirable for Chinese speakers even at their initial stage of L2 learning, because the L2 initial state is assumed to be derived from the L1 grammar in its entirety according to Schwartz and Sprouse's (1996) Full Transfer/Full Access model and L2 learning may benefit from UG when the L2 input diverges from the L1-based interlanguage.

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

a. None of the horses jumped over the fence. (Surface scope)

b. Some of the horses jumped over the fence. (Inverse scope)

By contrast, English-speaking learners of Chinese are predicted to have difficult in acquiring the native-like interpretation of (31) in its Chinese counterpart, because the successful acquisition would require English speakers to preempt or unlearn the L1 interpretation (the inverse scope one) that does not exist and is otherwise unacceptable in the L2; yet there is no such positive evidence in the input to show that (31b) is not permitted in Chinese. Therefore, English-speaking learners of Chinese are exactly confronted with a learning scenario in which the L2 is a subset of the L1 and L2 learners have to unlearn the L1-transferred interpretation in the absence of positive evidence, in which case some form of (direct) *negative evidence* presumably is needed to explicitly inform learners which interpretation is unacceptable in the L2. Given that this kind of evidence is available and effective, preempting L1 options is generally more problematic than adding interpretations to the L2, as noted by Gabriele (2009, p. 372) that 'restructuring is particularly difficult when learners need to unlearn certain aspects of their L1 in the absence of explicit input that indicates which properties are ruled out by the L2'.

Although the directional difference in the success of arriving at a target-like interpretation can be predicted based on positive or negative evidence, other things being equal, one assumption for this prediction to be borne out is that such evidence is reliable and robust. However, this evidence is often variable in terms of its reliability and robustness depending on a number of different factors, such as form-meaning mapping transparency, frequency, salience, semantic basicness, and communicative intent (DeKeyser, 2005; Ellis, 2002; Ellis & Wulff, 2020). Each factor is inspected with respect to scope phenomena. As for form-meaning mapping, one sentence (e.g., as in (31)) containing different logical operators accommodates two different meanings at least in English, which represents an instance in which the mapping between form and meaning is not transparent and the meaning that these operators bring to a sentence does not have a one-to-one correspondence with the surface syntax. Moreover, for learners to be directed to a particular interpretation in competition with the other, discourse context must be consulted. Although it is unknown how often discourse context exists in the naturalistic input for the interpretive ambiguity to be resolved, quantified sentences as those containing quantifier and negation are quite rare[22] in adult speech as found in a corpus analysis (Gennari & MacDonald, 2006). As such, it appears that scope phenomena are highly infrequent in the language input. As for saliency, due to the ambiguous nature of quantifier scope in English, the available interpretations of quantified sentences are often not immediately obvious, hence not salient particularly for less preferred interpretations. As far as semantic basicness and communicative intent are concerned, sentences like (31) are somehow superfluous, as whatever can be expressed with the form in (31) can be alternatively expressed with 'None of the horses jumped over the fence' if one prefers (31) to be

---

[22] However, it is important to note that certain sentences (e.g., used as proverbs) like 'All that glitters is not gold' do not make sense when interpreted with surface scope according to one's world knowledge. In such cases, the inverse scope interpretation is necessary. This sentence itself may serve as positive evidence for the availability of inverse scope interpretations for negatively quantified sentences in English.

interpreted for a surface scope reading. Therefore, it is hard to argue that the acquisition of a form like (31) is much driven by communicative needs. For the same reason, quantified sentences are not semantically basic.

In the context of classroom instruction, the interaction between logical operators for its interpretation is not explicitly taught in the classroom contexts. During the data collection, I interviewed some of the participants and language teachers who confirmed that students in language classes were not explicitly taught how to interpret sentences like (31), even though they were taught the distribution of the quantificational expressions. For example, it was taught and emphasized that ‘every NP’ in most cases must be accompanied by the free morpheme DOU in Chinese. In addition, direct negative evidence appears in the form of corrections in the classroom. Regarding quantifier scope, since the interpretation of this aspect is not explicitly taught, it is unlikely that any such correction would exist in the instructional settings. For example, Wu (2020) provides a guide for learners of Chinese at various levels, enumerating the most frequently misused and confusing words, as well as grammatical phenomena. Although this guide discusses the use of DOU in Chinese for when it is not required given certain grammatical contexts, it does not make explicit the potential restrictions on interpretation that may arise from the presence or absence of DOU in that context. As such, a POS problem arises for either learners of English or learners of Chinese who must acquire the L2 knowledge for which evidence is absent in quantity or degenerate in quality, be it from naturalistic input, L1 grammar, or L2 instruction. Comparatively speaking, English-speaking learners of Chinese should experience a POS problem to a larger extent than Chinese-speaking learners of English, because positive evidence can be used to support an interpretation even when it is ambiguous and unreliable, as is the case with Chinese-speaking learner of English.

Despite POS, L2 properties arguably are acquirable resorting to some so-called triggering input that would inform learners of the underlying property being acquired (White, 2003). The triggering input may be related in a quite indirect way to the acquired property that in fact 'may bear no obvious relation to the input that triggered it' as noted by Lidz and Gagliardi (2015, p.334). The reasoning is that by being sensitive to property B, which is clearly available in the L2 input, one may be able to acquire property A, which is otherwise unavailable. Triggering input for its role has been explicitly investigated in the acquisition of Japanese passives (Hara, 2007). For English speakers who are learning Japanese, the restriction of the *ni* direct passive to perfective readings cannot be directly derived from the L2 input, posing a POS problem for learners with English as the L1. It has been argued that such learners eventually acquired this restriction due to their sensitivity to the affectivity property associated with *ni* direct passives, which they encounter in their L2 input. In a more recent study, Kume and Marsden (2021) found that English-speaking learners of Japanese were found to acquire the definiteness properties on Japanese floating numeral quantifiers. They argued that this success partly was driven by the triggering input linked to the morphological status of Japanese numeral quantifiers as compound words.

In a similar vein, some properties available in the target language input may serve as the trigger for the interpretive properties in quantifier scope. For Chinese-speaking learners of English, their knowledge about definiteness of NPs in English may bear some relation to the extent to which they can access a particular scope interpretation for doubly quantified sentences. As in sentences such as (3) and (4) (repeated in (32) below), an indefinite NP occurs in the object position for (32a) and in the subject position for (32b).

(32) a. Every child climbed a tree.

b. A child climbed every tree.

Unlike English whose definiteness is overtly expressed through its article system, Chinese is a well-known article-less language in which definiteness cannot be realized overtly through the functional equivalents of the English articles. Chinese speakers may navigate through their knowledge about NP definiteness to decide the interpretations of sentences like (32). For example, if they come to know through English input that *a child* should be interpreted as indefinite regardless of where it occurs despite the fact that a subject NP in Chinese tends to be interpreted as definite (Yang & Wu, 2020), they should find it possible to access (32b) on its inverse scope reading in which *a child* does not refer to one specific child in context. Likewise, if Chinese learners of English are shown to possess the knowledge about NP definiteness in English, they should access (32a) on its surface scope reading in a way comparable to native speakers of English. For the Chinese quantifier scope, there are two cases where DOU is mandatory, as in sentences like (19) and (27) (repeated in (33) and (34) respectively below). The pattern arises that DOU must be present when Mei + NP is in the subject position.

(33) Mei-yi-ge haizi dou pa-le yi-ke shu.

Every-one-CL child DOU climb-PERF one-CL tree

'Every child climbed a tree.'

(34) Mei-yi-pi ma dou meiyou tiao guo liba.

Every-one-CL horse DOU not jump over fence

'Every horse didn't jump over the fence.'

As discussed in section 2.2, DOU in (33) and (34) is used as a universal quantifier distributer that quantifies and distributes over expressions preceding it. L2 input is available exemplifying the use of DOU as a universal quantifier distributer in cases where there is no interaction between logical operators. Such examples can be seen in (35). As such, when English-speaking learners of Chinese come to know that DOU is mandatory and can be used as a universal quantifier distributer from input like (35), it is likely that these learners would interpret sentences like (34) and (35) with surface scope readings.

(35) a. Mei-yi-ge xuesheng dou qu dalanqiu le.

Every-one-CL student DOU go-to play basketball

'Every student went to play basketball.'

b. Meitian dou xiayu.

Every day DOU rain

'It rains every day.'

The current study attempts to address the correlation between the sensitivity to these 'triggers' in the input and the acquisition of the interpretative properties associated with quantifier scope. However, it is important to note that there may not always be a direct correlation between the acquired properties and the input that triggers them, as they may not be directly related to each other, at least from the perspective of an L2 learner. Another consideration is the influence of pragmatics and is discussed in the following section.

3.4.2 Pragmatic factors

According to the Interface Hypothesis (e.g., Sorace 2011; Tsimpli & Sorace 2006), structures at the external interface pose particular challenges for L2 learners. External interfaces arise from the interaction between pragmatics/discourses (external to the grammar) and syntax or semantics (internal to the grammar). As argued previously, quantifier scope for its interpretation involves syntax, semantics, and pragmatics/discourse. The interpretation of quantifier scope by learners of English and learners of Chinese should consider in detail the interface factors, as in Özçelik (2018), the only study so far that has examined L2 quantifier scope attesting to the Interface Hypothesis, with a focus on negatively quantified sentences.

Before delving into the operation of external interfaces in Chinese and English quantifier scope and their impact on the interpretation of sentences containing quantifiers and/or negation, it is important to distinguish between pragmatics and discourse, although the two terms are often used interchangeably in the L2 literature, as noted by White (2011, p.581). Rothman and Slabakova (2011) provides a brief overview of how these two terms are distinct while related. According to them, pragmatics and discourse are in a superset-subset relationship as schematized in Figure 3.1, with pragmatics being the superset and discourse being the subset.

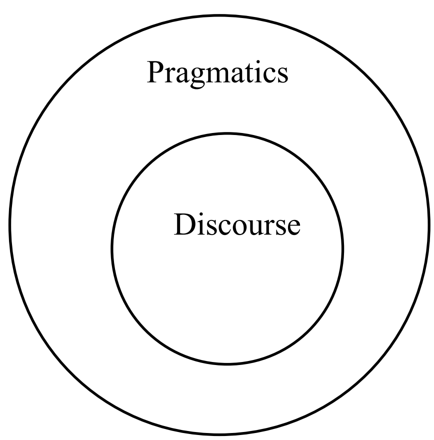

Figure 3.1 Relation between pragmatics and discourse (adapted from Rothman & Slabakova, 2011).

Discourse involves constructions whose interpretation depends on information from the previous discourse context. A much-studied topic is the interpretation of null subject pronouns in Italian or Spanish, for which to be achieved the discourse information must be integrated from the previous context. Pragmatics is much broader in scope including domains of study such as scalar implicature, presupposition, and aspectual coercion. The interpretation of pragmatics-related properties depends on knowledge of the world and universal pragmatic principles. Nevertheless, since the two terms form a superset-subset relationship, the wider term (pragmatics) asymmetrically entails the narrow term (discourse) such that wherever discourse is applicable, pragmatics is also applicable but not the vice versa. In this dissertation, I will use each term in its narrow sense.

The discourse context plays a crucial role in the setup of TVJT, where participants determine the degree of truthfulness of a test sentence with respect to the context provided, showcasing their underlying linguistic knowledge (Thornton et al., 2017). The link between this method and its intended purpose can be explained as follows. As argued by Scontras et al. (2017), the initial assumption for readers is that each sentence corresponds to an unambiguous interpretation, and they only abandon this assumption if compelled to do so in context. This is particularly evident in cases where one sentence permits two interpretations, but one interpretation is significantly preferred over the other. In this sense, the discourse sets up a meaningful scenario for participants to respond to sentences in felicitous contexts, enabling the detection of less privileged interpretations that may otherwise be undetachable without any contextual support. In the present study with experiments testing L2 learners, appropriate discourse contexts are created for test sentences of all kinds and participants would not be affected by the presence or absence of contextual support for the extent to which they can access all possible interpretations. Instead, the impact of pragmatics on scope interpretation may vary depending on the language group and the type of quantifier scope being considered.

In terms of pragmatics, to achieve full competence in quantifier scope interpretation, one needs to attain the knowledge of syntax and semantics in cases without scope ambiguity. In cases where scope ambiguity exists, one should additionally acquire the knowledge of pragmatics. One pragmatic principle that complicates scope interpretation in English is Principle of Charity, according to which whenever possible speakers will access an interpretation that makes the sentence true. For example, for (31), when it is presented following a context in which three horses were trying to jump over fence, but only two of them succeed, this sentence would be true on its inverse scope reading (31b) but false on its surface scope reading (31a) for a native speaker of English. Given the Principle of Charity, the speaker would choose the reading (31b) that makes (31) true if both readings are accessible (afforded by the discourse context) and would therefore endorse (31). The same reasoning can be applied to the case where a context in support of surface scope readings is presented following (31). As such, if Chinese-speaking learners of English were to fully acquire the inverse scope reading for sentences like (31) in English (which is absent in Chinese), they would accept (31) against the context in favor of the inverse scope reading if they can correctly employ the Principle of Charity. For English-speaking learners of Chinese, because sentences in the target language do not involve ambiguity with only one interpretation to be acquired, the Principle of Charity applies vacuously in this case (Gualmini et al., 2008; Özçelik, 2018).

To sum up, in the learning scenario where one must acquire a sentence with two possible interpretations, i.e., the acquisition of doubly quantified and negatively quantified sentences for both surface scope and inverse scope readings by Chinese-speaking learners of English, the Principle of Charity should be at play in determining which of the two available interpretations to select. By contrast, in the learning scenario where the task is to acquire the only interpretation in the target language, i.e., the acquisition of doubly quantified and negatively quantified sentences for the surface scope reading by English-speaking learners of Chinese, this principle, while involved (via vacuous application), does not play an essential role. The prediction for both groups in terms of success rate, if based on the Interface Hypothesis, would be that Chinese-speaking learners of English should be less successful than English-speaking learners of Chinese in arriving at target scope interpretation because the former group is tasked with addressing challenges posed by the external interface, while the latter group is not.

3.4.3 Task factors

Although there is some evidence that L2 learners can acquire target-like scope interpretation despite negative L1 transfer and absence of positive and negative evidence in the input, the results of the studies reviewed in sections 3.1 and 3.2 are inconsistent. The mixed results across studies may be attributed to differences in the types of experimental tasks used and aspects of their features. For example, the aforementioned studies used truth-value judgments involving written stories, pictures, or a combination of both. The test sentences were presented in either the written format or the aural format. The response scale for judgements includes binary options and a small number of finite options, i.e., Likert Scale. Currently, online tasks for measuring real-time interpretation of quantified sentences are restricted to self-paced reading.

Before discussing the potential different impacts of truth-value judgement tasks (TVJTs) in their written story format vs. picture format on experimental results, a justification is provided for the use of truth-value judgement tasks rather than traditional acceptability judgment tasks (AJTs) for testing quantifier scope interpretation. AJTs involve a metalinguistic judgment about form in that a sentence can be judged to have different degrees of grammaticality depending on whether that sentence violates grammatical constraints. While a quantified sentence remains grammatical, the real concern lies in its potential ambiguity between interpretations and the preference for one interpretation over another. Therefore, the question of ambiguity and preference can be methodologically at least partly addressed by virtue of truth-value judgment tasks in which a particular interpretation provided by the context is forced on learners for them to accept or reject. In this respect, TVJTs require one to make a judgment on meaning; this interpretation-based task arguably reduces the amount of conscious knowledge required and lowers processing demands due to the context it provides, hence better tapping into L2 learners' linguistic competence (Orfitelli & Polinsky, 2017).

Although both versions of TVJT have been argued to provide a window into leaners' linguistic competence, they differ in their ability to detect the less preferred interpretation for an ambiguous sentence (Slabakova, 2013; White et al., 1997). This task effect has been explicitly attested to in White et al. (1997). They investigated the interpretation of reflexives in French-speaking and Japanese-speaking learners of English. In English, the reflexive *herself* in sentences like 'Mary showed Susan a portrait of herself' can have either *Susan* or *Mary* as its antecedent, although the subject (i.e., *Mary*) antecedent is preferred. In their experiment, the test sentence was presented in the context of either a written story or a picture. The task for the participants was to decide whether that sentence matched the context. White et al. (1997) found that both native speakers and L2 learners accepted object antecedents more consistently on the story task than on the picture task. This indicated a task effect, in that the participants appeared to perform better on the story task than on the picture task[23] when dealing with the less preferred but still grammatically correct interpretation, despite both interpretations being available in the grammar.

Across existing studies, the task effect in L2 quantifier scope interpretation can be observed to some extent, albeit not consistently due to other factors in play. For example, participants seemed to accept inverse scope readings to a lesser extent in studies involving picture contexts (e.g., Chu et al., 2014; Wu & Ionin, 2019, 2022) than in studies involving story contexts (e.g., Özçelik, 2018). However, there has been no direct comparison of how learners perform on story tasks vs. picture tasks with regard to quantifier scope interpretation in a single study. The current study using both story and pictures tasks will shed light on the task effect.

Task effects may also be complicate by features of stimuli. First, verb type has been found to modulate the extent to which certain scope interpretations can be accessed both in English and Chinese (Kurtzman & MacDonald, 1993; Zhou & Gao, 2009). However, it has not been properly controlled for in some of the previous studies in which agentive verb (e.g., *read*) and psych verbs (i.e., *scold*) were mixed (e.g., Marsden, 2009; Wu & Ionin, 2022), making their results hard to interpret. Second, lexical difference in quantifiers may also make results across studies hard to compare. For instance, both 'all…not' and 'every…not' were found in previous studies as

[23] A different version of sentence-picture matching tasks is picture selection tasks in which participants typically read or listen to one sentence and look at a set of pictures (two to four) and they have to decide which picture goes with what is going on in that sentence. Tasks of this version require one to consider possible interpretations for an ambiguous sentence at the same time and may even prove more challenging when dealing with less preferred interpretations.

configurations for negatively quantified sentences. However, ‘all’ and ‘every’ are different universal quantifiers in that ‘all’ allows a collective interpretation while ‘every’ does not (Beghelli & Stowell 1997; Ioup, 1975, inter alia). Consequently, deriving the inverse scope reading for ‘every...not’ constructions is expected to be less challenging compared to ‘all...not’ constructions. The present study aims to mitigate the influence of verb type and lexical differences on the results.

In sentence-picture matching tasks, the sentence stimuli are often presented aurally, requiring participants to process the sentences in real-time as they hear them, thereby limiting their conscious introspection on the sentence’s meaning. In this respect, the way stimuli are presented may affect the extent to which explicit knowledge is involved. Among other ways, imposing time pressure on participants is one of the most effective ways to minimize their reliance on explicit knowledge; hence, reaction time data would directly speak to how learners utilize implicit knowledge for the online processing of language input (Ellis, 2005; Godfroid et al. 2015). The following section considers processing factors believed to complicate quantifier scope interpretation.

3.4.4 Processing factors

Arguably, a learner’s linguistic competence can be tapped if one gives linguistic judgments based on his/her implicit knowledge. As evidenced in Ellis (2005), tasks implemented in untimed settings elicit data reflecting learners’ explicit knowledge and learners tend to focus their attention on form in this case; learners are more likely to use implicit knowledge and tend to focus their attention on meaning when tested under time pressure. While the offline TVJT does not impose time pressure on participants and enables the use of explicit and metalinguistic knowledge, it is not as explicit as traditional AJTs since its emphasis is on meaning rather than form. Nevertheless, timed tasks have been argued to tap implicit knowledge to a greater extent than the untimed ones (Ionin, 2021). Moreover, auditory presentation of sentences, combined with time pressure, may work together to make the tasks more implicit (Ionin & Zyzik, 2014). In response to the call of making a direct comparison of how learners perform on timed vs. untimed tasks with constructions qualifying as POS problem (Ionin, 2021, p. 385), the present study includes an online TVJT with picture contexts and one offline TVJT with story contexts. By comparing the performance of the same group of learners offline and online, valuable insights can be gained into the extent to which implicit and explicit knowledge underlies their behavior. Evidence from different tasks would also allow us to fully understand the interlanguage competence.

In online tasks, participants' responses are measured as they process sentences that are unfolding over time. For quantifier scope interpretation to be accomplished, processes must involve the online derivation and processing of sentences in which representations from syntax, semantics, and/or pragmatics are combined for computing the relation between logical operators in a sentence. In L2 sentence processing, the focus of inquiry has primarily been on morpho-syntax (see Juffs & Rodríguez, 2014, for an overview). As reviewed in section 3.2, there has been a dearth of research on L2 processing of quantifier scope, particularly among Chinese learners of English and English learners of Chinese. To address potential differences between L1 and L2 processing of quantifier scope interpretation in real-time, I briefly examine relevant models of L2 sentence processing. The first model pertains to the current articulation of the Interface Hypothesis, primarily attributing interface difficulties to processing costs (Sorace, 2011). The earlier version of this hypothesis, which focused on the representational aspects of interlanguage development, has already been discussed in section 3.4.2 and will not be reiterated here. This hypothesis in its more recent articulation argues that learners' difficulties in (external) interface conditions largely stem from processing deficits in the integration of syntactic-semantic and pragmatic information.

In comparison to native speakers, L2 learners may experience greater costs in processing quantifier scope particularly pertaining to external interfaces. One explanation for these challenges is that L2 learners are limited in their cognitive resources in integrating different types of information for online sentence comprehension (Sorace, 2011). However, many of the studies attesting to the IH used offline measures that would not directly index processing (cf. Leal & Hoot, 2022). As such, the use of online measures would allow for weighing in directly on this hypothesis's claims about processing. Following the predictions of the IH, it should be the case that the interpretation of quantifier scope would be more challenging when pragmatic information must be consulted compared to when it does not, particularly during the online processing among L2 learners.

Other models that explain differences between L1 and L2 processing, taking resource deficits into account, include capacity-based models (Hopp, 2010; McDonald, 2006) and the Lexical Bottleneck Hypothesis (Hopp, 2018). According to capacity-based models, L2 processing is affected to a larger extent than L1 processing by resource limitations in language processing because the L2 parser relatively has lower automaticity, slower processing speed, and smaller working memory capacity. Hence, processing an L2 tends to be more cognitively demanding than processing an L1. Unlike models like the Shallow Structure Hypothesis (Clahsen & Felser, 2006), capacity-based models do not consider the differences between L1 and L2 processing in terms of the types of information that L2 learners tend to over- or underuse in comparison to native speakers during processing. However, this model still predicts protracted difficulty in integrating multiple types of information for processing in L2 learners due to their less efficient information integration capacities. As an extension of the capacity-based model, the Lexical Bottleneck Hypothesis claims that non-native-like L2 parsing derives from inefficiency and reduced automaticity in lexical access, due to two characteristics of the bilingual mental lexicon being that (a) lexical activation is slower and non-selective, and (b) demanding lexical processing exhausts resources that limit the capacity for syntactic processing. While these models predict aspects of L1-L2 processing differences, they suggest that L2 processing may not differ qualitatively from L1 processing, in that L2 learners can possibly progress to a stage where L1 and L2 processing converge. Although this study was not specifically designed to test these models, they can offer post hoc explanations for learners' processing behavior.

3.5 Research questions

The overall objective of this dissertation is to investigate the interpretation, processing, and acquisition of quantifier scope by Chinese-speaking learners of English and English-speaking learners of Chinese. The present study adds to previous work on L2 quantifier scope in the following ways. First, quantifier scope has received limited attention in studies involving Chinese-speaking learners of English and English-speaking learners of Chinese. Furthermore, to the best of my knowledge, the interpretation of quantifier scope by these two groups of L2 learners has not been investigated in the context of online processing. Second, the present study is one of the first to enable a direct comparison of L2 learners' performance on offline and online tasks concerning quantifier scope, qualifying as a POS phenomenon. This allows us to investigate the offline vs. online interpretation of quantifier scope, the roles of implicit vs. explicit knowledge in these processes, and potential task effects. Third, the bidirectional nature of this study allows for the examination of how L2 learners overcome cross-linguistic differences in quantifier scope interpretation and addressing the influence of language input in the L2 interpretation of quantifier scope. Fourth, given the difference in the extent to which Chinese and English allow both readings for a quantified sentence, it is possible that pragmatics (e.g., the Principle of Charity) plays a role in the interpretation of quantifier scope for English, while its influence on Chinese may not be as significant.

The present study thus contributes by adding an additional external interface property to evaluate the IH (with English learners of Chinese and Chinese learners of English), in response to the calls made by several researchers in recent years (e.g., Rothman & Slabakova, 2011; White, 2011). Finally, potential lexical effects due to the main verb type and quantifier items are controlled for in the present study by using agentive verbs across test sentences and 'every' instead of 'all' in negatively quantified sentences. Specifically, this dissertation addresses the following research questions (RQs).

RQ1) *How do English and Chinese speakers interpret quantifier scope in their respective native language in the absence of supportive context?* (Experiment 1)

(RQ1-1) Do native speakers of English allow for quantifier scope ambiguity in English sentences that involve the interaction of universal quantifiers with either existential quantifiers or negation, when no supportive context is given? If yes, what is their preferred interpretation for each sentence type? If not, what is the only interpretation available?

(RQ1-2) Do native speakers of Chinese allow for quantifier scope ambiguity in Chinese sentences that involve the interaction of universal quantifiers with either existential quantifiers or negation, when no supportive context is given? If yes, what is their preferred interpretation for each sentence type? If not, what is the only interpretation available?

RQ2) *Can Chinese-speaking learners of English and English-speaking learners of Chinese interpret quantifier scope in their respective target languages as an L2 in a manner comparable to the native speakers, when supportive context is given in the offline written story-based TVJT?* (Experiment 2)

(RQ2-1) Which learner group is more successful in arriving at the target interpretation: Chinese-speaking learners of English or English-speaking learners of Chinese?

(RQ2-2) How does L2 proficiency affect L2 learners' performance on quantifier scope interpretation?

RQ3) *Is the processing of quantifier scope in the target language by Chinese-speaking learners of English and English-speaking learners of Chinese similar to that of native speakers, when supportive context is given in the online TVJT with the visual-world paradigm?* (Experiment 3)

(RQ3-1) Which learner group is more successful in processing the target interpretation: Chinese-speaking learners of English or English-speaking learners of Chinese?

(RQ3-2) How does L2 proficiency affect L2 learners' performance on the online processing of quantifier scope interpretation?

4.0 Experiment 1: Quantifier scope interpretation in native English and Chinese

4.1 Introduction

Experiment 1 establishes a baseline for how doubly quantified and negatively quantified sentences are interpreted in English (Experiment 1a) and Chinese (Experiment 1b) in the absence of supportive context. The results of this experiment will provide empirical support for the cross-linguistic differences in the quantifier scope interpretation by native speakers of English and Chinese. This experiment is necessary as the findings in the L1 literature are inconclusive. Moreover, the data from this experiment are also to be interpreted with reference to the role of discourse context that would otherwise be present for eliciting all possible interpretations especially the less preferred ones. As has been argued in section 3.4.2, the purpose of including appropriate contexts in tasks such as TVJT is to mitigate any unusual pragmatic factors that may hinder a speaker from accessing readings that would otherwise be allowed by their grammar. In short, possible interpretations are 'forced' in pragmatically supportive contexts. Methodologically, presenting test sentences with context has become the standard practice in many of the studies concerning interpretive properties of the linguistic phenomena in question (e.g., Anderson, 2004; Guasti et al., 2005; Lee, 2009; Slabakova, 2010). Experiment 1a tests whether native speakers of English allow both the surface scope and inverse scope readings across all types of quantified sentences. Additionally, it aims to establish the preferred interpretation out of all possible interpretations in English. The purpose of Experiment 1b was to test native speakers of Chinese using the Chinese version of the experiment, in which the Chinese sentences were translated from English.

4.2 Experiment 1a: Sentence-interpretation matching task in L1 English

4.2.1 Methods

4.2.1.1 Participants

A total of 62 adult monolingual English speakers recruited via Prolific Academic (www.prolific.co) took part in this experiment (mean age = 32.4 years, *SD* = 7.09, 40 females). The sample size for this experiment and the subsequent experiments was determined based on a *prior* power analysis using G*power (version 3.1.9.6; Faul et al., 2009). Anticipating a small effect size[24] (*d*=.20) for a within-factor repeated measures ANOVA,[25] the analysis indicated that at least 36 participants for each group should be recruited to achieve 80% power (alphas = .05). The planned mixed-effects model that accounts for random effects on subjects and items is expected to yield higher power than the corresponding by-subjects ANOVA, if not more (Baayen et al., 2008). Since the design is consistent (2*2 within-subject) across participant group and experiment

---

[24] Plonsky and Oswald (2014) went so far as to propose a higher threshold (*d* = .40) for the small effect size range in L2 research.

[25] At the stage of data collection, a mixed-effects model was fit to the data from the first 36 participants by including both independent variables and their interactions as the fixed effects and subjects and items as random effects. The distribution of coefficients of the fixed effects obtained from this modeling served as input for conducting simulations. 1000 simulations were run on each participant for the same data structure, confirming that a sample size ranging from 35 to 45 would consistently yield a statistical power of 80% or higher. The statistical consultation services provided by the Statistics Consulting Center at Pitt were instrumental in providing expert guidance on using simulations to approximate sample size.

(experiments 1 and 2)[26], each group included a minimum of 36 participants considering the potential loss of data. The participants gave written informed consent and were compensated for their participation. All participants had normal or corrected-to-normal vision and reported no history of language deficits.

4.2.1.2 Design and materials

In the present study, two types of quantifier scopes were tested: one in which the universal quantifier and the existential quantifier interact, resulting in doubly quantified (DQ) sentences, and the other in which the universal quantifier and negation interact, resulting in negatively quantified (NQ) sentences. Specifically, two kinds of DQ sentences were constructed, all illustrated in Table 4.1: One in which the universal quantifier precedes the existential quantifier (i.e., UE sentences), and the other in which the existential quantifier precedes the universal quantifier (i.e., EU sentences). Similarly, two kinds of NQ sentences were constructed: One in which the universal quantifier precedes negation (i.e., UN sentences), and the other in which negation precedes the universal quantifier (i.e., NU sentences). Main verbs in all test sentences were agentive verbs. Each sub-type of quantified sentences was evaluated against its corresponding surface scope (SS) and inverse scope (IS) readings. For example, test sentences like (3) (e.g., Every child climbed a tree) and (4) (e.g., A child climbed every tree) were presented following either the surface scope reading or the inverse scope reading. As such, the participants were tested on four conditions generated by

[26] Experiment 3 utilized a within-subject design with a 2x3 factorial arrangement. G*power analysis indicated that a sample size of 28 participants would be sufficient to achieve 80% power.

crossing word order and interpretation in both DQ sentences and NQ sentences, as exemplified in Table 4.1. Appendix A contains a list of all items in Experiment 1.

Table 4.1 Example sentences across conditions.

Quantifier scope	*Word order*	*Interpretation*	
		Surface Scope	*Inverse Scope*
DQ	Every child climbed a tree.	Every child climbed a different tree.	Every child climbed the same tree.
	A child climbed every tree.	The same child climbed every tree.	A different child climbed every tree.
NQ	Every horse didn't jump over the fence.	None of the horses jumped over the fence.	Some of the horses jumped over the fence.
	The horse didn't jump over every fence.	The horse jumped over some of the fences.	The horse jumped over none of the fences.

For each type of quantifier scope, 12 lexically matched sets of items across the four conditions were created, for a total of 48 target items. The target items were distributed into four lists using a Latin Square design. Each list comprises a total of 24 distinct target items, 12 from DQ sentences and 12 from NQ sentences. Despite being presented in the same experiment, DQ and NQ sentences were treated as separate, between-item manipulations and therefore analyzed independently. In addition to the target items, 30 filler items spanning the full range of possible degrees to which readings can match test sentences were created and were intermixed among target items. To verify the test sentences for acceptability, I also conducted a norming test[27] using Prolific with 21 English monolinguals who did not participate in the main experiments. The norming test was implemented as a Qualtrics Survey. In this test, all 12 sets of DQ sentences and 12 sets of NQ sentences were evenly distributed across four lists. Individual participants were randomly assigned a list of 24 test sentences interspersed with another 24 filler sentences.[28] Participants were asked to rate the acceptability of each test sentence on a scale from 1 – 7 (1 = *completely unacceptable and not making sense*; 7 = *completely acceptable and making sense*). The target sentences of all conditions (M = 5.70, SD = 1.61) were rated significantly higher than the fillers (M = 1.99, SD = 1.32) (F (1, 1002) = 1587, p < .001). In addition, among target sentences, ratings did not differ between UE (M = 6.13, SD = 1.41) and EU sentences (M = 6.12, SD = 1.25) (F (1, 249) = 0.002, p = .967). NU sentences (M = 5.98, SD = 1.51) were rated significantly higher than UN sentences

[27] The norming test results were applicable across experiments because the same test sentences were used in all experiments.

[28] To avoid inflating differences between sentences, I included the ungrammatical fillers among the test sentences that were generally grammatical. The ungrammatical filler items were created using ChatGPT (April 4, 2023, OpenAI, https://chat.openai.com/chat.), a state-of-the-art language model trained by OpenAI.

($M = 4.59$, $SD = 1.73$) ($F(1, 249) = 46.11$, $p < .001$). As such, the target sentences overall were acceptable. It has been suggested that acceptability judgments can be affected by factors that influence sentence processing, such as ambiguity and frequency (Sprouse, 2018). Therefore, lower ratings of UN sentences compared to NU sentences may be due to their ambiguity[29] to English speakers and infrequent occurrence in the input.

To explore whether one's definiteness knowledge may inform interpretive scope with quantifiers, a definiteness test was created to assess learners' definiteness knowledge in English. Twenty multiple-choice questions were constructed, each with four possible answers. The 20 test items were counterbalanced across the four possible answers. Each question has only one correct answer. Below is an example test item, as in (36).

(36) Sometimes ____ moon is closer to the earth than other times.

1. a (an)
2. the
3. one
4. no article

4.2.1.3 Procedure

On each trial, participants saw a test sentence which was followed by a possible interpretation, being either the surface scope reading or the inverse scope reading, as in Figure 4.1.

---

[29] In Chapter 5, the analysis will show that UN sentences even in supportive contexts tend to be more ambiguous than NU sentences.

In this sentence-interpretation matching task,[30] participants were instructed to read a series of sentences said by Kris[31], and then determine how well each sentence conveyed Kris's intended message by rating the degree to which that sentence matched the given unambiguous interpretation on a scale from 1 (indicating that the sentence did not match the unambiguous interpretation at all) to 7 (indicating that the sentence completely matched the interpretation). At the beginning, the participants were asked for their basic biographic information such as age, gender, and language background. A block of five practice trials preceded the main experimental session to familiarize participants with the task. After the sentence-interpretation matching task, the participants completed the test targeting definiteness knowledge. This experiment was implemented as a Qualtrics Survey and completed individually by participants on their own. This experiment took roughly 20 min.

[30] During the experiment's pilot phase, both possible interpretations were presented at opposite ends of the scale, implying a trade-off between them. However, this setup may not accurately reflect participants' actual interpretations. Therefore, the task was revised to its current version, aiming to capture each sentence's true interpretation more effectively.

[31] This is an imaginary figure who is supposedly the speaker of the test sentences.

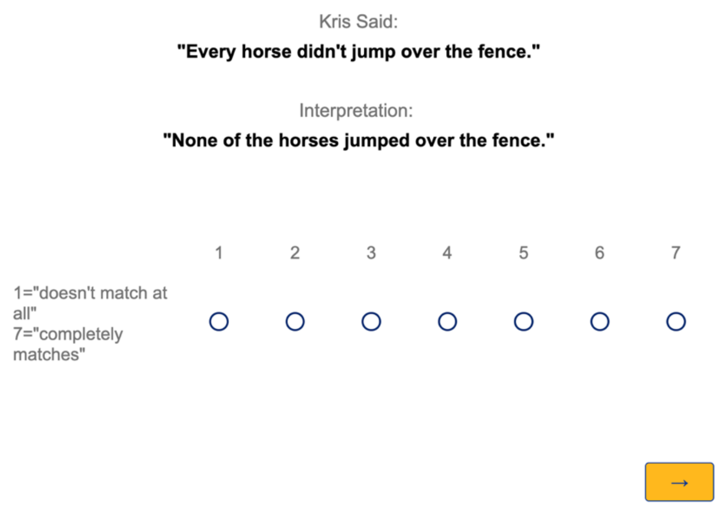

Figure 4.1 Sample trial for the sentence-interpretation matching task.

4.2.1.4 Analysis

Data preprocessing prior to statistical analysis included the following steps. First, to assess participant engagement and comprehension, I calculated the difference between the average rating of four filler items that were clearly acceptable and four filler items that were clearly unacceptable. Participants whose mean differences between unacceptable and acceptable filler items were greater than 0 were excluded from the study, as this indicated either a lack of understanding of the task or disengagement from the experiment. Accordingly, one participant was removed before further analysis. Second, raw ratings from the retained participants were z-score transformed. Z-scoring by participant was to reduce biases from potential differences in individual participants' perception and use of the 7-point scale (Schütze & Sprouse, 2013; Spinner & Gass, 2019).

All statistical analyses were conducted with the R programming language (R Core Team, 2021). The 'tidyverse' package (version 1.3.1) was used for visualization (Wickham et al., 2019). The 'lme4' package (version 1.1.27.1) in R was used to analyze the data collected from each task

(Bates et al., 2015). Specifically, I employed mixed-effects modeling procedures that account for the variability in both participants and stimuli simultaneously (Linck & Cunnings, 2015). With respect to statistical modeling, z-score transformed ratings were fit using linear mixed-effects models (Baayen et al., 2008). Fixed effects were word order and interpretation for each type of quantifier scope. Sum-coding contrast (−0.5, 0.5) was adopted for these two-level variables to obtain ANOVA-style main effects and interactions.[32] To examine the interactions among variables, post-hoc comparisons of the estimated marginal means were performed using the emmeans package (Lenth 2021; version 1.7.0), and Tukey adjustments were applied for pairwise comparisons. The random-effects structure was kept maximal for the initial model allowed by the experimental design (Barr et al., 2013), for which I included by-participant and by-item intercepts, by-participant random slopes for within-subject factors and their interactions, and by-item random slopes for between-subject factors wherever possible. In cases where models failed to converge, the random-effects structures were simplified by iteratively removing the correlation between random effects and the random effect contributing to the least variance until models converged. *P*-values for the mixed-effects model were obtained using the 'lmerTest' package (version 3.1.3; Kuznetsova, et al., 2017), which employs Satterthwaite approximations to estimate degrees of freedom.

[32] With this coding scheme, the intercepts in a model represent the grand mean of the response variable, rather than the mean of the baseline level determined by the default treatment contrasts in R. Using treatment contrasts in models with interactions is not recommended, as it can complicate result interpretation (Singmann & Kellen, 2019).

4.2.2 Group results

As planned, DQ and NQ sentences were analyzed and reported separately. Table 4.2 summarizes the mean ratings across conditions for DQ and NQ sentences, which are also illustrated in boxplots[33] as in Figure 4.2 and 4.3, respectively.

Table 4.2 Exp1a: Mean ratings across conditions and quantifier scope types in L1 English (standard deviations in paratheses)

Quantifier scope	Word order	Interpretation	
		SSR	ISR
DQ	UE	4.35 (1.64)	3.36 (1.68)
	EU	4.89 (1.71)	3.29 (1.65)
NQ	UN	3.53 (2.20)	4.81 (1.88)
	NU	5.65 (1.40)	2.17 (1.45)

[33] Compared to barplots, boxplots are considered to be more effective for visualizing the distribution of data (Larson-Hall & Plonsky, 2015; Larson-Hall, 2017). The plots for this experiment and subsequent ones display the following information: the median, the ends of the first and third quartiles of data, the upper and lower whiskers extending out to the minimum and maximum values in the data, mean ratings by condition (i.e., the blue rhombus in the middle of each plot), and overlaid dots representing mean ratings by subject and condition.

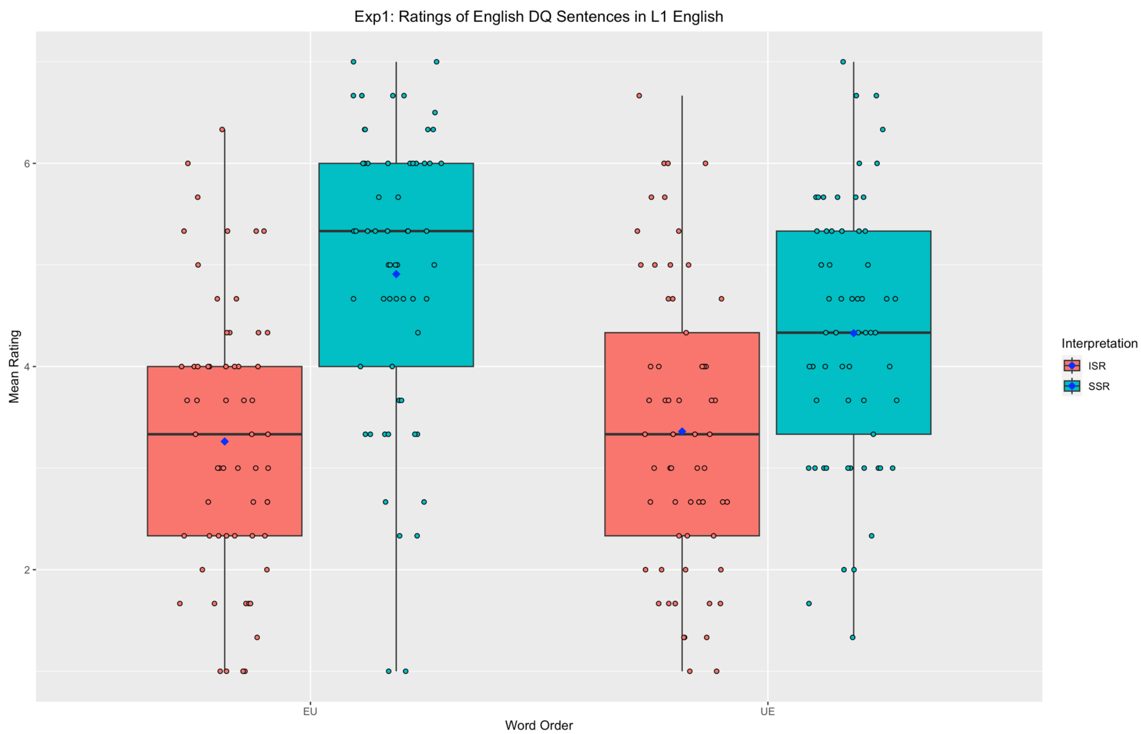

Figure 4.2 Exp1a: Mean ratings of DQ sentences by condition in L1 English

Table 4.3 presents the outputs of linear-mixed effects models with word order (NU: -0.5, UN: 0.5; EU: -0.5, UE: 0.5) and interpretation (ISR: -0.5, SSR: 0.5) as centered fixed effects for both types of quantifier scope. Plots of model residuals against fitted values and Q-Q plots based on z-scores revealed no obvious deviations of normality and homoscedasticity, meeting assumptions required by the statistical tests. I report the statistical analysis results as follows. For DQ sentences, the model returned a significant main effect of word order (b = -0.10, p = .022), reflecting lower ratings for sentences in the UE word order compared to those in the EU word order. Interpretation also emerged as a main effect (b = 0.58, p < .001), indicating that sentences followed by surface scope readings (SSRs) were more acceptable than sentences followed by inverse scope readings (ISRs). Crucially, a significant interaction was found between word order and interpretation (b = -0.28, p = .002). Post-hoc comparisons used to explore this interaction demonstrated that sentences in either the EU (b = -0.72, $p < .001$) or UE (b = -0.44, $p < .001$) word order was more acceptable when they were followed by SSR compared to when they were followed by ISR.

For NQ sentences, the statistical analysis showed main effects of word order (b = 0.11, p = .043) and interpretation (b = 0.50, $p < .001$), indicating that sentences in the UN word order were more acceptable than those in the NU word order, and that sentences followed by SSR were more acceptable than those followed by ISR. A significant interaction emerged between word order and interpretation (b = -2.15, $p < .001$). Accordingly, in line with the descriptive results, the ISR was more accepted than SSR when sentences are in the UN word order, and the SSR was accepted more than ISR when sentences are in the NU word order.

Figure 4.3 Exp1a: Mean ratings of NQ sentences by condition in L1 English

Table 4.3 Exp1a: Model output for ratings of DQ and NQ sentences in L1 English (*n*=61).

Quantifier scope	Fixed effects	*B*	*SE*	*t*	*p*
DQ	(Intercept)	-0.03	0.06	-0.53	.578
	Word order	**-0.10**	**0.05**	**-2.30**	**.022**
	Interpretation	**0.58**	**0.05**	**12.86**	**< .001**
	Word order*Interpretation	**-0.28**	**0.09**	**-3.06**	**.002**
NQ	(Intercept)	-0.001	0.04	-0.03	.980
	Word order	**0.11**	**0.06**	**2.03**	**.043**
	Interpretation	**0.50**	**0.06**	**8.82**	**< .001**
	Word order*Interpretation	**-2.15**	**0.11**	**-19.04**	**< .001**

4.2.3 Individual results

In the visualization, individual ratings across conditions were variable and inconsistent, as also evidenced by large standard deviations in Table 4.2. Moreover, group analysis does not reveal the extent to which less preferred readings are available compared to preferred readings. Therefore, it is important to examine individual performance on the experimental task. Based on individuals' relative preferences for surface scope and inverse scope readings depending on the type of quantified sentences encountered, participants were classified into five different patterns,[34] as described in (37).

(37) Patterns of ratings in L1 English

Pattern 1: Preference for the surface scope (SurPre)

Subjects showed an overall preference for the surface scope reading if the individual mean ratings of SSR were equal to or greater than 4 and greater than ISR by 1 point or more.

Pattern 2: Preference for the inverse scope (InvPre)

Subjects showed an overall preference for the inverse scope reading if the individual mean ratings of ISR are equal to or greater than 4 and the difference between ISR and SSR is greater than 0.9.

Pattern 3: Ambivalence with no scope preference (Ambi)

Subjects were ambivalent with no preference for either reading if the mean ratings of SSR are between 3 and 4.5, and the mean ratings of ISR are also between 3 and 4.5, and the absolute difference between the two readings is less than 1.1.

Pattern 4: Acceptance of both readings (Both_accept)

[34] I loosely modeled the classification of scope preference patterns after Chung (2013), with two modifications. First, the pattern of 'Rejection of both readings' was additionally included to cover the full range of scope preference patterns. Second, the range of mean ratings for each pattern was determined by considering both the extent to which the patterns can be discerned and the full range of patterns that can be covered.

Subjects accepted both readings if the ratings of both readings are 4 or higher.

Pattern 5: Rejection of both readings (Both_reject)

Subjects rejected both readings if the ratings of both readings were 3.9 or lower.

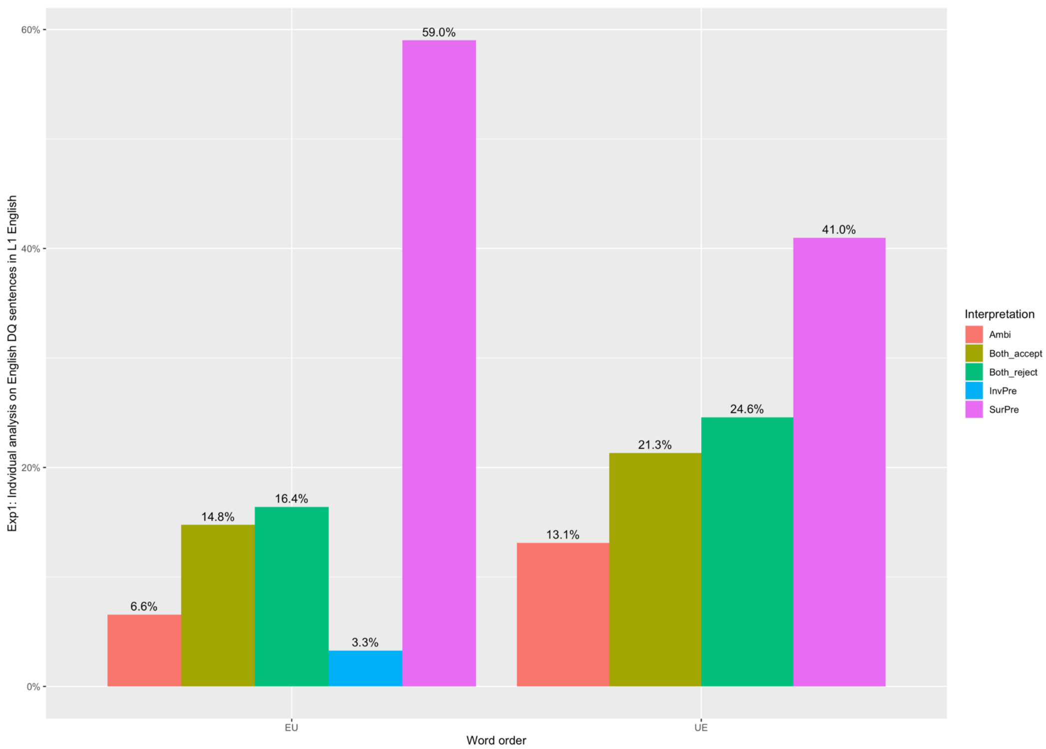

Figure 4.4 Exp1a: Percentage distribution of individual analysis for DQ sentences in L1 English.

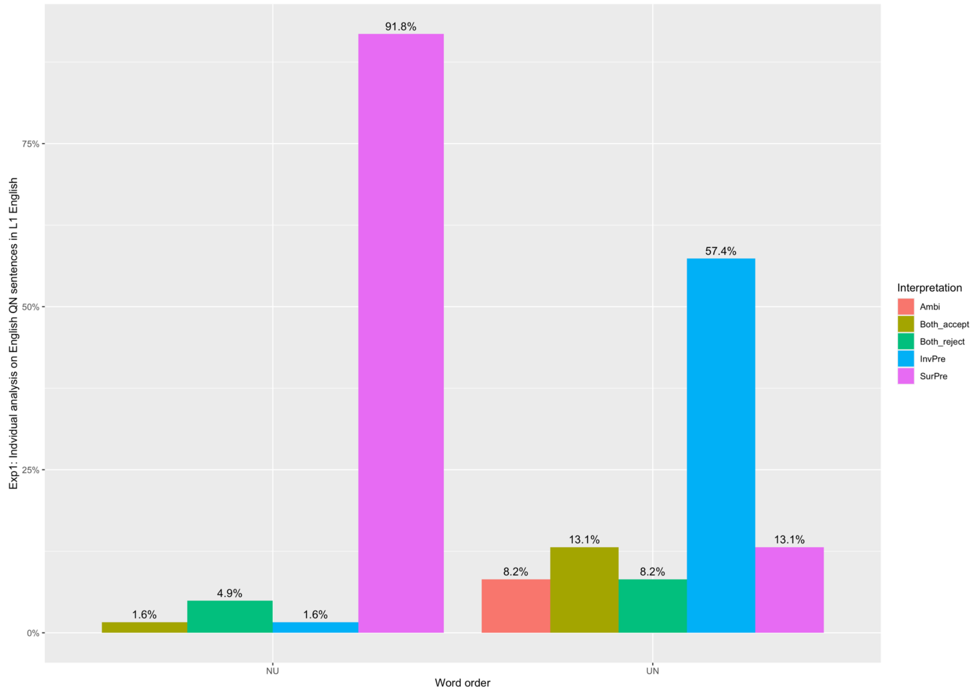

Figure 4.5 Exp1a: Percentage distribution of individual analysis for NQ sentences in L1 English.

Figures 4.4 and 4.5 show the percentage distribution of these five patterns of individuals for DQ (e.g., Every child climbed a tree) and NQ (e.g., Every horse didn't jump over the fence) sentences. For DQ sentences, more than half of the English speakers (59%) preferred the surface scope reading when the existential quantifier precedes the universal quantifier in a sentence. While the preference for the surface scope reading decreased when the existential quantifier is in the subject position (41%), this tendency prevails compared to other patterns. The results of NQ sentences showed a different picture in which more than half of the participants (57.4%) displayed a preference for the inverse scope reading in sentences with the UN word order. In the case of sentences with the NU word order, the majority of the participants (91.8%) fell into Pattern 1, in which the surface scope reading was preferred. Comparing DQ and NQ sentences, the judgments in DQ sentences appear to exhibit higher variability than those in NQ sentences. This is evident from the varying percentages of different patterns, with only some of the dominant patterns showing notably higher percentages in DQ sentences.

4.2.4 Definiteness test results with L1 English

Materials for the definiteness test were checked with native speakers of English. The results revealed reasonably high mean accuracy rates across participants ($M = 0.90$, $SD = 0.0998$) and test items ($M = 0.87$, $SD = 0.12$), suggesting that the test was appropriate and reliable. The correlation between performance on the definiteness test and that on the experimental task is only reported for Experiments 2 and 3, which are related to learner data.

4.3 Experiment 1b: Sentence-interpretation matching task in L1 Chinese

4.3.1 Methods

4.3.1.1 Participants

Sixty-seven Chinese speakers participated in experiment 1 (mean age = 20.6 years, SD = 1.58, 36 females). To minimize the interference from English, these participants were recruited in colleges of Mainland China. They were non-English majors and reported no history of staying in English-speaking countries. Recruiting Chinese students who have not been exposed to English in instructional settings can be extremely difficult, given that English classes are mandatory for all levels of education in China. Nevertheless, these participants demonstrated a low level of proficiency in English as assessed by the LexTALE task[35] (*M* = 55.71, *SD* = 9.63). According to Lemhöfer and Broersma (2012), those who scored at 59 (out of 100) or lower fall into the low range on this task. Participants were paid for their participation.

4.3.1.2 Design and materials

The experimental stimuli (including the target items and fillers) for the Chinese sentence-interpretation matching task were translated from the English items used in Experiment 1a. The translation was completed by two bilingual speakers who are fluent in both Chinese and English. They ensured that the meaning of the sentences in the source language was accurately conveyed in the target language while also ensuring that the grammatical forms of sentences translated into the target language were correct. Two aspects were paid particular attention to when translating and checking the Chinese stimuli. First, the DOU (都) particle was added when 'every + NP' was present as in sentences with the UE and UN word orders. Second, the predicate 'you' (有) is added right before the numeral phrase patterned as [numeral + classifier + NP] for sentences with the EU word order.

Despite being obtained through the translation of English sentences, the Chinese stimuli underwent a norming test for acceptability, which was implemented as a Qualtrics Survey and completed by native speakers of Chinese. An independent group of 38 Chinese speakers rated on a 7-point scale a total of 48 sentences (including 24 target items and 24 unacceptable fillers) for their acceptability (1 = *completely unacceptable and not making sense*; 7 = *completely acceptable and making sense*). Not surprisingly, the target sentences received a favorable response from

[35] This task will be introduced in detail in Chapter 5 for its implementation and score calculation.

Chinese native speakers (M = 4.78, SD = 2.32), demonstrating their overall acceptability. Inspecting on the between-condition contrasts, ratings did not differ between UE and EU sentences (F (1, 454) = 3.804, p = .0517)[36], and UN and NU sentences (F (1, 454) = 0.644, p = .423).

A test was created to assess the extent to which learners would know that DOU is mandatory in certain linguistic contexts. The knowledge might possibly serve as a kind of triggering input that learners may rely on to arrive at the target interpretation for which there is no sufficient evidence in the L2 input. This task was a forced-choice preference judgment task, in which sentences appeared in pair. The sentences in each pair differed only as to whether DOU was present, as exemplified in (38). In (38), DOU must be present when a universally quantified NP occurs in the subject position. As a result, only sentence with DOU would be accepted.

(38) A. 每个孩子都长得很高。
Mei-ge haizi dou zhangde hen gao.
Every-CL DOU grow very tall
'Every child is very tall.'

B. *每个孩子长得很高。
Mei-ge haizi zhangde hen gao.
Every-CL grow very tall
'Every child is very tall.'

---

[36] This marginal difference suggests that UE sentences are somewhat more acceptable than EU sentences. EU sentences configured in (20) have been argued to be bi-clausal, with the predicate 'you' composing with an indefinite object modified by a relative clause containing a universally quantified NP (Huang, 1987; Liu, 2011). This complex syntactic composition likely resulted in increased processing costs, leading to decreased acceptability, assuming that the more difficult a sentence is to process, the less acceptable it is to readers (Sprouse, 2018).

Five different test items were constructed in which DOU arguably must be present (Li, 2012). DOU modifies various elements across the test items, including grammatical subjects, topics, and place adverbials, all of which express universal meanings. Out of the five items, one includes a free choice item (i.e., *wh*-word), which must be licensed by DOU if the *wh*-word is interpreted as a universally quantified expression. Each participant received 15 test items, consisting of 5 critical items and 10 filler items with diverse linguistic structures (e.g., tense/aspect, classifier, negation).

4.3.1.3 Procedure

This experiment followed a similar procedure to Experiment 1a, with the exception that participants' English proficiency was assessed with the LexTALE task (Lemhöfer & Broersma 2012). This experiment took roughly 25 min. A separate group of 33 native Chinese speakers in Mainland China completed the DOU test. The participants were non-English major college students. In the DOU test, participants were given pairs of sentences like (38). They were asked to compare each pair of sentences and circle one of the four responses: 'only A'; 'Only B'; 'A and B'; 'None of the above', given the question 'Which sentence is in accordance with the Chinese grammar?'.

4.3.1.4 Analysis

The analysis procedures were identical to Experiment 1a. Specific to this experiment, one participant who failed to properly perform on filler items was excluded for further analysis.

4.3.2 Group results

The descriptive results were summarized in Table 4.4. Figures 4.6 and 4.7 illustrate the mean ratings of judgments by condition and participant for DQ sentences and NQ sentences, respectively. The figures demonstrate that Chinese participants generally accepted the SSR in both the DQ and NQ sentences, with the mean ratings exceeding 4 in all conditions. By contrast, the ISR received mean ratings lower than 3.5 across the board, indicating that these conditions were almost unacceptable.

Table 4.4 Exp1b: Mean ratings across conditions and quantifier scope types in L1 Chinese (standard deviations in paratheses)

Quantifier scope	Word order	Interpretation	
		SSR	ISR
DQ	UE	4.78 (1.88)	2.81 (1.67)
	EU	5.25 (2.10)	1.82 (1.33)
NQ	UN	6.35 (1.34)	1.61 (1.31)
	NU	5.17 (2.06)	3.36 (2.48)

Table 4.5 presents the outcomes of the statistical models for Experiment 1b. For DQ sentences, the model showed a main effect of interpretation (b = 0.11, p < .001), with higher acceptance rates for the SSR than for the ISR. In addition, word order significantly interacted with interpretation (b = -0.59, p < .001). A post hoc comparison revealed that the SSR was more acceptable than the IRS for both UE (b = -0.81, p < .001) and EU (b = -1.40, p < .001) sentences.

For NQ sentences, the statistical modeling yielded a main effect of interpretation (b = 1.35, p < .001), driven by higher acceptance rates for the SSR than for the ISR. Moreover, an interaction occurred between word order and interpretation (b = 1.19, p < .001), which was qualified by the post hoc analysis indicating that the SSR received higher mean ratings than the ISR, regardless of whether they were followed by UN (b = -1.94, p < .001) or NU (b = -0.75, p < .001) sentences.

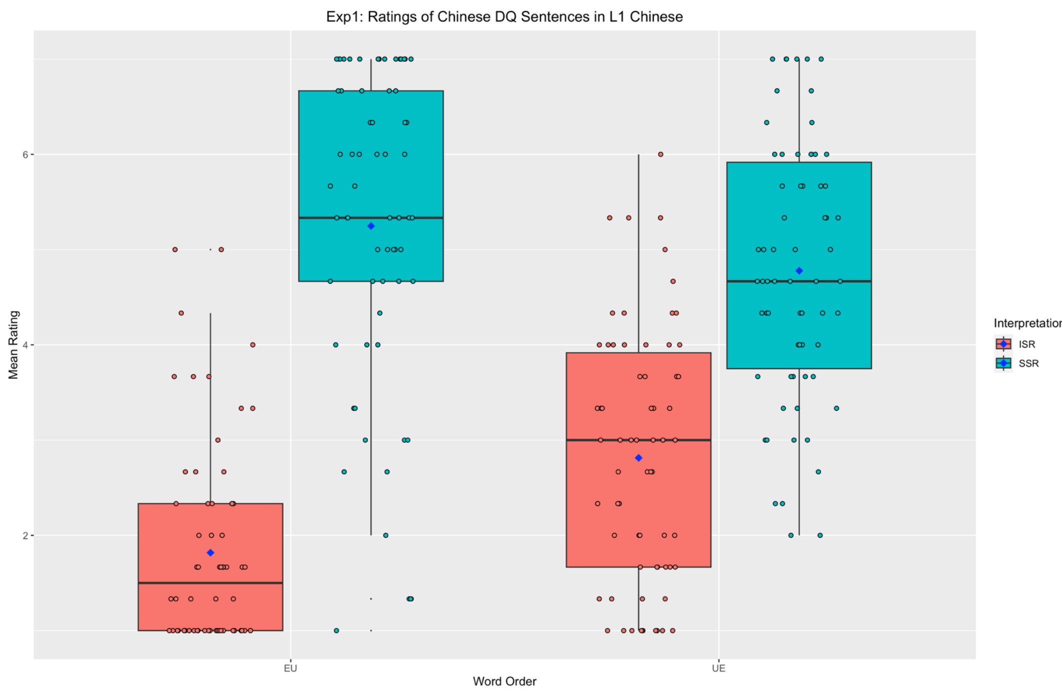

Figure 4.6 Exp1b: Mean ratings of DQ sentences by condition in L1 Chinese.

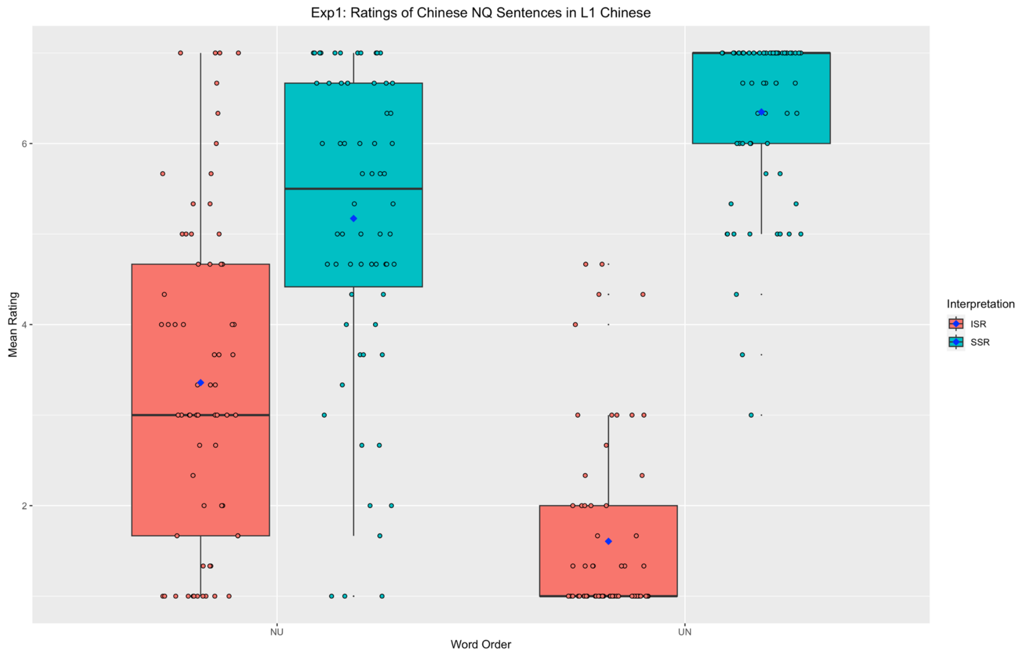

Figure 4.7 Exp1b: Mean ratings of NQ sentences by condition in L1 Chinese.

Table 4.5 Exp1b: Model output for ratings of DQ and NQ sentences in L1 Chinese (*n*=66).

Quantifier scope	Fixed effects	*B*	*SE*	*t*	*p*
DQ	**(Intercept)**	**-0.11**	**0.05**	**-2.18**	**.033**
	Word order	0.11	0.06	1.82	.083
	Interpretation	**0.11**	**0.03**	**34.34**	**< .001**
	Word order*Interpretation	**-0.59**	**0.06**	**-9.22**	**< .001**
NQ	**(Intercept)**	**0.08**	**0.03**	**2.29**	**.028**
	Word order	-0.12	0.06	-1.92	.069
	Interpretation	**1.35**	**0.04**	**36.02**	**< .001**

Word order*Interpretation	**1.19**	**0.07**	**15.94**	**< .001**

4.3.3 Individual results

Individual performance by Chinese speakers was examined. Based on each individual's mean ratings for surface scope and inverse scope readings of different kinds of quantified sentences, participants were categorized into five scope preference patterns, following the procedure used in Experiment 1a. Figures 4.8 and 4.9 illustrate the percentage distribution of each pattern across sentence types. As clearly shown in the figures, most of the Chinese speakers fell into Pattern 1, in that more than half of the participants preferred the surface scope reading in each type of quantified sentence. Quite noticeable are the UN sentences in which 93.9 % of the participants showed a preference for the surface scope reading, while only a small minority exhibited other patterns. Overall, few participants showed a preference for inverse scope readings in each sentence type, except for EU sentences, where 16.7% preferred it.

Based on individual analysis alongside group analysis results and considering the potential influence of language contact on native language interpretation and processing (Dussias & Sagarra, 2007), an exploration was conducted to investigate whether variability in L1 Chinese speakers' acceptance of inverse scope interpretations was affected by their L2 proficiency in English, which serves as an indicator of their exposure to English. The findings for doubly quantified sentences revealed an interaction between L2 English proficiency and word order (b = -0.023, p < .001). As a result of this interaction, two patterns of results arise: 1) their acceptance of inverse scope readings for UE sentences increased as their L2 English proficiency increased, and 2) their acceptance of inverse scope readings for EU sentences decreased as their L2 English proficiency increased. This novel finding indicates the impact of language contact on scope interpretation, as L1 Chinese speakers who had more exposure to English were more inclined to accept inverse scope readings allowed in English for UE sentences. Regarding EU sentences, it is possible that with an increase in their proficiency in L2 English, they became more attuned to the Single Reference principle, leading them to be less receptive to accepting inverse scope readings. Similarly, for negatively quantified sentences, the findings showed an interaction between L2 English proficiency and word order ($b = 0.028$, $p < .001$). Consequently, it is evident that L1 Chinese speakers' acceptance of inverse scope readings for NU sentences decreased as their L2 English proficiency increased. The acceptance of inverse scope readings for NU sentences was higher among L1 Chinese speakers (3.36 on ISR) compared to L1 English speakers (2.17 on ISR). However, as the English proficiency of L1 Chinese speakers increased, their overall acceptance of such readings decreased. This shift in acceptance can be attributed to their tendency to align more with the preferences of native English speakers, who generally dislike inverse scope readings for NU sentences due to pragmatic reasons.

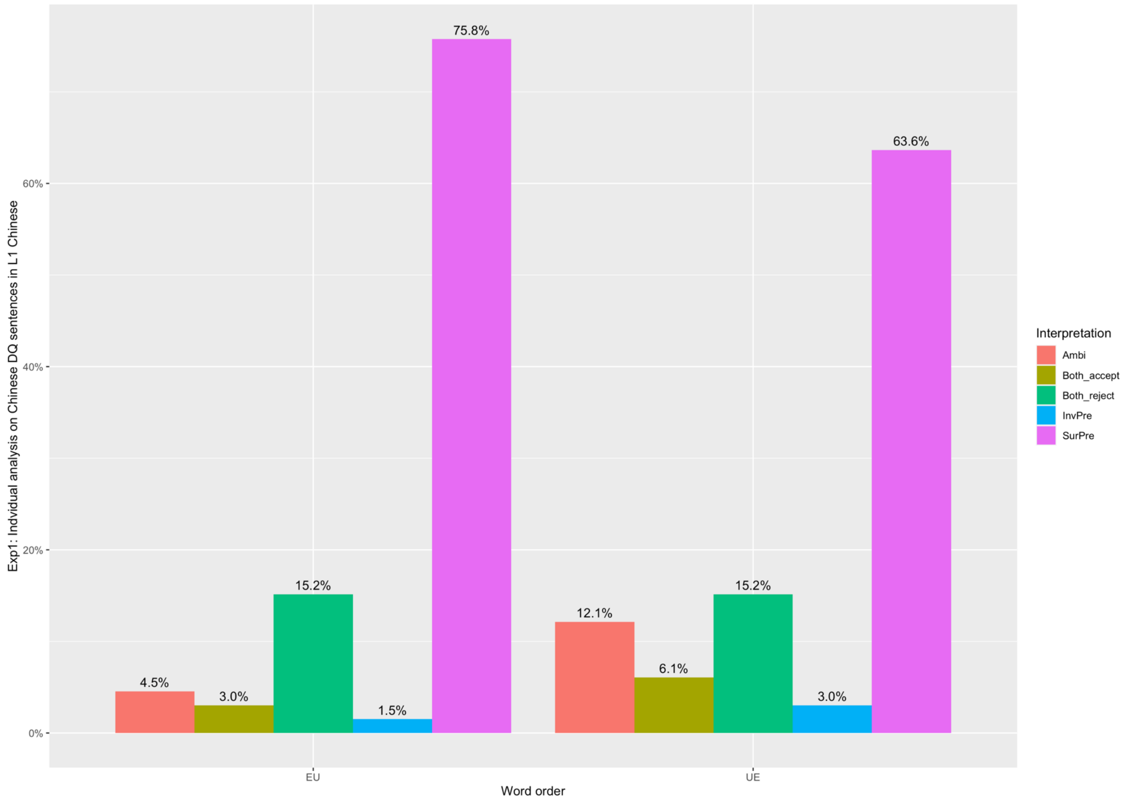

Figure 4.8 Exp1b: Percentage distribution of individual analysis for DQ sentences in L1 Chinese.

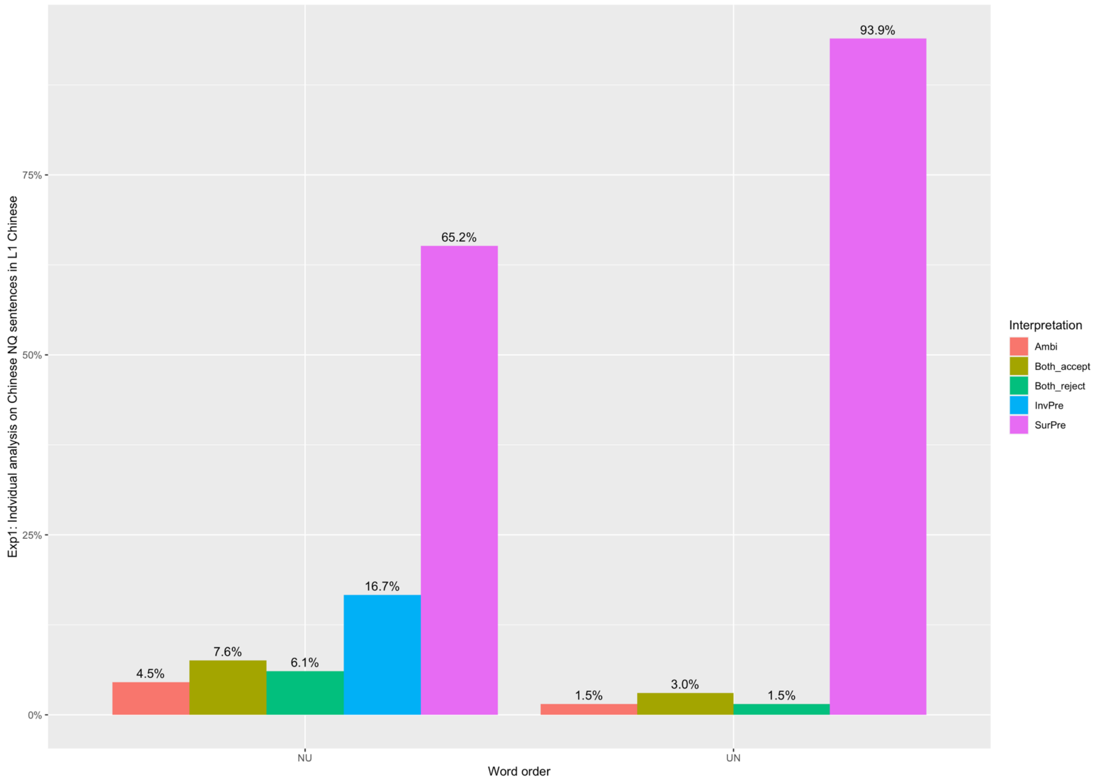

Figure 4.9 Exp1b: Percentage distribution of individual analysis for NQ sentences in L1 Chinese.

4.3.4 DOU test results with L1 Chinese

The DOU test results showed high mean accuracy rates across both participants ($M = 0.85$, $SD = 0.17$) and test items ($M = 0.89$, $SD = 0.12$), indicating the test's appropriateness and reliability. However, the correlation between performance on the DOU test and that on the experimental task is only reported for Experiments 2 and 3, which collected data from learners.

4.4 Comparing L1 English and L1 Chinese

To determine if Chinese has a more restricted distribution of scope ambiguity compared to English, it is crucial to focus on comparing the extent to which each language permits inverse scope readings, given that surface scope readings were reasonably acceptable across the board. As such, linear mixed-effects models were constructed to predict z-scored ratings as a function of word order and language for each type of quantified sentences. Not surprisingly, DQ sentences on the ISR overall received higher acceptance rates in English than in Chinese, qualified by a main effect of language ($b = 0.33$, $p < .001$). Interestingly, the interaction between word order and language was obtained ($b = 0.19$, $p < .001$), due to a main effect of language being present in the EU condition but not in the UE condition. The model output is visualized in Figure 4.10, illustrating that the English speakers accepted the ISR to a larger extent than Chinese speakers for EU sentences and this difference did not reach significance for UE sentences.

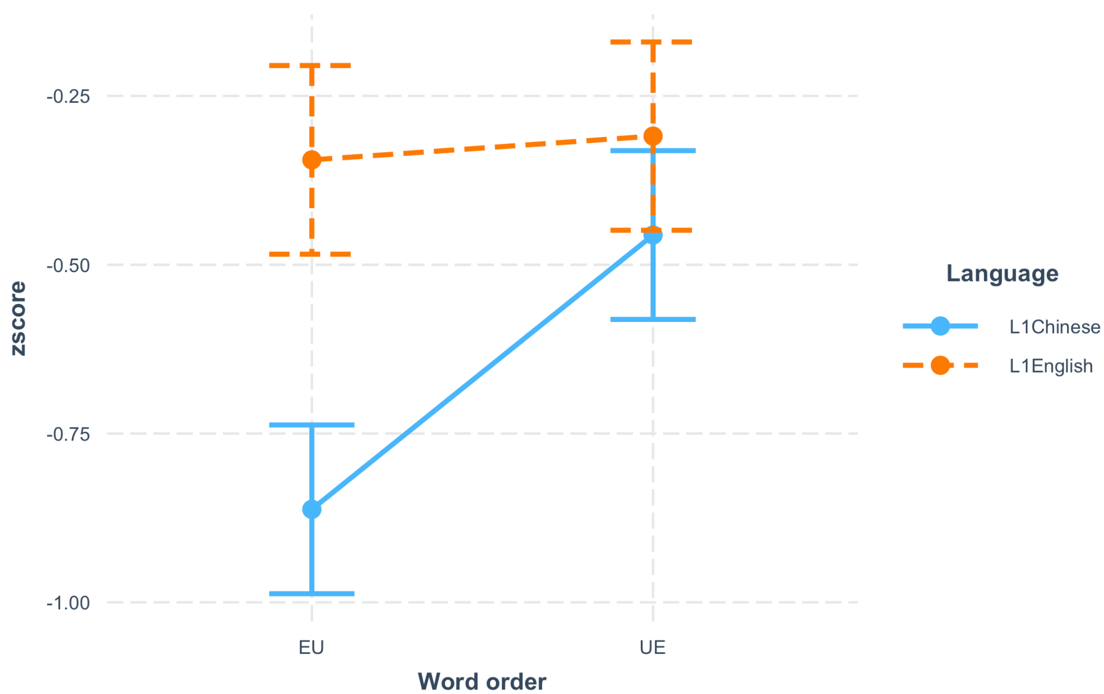

Figure 4.10 Interaction plot for DQ sentences on the ISR by L1 English and L1 Chinese.

As for NQ sentences, a similar modeling procedure also returned a main effect of language ($b = 0.19$, $p < .001$), indicating that NQ sentences on the ISR generally were more acceptable for English speakers than for Chinese speakers. In addition, word order significantly interacted with language ($b = -0.95$, $p < .001$), as depicted in Figure 4.11. A post hoc comparison analysis revealed that UN sentences on the ISR were more acceptable in English than in Chinese ($b = -1.29$, $p < .001$); by contrast, NU sentences on the ISR were less acceptable in English than in Chinese ($b = 0.61$, $p < .001$).

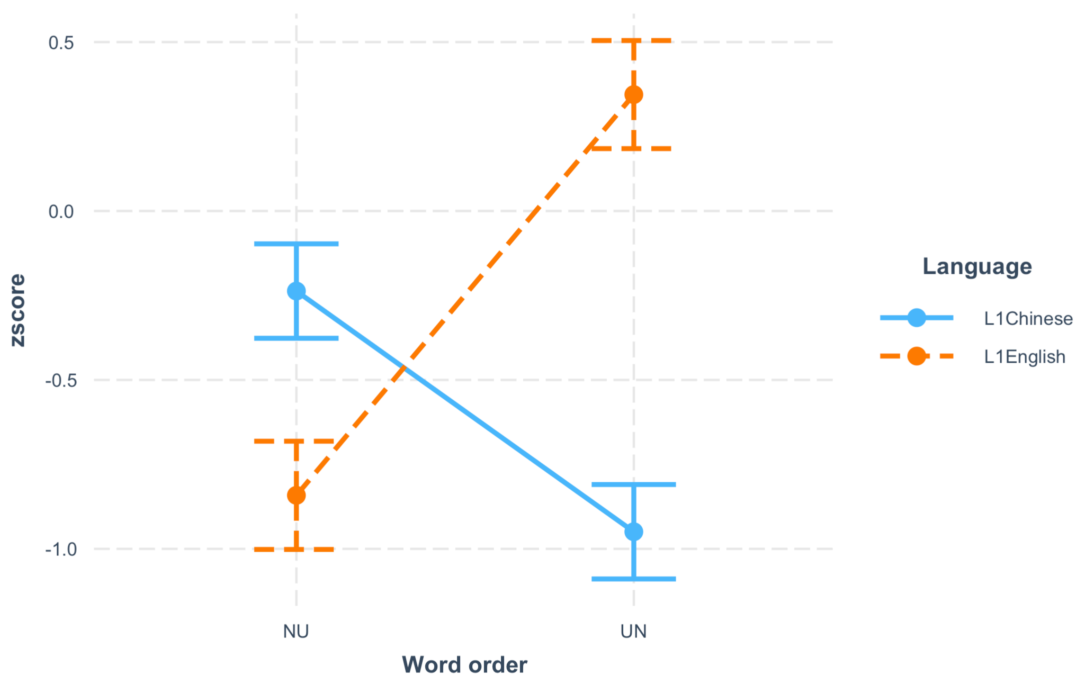

Figure 4.11 Interaction plot for NQ sentences on the ISR by L1 English and L1 Chinese.

4.5 Discussion for Experiment 1

Experiment 1 was designed to address the interpretation of quantifier scope by native speakers of English and Chinese in the absence of supportive context. To summarize, this experiment yielded two major findings. First, while the interpretation of quantifier scope in English can be ambiguous in some cases (i.e., for UE and UN sentences, both readings close to 3.5 or above out of 7), there is a general preference for the surface scope interpretation in all types of quantified sentences, with the exception of negatively quantified 'every...not' sentences, where the preferred reading is the inverse scope interpretation. The inconsistent acceptance of the inverse scope reading can be attributed to the absence of supportive contexts that would facilitate its accessibility. Second, Chinese participants consistently exhibited strong resistance to inverse scope interpretations, particularly as indicated by low ratings (reaching the floor level for EU and UN sentences) for such readings. Overall, these findings support the hypothesis that Chinese is more rigid in permitting scope ambiguity compared to English, with Chinese speakers significantly less receptive to inverse scope readings. Group analysis and individual analysis converge in these respects. Overall, this experiment provides empirical evidence for the cross-linguistic difference in quantifier scope interpretation between English and Chinese.

While cross-linguistic patterns for quantifier scope interpretation are generally clear, it is important to note that some exceptional patterns may require closer attention. First, the findings for doubly quantified sentences in English suggested that the surface scope reading was preferred, consistent with Anderson (2004) where the participants were asked to select an interpretation among others without a supportive context. Second, in the current study, the native English speakers showed a preference for the inverse scope reading in negatively quantified 'every…not' sentences, which differs from the findings of Lee (2009), where English speakers did not display a preference for either reading. This may reflect a task effect. In Lee (2009), participants were presented with each test sentence and given a choice between two possible readings. They had to select the interpretation that corresponded to their initial interpretation of the sentence. In contrast, the current study presented only one possible reading at a time following each sentence. This approach allowed participants to focus on the available reading and process it more deeply. Consequently, the reading that received adequate attention and deeper processing was more likely to stand out as the preferred reading in the current experiment. Third, a notable exception was observed among Chinese speakers, who rated NU sentences for the ISR even higher (mean rating: 3.36) than English speakers (mean rating: 2.17), suggesting that the assumption that inverse scope interpretations are universally unavailable in Chinese might not hold for all quantified structures and warrants further investigation on a structure-by-structure basis.

5.0 Experiment 2: Offline interpretation of quantifier scope in L2 English and Chinese

5.1 Introduction

The findings from Experiment 1 indicate that, in the absence of supportive contexts, both English and Chinese, to some extent, seem to display challenges when it comes to achieving inverse scope interpretations across different types of quantifier scope. However, English speakers showed a greater acceptance of inverse scope interpretations in English compared to Chinese speakers' acceptance of inverse scope interpretations (except for NU sentences) in Chinese. The findings from Experiment 1 have three implications that establish the background for Experiment 2. First, the absence of support from discourse contexts for English speakers, who are known to allow scope ambiguity, may have obscured possible scope interpretations, particularly the less accessible ones such as inverse scope interpretations. In other words, the emergence of less accessible interpretations permitted by the grammar of English may necessitate support from discourse contexts in addition to the interplay between syntax and semantics. Second, Chinese and English exhibit cross-linguistic differences in the interpretation of quantifier scope, particularly in their varying acceptance of inverse scope interpretations. Third, there was variability in quantifier scope interpretation across different orderings and among participants. Notably, a novel finding emerged that Chinese permits inverse scope interpretations to a greater extent for NU sentences than for UN sentences and even permits more extensive inverse scope interpretations for NU sentences compared to English. In Experiment 2, it will be investigated whether inverse scope interpretations for NU sentences in Chinese would see an increase in acceptance when supportive contexts are available. If this is the case, sentences with the ISR may not be regarded as a typical learnability problem for English-speaking learners acquiring Chinese.

Against the backdrop of the role of discourse context and the cross-linguistic difference, a series of truth-value judgment tasks (TVJT) in Experiment 2 was employed to enable the detection of all possible scope interpretations by presenting sentences along with appropriate discourse contexts. This approach aimed to reveal the interpretation that may otherwise be obscured for speakers of the target language grammar in the absence of supportive contexts. Furthermore, a bidirectional experimental design was adopted to investigate learners of L2 English and Chinese, two languages shown in the previous chapter to exhibit cross-linguistic differences in quantifier scope interpretation. This design takes into account factors such as L1 transfer and L2 input, which may influence quantifier scope interpretation in L2 learners. The main objective of Experiment 2 was to examine the offline interpretation of doubly quantified and negatively quantified sentences among L2 learners of English and Chinese. Additionally, the potential impact of L2 proficiency on the interpretation was explored. Therefore, Experiment 2 addresses the following research question, accompanied by two specified sub-questions (RQ2 restated below).

RQ2) *Can Chinese-speaking learners of English and English-speaking learners of Chinese interpret quantifier scope in their respective target languages as an L2 in a manner comparable to the native speakers, when supportive context is given in the offline written story-based TVJT?*

(RQ2-1) Which learner group is more successful in arriving at the target interpretation: Chinese-speaking learners of English or English-speaking learners of Chinese?

(RQ2-2) How does L2 proficiency affect L2 learners' performance on quantifier scope interpretation?

A written TVJT was conducted separately with Chinese-speaking learners of English and English-speaking learners of Chinese, respectively. Additionally, the task was administered to two control groups: native English speakers and native Chinese speakers. The data collected from these control groups served as the baseline for the L2 experiments.

5.2 Experiment 2a: L2 English Truth Value Judgment task

5.2.1 Methods

5.2.1.1 Participants

This experiment included 64 Chinese-speaking learners of English (mean age = 22 years, *SD* = 1.89, 40 females) and 46 native English speakers (mean age = 34 years, *SD* = 6.19, 18 females). None of these participants took part in Experiment 1. Native speakers of English were recruited via Prolific. Participants who met the following criteria specified in Prolific were included in the study as native speakers: between the ages of 18 and 45, monolingual English speakers, residing in either the US or the UK, and having completed at least a bachelor's degree or higher. The L2 participants in this experiment were undergraduate or graduate students from a university in eastern China who were recruited through WeChat groups. WeChat is a widely used social media platform among Chinese-speaking communities worldwide, and the groups used for this recruitment were restricted to members from that university as the participants. These L2 participants reported that Chinese was their first language and dominant in their daily use since childhood. None had the experience of living in any English-speaking countries. All participants were paid for their participation. They had normal or corrected-to-normal vision and reported no history of language deficits. Table 5.1 summarizes L2 participants' demographic and language proficiency information.

Table 5.1 Information for the L2 group with Chinese learners of English

	L2 group (*n* = 64)
	M (*SD*)
Age (years)	22 (1.89)
Onset age of L2 learning (years)	6.4 (2.5)
Length of instruction (years)	15.1 (2.9)
Self-ratings (1-10)	
Speaking	5.8 (2.0)
Listening	6.3 (2.1)
Reading	6.9 (1.5)
Writing	5.7 (1.6)
LexTALE (0-100 %)	69.3 (12.1)
Gender (male/female)	24/40

The proficiency of the L2 participants was measured using the LexTALE English Test (Lemhöfer & Broersma 2012), which has demonstrated a strong correlation with standard proficiency assessments such as the Quick Placement Test (QPT). Based on the Common

European Framework (CEF) for language levels, participants were categorized into intermediate (or lower) (score equal to 59 or below), upper intermediate (score of 60–80), and advanced (score of 80–100) levels, depending on their performance in the LexTALE task. In addition, the L2 participants self-rated their abilities in four English skills (speaking, listening, reading, and writing) using a 10-point Likert Scale. As depicted in Figure 5.1, the proficiency levels of these participants exhibit sufficient variability spanning a wide range from the low level to the advanced level. The LexTALE scores significantly correlated with the average scores of the four language skills (r = 0.59, $p < .001$).

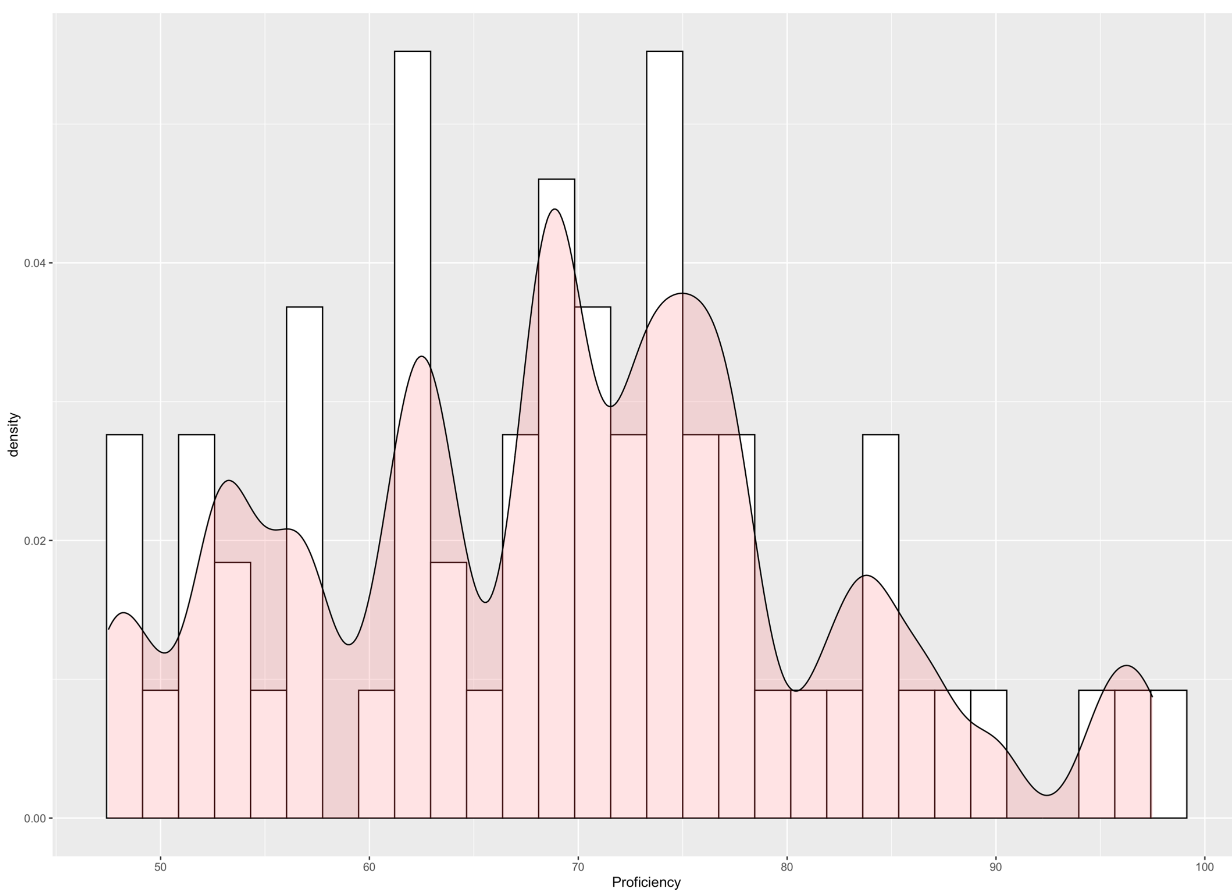

Figure 5.1 Distribution of LexTALE scores (range: 47.5 - 97.5) within the L2 group.

5.2.1.2 Design and materials

The task was a written TVJT. Each item was comprised of a story context and a corresponding test sentence. For example, the test sentence 'Every child climbed a tree' was presented following a story where each of the three children had a unique tree to climb in a climbing race. In this story context, on its surface scope reading, the test sentence would be true since each child climbed a different tree. However, on its inverse scope reading, the sentence would be false because it is not the case that each child climbed the same tree. Hence, the reasoning behind one's decision is that if participants access the inverse scope reading of this sentence, they are likely to reject it, whereas if they access this sentence on its surface scope reading, they are likely to accept it. The test sentence was presented with an alternative story context, which resulted in the inverse scope reading. In addition to the 'interpretation' factor, another manipulated factor was word order, where test sentences such as 'A child climbed every tree', with the existential quantifier preceding the universal quantifier, were presented following contexts for either of the readings. As a result, participants were tested in four different conditions for both doubly quantified and negatively quantified sentences. As in Experiment 1, doubly quantified (DQ) and negatively quantified (NQ) sentences were also manipulated between items in this experiment. Therefore, their results were analyzed separately.

For this experiment, 48 quadruplets of target items were constructed by crossing Interpretation (Surface scope vs. Inverse scope) and Word Order (UE vs. EU; UN vs. NU) for DQ and NQ sentences, respectively. A quadruplet of the target items for Experiment 2a is presented in Table 5.2. These stimuli were distributed evenly across four lists according to a Latin Square design, such that no participant saw the same test sentence twice and the lists were counterbalanced across participants. Eventually, each participant was assigned a list containing 12 target items (three items per condition) for DQ sentences and 12 target (also three items per condition) items for NQ sentences, along with 31 filler items with varying linguistic structures, some of which included quantifiers (e.g., Every apple fell from the tree) or negation (e.g., Nobody is wearing a hat). The test sentences used in this experiment were identical to those in Experiment 1a, which had been previously checked for their acceptability and plausibility. A native English speaker created the story contexts, which were then reviewed by another native English speaker for plausibility and appropriateness. Coh-Metrix (version 3.0) (Graesser et al., 2011), an automated text analysis tool, was used to generate indices that characterize texts at different levels of language, which include type-token ratio, number of sentences, number of words, text easability principal component z scores (for syntactic simplicity)[37], and CELEX Log frequency for all words. For DQ and NQ sentences, conditions were similar for all indices (all *ps* > .05), except the number of sentences which differed by Interpretation in DQ sentences ($F\ (1, 44) = 4.737$, $p = .0349$)[38]. This nevertheless indicates that texts across conditions were generally comparable in terms of these textual features. Appendix B lists the items for Experiments 2 and 3.

Table 5.2 Example conditions and items in Experiment 2a

Quantifier scope	Condition	Test sentence	Story context

[37] It provides scores on the ease or difficulty of the linguistic characteristics of a given text.

[38] The main effect of the number of sentences was driven by a higher number of sentences in the SSR condition compared to the ISR condition. Supposedly, the number of sentences should have a minimal impact on how a text is read, particularly in offline experimental settings.

DQ	UE-SSR	Every child climbed a tree.	At break one day, three kids decided to have a race and see who could climb to the top of a tree the fastest. There were three trees in the playground that were the same height and so they decided to each climb a different tree so it would be fair. They counted down and then each kid raced to the top of their tree.
	UE-ISR	Every child climbed a tree.	At break one day, three kids decided to have a race and see who could climb to the top of a tree the fastest. There was only one tree in the playground and so they decided they would take turns climbing it and see how long it took for each of them. So, one by one each of the kids took a turn climbing the same tree.
	EU-SSR	A child climbed every tree.	In the second grade, there was a boy who loved to climb trees. There were three trees in the

			playground in a row and one day he decided to see if he could climb them all in one day. So, he climbed them all, one after the other, and he succeeded in climbing all three before break ended.
	EU-ISR	A child climbed every tree.	In the second grade, there were three kids who loved to climb trees and there were three trees in the playground. One day they decided to have a race and see who could climb the fastest. They each started at the bottom of a tree, counted down and then each kid raced to the top of their tree.
NQ	UN-SSR	Every horse didn't jump over the fence.	Three horses looked at a fence in their field. One horse suggested jumping over the fence, but the other horses said the fence was too tall to jump over, so the plan was abandoned.

UN-ISR	Every horse didn't jump over the fence.	Three horses looked at a fence in their field. One horse suggested jumping over the fence. The first two horses succeeded. The third one, who had hurt its leg the day before, decided not to jump.
NU-SSR	The horse didn't jump over every fence.	One day, a horse decided to test his jumping skills. He found a row of fences and then decided to challenge himself by jumping over all of them. But some of the fences were too high and he only succeeded in jumping over some of the fences.
NU-ISR	The horse didn't jump over every fence.	One day, a horse decided to test his jumping skills. He found a row of fences and then decided to challenge himself by jumping over all of them. But the fences were too high and so he failed in each attempt.

5.2.1.3 Procedure

Both the L1 and L2 participants completed the written TVJT. In the same experimental session, they also completed the visual-world eye-tracking experiment. The specifics of the eye-tracking experiment task will be introduced in chapter 6. In the TVJT, participants were given sufficient time to read the story context. Following the story presented on the same screen was a test sentence for which the participants were asked to decide the extent to which the sentence matched the description of the story context on a 7-point Likert Scale (1 = 'doesn't match at all'; 7 = 'completely matches'). Participants were explicitly told to focus their attention on the interpretive aspects of the stories and the sentences. The L1 participants completed all the tasks in the following order: eye-tracking task, TVJT, and language background questionnaire. These tasks were administered using the Gorilla Experimental Builder (Anwyl-Irvine et al., 2020) (https://gorilla.sc/). Participants completed the tasks on their own computers via Prolific at home.

The L2 participants completed a battery of tasks in a specific order, which included an eye-tracking task, definiteness test, background questionnaire, TVJT, and LexTALE. This task order was deemed most effective to avoid a repetition effect. Although a Latin Square design was used, some test sentences may still repeat across tasks. The eye-tracking task was administered prior to the TVJT, because participants had brief exposure to each sentence during the online task and were unlikely to retain clear memories of them in the offline TVJT. Additionally, several other surveys were inserted between the two tasks to further minimize any potential repetition effects. The eye-tracking task, definiteness test, and background questionnaire were administered through Gorilla, while TVJT and LexTALE were administered using Qualtrics. The LexTALE is a lexical decision task that does not have a time limit, and it involves participants making a Yes or No decision on whether a given sequence of letters is a real English word or not. Even if they did not understand the meaning of a string, they had to respond with Yes if they believed that it was a legitimate word. In total, there were 63 trials, out of which 42 were real words. The L2 participants completed all tasks in a spacious and quiet office at the institution where the experiment was conducted in China. The participants were closely supervised by a trained researcher throughout the duration of the experiment.

Prior to these tasks, the participants filled out a consent form and were given detailed instructions for each task. Preceding the main experimental session of TVJT was a block of 5 practice items. Instructions to each task were given in English for the L1 participants and in Chinese for the L2 participants. The L1 participants completed the experiment in approximately 30 min, while the L2 group took around 45 min to finish.

5.2.1.4 Analysis

Given that the design and measurements of this experiment were similar to those of Experiment 1, the statistical procedures used in the previous experiment were applied here and thus will not be reiterated in this section. During data trimming, out of the total participants, five individuals were removed from further analysis - one L1 participant and four L2 participants - as they did not pass the required performance filters on the filler items. Details on how the LexTALE scores were handled are elaborated here. After the removal of three dummies, the test score (0–100) was calculated out of 60 trials using the formula provided by Lemhöfer and Broersma (2012):

((number of words correct/40*100) + (number of nonwords correct/20*100)) / 2

To address the gradient nature of L2 proficiency and mitigate the limitations of creating categorical groups, which can reduce statistical power and produce spurious results in certain datasets (Leal, 2018; Plonsky & Oswald, 2017; Van Hell & Tanner, 2012), the regression model included LexTALE scores as a continuous variable, rather than grouping learners based on cutoff points to determine their proficiency levels. Before being included in the L2 data as a fixed effect, the LexTALE scores were scaled and centered. Group was included as a fixed factor, which was sum coded (-0.5 for L1 vs. 0.5 for L2).

5.2.2 Group analysis results

Mean ratings across conditions for the experimental items are summarized in Table 5.3. The descriptive results for DQ sentences and NQ sentences are visualized in Figure 5.2 and Figure 5.3, respectively. Results of DQ sentences are presented first and then those for NQ sentences. While the statistical analyses returned the information about all fixed effects and their interactions, I will only focus on the results that are related to the research questions. As such, results concerning the main effects of word order, language, and their interaction will not be discussed, because these results are of less theoretical interests. The linear mixed-effects models were fit to ratings of both groups and the model output is summarized in Table 5.4.

Table 5.3 Mean ratings of the TVJT by L1 and L2 English speakers in Experiment 2a

Quantifier scope	Group	Word order	Interpretation	
			SSR	ISR
DQ	L1	UE	6.58 (0.89)	5.41 (1.92)
		EU	6.32 (1.29)	4.64 (2.47)

	L2	UE	5.83 (2.08)	4.48 (2.42)
		EU	6.26 (1.60)	1.59 (1.48)
NQ	L1	UN	5.99 (1.55)	5.25 (2.17)
		NU	6.16 (1.46)	5.95 (1.77)
	L2	UN	5.86 (2.03)	2.84 (2.31)
		NU	6.24 (1.55)	5.51 (2.06)

The descriptive statistics for DQ sentences show that both L1 and L2 English speakers rated surface scope readings higher than inverse scope readings. Upon examining DQ sentences with a focus on their inverse scope readings, L1 English speakers, who were tested on UE and EU sentences, gave an average rating exceeding 4.5 out of 7. These ratings support the availability of inverse scope readings in these sentence structures. Except for EU sentences with inverse scope readings (1.59 out of 7), L2 learners generally accepted inverse scope readings for DQ sentences (average ratings larger than 4.5 out of 7). Remarkably, L1 speakers accepted the inverse scope readings to a larger extent than L2 speakers for each type of quantified sentences even though a supportive context had been provided for both groups of participants. [39]

The statistical modeling returned a main effect of interpretation ($b = 0.95$, $p < .001$), driven by higher ratings for surface scope readings than for inverse scope readings. There was a significant interaction between word order and interpretation ($b = -0.77$, $p < .001$), with higher

[39] Although L2 speakers, in general, accepted the SSR readings, L1 speakers showed a higher acceptance of the SSR readings for each type of quantified sentence, reflected in their overall higher ratings. This finding is not surprising since L2 learners are presumably limited in their proficiency and have not yet reached native-like proficiency in quantifier scope interpretation.

ratings for surface scope readings than for inverse scope readings in either UE or EU sentences. Interpretation significantly interacted with language group (b = 0.36, p = .005), indicating that surface scope readings received higher ratings than inverse scope readings in both the L1 and L2 groups. Interestingly, the interaction between interpretation and language group revealed a main effect of language group that was only present in surface scope readings (b = -0.19, p = .014), but not in inverse scope readings (b = 0.17 p = .148), suggesting that L1 and L2 speakers did not differ in their acceptance of DQ sentences on their inverse scope readings.

Crucially, a three-way interaction among word order, interpretation, and group was significant (b = -0.98, p < .001). Based on the results of post-hoc comparisons, for both UE and EU sentences, the SSR received higher ratings compared to the ISR in both the L1 and L2 groups (L1: b = -0.64, p < .001; L2: b = -0.50, p < .001 for UE sentences; L1: b = -0.92, p < .001; L2: b = -1.76, p < .001 for EU sentences). Furthermore, regarding UE sentences, no significant language group effect was observed for either the SSR or the ISR. This suggests that L1 and L2 English speakers did not differ in their acceptance of the SSR and the ISR (all ps > .5). Regarding EU sentences, the results showed that L1 English speakers rated the SSR lower than L2 English speakers (b = -0.34, p = .002)[40], and that L1 English speakers rated the ISR higher than L2 English speakers (b = 0.51, p = .004).

---

[40] While L1 speakers rated SSR slightly higher than L2 speakers for EU sentences, as shown in Table 5.3 with raw ratings, L2 learners gave significantly higher ratings for SSR than L1 English speakers when considering z-transformed ratings.

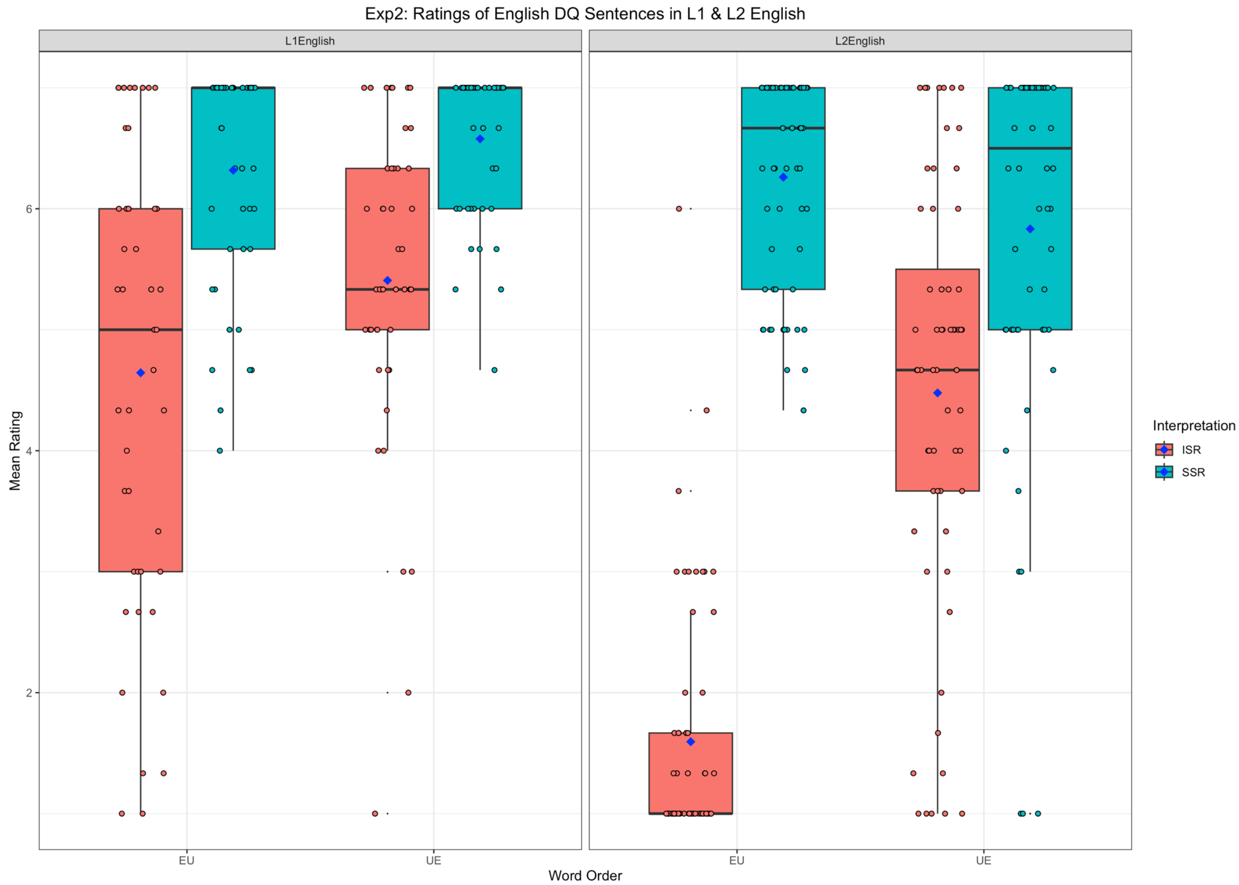

Figure 5.2 Exp2a: Mean ratings of DQ sentences by condition in L1 & L2 English.

Turning to the results for NQ sentences, surface scope readings were rated higher than inverse scope readings across the board. Descriptively, L2 learners were less likely to accept NQ sentences on their inverse scope readings compared to L1 speakers. This was especially true for UN sentences, as shown in the visualization (See Figure 5.3). Overall, both L1 and L2 speakers demonstrated the ability to access inverse scope readings in the UN and NU sentences, with average ratings above 4.5 out of 7. However, L2 learners had difficulties with UN sentences, rating the inverse scope readings significantly lower (mean rating: 2.84). This suggests that L2 learners encountered challenges in accessing the inverse scope reading in UN sentences.

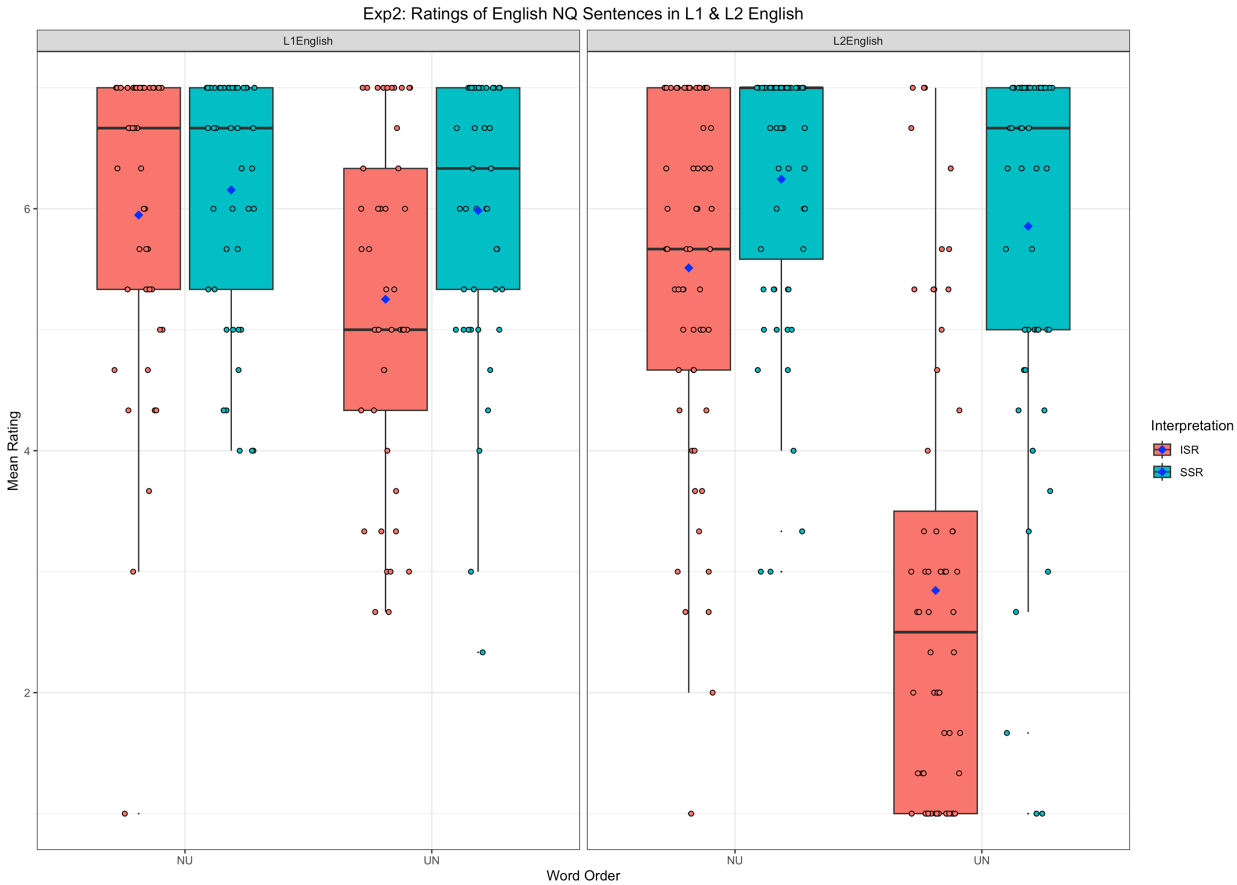

Figure 5.3 Exp2a: Mean ratings of NQ sentences by condition in L1 & L2 English.

The statistical analyses indicated a main effect of interpretation (b = 0.95, p < .001), induced by higher ratings for surface scope readings than for inverse scope readings. An interaction emerged between word order and interpretation (b = -0.77, p < .001), due to surface scope readings being rated higher than inverse scope readings across both UN and NU sentences. The interaction between word order and interpretation was also driven by lower ratings of ISRs for UN than for NU sentences (b = 0.69, p < .001), but comparable ratings between UN and NU sentences for SSRs (b = 0.12, p = .113). An interaction also emerged between interpretation and language group (b = 0.36, p < .001), resulting from overall higher ratings for surface scope readings than for inverse scope readings in both language groups. This interaction further indicated that L1 and L2 speakers differed in their acceptance of NQ sentences only for surface scope readings (*b* = -0.40, *p* < .001), but not for inverse scope readings (*b* = 0.05, *p* = .577).

More importantly, a three-way interaction was observed among word order, interpretation, and group (*b* = -0.98, *p* = .036). Based on the post-hoc comparisons, both language groups rated the SSR higher than the ISR in all quantified sentence types, except for NU sentences in L1 English where the difference was not significant (*b* = -0.11, *p* = .756). Specifically, for L1 UN sentences, the SSR was rated higher than the ISR with significance (*b* = -0.40, *p* = .0497), while for UN (*b* = -1.13, *p* < .001) and NU (*b* = -0.28, *p* = .012) sentences by L2 learners, the SSR received significantly higher ratings compared to the ISR. Additionally, L2 learners gave significantly higher ratings to the SSR for NU sentences compared to L1 speakers (*b* = -0.42, *p* < .001). However, no significant difference was found between L1 speakers and L2 learners in their ratings of the ISR for NU sentences (*b* = -0.26, *p* = .184). In relation to UN sentences, L2 learners rated the SSR significantly higher than L1 speakers[41] (*b* = -0.37, *p* = .022). The difference in ratings of the ISR between L1 and L2 speakers was marginal (*b* = 0.36, *p* = .071). As a note, in cases where the *p*-values were slightly greater than 0.05 (which are conventionally taken to be marginally significant without additional justification), a Bayesian analysis was performed to statistically corroborate the presence or absence of certain effects.

For the marginal difference in ratings of the ISR between L1 and L2 speakers in UN sentences, a Bayesian analysis was conducted. The Bayes factors for the 'group' term for the fixed

[41] The raw ratings for L1 English speakers were slightly higher than those for L2 English speakers in UN sentences on their SSRs, as in Table 5.3. However, the z-scores for SSR with UN sentences were higher for L2 learners than for L1 speakers (0.48 vs. 0.11), which explains why L2 learners outperformed L1 speakers in terms of SSR ratings with UN sentences.

effects in the linear mixed-effects models were calculated using the BayesFactor package (Morey & Rouder 2018; version 0.9.12). The Bayes factors reported here are of BF10 type, which is the ratio of the likelihood of the data given one hypothesis (H_1) that the 'group term is present compared to the likelihood of the data given another hypothesis (H_0) that there is no 'group' term, where H_1 is the alternative hypothesis and H_0 is the null hypothesis. Following Jeffreys et al. (1939/1961), a Bayes factor greater than 3 (H_1 accepted) or else less than 1/3 (H_0 accepted) represents substantial evidence; anything between 1/3 and 3 represents inconclusive evidence. The Bayes factor here is 1.42, suggesting that the result is statistically inconclusive and fails to rule out the possibility that L1 and L2 speakers differ in ratings of the ISR with UN sentences.

The potential influence of L2 proficiency on the pattern of results was examined, but no significant influence was found in any of the cases. One's definiteness knowledge, which might serve as the triggering input for L2 learners, was explored for its correlation[42] with the inaccessible scope readings, namely EU sentences on their inverse scope readings (M = 1.59) and UN sentences on their inverse scope readings (M = 2.84). Chinese-speaking learners of English achieved the accuracy rate of 0.43 (SD = 0.08) on the definiteness test. Chinese learners of English often struggle with the definiteness properties of NPs because Chinese lacks an article system (Feng, 2019). As a result, it is reasonable that L2 learners tend to perform poorly on the definiteness test.

Since the triggering input should primarily play a role in interpretation that is not easily informed from the L2 input, definiteness knowledge was explored to determine its potential

[42] Regression analysis was not performed in this case, as it was not expected that individual differences in definiteness knowledge would fully account for the variation in ratings observed for a particular scope. In other words, it was not assumed that definiteness knowledge was the causal factor responsible for the differences in ratings among individuals.

influence on the inverse scope interpretation across different types of quantifier scope. To investigate the relationship between ISR ratings and individuals' definiteness test scores for each type of quantified sentence, I conducted a series of correlation analyses. The results showed that for the UE sentences, there was no significant correlation ($r = .06$, $p = .242$). For the EU sentences, a small[43] negative correlation was found ($r = -.26$, $p < .001$). A weak positive correlation was observed for UN sentences ($r = .14$, $p = .008$), whereas no significant correlation was found for NU sentences ($r = -.05$, $p = .363$).

Table 5.4 Exp2a: Model output for ratings of DQ and NQ sentences in L1(n=45) and L2 (n=60) English.

Quantifier scope	Fixed effects	*B*	*SE*	*t*	*p*
DQ	(Intercept)	-0.02	0.05	-0.4	.690
	Word order	**0.37**	**0.08**	**4.42**	**<.001**
	Interpretation	**0.95**	**0.06**	**15.49**	**<.001**
	Group	0.01	0.07	0.16	.874
	Word order*Interpretation	**-0.77**	**0.11**	**-6.85**	**<.001**
	Word order*Group	0.19	0.11	1.81	.079
	Interpretation*Group	**0.36**	**0.12**	**2.89**	**.005**
	Word order*Interpretation*Group	**-0.98**	**0.22**	**-4.36**	**<.001**

[43] According to Plonsky and Oswald (2014, p.889), correlation coefficients close to .25 were considered small, .40 medium, and .60 large.

	(Intercept)	**-0.02**	**0.05**	**-0.4**	**.001**
	Word order	**0.37**	**0.08**	**4.42**	**<.001**
	Interpretation	**0.95**	**0.06**	**15.49**	**<.001**
	Group	**0.01**	**0.07**	**0.16**	**.018**
NQ	**Word order*Interpretation**	**-0.77**	**0.11**	**-6.85**	**<.001**
	Word order*Group	**0.19**	**0.11**	**1.81**	**.006**
	Interpretation*Group	**0.36**	**0.12**	**2.89**	**<.001**
	Word order*Interpretation*Group	**-0.98**	**0.22**	**-4.36**	**.036**

5.2.3 Individual analysis results

The group analysis indicates that English speakers, on the whole, exhibited greater tolerance for inverse scope readings compared to Chinese learners of English. Nevertheless, Chinese-speaking learners of English demonstrated considerable ability in acquiring inverse scop readings in most cases. To better contextualize these patterns, individual data were examined. The five distinct patterns (repeated in (39)) used in Experiment 1 were adopted here for analysis since the same measurement method was employed in this experiment.

(39) Patterns of ratings in L1 and L2 English for Experiment 2a

Pattern 1: Preference for the surface scope (SurPre)

Pattern 2: Preference for the inverse scope (InvPre)

Pattern 3: Ambivalence with no scope preference (Ambi)

Pattern 4: Acceptance of both readings (Both_accept)

Pattern 5: Rejection of both readings (Both_reject)

Figure 5.4 and Figure 5.5 display the percentage distribution of these patterns among individual L1 and L2 participants' DQ sentences, respectively. As clearly shown in Figure 5.4, most of the L1 English participants preferred the surface scope reading for DQ sentences, but a significant proportion of them also accepted both readings. This is especially true in UE sentences, for which 48.9% of the L1 English participants preferred the surface scope reading and 46.7% of them accepted both readings. In the L2 group, the pattern for the EU sentences was consistent: Most L2 learners preferred the surface scope reading. Although L2 learners generally displayed a surface scope preference for the UE sentences, they also exhibited a wider range of preferences, including acceptance of both readings (18.3%) and a preference for the inverse scope readings (16.7%), albeit in much smaller proportions for the latter two cases.

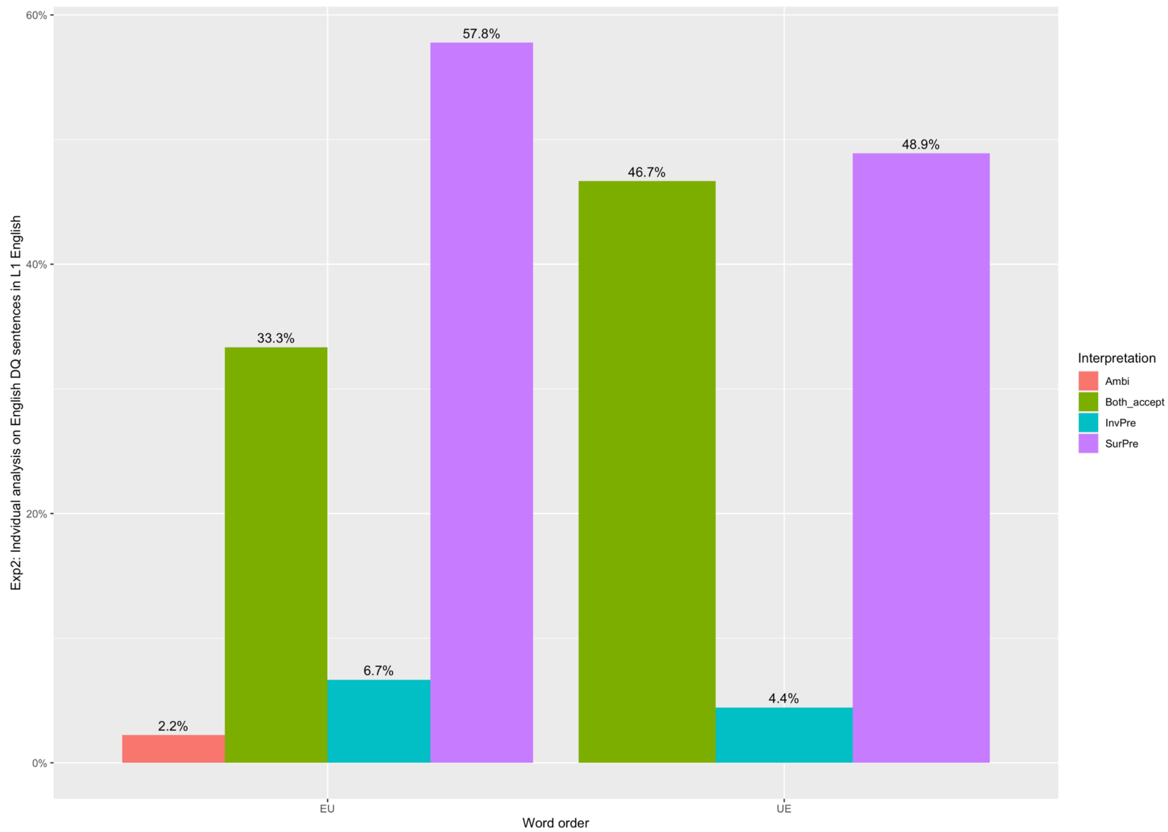

Figure 5.4 Exp2a: Percentage distribution of individual analysis for DQ sentences in L1 English.

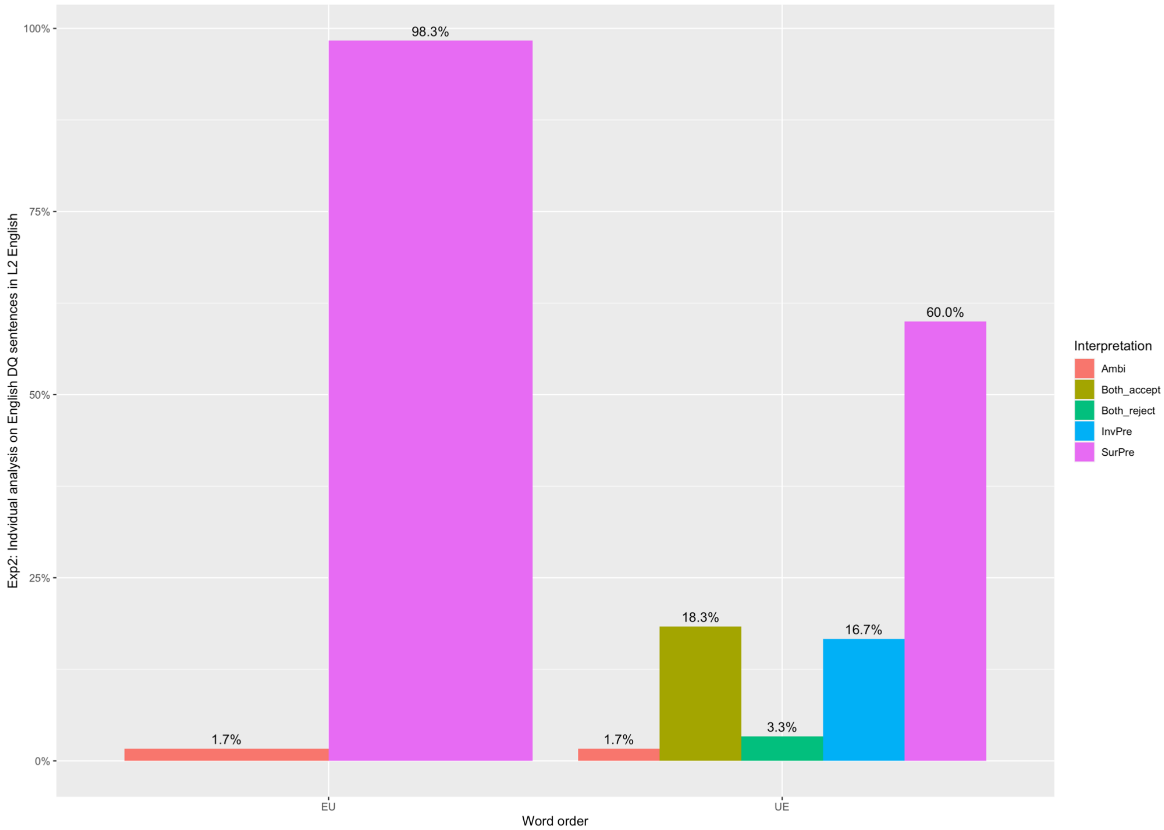

Figure 5.5 Exp2a: Percentage distribution of individual analysis for DQ sentences in L2 English.

Figure 5.6 and Figure 5.7 display the percentage distribution of these five patterns for NQ sentences by L1 and L2 participants, respectively. As for L1 English speakers, most of them allowed a wide range of interpretations. More than half of the L1 participants fell into Pattern 4 for NU sentences, in that these participants accepted both readings. For UN sentences, the majority fell into either Pattern 1 or Pattern 4, demonstrating the acceptance of both readings and the preference for the surface scope reading. For the L2 participants, NU sentences permit a broad range of interpretations ranging from 'acceptance of both readings' to 'the preference for the surface scope reading'. The majority of L2 learners in the UN sentences showed a marked preference for the surface scope reading.

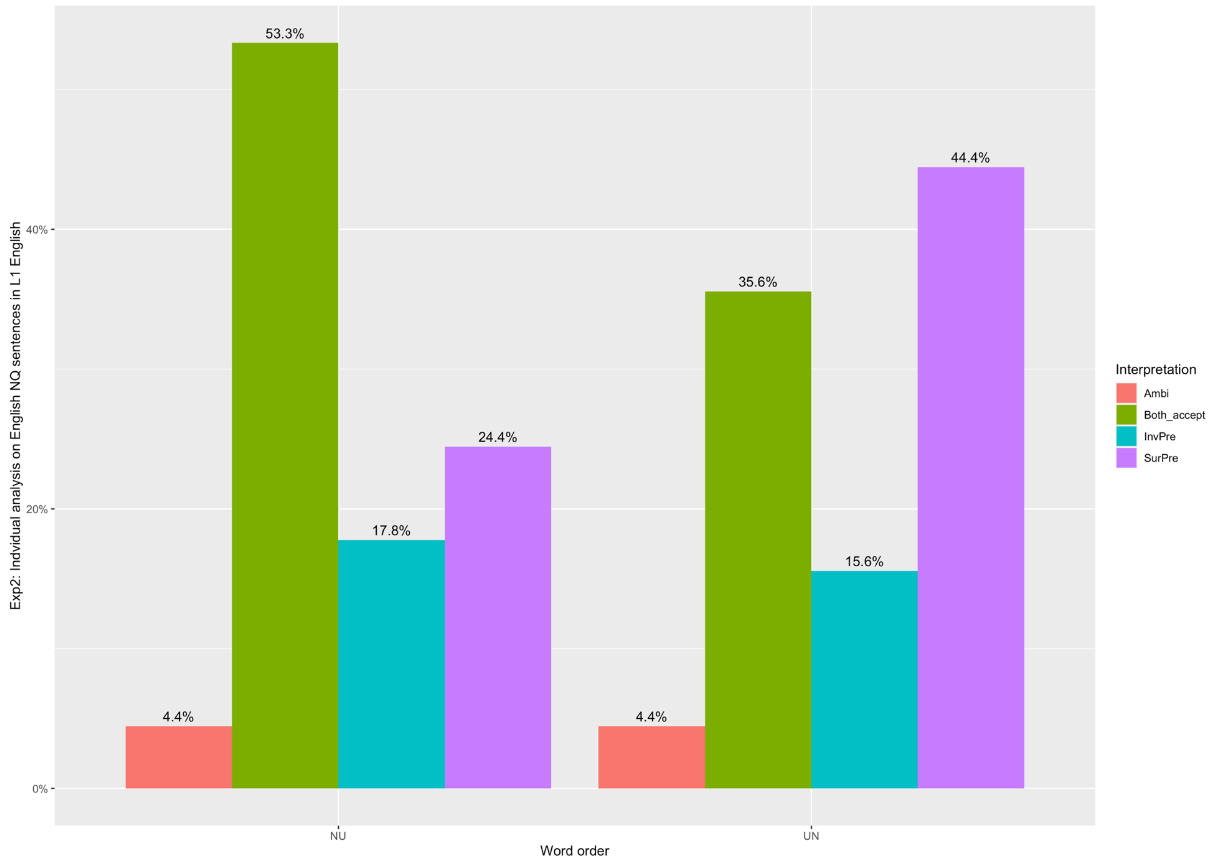

Figure 5.6 Exp2a: Percentage distribution of individual analysis for NQ sentences in L1 English.

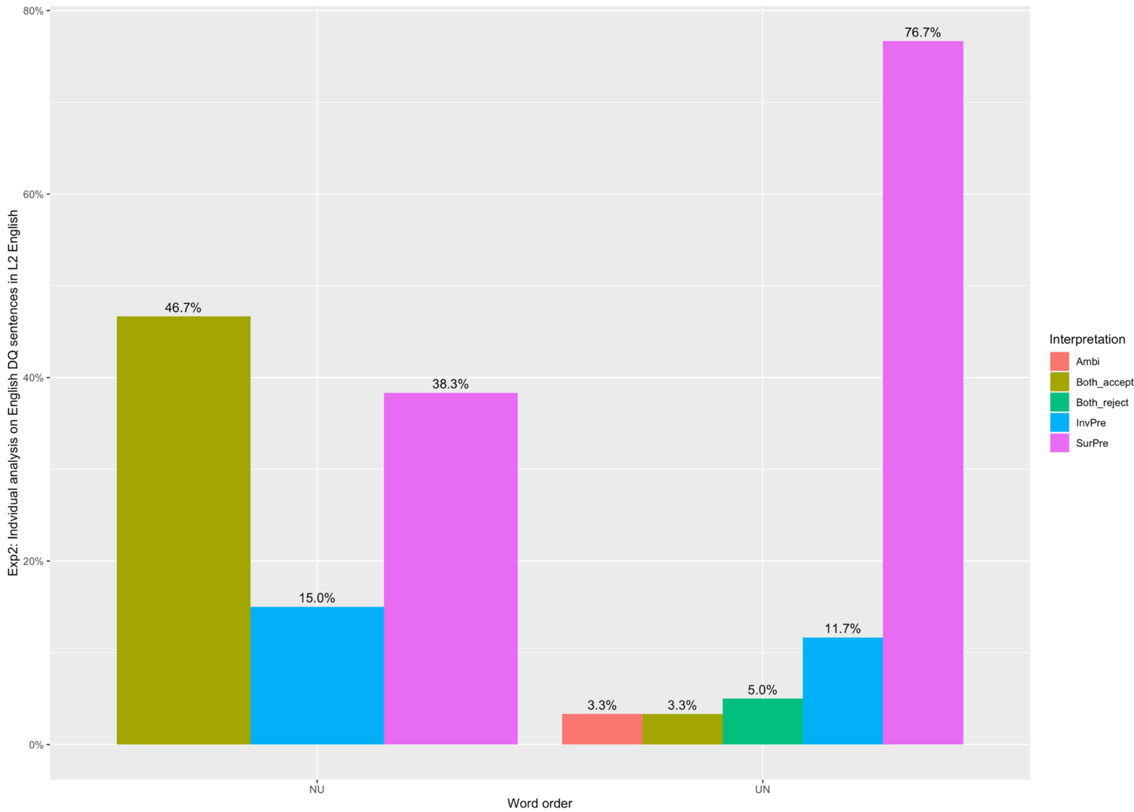

Figure 5.7 Exp2a: Percentage distribution of individual analysis for NQ sentences in L2 English.

5.3 Experiment 2b: L2 Chinese Truth Value Judgment task

5.3.1 Methods

5.3.1.1 Participants

An experimental group of 43 English-speaking learners of Chinese[44] (mean age = 25 years, *SD* = 10.78, 22 females) and a control group of 53 native Chinese speakers[45] (mean age = 19 years, *SD* = 1.19, 25 females) participated in Experiment 2 for the written TVJT and Experiment 3 for the eye-tracking task. These L1-Chinese participants were undergraduate students at an eastern Mainland Chinese university during the testing phase. They were born and raised in China and had no previous exposure to English-speaking countries or regions. The native Chinese participants[46] achieved a mean score of 53 (*SD* = 7.47) on LexTALE, which places them in the range of low proficiency speakers according to Lemhöfer and Broersma (2012). As such, their English knowledge should have little impact on their performance in these experimental tasks (in the

[44] The study included an eye-tracking task and a subsequent TVJT, which 43 English-speaking learners of Chinese were asked to complete. However, only 42 participants finished both experiments; one participant only completed the eye-tracking task.

[45] Out of 53 native Chinese participants asked to complete both experiments, eye-tracking data from two participants were not recorded for unknown reasons. As a result, data from only 51 participants were available for eye-tracking data analysis.

[46] It is unavoidable to include L1-Chinese participants who have learned English for some time, because universities in China require all students to take English courses.

Chinese version)[47]. Regarding English-speaking learners of Chinese, these participants were recruited through various avenues, and thus had a diverse range of language backgrounds. A general criterion for recruiting L1-English L2-Chinese participants was that they must have at least two years of prior experience with the Chinese language by the time of testing, so that they would be able to handle quantifier scope, a linguistic phenomenon that is quite sophisticated in Chinese. This is particularly important since the TVJT involves a written story, which can be challenging for those with limited Chinese proficiency.

The L2 participants were asked to complete an extensive language background questionnaire, which elicited information about age, gender, dominant/first language, experience of living in Chinese-speaking countries or regions, if family members speak Mandarin Chinese or a dialect of Chinese at home, birthplace, age of first exposure to Chinese (age of onset), whether they were enrolled in or had completed any Chinese courses, length of instruction (in semester of learning at college), years of learning Chinese in secondary and/or high schools, self-ratings for proficiency in Chinese on a 10-point Likert scale[48], among other questions. Out of the 43 English-speaking learners of Chinese, all reported English as their first[49] and dominant language, with their exposure to Chinese primarily through classroom instruction; most of them (79%; 34/43) were

---

[47] The control group for the L2 Chinese study consisted of Chinese L1 speakers with English training, while the control group for the L2 English study comprised English L1 speakers with no Chinese training. Future research can consider including a group of L1 Chinese speakers with no English learning experiences as the control group for L2 English learners. By implementing this approach, the control groups for both L2 experiments can maintain a high level of homogeneity.

[48] 1 = very poor, 10 = very good

[49] Among them, four reported both English and Chinese as the first languages that they learned.

enrolled in or had completed different levels of Chinese courses at universities in the US or China; 40% (17/43) of them had experience living in Chinese-speaking countries or region (mostly Mainland China); 35% (15/43) of them reported that they have one or two parent(s) who speaks at least one dialect of Chinese at home[50]; 93% of them[51] (40/43) reported to be born and raised in the English-speaking countries. These participants' Chinese proficiency was independently measured by a Mandarin proficiency test, which is based on HSK[52] and adopted from Xu and Yuan (2022). This test has a maximum score of 40. Table 5.5 summarizes the basic descriptive information about the L1-English L2-Chinese participants.

Table 5.5 Information for the L2 group with English-speaking learners of Chinese

	L2 group (*n* = 43)
	M (*SD*)
Age (years)	25 (10.78)
Age of onset (years)	15.4 (12.7)
Length of college instruction (semesters)	2.9 (2.5)
Total years of learning	2.6 (2.7)
Self-ratings (1-10)	

[50] These participants are presumably heritage speakers of Chinese.

[51] Of the three participants, all of whom were born in China, two were adopted and brought to the US before the age of 2, while the third immigrated to the US at the age of 7.

[52] HSK (*Hànyǔ shuǐpíng kǎoshì*) is an international standardized exam that assesses the Chinese language proficiency of non-native speakers. This test is administered by *Hanban*, an agency of the Ministry of Education of the People's Republic of China (PRC).

Speaking	5.7 (1.8)
Listening	6.8 (1.8)
Reading	5.5 (2.1)
Writing	4.2 (1.9)
Mandarin proficiency score (out of 40)	20.8 (10)
Gender (male/female/non-binary)	20/22/1

As a point of clarification, those who reported having family members who speak Chinese may have been exposed to both English and Chinese languages in their home environment and are commonly referred to as heritage language learners. In contrast, non-heritage learners did not have the benefit of such exposure to Chinese language at home. However, heritage learners were not treated as a separate group from non-heritage learners (see Fang & Xu, 2022, for a similar practice), since both groups exhibited variability in the aforementioned language characteristics. As such, the language characteristics of all participants, including whether they were heritage or non-heritage learners, are included to model the general proficiency in the L2 Chinese group. To assess the general proficiency in the L2 Chinese group, a composite L2 proficiency measure as reported in previous L2 studies (Bice & Kroll, 2021; Luque & Morgan-Short, 2021) was created. This measure was generated through the following procedures: (a) z-scoring participants' scores on each of the language proficiency related measures, including birthplace[53], experience of living in

[53] In order to conduct statistical analysis on the impact of birthplace, experience of living in Chinese-speaking countries or regions, and Mandarin Chinese or dialect spoken at home on language proficiency, the three categorical variables were transformed into numerical ones. Birthplace was assigned a value of +1 for participants born in Chinese-speaking countries or regions and otherwise -1. Similarly, experience of living in Chinese-speaking countries

Chinese-speaking countries or regions, if family members speak Mandarin Chinese or a dialect of Chinese at home, length of college instruction, total years of learning, self-rated proficiency in listening, speaking, reading, and writing, age of acquisition, Mandarin proficiency score, and (b) combing the z-scores: z(birthplace) + z(experience of living in Chinese-speaking countries or regions) + z(if family members speak Mandarin Chinese or a dialect of Chinese at home) + z(length of college instruction) + z(total years of learning) + z(listening) + z(speaking) + z(reading) + z(writing) – z(age of acquisition) + z(Mandarin test score). Remarkably, the z-scored measures of language characteristics, other than Mandarin proficiency scores, exhibited a significant positive correlation with Mandarin proficiency scores ($r = 0.7$, $p < .001$), as illustrated in Figure 5.8. The distribution of the composite measure is illustrated in Figure 5.9.

All participants reported normal or corrected-to-normal vision. They were compensated for their participation.

---

or regions was assigned a value of +1 for those who had lived in Chinese-speaking areas and -1 for those who had not. Finally, if participants reported that Mandarin Chinese or a dialect of Chinese was spoken at home, it was assigned a value of +1; if not, it was assigned a value of -1.

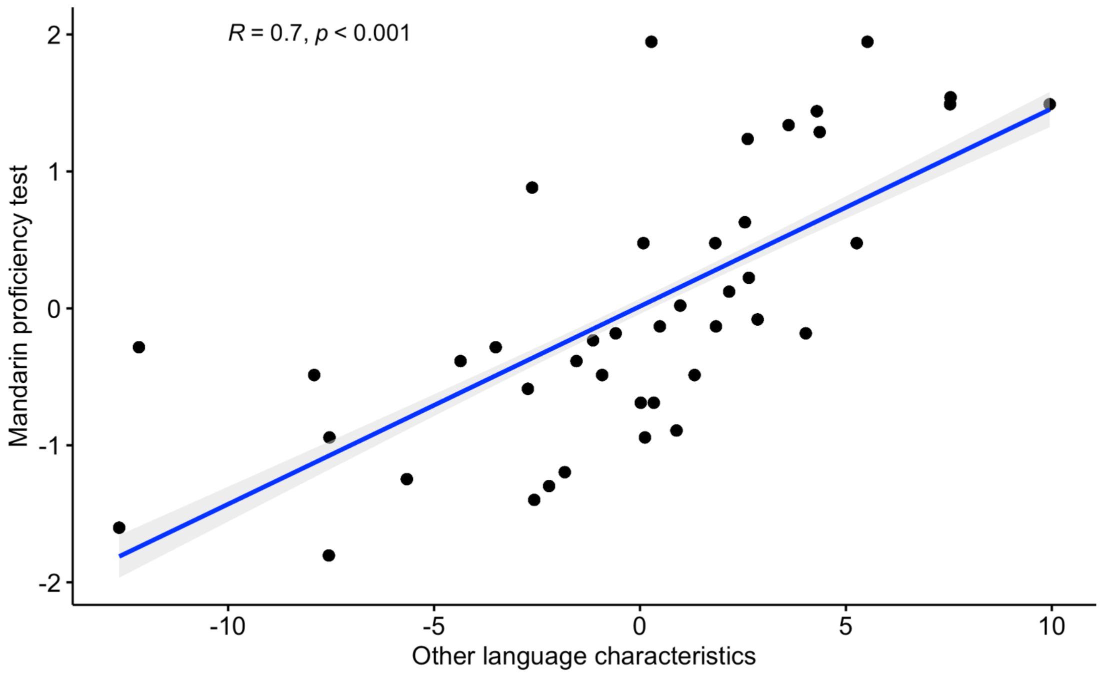

Figure 5.8 Correlation between Mandarin proficiency test and other language characteristics.

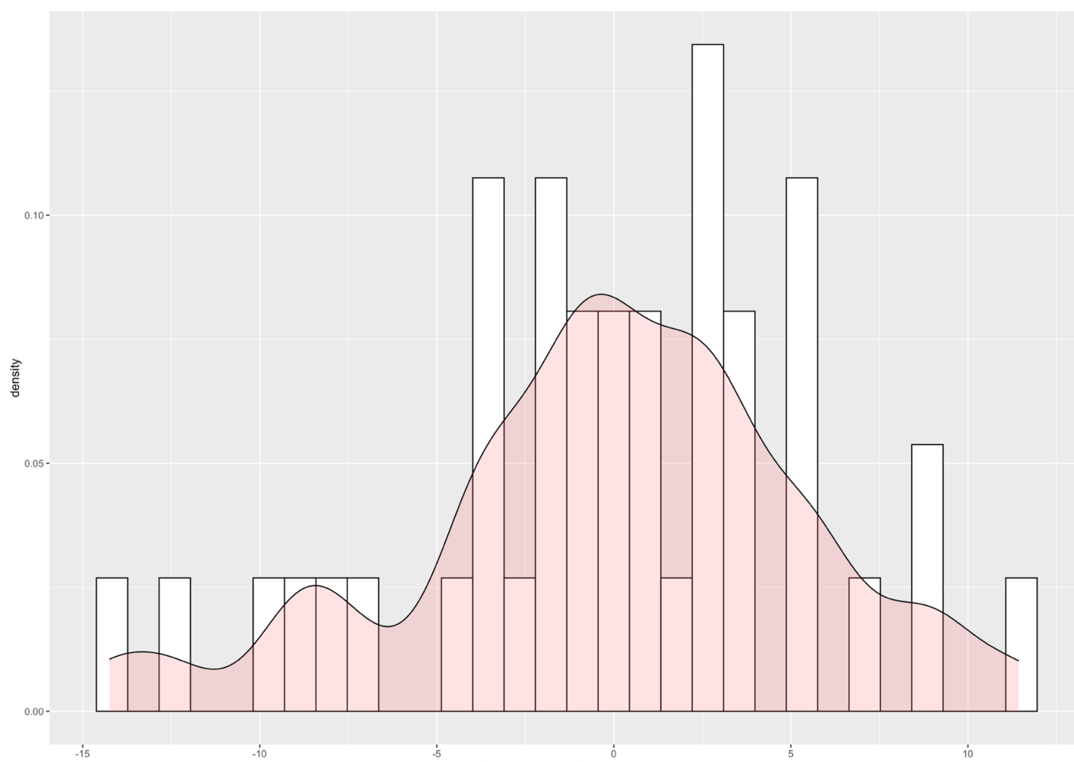

Figure 5.9 Distribution of composite z-scores within the L2 Chinese group.

5.3.1.2 Design and materials

The experimental stimuli including the target items and fillers for the Chinese TVJT were translated from the English stimuli used in Experiment 2a. The translation was conducted by two bilingual speakers proficient in both English and Chinese. To accommodate the different proficiency levels of the target L2 participants, the lexical items and sentences were modified as needed[54]. After modification, an intermediate student who was selected from the participant pool but did not take part in the experiment reviewed all the experimental materials to ensure that the lexical items were familiar to participants. In the TVJT for L1 English L2 Chinese participants[55], all the test sentences and written stories were presented with simplified Chinese characters along with *pinyin* (i.e., a Chinese phonetic transcription system) on top of each character. This was done to facilitate text comprehension for learners of Chinese (Bassetti & Lu, 2016). An example item for this experiment is illustrated in (40).

[54] The modifications primarily focused on lexical items assumed to pose challenges for English-speaking learners of Chinese. Throughout this process, Chinese sentences were carefully refined to ensure acceptability and plausibility. Overall, the norming conducted for the English stimuli is applicable to the Chinese stimuli as well.

[55] For the TVJT, only texts for the L2 group were complemented with *pinyin*. Native Chinese speakers read plain Chinese texts.

40. [**Story**]

yǒuyītiān sān gè háizi juédìngjìnxíng yī chǎng bǐsài kànkanshuínéngzuìkuài dì pá dàoshùdǐng cāochǎng shàngyǒusān
有一天，三个孩子决定进行一场比赛，看看谁能最快地爬到树顶。操场上有三

kē shù tāmen de gāodù xiāngtóng suǒyǐ tāmen juédìng měirén pá yī kē bùtóng de shù zuìhòu měigè háizi dōu pá dào le
棵树，它们的高度相同，所以他们决定每人爬一棵不同的树。最后，每个孩子都爬到了

shù dǐng
树顶。

'At break one day, three kids decided to have a race and see who could climb to the top of a tree the fastest. There were three trees in the playground that were the same height and so they decided to each climb a different tree so it would be fair. They counted down and then each kid raced to the top of their tree.'[56]

[**Test sentence**]

měi yí gè hái zi dōu pá le yī kē shù
每一个孩子都爬了一棵树。

'Every child climbed a tree.'

[56] For the sake of readability, English texts were added here, although they were not included in the experiment.

The design of this experiment is identical to that of Experiment 2a. Given that the Chinese stimuli were almost directly translated from the English version, the textual characteristics of the Chinese stimuli should be comparable to those of the English version reported in Experiment 2a.

5.3.1.3 Procedure

Unless otherwise noted, other procedures were similar to those in Experiment 2a. The Chinese L1 participants who participated in the study completed several tasks in a specific order. First, they performed an eye-tracking task using Gorilla. Next, they completed three additional tasks, namely the TVJT, LexTALE, and language background questionnaire, which were administered through Qualtrics. The Chinese L2 participants completed a series of tasks in a predetermined order, using different online platforms. Specifically, they completed an eye-tracking task, the DOU test, and a language background questionnaire on Gorilla. In addition, they completed the TVJT and Mandarin proficiency test on Qualtrics. Both groups of participants individually completed these tasks on their own at home. Right before the TVJT, a glossary of words (used in the stories for TVJT) considered difficult was listed for the L2 participants. The glossary included *pinyin* and English translations for each word. The participants were encouraged to review the glossary and test themselves with follow-up multiple-choice questions[57] related to these words. This helped them prepare for the TVJT and improve their understanding of the material. Instructions to each task were given in Chinese for the L1 participants and in English for the L2 participants. The L1

[57] An example of this test is a multiple-choice question 'What's the meaning of '篱笆(líba)' along with four choices: a. candle, b. umbrella, c. clean, d. fence. This test was not graded, but simply to consolidate the participants' understanding of the specific words and eliminate any extraneous effects resulting from unfamiliarity with the words.

participants completed the experiment in approximately 45 min, whereas the L2 group required around 60 min to finish.

5.3.1.4 Analysis

The analysis procedures were followed, as in Experiment 2a, resulting in the exclusion of one Chinese L1 participant during the data trimming stage. No Chinese L2 participant was excluded. Overall, data from both groups were fit using linear mixed-effects models, which included group, word order, and interpretation as fixed effects. The L2 data were analyzed using linear mixed-effects models with the composite scores included as a continuous fixed factor. The group analysis and individual analysis results are reported in the following sections, respectively.

5.3.2 Group analysis results

Mean ratings of DQ and NQ sentences across conditions by L1 and L2 Chinese speakers are summarized in Table 5.6. These results are visualized in Figure 5.10 for DQ sentences and in Figure 5.11 for NQ sentences[58]. The output for statistical analysis is presented in Table 5.7.

---

[58] Considering that Chinese speakers maintain contact with English through classroom learning and the potential influence of this contact on their native language processing (Dussias & Sagarra, 2007), an investigation was conducted to explore whether L2 English proficiency could predict any variance in L1 Chinese acceptance of inverse scope readings. The findings revealed that L2 English proficiency did not have any impact on the variance in L1 Chinese speakers' acceptance of inverse scope readings in Chinese for both doubly and negatively quantified sentences. Systematic research could be done in the future in the domain of quantifier scope interpretation by including native Chinese speakers spanning a wider range of English proficiency.

Table 5.6 Mean ratings of the TVJT by L1 and L2 Chinese speakers in Experiment 2b

Group	Quantifier scope	Word order	Interpretation	
			SSR	ISR
DQ	L1	UE	6.12 (1.76)	3.97 (2.48)
		EU	6.58 (1.20)	1.76 (1.80)
	L2	UE	5.17 (2.07)	4.76 (2.09)
		EU	5.78 (1.64)	2.93 (1.94)
NQ	L1	UN	6.49 (1.42)	1.72 (1.67)
		NU	6.19 (1.85)	5.74 (1.94)
	L2	UN	5.16 (2.15)	2.72 (1.80)
		NU	4.94 (2.34)	5.02 (1.99)

As for DQ sentences, both L1 and L2 Chinese speakers rated surface scope readings higher than inverse scope readings. It is not surprising to observe the larger discrepancy between SSR and ISR for L1 speakers compared to L2 speakers. Specifically, the average ratings provided by native Chinese speakers for ISR hovered around or below 3.5 (UE: 3.97; EU: 1.76), suggesting a potential limitation in the availability of inverse scope readings for DQ sentences. For English-speaking learners of Chinese on DQ sentences, they have acquired the absence of ISR for EU sentences, but accepted ISR for UE sentences.

A significant main effect of interpretation emerged ($b = 0.98$, $p < .001$), with higher ratings for SSR than for ISR with DQ sentences. The interaction between word order and interpretation was also significant ($b = -1$, $p < .001$), due to the fact that SSR was rated higher than ISR for both UE ($b = -0.48$, $p < .001$) and EU ($b = -1.48$, $p < .001$) sentences. Another significant interaction was observed between interpretation and group ($b = -0.61$, $p < .001$), because (a) ISR was given lower ratings compared to SSR by both L1 and L2 Chinese speakers, and (b) ISR was rated significantly lower by L1 Chinese speakers than L2 Chinese speakers. The lack of a significant three-way interaction between word order, interpretation, and group suggests that L1 and L2 Chinese speakers exhibit a similar pattern in how word order and interpretation interact.

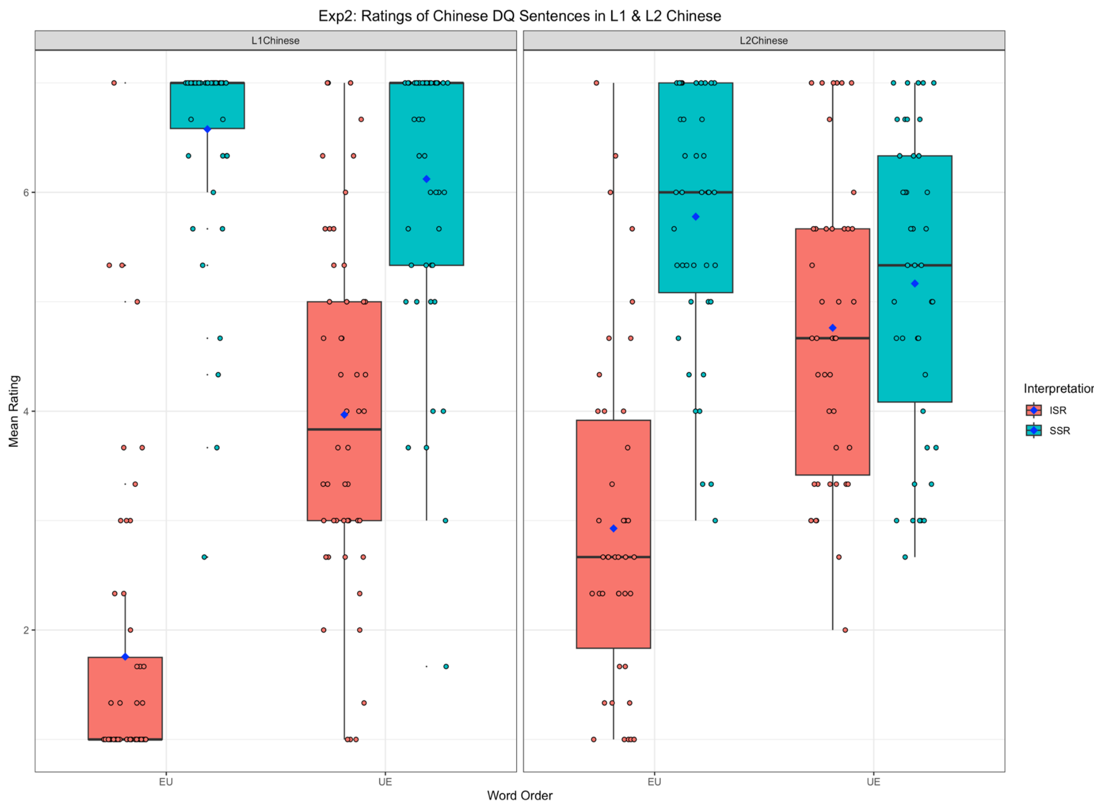

Figure 5.10 Exp2b: Mean ratings of DQ sentences by condition in L1 & L2 Chinese.

As for NQ sentences, both L1 and L2 Chinese speakers accepted SSRs. However, ratings of ISRs varied depending on the sentence type. Specifically, L1 and L2 speakers found ISR for UN sentences unacceptable, while ISR for NU sentences was deemed acceptable. The statistical analysis returned a main effect of interpretation ($b = 0.73$, $p < .001$), induced by higher ratings for SSR than for ISR. A significant interaction arose between word order and interpretation ($b = 1.33$,

$p < .001$), which indicated that SSR received higher ratings than ISR for UN sentences ($b = -1.39$, $p < .001$), but not for NU sentences. An interaction between interpretation and language group yielded significant results ($b = -0.47$, $p < .001$). Chinese speakers in both L1 ($b = -0.97$, $p < .001$) and L2 ($b = -0.49$, $p < .001$) groups rated SSR higher than ISR, with a more pronounced difference observed in L1 Chinese speakers (as reflected in different effect sizes between groups). However, no significant difference was found between L1 and L2 Chinese speakers in their (low) ratings of ISR ($b = -0.09$, $p = .313$). There was a three-way interaction among word order, interpretation, and group ($b = -0.54$, $p = .015$). Based on the post-hoc comparisons, for L1 Chinese speakers, UN sentences on SSRs were rated higher than ISRs ($b = -1.76$, $p < .001$). NU sentences did not differ in the ratings between SSRs and ISRs among L1 Chinese speakers ($b = -0.17$, $p = .428$). The pattern for L2 Chinese speakers was similar to that for L1 Chinese speakers: SSRs for UN sentences were rated higher than ISRs ($b = -1.02$, $p < .001$), and SSRs did not differ from ISRs with NU sentences in their ratings ($b = 0.04$, $p = .99$). More importantly, L1 and L2 Chinese speakers did not differ in their ratings of NU sentences on the ISR ($b = 0.17$, $p = .51$). L1 Chinese speakers rated the SSR higher than L2 Chinese speakers in NU sentences ($b = 0.38$, $p = .024$). Regarding UN sentences, L1 Chinese speakers rated the ISR lower than L2 Chinese speakers ($b = -0.35$, $p = .024$). In contrast, L1 Chinese speakers rated the SSR higher than L2 Chinese speakers ($b = 0.39$, $p = .006$).

Table 5.7 Exp2b: Model output for ratings of DQ and NQ sentences in L1(n=52) and L2 (n=43) Chinese

Quantifier scope	Fixed effects	*B*	*SE*	*t*	*p*
DQ	(Intercept)	0.05	0.04	1.2	.237

	Word order	**0.29**	**0.08**	**3.64**	**.001**
	Interpretation	**0.98**	**0.06**	**17**	**<.001**
	Group	0.09	0.07	1.35	.184
	Word order*Interpretation	**-1**	**0.11**	**-8.99**	**<.001**
	Word order*Group	-0.07	0.12	-0.59	.561
	Interpretation*Group	**-0.61**	**0.12**	**-5.28**	**<.001**
	Word order*Interpretation*Group	-0.03	0.22	-0.14	.890
NQ	**(Intercept)**	**0.09**	**0.04**	**2.38**	**.021**
	Word order	**-0.56**	**0.06**	**-9.09**	**<.001**
	Interpretation	**0.73**	**0.06**	**12.32**	**<.001**
	Group	**-0.15**	**0.07**	**-2.06**	**.045**
	Word order*Interpretation	**1.33**	**0.11**	**12.14**	**<.001**
	Word order*Group	**0.25**	**0.12**	**2.16**	**.039**
	Interpretation*Group	**-0.47**	**0.12**	**-4**	**<.001**
	Word order*Interpretation*Group	**-0.54**	**0.22**	**-2.47**	**.015**

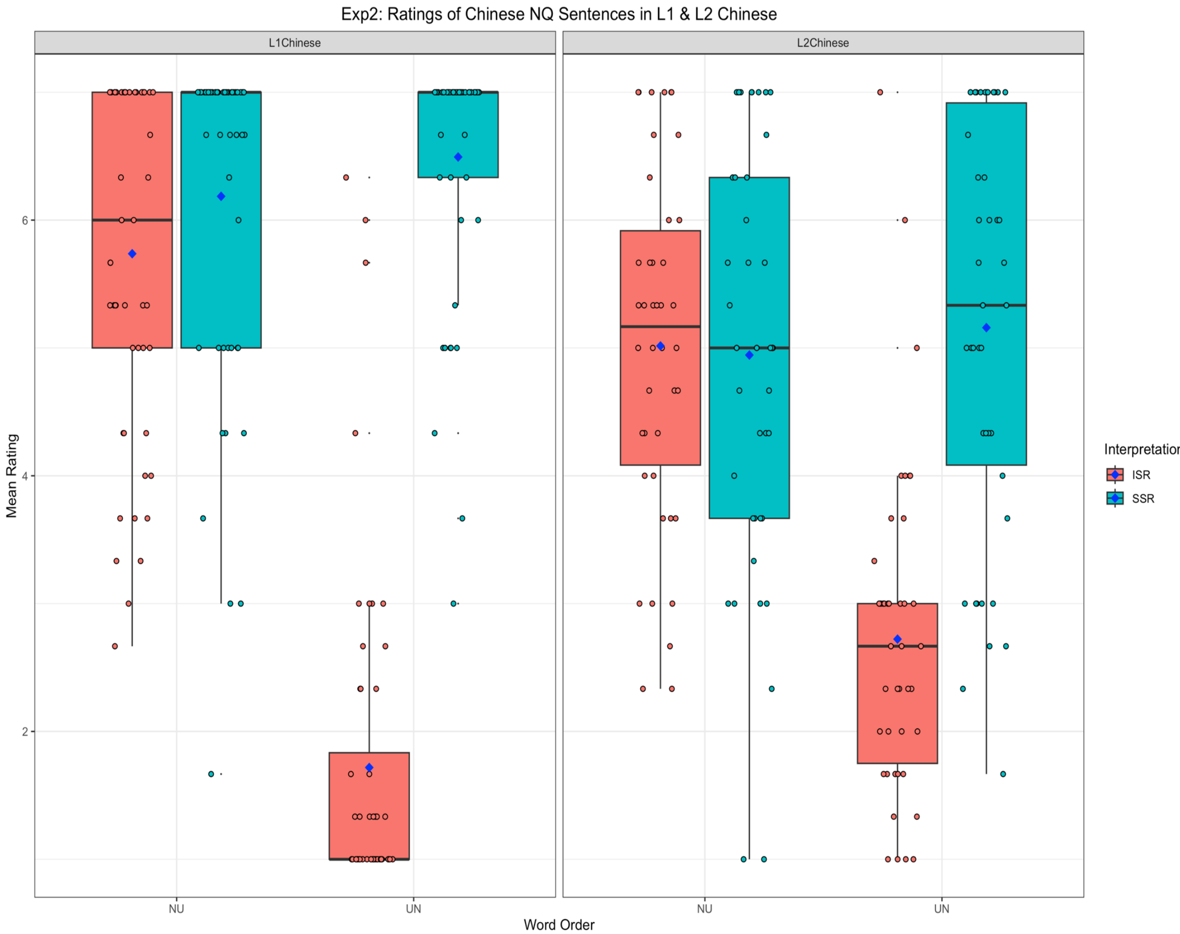

Figure 5.11 Exp2b: Mean ratings of NQ sentences by condition in L1 & L2 Chinese.

Proficiency was examined for its influence on the ratings of DQ and NQ sentences by English-speaking learners of Chinese. It turned out that L2 proficiency significantly interacted with interpretation for both DQ (b = 0.07, p = .008) and NQ (b = 0.04, p = .049) sentences. As illustrated in Figure 5.12[59], English L1 Chinese L2 learners tended to lower their preference for DQ sentences on their ISRs as their proficiency in Chinese increased. Similarly, as Chinese L2 learners became more proficient in the language, they reduced their preference for NQ sentences

[59] Both Figure 5.12 and Figure 5.13 were plotted using the linear regression model fit.

on ISRs. However, unlike DQ sentences, the proficiency effect on the inverse scope reading for NQ sentences may differ in its directionality between UN and NU sentences, as can be seen from the ISRs with UN and NU sentences in Table 5.6. The proficiency effect was examined separately for UN and NU sentences. It was found that proficiency only modulated quantifier scope interpretation for UN sentences (b = 0.06, p = .007), as illustrated in Figure 5.13, but not for NU sentences (b = 0.04, p = .123). In either case, it is not surprising that learners' preference for SSRs increased as their proficiency went up.

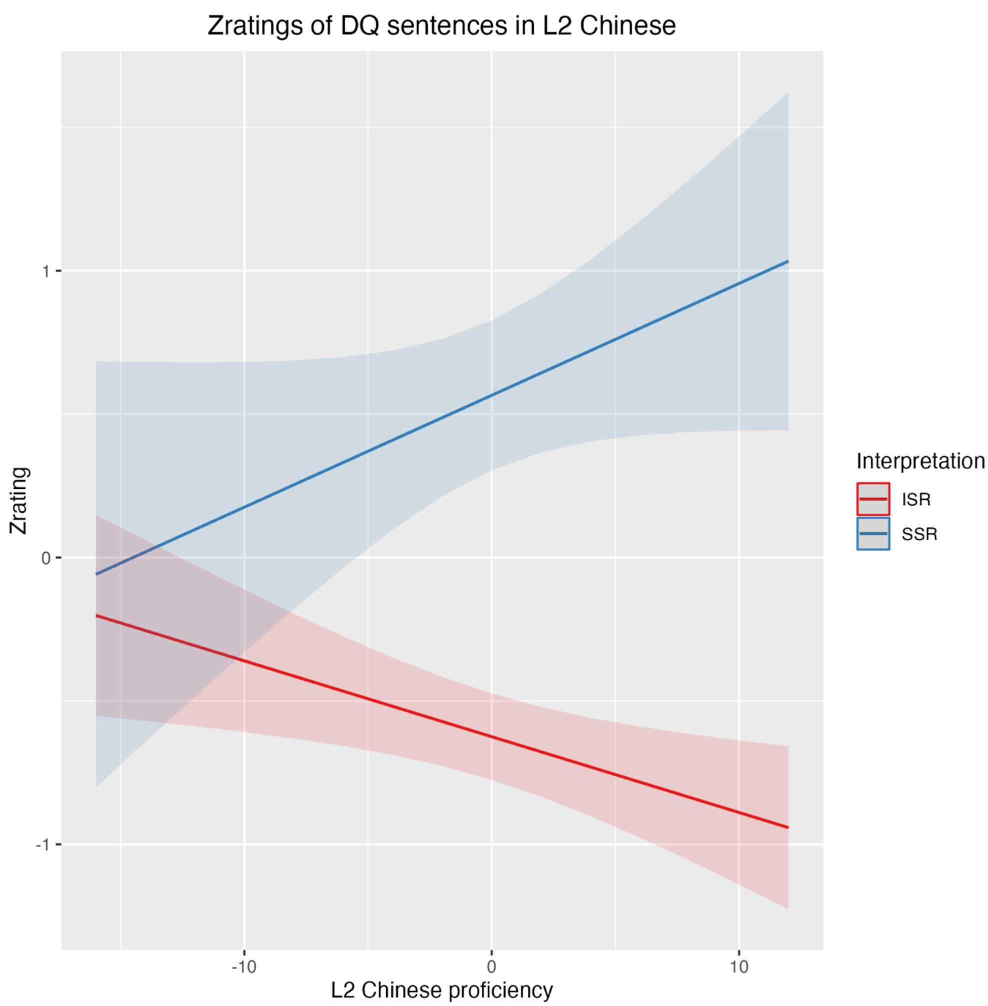

Figure 5.12 The role of proficiency in rating DQ sentences by L2 Chinese speakers.

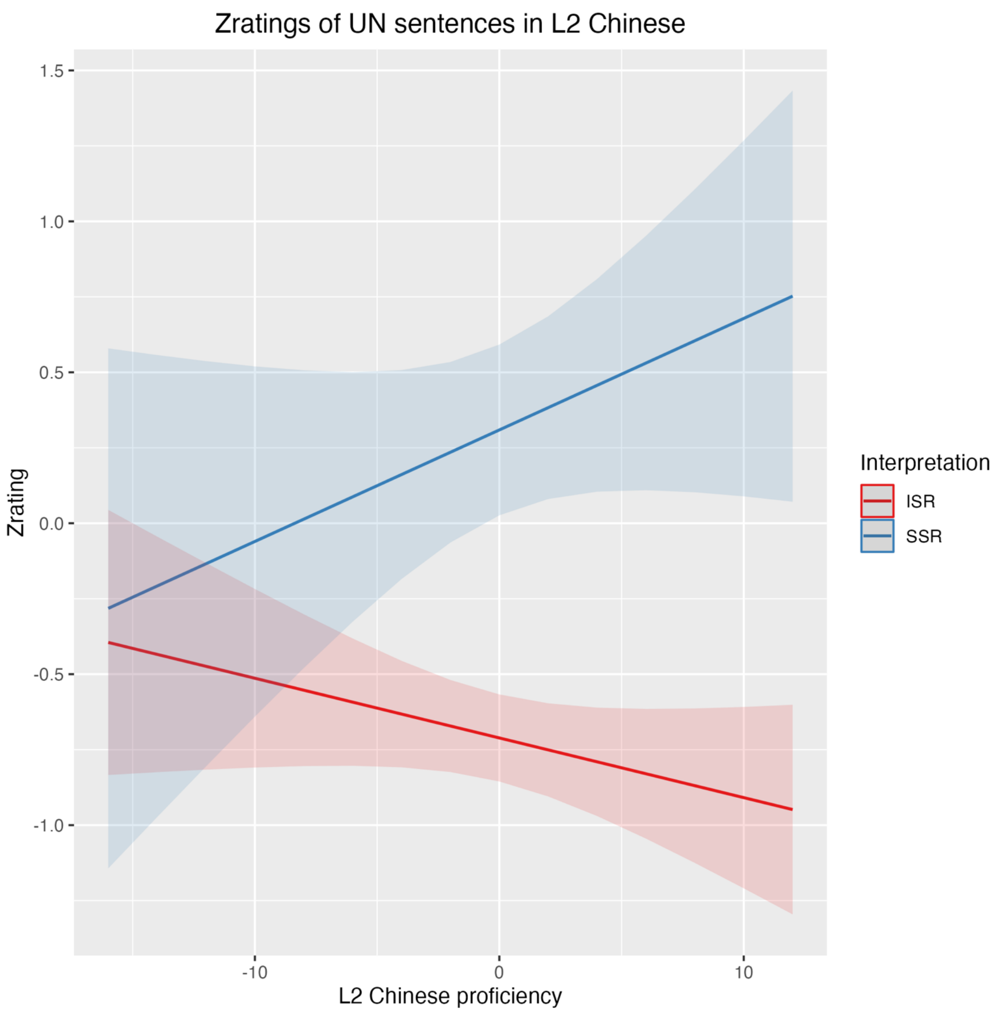

Figure 5.13 The role of proficiency in rating UN sentences by L2 Chinese speakers.

L2 Chinese speakers achieved an accuracy rate of 0.54 (*SD* = 0.7) on the DOU test. This result is in line with a study by Li (2013), which found that English-speaking learners of Chinese experienced challenges in acquiring the correct use of the particle DOU in certain syntactic contexts where it is obligatory. The correlation between ratings of ISR for UE and UN[60] sentences

[60] Chinese EU and NU sentences were not examined here, because no DOU was present in these sentences.

and individual learners' knowledge about DOU was explored. The analysis demonstrated no significant correlation for UE sentences (r = .03, p = .8). For UN sentences, a small negative correlation was found (r = -.27, p = .005). As L2 Chinese learners' accuracies with DOU improved, their acceptance of the inverse scope reading for UN sentences declined.

5.3.3 Individual analysis results

As in Experiment 2a, it is also important to examine individual performance on the TVJT for both the L2 Chinese group and the L1 Chinese group on the TVJT, considering the latter as the baseline. The participants would be classified into five different patterns, as described in (39). As for DQ sentences (as in Figure 5.14), most of the L1 Chinese participants fell into Pattern 1 in which the SSR was strongly favored. Nevertheless, in the case of UE sentences, a few L1 participants still fell into Pattern 2 and Pattern 4, either accepting both readings or preferring the ISR. This finding is largely in line with the results of the group analysis, which supports the claim that Chinese DQ sentences generally limit their acceptable interpretations to the SSR. In the L2 Chinese group as in Figure 5.15, while most learners fell into Pattern 1 (preference for SSR), a small proportion of L2 learners accepted both readings for DQ sentences (EU: 9.5%; UE: 14.3%).

31 % of the L2 learners even fell into Pattern 2 in which the ISR was preferred in the case of UE sentences.

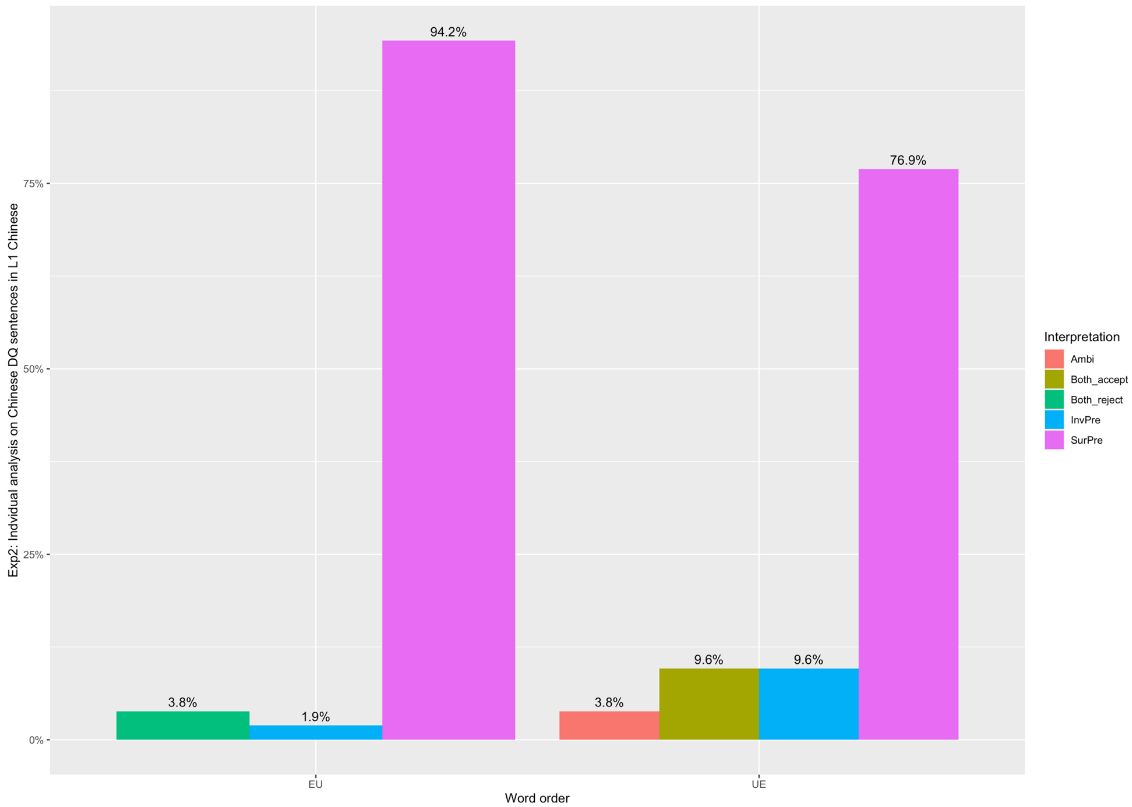

Figure 5.14 Exp2b: Percentage distribution of individual analysis for DQ sentences in L1 Chinese.

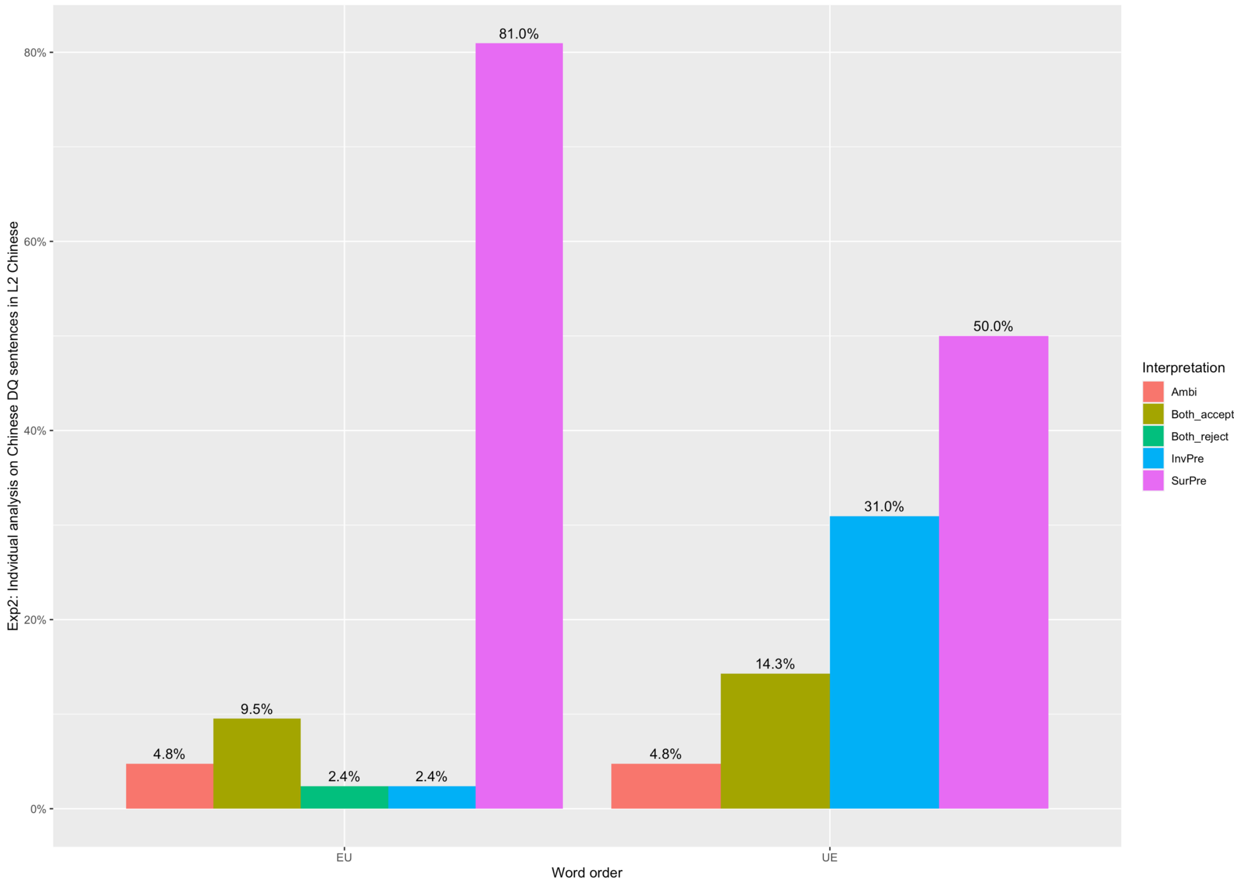

Figure 5.15 Exp2b: Percentage distribution of individual analysis for DQ sentences in L2 Chinese.

As for NQ sentences, the individual results from the L1 and L2 groups showed a similar pattern that mirrored the results of the group analysis. The interpretation pattern for UN sentences in the L1 group was quite consistent (as in Figure 5.16), in that the majority of the L1 participants showed a preference for the SSR. As in Figure 5.17, while more than half of the L2 participants exhibited a preference for SSR in UN sentences (Pattern 1), other patterns were also observed among these participants. Notably, 21.4% of the L2 participants completely rejected ISR for UN sentences. Compared to UN sentences, NU sentences displayed a greater degree of variability in their ratings, as observed in both the L1 and L2 groups. Specifically, in both L1 and L2 groups, a comparable proportion of participants showed a preference for either the SSR or the ISR.

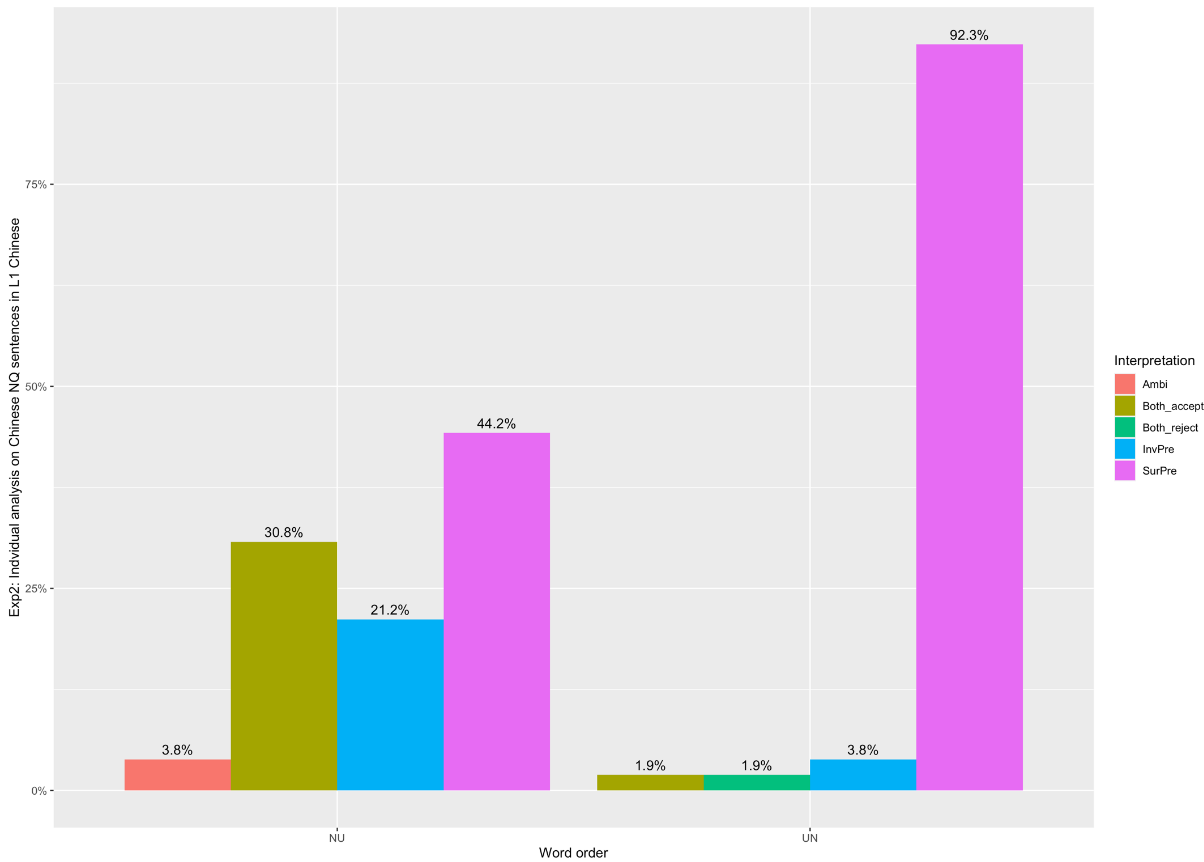

Figure 5.16 Exp2b: Percentage distribution of individual analysis for NQ sentences in L1 Chinese.

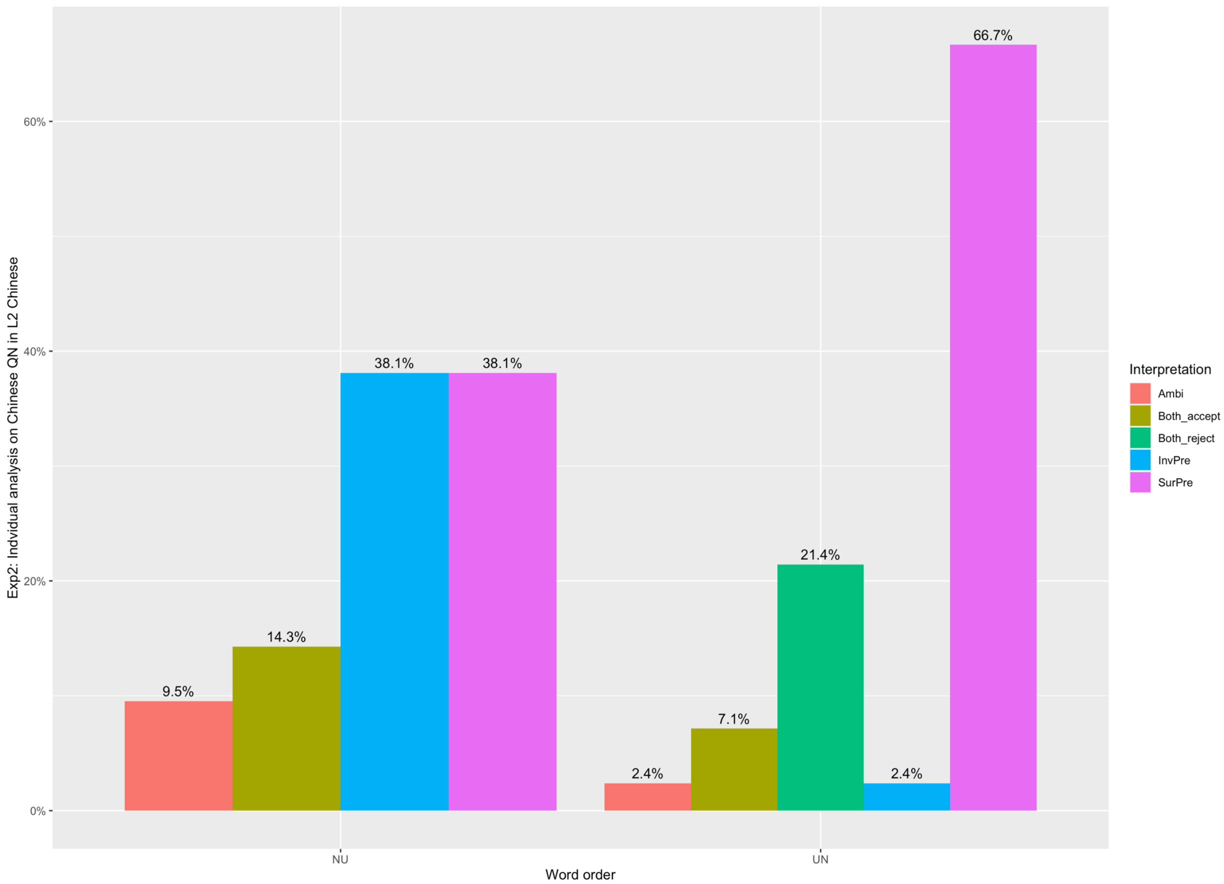

Figure 5.17 Exp2b: Percentage distribution of individual analysis for NQ sentences in L2 Chinese.

5.4 Discussion for Experiment 2

The current experiment investigated the interpretation of quantifier scope as represented by doubly quantified and negatively quantified sentences by Chinese-speaking learners of English and English-speaking learners of Chinese, respectively. Data from native speakers of English and Chinese were also gathered and served as the control groups for the L2 data. This investigation was conducted in the context of offline interpretation using the TVJT. This method allows for forcing interpretations in pragmatic contexts, which is particularly useful in detecting less accessible interpretations and hence better tapping into the learners' knowledge representation of the target language. In addition, the study examined how L2 proficiency and triggering input may impact the L2 group's performance in the TVJT.

For English DQ and NQ sentences, the results showed that native English speakers had access to both surface scope and inverse scope readings but preferred the former. The finding that surface scope readings were preferred across the board in native speakers of English was in line with prior literature (Anderson, 2004; Chung, 2009; Lee, 2010; Wu & Ionin, 2019, 2022). The observed preference for surface scope readings in native English speakers can be attributed to the Processing Scope Economy principle, as proposed by Anderson (2004). This principle suggests that the extra syntactic displacement required for inverse scope interpretations imposes processing costs at LF, leading to a less efficient and less preferred reading. For Chinese-speaking learners of English, it is not surprising that learners can easily acquire surface scope readings across different quantifier scope types. This is because surface scope readings are generated through simpler syntax and semantic operations compared to inverse scope readings. Additionally, the learners' L1 always permits surface scope readings with both DQ and NQ sentences. In addition, they exhibit a preference for surface scope readings over inverse scope readings, similar to L1 speakers. This finding suggests that L2 learners are sensitive to the interplay between syntax and semantics required for scope interpretations.

The evidence regarding Chinese speakers' ability to acquire the English inverse scope readings appears to be mixed and dependent on the syntactic structure involved. For Universal-Existential (UE) quantifier sentences, L1 and L2 English speakers did not differ in their ratings of ISRs (L1 English: 5.41 vs. L2 English: 4.48), suggesting that L2 learners of English access inverse scope interpretations in this case to the same extent as native speakers of English. In contrast, L2 learners' ratings for Existential-Universal (EU) sentences on their ISRs were significantly lower than L1 English speakers' (L1 English: 4.46 vs. L2 English: 1.59). This suggests that L2 learners relative to L1 speakers failed to access the ISR for EU sentences.

Regarding NQ sentences, including both Negative-Universal (NU) and Universal-Negative (UN) sentences, L2 learners of English were found to access inverse scope readings in a way similar to native speakers of English. However, given the current evidence, it remains inconclusive whether L1 and L2 English speakers differed in their (relatively high) acceptance of the ISR with UN sentences.

Overall, the big picture that emerges suggests that Chinese speakers of English can generally acquire inverse scope readings in most cases, including DQ and NU sentences, except for EU sentences, which apparently remain a challenge for them. Our experiment is one of the first to show that Chinese-speaking learners are capable of acquiring inverse scope readings for UE and NU sentences to a degree that is similar to that of native speakers of English. Besides L2 input, which informed learners of the possibility of ISRs for UE and NU sentences in English (L1 English - UE on ISR: 5.41 and NU on ISR: 5.95), input from L1 via positive transfer may have also contributed to the learners' awareness of the availability of ISRs, especially for NU sentences (L1 Chinese - UE on ISR: 3.97, NU on ISR: 5.02). The results for EU sentences are largely consistent with previous studies by Chu et al. (2014) and Wu and Ionin (2022), which tested only 'a-every' sentences in Chinese-speaking learners of English and had some methodological issues, such as the use of a mixture of agentive and psych verbs.

There are several potential explanations for why acquiring ISRs for EU sentences was challenging for Chinese-speaking leaners of English. One possibility is that negative transfer from the L1 prevented L2 learners from acquiring the target grammar successfully. However, this explanation seems unlikely, as it cannot account for the successful acquisition of ISRs in other cases where L1 negative transfer did not play a role. The second possibility is that learners' performance was affected by the Processing Scope Economy, but this explanation again is unsatisfactory. As discussed above, native English speakers dispreferred but accepted the ISR across different types of quantifier scope. Although L2 learners are influenced by the Processing Scope Economy and tend to give higher ratings for the SSR than for the ISR, they rated the ISR with EU sentences significantly lower than L1 English speakers did. This suggests that the failure of L2 learners to acquire the ISR with EU sentences cannot be solely attributed to the processing-based theory. Under similar conditions, they would otherwise accept the ISR to a degree comparable to native English speakers. Another possible explanation is that the positive evidence, albeit available in the L2 input, might not be robust enough for L2 learners to add to their grammar this new interpretation (i.e., ISR) in the case of EU sentences. At first glance, this account may seem inadequate to explain this pattern of results, as the ISR was acquirable in other cases (i.e., UE, UN, NU) with the availability of positive evidence. However, upon closer inspection, this explanation cannot be entirely dismissed. Logically, it is always possible for learners to acquire target properties given that such properties are available in the L2 input, because this learning scenario does not lead to a severe learnability problem.

However, the availability of positive evidence may not suffice for learners to eventually acquire the target property. The frequency with which learners encounter contexts that support a particular interpretation is another important factor in the acquisition of that interpretation (Hopp et al., 2020). For example, Montrul (2001) found that L1-Spanish L2-English learners had difficulty acquiring the transitive structure of certain English manner-of-motion verbs (e.g., 'The captain marched the soldiers to the tent') because this structure is less typical and frequent than the intransitive one (e.g., 'The soldiers marched'). Although both the SSR and ISR for EU sentences are available in the L2 input, L2 learners are less likely to encounter contexts that favor the ISR than those that favor the SSR, as evidenced in the fact that the ISR was much less preferred than the SSR with EU sentences for native English speakers. Native English speakers were more likely to accept the ISR in UE sentences compared to EU sentences. Similarly, L2 learners found it easier to acquire the ISR in UE sentences than in EU sentences, likely because the magnitude of the interpretation effect (i.e., difference between SSR and ISR) was smaller for UE sentences (b = -0.64) than for EU sentences (b = -0.92). In other words, L2 learners are more likely to encounter contexts that support the ISR in UE sentences compared to EU sentences in the L2 input. This is because the ISR was much less preferred for EU sentences than for UE sentences. Consequently, L2 learners would only interpret EU sentences as the ISR in rare situations where specific contexts lead to such an interpretation.

Another plausible explanation is that the Single Reference principle of Kurtzman and MacDonald (1993) was actively at play in L2 learners' rejection of the ISR for EU sentences. As readers parse sentences incrementally, encountering an indefinite NP at the beginning of an EU sentence leads them to associate the NP with a single referent, which can be incompatible with the ISR for EU sentences. This principle is also active in L1 speakers' interpretation of DQ sentences as well, because L1 speakers gave lower ratings for the ISR with EU sentences (mean rating: 4.46) than for the ISR with UE sentences (mean rating: 5.41). However, the significantly lower ISR scores given by L2 learners on EU sentences compared to L1 speakers could suggest that L2 learners were more affected by this principle. In fact, the negative correlation found between definiteness test scores and ISR ratings for EU sentences (r = -.26, p < .001) provides additional evidence for the role of this principle in relation to the definiteness property of NPs in the L2 group. As the L2 learners became more sensitive to NPs' definiteness property in the context of EU sentences, this principle became increasingly active, resulting in a greater reluctance to accept EU's ISR.

The present results for UN sentences suggest that Chinese-speaking learners of English can access ISR to a degree similar to that of L1 speakers in such sentences. This finding diverges from previous studies, particularly the one conducted by Wu and Ionin (2019) with Chinese-speaking learners of English, which found that even advanced learners failed to perform in a native-like manner with 'every-not' sentences. Wu and Ionin (2019) primarily attributed the learners' failure to acquire the inverse scope interpretation for UN sentences to the infrequent occurrence of relative positive evidence. Contrary to this explanation, I argue that the inverse scope interpretation stands out, and may even be more prominent than the surface scope interpretation, for English UN sentences, as found in Musolino et al. (2000) and Musolino and Lidz (2006). In the current experiment, the ISR ratings for UN sentences were slightly lower than the SSR among L1 English speakers, yet the difference between them was only marginally significant. Moreover, as shown in the individual analysis results, about half of the participants fell into the patterns of 'accept both readings' and 'prefer the surface scope reading'. Therefore, it is likely that the L2 input provides some opportunities for learners to encounter contexts that support the ISR for UN sentences (e.g., 'every lawyer's not a crook'; 'every student isn't good at syntax'), and Chinese-speaking learners of English were thus able to acquire this interpretation. This finding is consistent with previous research conducted with learners of L1 Korean (Chung, 2012; Lee, 2009). It is also worth noting that Wu and Ionin (2019) provided participants with both written stories and pictures, but this experiment only had written stories. Consequently, it is difficult to directly compare the results of the two studies.

For NU sentences, unlike previous studies with L1 English speakers (Chung, 2012; Musolino & Lidz, 2006), which showed that native speakers of English had low acceptance of the ISR for NU sentences, this experiment is the first to find that L1 English speakers accessed their ISRs as well as their SSRs equally well. This finding suggests that English speakers are able to access the inverse scope readings for NU sentences when provided with supportive discourse contexts, contradicting the prediction that the Principle of Cooperation would lead to rejection of the ISR as the informationally stronger statement, particularly in the absence of supportive contexts (as can be seen from Experiment 1 in which the mean rating of ISR for NU sentences was 2.17 out of 7). L2 learners were found to be target-like with respect to the acquisition of the ISR for NU sentences. The availability of contexts that support ISRs for NU sentences in the L2 input may explain why learners would be able to acquire this interpretation. Alternatively, it is possible that learners are informed of the existence of the inverse scope interpretation for NU sentences through their L1, i.e., Chinese, as Table 5.6 shows that L1 Chinese speakers well accepted the ISR for NU sentences (mean rating: 5.02).

Compared to Chinese-speaking learners of English, English-speaking learners of Chinese are expected to face greater challenges in acquiring native-like quantifier scope interpretation, particularly the inverse scope interpretations. This is because English speakers need to unlearn their L1 interpretation of inverse scope, which does not exist in the L2 language, i.e., Chinese. However, there is no positive evidence in the Chinese input to show that inverse scope is not permitted. Nevertheless, it is important to point out that evidence for the absence of inverse scope interpretations across different types of quantifier scope in Chinese has been scarce and mixed, as reviewed in Chapter 2. Based on this experiment, it appears that surface scope interpretations tend to be more prominent than inverse scope interpretations in Chinese. Additionally, inverse scope interpretations seem to be rarely available, except in NU sentences. However, caution is warranted when drawing definitive conclusions due to the scores' variability across participants and their proximity to the scale's midpoint (UE on ISR: 3.97 for L1 Chinese). Therefore, this experiment provides novel evidence for the presence or absence of inverse scope interpretations in (L1) Chinese.

The results for English-speaking learners of Chinese revealed that they successfully acquired the surface scope interpretations for Chinese across the board. This finding is not unexpected, as these interpretations are generally simple and exist as the prevalent ones in their L1, namely English. Of particular interest are the L2 results with inverse scope interpretations, which demonstrated that the learners performed at a level comparable to that of native speakers in their ability to acquire such interpretations in Chinese, except for UE sentences. Regarding DQ sentences (including UE and EU sentence) on ISR, although native Chinese speakers gave significantly lower ratings than English-speaking learners of Chinese, EU sentences themselves were quite low for their ISR ratings by L2 learners with a mean rating of 2.93, which suggests that this interpretation was almost always rejected by the L2 learners. The extent to which the L2 grammar converges with the L1 grammar may be modulated by L2 Chinese proficiency, as evidenced by a proficiency effect for DQ sentences (see Figure 5.12): The rejection of ISR became more pronounced as learners' proficiency increased. To put it differently, if learners were reaching the end state of their L2 proficiency, they should be able to acquire the absence of inverse scope interpretations for Chinese in a way comparable to L1 speakers. Regarding UE sentences, it appears that L1 Chinese speakers achieved an average rating of approximately 3.97 on ISR, which may indicate that this interpretation is considerably challenging for them. In contrast, the findings suggest that L2 learners in this study encountered difficulty in acquiring the absence of ISR for UE sentences, as indicated by their relatively higher mean rating of 4.76.

The results obtained from NU sentences in Chinese on their ISRs were intriguing and also somewhat unexpected. The fact that both L1 speakers (mean rating: 5.74) and L2 learners (mean rating: 5.02) gave this interpretation high ratings is noteworthy, especially considering that there was no significant difference in their ratings. Such results were also confirmed in the individual analysis. To fully contextualize the L2 results, it is worth first taking a closer look at the availability of ISR in L1 Chinese for NU sentences. To the best of my knowledge, there have only been two studies that directly investigated the preferred interpretation of such sentences in Chinese: Zhou and Crain (2009) and Fan (2017). Typologically, Mandarin Chinese, in contrast to English, has been claimed to exhibit scope rigidity (Aoun & Li, 1989; Huang, 1982; Lee, 1986). In general, the claim of scope rigidity holds true for most cases, including UE, EU, and UN sentences, as the current experiment with native Chinese speakers demonstrates that surface scope readings are the only possible interpretations. However, the data from this dissertation reveals an exception: ISR is in fact available for NU sentences in L1 Chinese, challenging this claim.

Zhou and Crain (2009) tested NU sentences in which negation precedes the universal quantifier, as in (41).

(41) Bushi mei-pi ma dou tiaoguo-le liba.

Not-be every-CL horse DOU jump-over-PERF fence

‘Not every horse jumped over the fence.’

They argued that such sentences only permit the surface scope interpretation in Chinese, and the lack of inverse scope interpretations for such sentences was not due to a pragmatic implicature. To support their argument, they conducted a survey in which 15 Chinese speakers were asked to indicate whether sentences like (42) were acceptable or not. The results showed that most of the participants (66.7%) rejected this sentence. Their reasoning is that if the lack of inverse scope reading is due to a pragmatic implicature, the statement added following (41) to become (42) should cancel the implicature without contradiction.

(42) Bushi mei-pi ma dou tiaoguo-le liba; shishishang, meiyou ma tiguo liba.

Not-be every-CL horse DOU jump-over-PERF fence; in fact, not-have horse jump-over fence

'Not every horse jumped over the fence; in fact, none of them did.'

They thus concluded that the lack of inverse scope interpretations for sentences such as (41) resulted from the focus-sensitive property of *shi*. In other words, Chinese speakers represent such sentences as cleft structures, corresponding to the English counterpart 'It wasn't every horse that jumped over the fence'. However, the NU sentences tested in the current experiment (as shown in (28) and repeated in (43)) do not contain the focus operator *shi*. As a result, the absence of certain focus-sensitive elements in the sentence allows for an inverse scope reading to arise.

(43) Zhe-pi ma meiyou tiao guo mei-shan liba.

This-CL horse not jump over every-CL fence.

'The horse didn't jump over every fence.

Fan (2017) tested Chinese NU sentences in the configurations such as 'NP bu/mei V suoyoude N (NP not V all N)', and found that 51.1% of the adult Chinese participants accepted such sentences on inverse scope readings. My test stimuli for Chinese NU sentences[61], which feature every-NP in the object position, yielded similar results in that these sentences for their inverse scope readings are available in Chinese (mean rating: 5.74). The question arises as to why the inverse scope interpretation stands out as a prominent interpretation for Chinese NU sentences. One possible explanation is that Chinese speakers (if not all) may permit a reading of (43) that includes the truth condition on which none of the horses jumped over the fence[62], if supportive discourse contexts are provided, as was the case in the current study in which Chinese speakers allow the inverse scope reading for sentences like (43) in the presence of discourse contexts. The lack of inverse scope reading in sentences like (43) may simply be caused by a pragmatic implicature, which was confirmed in a survey I conducted with 18 native speakers of Chinese. In this survey, I asked the Chinese speakers to rate the sentence consisting of (43) and the additional statement *shijishang, meiyou ma tiaoguo liba* (in fact, none of them did) on a 7-point Likert Scale. The mean rating of this sentence across participants was 4.72 (*SD* = 2.14), indicating acceptance of this sentence. The finding suggests that the implicature for the surface scope ('some') reading

---

[61] Fan (2017) noted that the configuration of NU sentences with every-NP in the object position (the kind of sentences used in this dissertation) would be ungrammatical in Chinese. However, this claim was not supported by the results of the norming study reported in section 4.3.1.2. In fact, the mean rating for these sentences was higher than 4.5, indicating that they were generally acceptable to Chinese speakers. Therefore, it appears that these sentences are in fact grammatical in Chinese.

[62] This differs from Zhou and Crain (2009) in which they argued that Chinese adults did not accept the truth condition for the inverse scope interpretation on sentences such as (40).

to prevail can be canceled by Chinese speakers without contradiction. Hence, Chinese speakers can accept the truth condition for the inverse scope (‘none’) reading in relation to (43).

Turning to the L2 results for NU sentences, English-speaking learners of Chinese have been shown to be able to acquire the presence of inverse scope readings. This is likely because L2 learners had the opportunity to encounter L2 input that supported such interpretations, as evidenced by the high acceptance of the ISR for NU sentences among native Chinese speakers. As for UN sentences, both the L1 speakers and L2 learners gave quite low ratings for the ISR, and more importantly there was no significant difference between their ratings. Therefore, these findings suggest that L2 learners of Chinese are able to acquire the absence of inverse scope interpretations for Chinese UN sentences, despite the absence of positive evidence in the input for such interpretations with UN sentences. One way to account for their successful acquisition in UN sentences is to assume that UG guides the learners to detect the abstract interpretive properties of the L2. It is also likely that the learners navigated through indirect/triggering L2 input to deduce which interpretations were or were not possible in the L2. This speculation receives support from the negative correlation between learners’ performance on the DOU test and their ISR ratings for UN sentences (r = -.27, p = .005). The negative correlation suggests that learners who performed better on the DOU test were less likely to rate UN sentences with inverse scope readings highly. In other words, learners who are more sensitive to knowledge of DOU are less likely to accept inverse scope interpretations, and more likely to accept surface scope interpretations. This finding confirms the role of DOU in enhancing sensitivity to SSR for UN sentences (Zhou & Crain, 2009). L2 proficiency in Chinese was also found to play a role in modulating the extent to which L2 learners could acquire the absence of ISR for Chinese UN sentences (see Figure 5.13).

In summary, the results of Experiment 2, which utilized the TVJT, consistently demonstrated that both English and Chinese L2 learners were able to acquire the surface scope interpretations of the L2. In most cases, Chinese-speaking learners of English were able to acquire inverse scope readings, primarily due to the positive evidence from L2 input. On the other hand, English-speaking learners of Chinese faced learnability problems, but in some cases, they were able to acquire ISRs, potentially guided by triggering input. Furthermore, the role of L2 proficiency in modulating the extent to which the less preferred interpretation (i.e., ISR) could be acquired was mainly observed in L2 learners of Chinese. Overall, this study makes a valuable contribution to the existing literature by providing novel evidence on L2 learners' (with Chinese and English as L1) scope interpretation of quantified sentences, particularly UE and NU sentences that have received little attention in previous research, and also exploring the potential influences of triggering input and L2 proficiency on L2 learners' scope interpretation. Additionally, this study sheds new light on the scope interpretation of quantified sentences, particularly DQ and NU sentences, among native speakers of Chinese, for which the existing evidence has been scarce and conflicting.

6.0 Experiment 3: Online processing of quantifier scope in L2 English and Chinese

6.1 Introduction

Experiment 3 aims to examine if the patterns of results observed in offline settings during Experiment 2 extend to the L2 processing of quantifier scope using eye-tracking in the visual-world paradigm (VWP; Tanenhaus et al., 1995). Different from the offline methods such as TVJT, the VWP allows us to investigate how quantifier scope is computed online and accessed by listeners during spoken language comprehension. The use of this paradigm rests on the *linking hypothesis*, which posits a connection between auditory-linguistic processing and visual processing (Tanenhaus & Trueswell, 2006). As reviewed in Chapter 3, limited research has been done on the acquisition and processing of doubly quantified and negatively quantified sentences by Chinese-speaking learners of English and English-speaking learners of Chinese. Furthermore, the exiting studies predominantly relied on offline tasks without time constraints (e.g., Chu et al., 2014; Wu & Ionin, 2019, 2022). Moreover, the few studies conducted on L2 processing of quantifier scope involved only L1 speakers of Korean as participants (Chung & Shin, 2022; Lee, 2009, 2010). These studies utilized self-paced reading as the primary method, which is considered to have lower ecological validity compared to eye-tracking. In addition, given that different measures may differ in their sensitivity to linguistic knowledge (e.g., Ellis, 2005; Tokowicz & MacWhinney, 2005), claims about the linguistic knowledge represented in L2 learners for quantifier scope interpretation should benefit from the triangulation of online eye-tracking data (e.g., eye fixations) with offline measures, permitting the understanding of the process of interpretation in its full form.

L2 processing differs from L1 processing due to various reasons and to varying degrees (cf. Clahsen & Felser, 2006; Cunnings, 2017; Hopp, 2006, 2010; McDonald, 2006). Most research has focused on morpho-syntax processing in L2, leaving a significant gap in understanding the L2 processing of quantifier scope. This linguistic phenomenon can lead to interpretive ambiguity attributed to covert movement at LF in formal approaches, requiring the human processor to automatically compute its resulting interpretation, i.e., inverse scope. Therefore, it is crucial to investigate how L2 learners differ from L1 speakers in this aspect, particularly considering the challenges of integrating multiple sources of information such as syntax, semantics, pragmatics, and visual contexts in the online processing of quantified sentences using the visual-world paradigm. Moreover, while some approaches suggest that L2 processing is not fundamentally different from L1 processing (Hopp, 2006, 2010; Juffs & Harrington, 1996; McDonald, 2006), it is often influenced by individual proficiency differences (Hopp, 2010; Hoshino et al., 2010) and cross-linguistic effects (Juffs & Rodríguez, 2014; Park & Kim, 2022; Tolentino & Tokowicz, 2011). Through the lens of quantifier scope, the present investigation seeks to provide novel evidence by investigating learners' processing of the target language (English vs. Chinese), which differs from their L1, while considering the influence of L2 proficiency.

This study also addresses a methodological issue. In many of the previous studies, the distinction between access and preference was not explicitly made. For example, in the study conducted by Musolino and Lidz (2006), children strongly rejected sentences like 'Every horse didn't jump over the fence' as an accurate description for the scenario where only two out of three horses jumped over the fence. This rejection has led researchers to interpret it as a failure to access the sentence's inverse scope reading. However, it is possible that individuals can access both readings but simply prefer one over the other due to factors such as ease of processing and frequency of encounter. The current experiment adopts the paradigm introduced by Lohiniva and Panizza (2016). For each trial, it introduces two contextual scenarios, contrasting with classical TVJT setups that typically feature only one scenario. Additionally, each test sentence undergoes three distinct visual presentations, namely 1) surface scope reading vs. false reading, 2) inverse scope reading vs. false reading, 3) surface scope reading vs. inverse scope reading. The first two sets of presentations are designed to detect the access of SSR and ISR, respectively. The third set aims to identify individuals' preference for one of the two possible interpretations.

Furthermore, Gualmini et al. (2008) argued that readers[63] interpret quantified sentences as answering a particular question known as the Question Under Discussion (QUD), which can be inferred from the discourse context. In contrast to previous studies (e.g., Musolino & Lidz, 2006; Scontras et al., 2017), the current experiment makes salient two possible interpretations, which are considered as relevant questions to QUD, by presenting them simultaneously in a single visual scene.

In sum, Experiment 3 addresses the following research question, accompanied by two sub-questions (RQ3 restated below).

RQ3) *Is the processing of quantifier scope in the target language by Chinese-speaking learners of English and English-speaking learners of Chinese similar to that of native speakers, when supportive context is given in the online TVJT with the visual-world paradigm?* (Experiment 3)

---

[63] The operation of the QUD when participants are listeners but not readers is currently unknown. Therefore, Experiment 3 of the current study would shed light on how the QUD operates in listeners.

(RQ3-1) Which learner group is more successful in processing the target interpretation: Chinese-speaking learners of English or English-speaking learners of Chinese?

(RQ3-2) How does L2 proficiency affect L2 learners' performance on the online processing of quantifier scope interpretation?

6.2 Experiment 3a: L2 English visual-world eye-tracking experiment

6.2.1 Methods

6.2.1.1 Participants

The participants for the eye-tracking experiment also participated in Experiment 2a. For participant details, refer to sections 5.2.1.1 of Experiment 2a.

6.2.1.2 Design and materials

Each experimental trial consists of a recorded sentence and a visual scene. The target sentences for this study consist of the sentences used in Experiments 1 and 2. There are a total of 48 target sentences, with 12 sentences for each type of quantifier scope. A female native speaker of American English, who was naïve to the experimental purposes, recorded the English stimuli. The recordings were made using Praat (Boersma & Weenink, 2018) in a sound-attenuated booth at a sampling rate of 44100 Hz. To minimize the confounding effect of prosody[64] on quantifier scope interpretation, the speakers were trained to read the sentences with a neutral and unbiased intonation. They were instructed not to emphasize any specific NP in the sentences during production. No significant differences were found in the total duration of the target English sentences across conditions, as indicated by the results of one-way ANOVA tests (all *ps* > .05) (UE: 1.86s, EU: 1.86s, UN: 2.30s, NU: 2.30s). In other words, the auditory stimuli across conditions were matched in terms of total duration.

Three black and white line drawings[65] were created for each sentence, representing potential interpretations. For instance, for the sentence 'Every horse didn't jump over the fence', three distinct pictures were generated, each illustrating a specific interpretation, as exemplified in Figure 6.1. The pictures were drawn using an iPad and exported at the highest resolution for stimuli presentation. To ensure the reliability of the drawings, two Chinese native speakers who were learners of English as L2 and not involved in the main experiment, were asked to match the pictures with their corresponding interpretations. There were no instances of disagreement or confusion. Take again 'Every horse didn't jump over the fence' as an example. Participants were tested with this sentence for three different visual scenes: A-C, B-C, A-B. In the A-C condition, correctly choosing A indicated the access of the surface scope reading for this sentence. In the B-C condition,

[64] The role of prosody in quantifier scope interpretation has yielded mixed evidence in the literature. Studies have shown that prosody influences quantifier scope interpretation in adults (Syrett et al., 2014), but not in children (McMahon et al., 2004). Wu and Ionin (2020) found that native English speakers were affected by prosody only in negatively quantified sentences, but not in doubly quantified sentences. However, the influence of prosody on quantifier scope interpretation in L2 learners remains unknown and is an area of future inquiry.

[65] All sentences used agentive verbs, allowing for the creation of pictorial representations of events.

correctly choosing B indicated the access of the inverse scope reading for this sentence. In the A-B condition, one's choice would indicate their preference for one of the two possible readings. Each visual scene had two areas of interest (AOIs), each representing a possible interpretation of the test sentence in question.

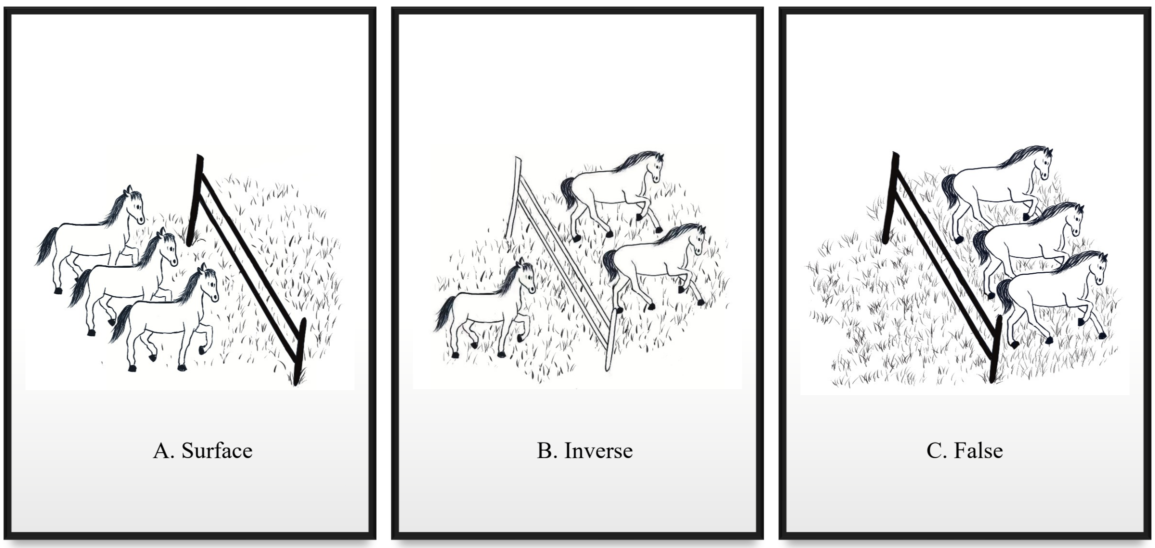

Figure 6.1 Sample pictures illustrating different interpretations of the sentence 'Every horse didn't jump over the fence' are shown as follows: A) Surface scope reading, B) Inverse scope reading, and C) False reading.

Three pictures were created for each sentence to generate three different presentations. In total, 144 pictures were created for the 48 target sentences, resulting in 144 visual scenes consisting of two pictures each. The 48 prerecorded target sentences were randomly assigned to six experimental lists, such that each list comprised 24 items, with 6 items per quantifier type across the three distinct visual scenes. The position of the pictures for scope interpretation was counterbalanced. 31 filler items were interspersed with 24 target items, such that individual participants saw 55 items for the eye-tracking experiment. The filler sentences used in the experiment shared similar complexities and lengths with the target items. Additionally, some of the filler sentences in this experiment included quantifiers or negations, which were largely unambiguous (e.g., 'Two apples on the plate are green'). This served the purpose of further diverting participants' attention away from the main objective of the experiment.

6.2.1.3 Procedure

The participants from Experiment 2 also participated in the current experiment. The experimental procedures for the task sequence were outlined in section 5.2.1.3. Importantly, the eye-tracking experiment was conducted before the TVJT to mitigate any potential priming effects between the experiments. This section focuses on the specifics of the eye-tracking experiment. This visual-world experiment was conducted using a novel webcam-based eye-tracking paradigm, which was programmed and implemented via Gorilla (Anwyl-Irvine et al., 2020). Eye-tracking measures were assessed using the Webgazer.js library (Papoutsaki et al., 2016), a webcam-based eye-tracking JavaScript library available as open source. This library employs dynamic calibration based on mouse clicks to map eye characteristics onto screen positions. It takes advantage of the rule of thumb that users tend to gaze directly at the areas they click while navigating a web page (Huang et al., 2012). WebGazer.js does not provide a consistent sampling rate due to slight variable delays in generating predictions, which can be influenced by the participant's computer and browser power. Research has demonstrated that webcam eye-tracking can yield experimental results that are comparable to those obtained in in-lab settings (Ovans et al., 2021; Vos et al., 2022). The current study is among the first studies in the field of second language acquisition to utilize webcam-based eye-tracking experiments (cf. Phillips, 2022; Wang, 2022).

Access to the experiment was restricted to desktop and laptop users only, while mobile phones and tablets were not allowed. To maximize favorability for experimental conditions, participants were prompted to utilize the settings believed to be optimal for the purposes of the experiment: 1) use earphones or headphones equipped with a microphone, 2) employ a mouse for clicking actions, and 3) ensure a quiet and well-lit environment. The L1 English participants were recruited via Prolific and completed the experiment on their own computers. The L2 English participants completed the experiment in an office under supervision. All participants were given a link to access the experiment hosted by Gorilla.

First, upon clicking the link, the participants were given the informed consent for them to read and sign in. Second, they were promoted to use the optimal settings during the eye-tracking experiment. Third, during the instruction phase, the participants were told that they would be presented with a pair of pictures displayed side by side on the screen while simultaneously hearing a spoken sentence. Their task was to choose the picture that best matched the heard sentence in terms of interpretation. It was emphasized that participants should make their decision as fast and accurately as they could, considering that they well understand each sentence. The main session started with five practice trials and a 5-point calibration, after which participants completed 55 experimental trials. A series of purposeful pictorial instructions [66] guided participants on maintaining the appropriate head position during calibration and experimental tasks.

If the estimate for one of the calibration points was too close to another, it triggered an automatic repetition of the calibration procedure. Additionally, to address any potential head drift or body repositioning by participants, recalibration was conducted midway through each

[66] I would like to express my gratitude to Simone Lira Calabrich for sharing her expertly crafted pictorial instructions.

experimental block, specifically after every 18 trials. In the analysis phase, eye-tracking estimates with face confidence values lower than 0.5 (a score ranging from 0 to 1 that represents the webcam-based eye-tracking machine learning model's confidence in detecting a human face) were removed from the analyses. Each trial consisted of the following events in sequence (See Figure 6.2). First, a fixation cross appeared in the center of the screen for 2s. Following the fixation cross was a 2s preview[67] of the visual scene, which preceded the presentation of each sentence. I selected a 2s preview duration based on a synthesis that revealed the prevailing duration across a majority of L2 studies with visual-world eye-tracking experiments, where the mean preview time was 2.4s and the median was 2s (Godfroid, 2020). The pictures stayed on the screen until one of them was selected or until 5s elapsed after the offset of the sentence, at which point the visual scene disappeared. The eye-tracking session lasted for about 20 min.

[67] Including a preview is based on the rationale that when individuals see an image, they create a mental representation, which interacts with the subsequent auditory information (Altmann & Kamide, 2007). Furthermore, certain types of previews have the potential to mitigate unwanted variability arising from tasks related to object recognition and visual search, thereby enhancing the sensitivity of assessing linguistic processing (Apfelbaum et al., 2021).

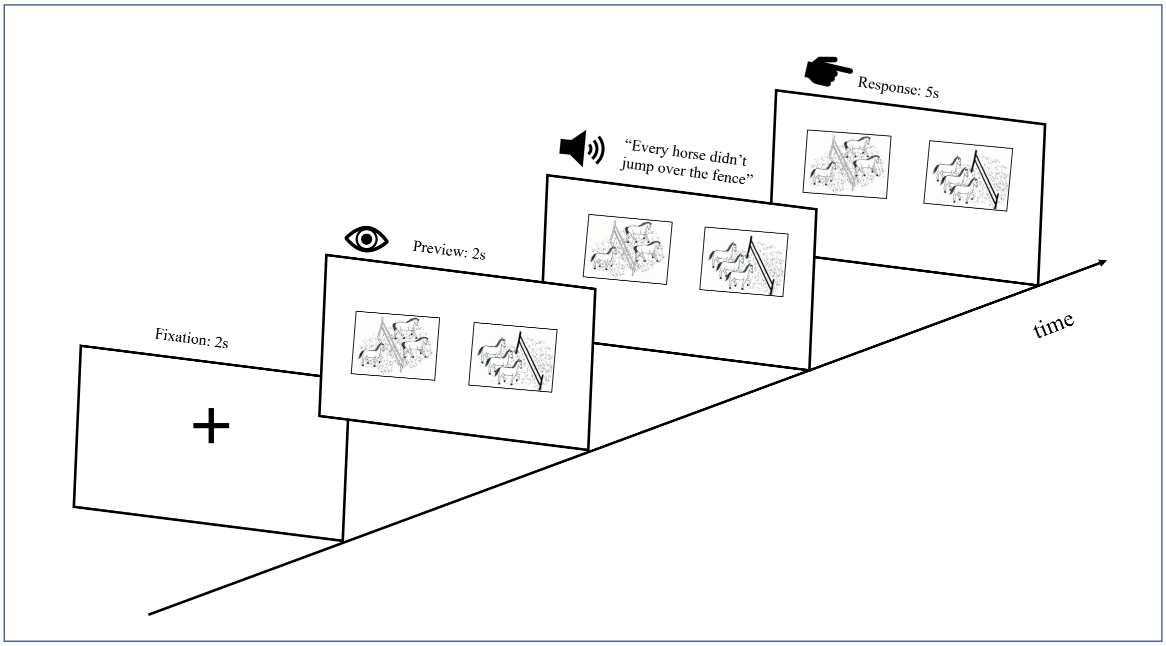

Figure 6.2 Schematic illustration of a single trial in the visual-world eye-tracking

6.2.1.4 Analysis

Two types of data were collected: sentence-picture matching data, which involved participants selecting pictures that represented their interpretation of the heard sentences, and eye-tracking data, which measured participants' eye fixations to the AOI. I first report the sentence-matching data and then the eye-tracking data, with the former focusing on the proportions of picture choices, and the latter focusing on the proportion of eye fixations to each AOI. Note that, for the current analysis, eye fixation proportion refers to the proportion of trials where the fixation falls within an AOI (if a fixation falls in this area, it is considered as a fixation on the object corresponding to that AOI), rather than the proportion of time spent fixating an AOI within a time window. Before detailing the analysis procedures for both types of data, I provide an overview of how the eye-tracking data were preprocessed. Eye fixations were analyzed only within the time frame (as in the fourth screen in Figure 6.2) from the end of the sentence until the participant made a picture choice by clicking. This decision was made because quantified sentences are globally ambiguous, and a complete interpretation was obtained only after the participant finished listening to the sentence. Previous research supports this claim, as studies have shown that the scope effect was not pronounced in any of the regions within doubly quantified sentences. Instead, the effect primarily arose in the continuations that served to disambiguate the globally ambiguous context sentence. These findings were obtained through self-paced reading (Anderson, 2004; Dwivedi, 2013) and eye-tracking in the reading paradigm (Zhou & Gao, 2009).

The full eye-tracking data were stored in separate files in Gorilla, with each participant and trial having its own downloadable file. These data were compiled into one large data frame using Python scripts. Each file contains information regarding participants' eye-gaze locations and the (predicted) coordinates of the screen zones. Normalized coordinates and gaze locations were used for analysis to accommodate variations in participants' computer screen sizes. The boundaries of each AOI were defined using zone_x, zone_y, zone_width, and zone_height coordinates, allowing for predictions of where the participant was looking at. Each predicted eye gaze was assigned a score indicating the fit to the face model. This score is generated by the Support Vector Machine (SVM) classifier, which evaluates the resemblance of the image to a face. The score ranges from 0 (no fit) to 1 (perfect fit), with values above 0.5 considered optimal. For both L1 and L2 English speakers, less than 0.2% of the data were excluded due to suboptimal face detection values (face_conf < 0.5). An eye gaze was recorded as an eye fixation to a specific AOI if it fell within the AOI boundaries within the time frame from the end of the sentence to the picture selection (1 = fixated, 0 = not fixated). Fixation detection was performed using the 'add_aoi' function in the eyetrackingR package (version 0.2.0) (Dink & Ferguson, 2015) in R (R Core Team, 2021).

Statistical analyses for the current experiment followed similar procedures employed in Experiments 1 and 2. As for the TVJT task in the form of sentence-picture matching, analyses focused on the trials where a quantified sentence was presented along with two pictures, one depicting the surface scope reading and the other depicting the inverse scope reading. This is because trials from this condition allowed us to reveal participants' preference for one interpretation over the other, quantifying the variability and statistically differentiating their choices. Additionally, the trials from the other two conditions (SSR/ISR together with false readings) yielded results in which participants consistently favored the pictures representing the SSR/ISR interpretations over the pictures representing false readings. This was evident across different groups and conditions. These findings will be presented in detail in the results section, providing descriptive results that demonstrate participants' strong performance in selecting the correct interpretations (surface scope or inverse scope readings). To examine the potential differences between L1 speakers and L2 learners in their preferences for SSR or ISR interpretations, a series of logistic mixed-effects models were fit to the data with trials targeting SSR/ISR preferences. The dependent variable in these models was participants' (binary) responses in selecting the pictures, while the fixed effect of interest was language group (L1 speakers vs. L2 learners). The analyses were conducted separately for each linguistic structure, allowing for a structure-by-structure examination of the effects. Furthermore, the potential influences of L2 proficiency and definiteness knowledge were explored within the L2 learner data. Individual analysis results were also reported for the data of trials simultaneously tested on SSR and ISR to further reveal data variability among individuals.

In the analysis of eye fixation data, eye fixations were aggregated across samples (60 samples per second under ideal conditions in Gorilla) collected within the time frame of interest (from sentence offset until the click made). By aggregating the eye fixation data within this timeframe, I calculated the mean proportions of eye fixations directed towards specific AOIs for each trial. Linear mixed-effects models were constructed for the combined L1 and L2 data of each condition of trials across different types of quantifier scope. As a result, for each type of quantifier scope, separate analyses were conducted for each type of visual scene (SSR-false reading (FR); ISR-FR; SSR-ISR). These analyses also considered the potential differences in interpreting a given sentence between L1 speakers and L2 learners by including language group as a fixed factor. Different from the sentence-picture matching task, the statistical analyses for the eye-tracking experiment included not only the condition targeting interpretive preference but also the two conditions related to access. This inclusion was motivated by the fact that eye tracking data provided more fine-grained information about participants' fixation patterns towards specific AOIs at different time points within each trial. Consequently, the data patterns exhibited higher variability compared to the data collected from the sentence-picture matching task with a binary choice, as demonstrated by the visualized figures presented in the results section. Individual analysis results were exclusively presented for the sentence-picture matching task, with the omission of eye fixation data. This choice was driven by the intention to concentrate on scrutinizing aggregated data, which could effectively unveil patterns in the collective (across samples) eye fixation behavior of participants.

6.2.2 Results

6.2.2.1 Results of picture selection from sentence-picture matching

The examination of fillers resulted in an accuracy rate of 93% for L1 English speakers and 91% for L2 English speakers, indicating a high level of understanding and engagement among the participants. The mean percentages for experimental items across conditions are shown in Figure 6.3 for UE sentences, Figure 6.4 for EU sentences, Figure 6.5 for UN sentences, and Figure 6.6 for NU sentences. As can be seen clearly from Figures 6.3 - 6.5, both L1 and L2 participants chose surface scope readings for SSR_FR scenarios and inverse scope readings for ISR_FR scenarios most of the time (except for UN_ISR_FR in L2 English, all percentages being higher than 90%). This is true across different types of quantifier scope.

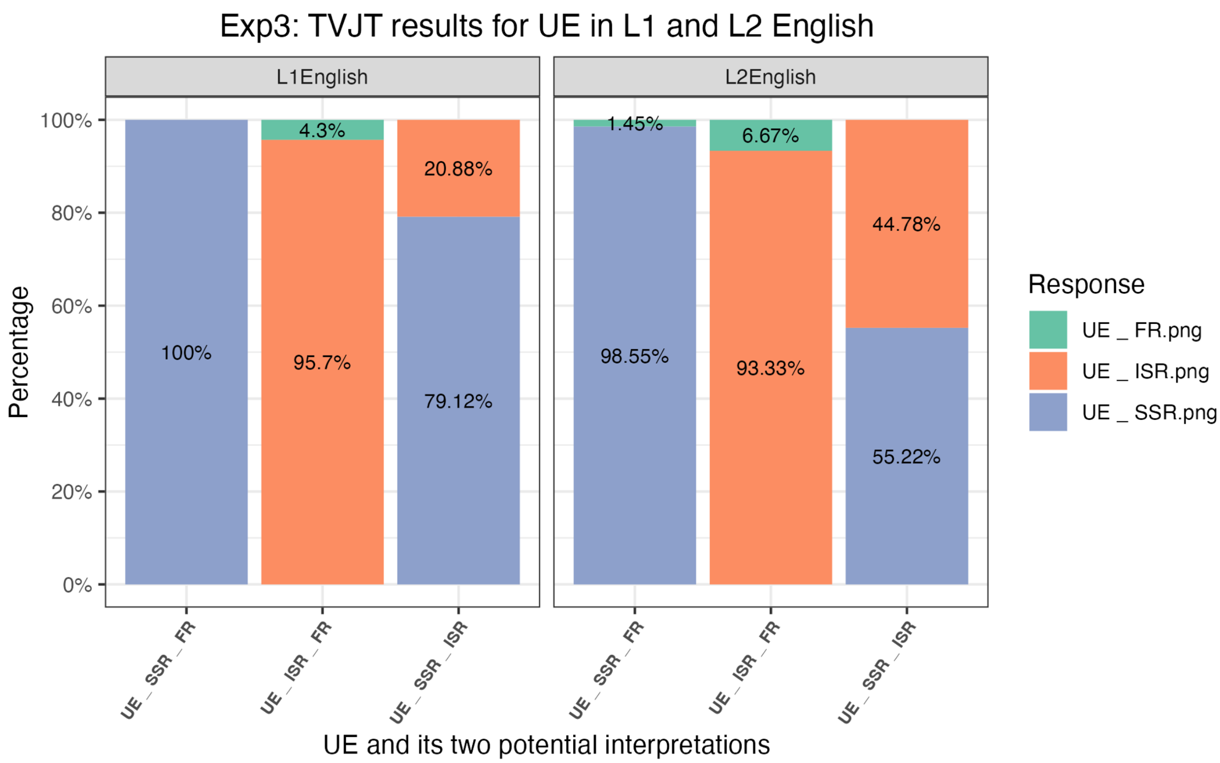

Figure 6.3 Exp3a: Mean percentages of UE sentences from TVJT in L1 & L2 English

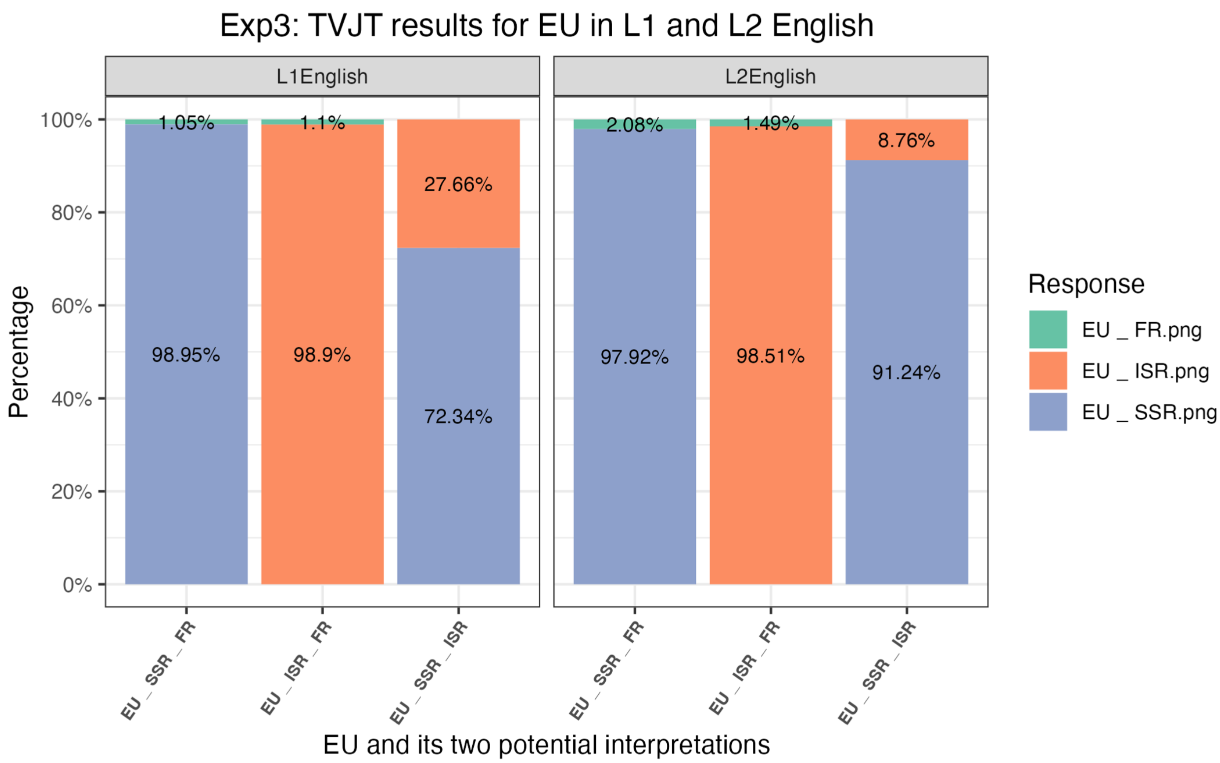

Figure 6.4 Exp3a: Mean percentages of EU sentences from TVJT in L1 & L2 English

In fact, their performance in the two scenarios reflects the accuracy of their interpretation of the given sentences. In both cases, the SSR or ISR was the only correct and possible interpretation for the sentences they heard, assuming they had access to this interpretation. In this regard, they displayed a rather high accuracy for the access to SSRs and ISRs across structures. Besides, the participants generally found the test sentences to be more felicitous when evaluated in the scenario where SSR was the only interpretation compared to the scenario where ISR was the only interpretation (SSR_FR vs. ISR_FR), which was evidenced by generally higher percentages for SSR than for ISR in these two scenarios (except for EU sentences with L2 English participants).

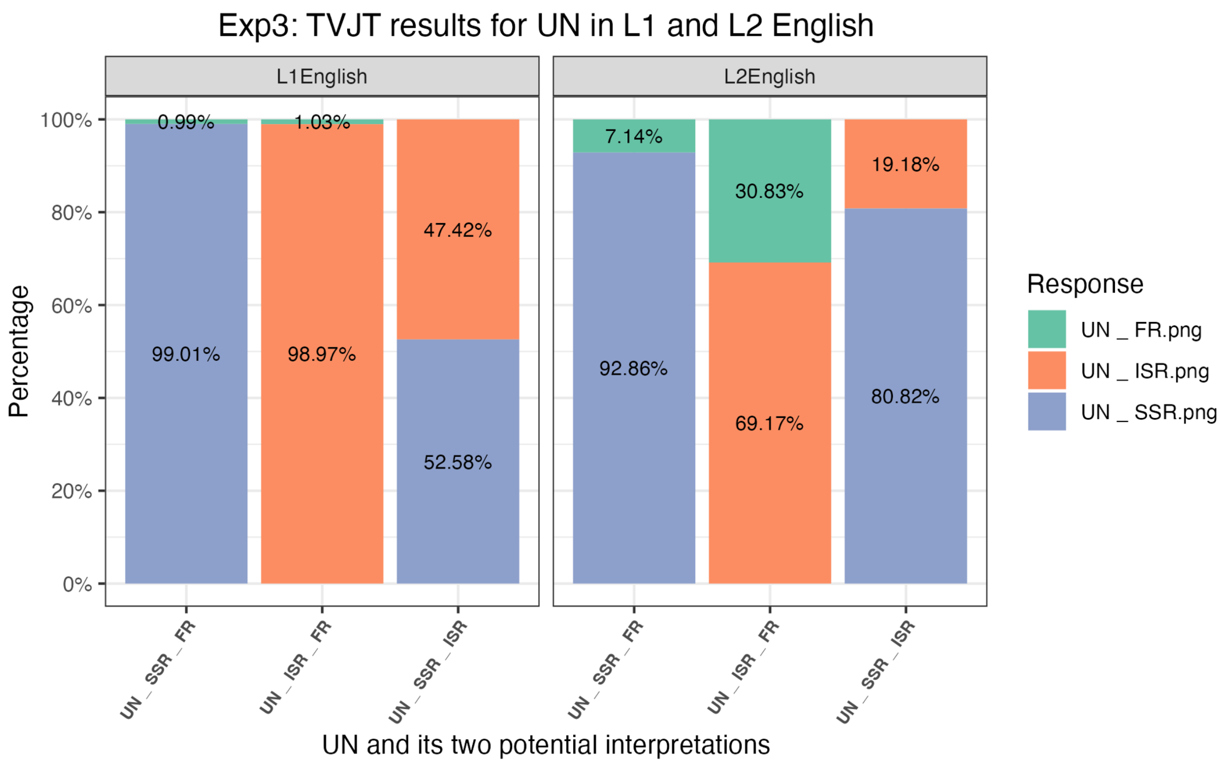

Figure 6.5 Exp3a: Mean percentages of UN sentences from TVJT in L1 & L2 English

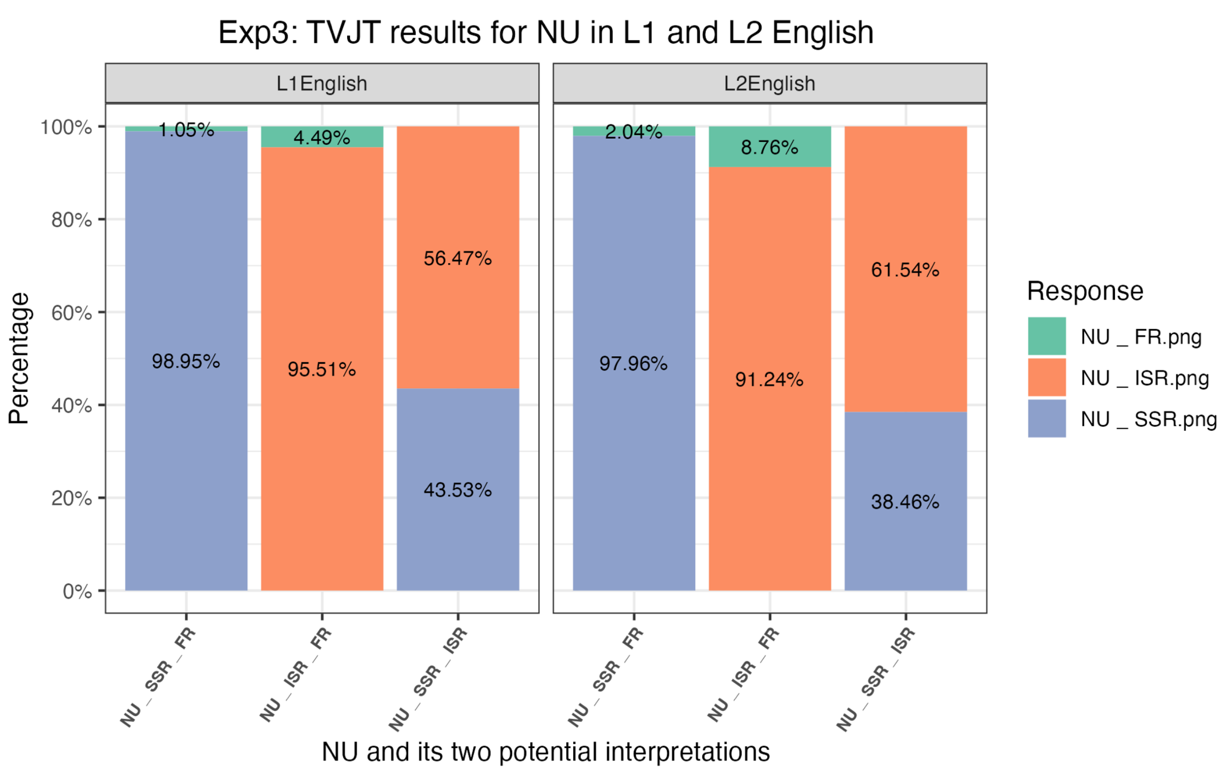

Figure 6.6 Exp3a: Mean percentages of NU sentences from TVJT in L1 & L2 English

Turning to the results for the scenario of SSR_ISR, L1 English speakers selected pictures representing the SSR over 70% of the time for UE (e.g., Every child climbed a tree) and EU (e.g., A child climbed every tree) sentences. When it comes to UN and NU sentences, L1 English speakers selected pictures representing both SSR and ISR to a similar extent. Regarding L2 learners, although they were more inclined to interpret UE (SSR: 55.22% vs. ISR: 44.78%), EU (SSR: 91.24% vs. ISR: 8.76%), and UN (SSR: 80.82% vs. ISR: 19.18%) sentences as SSR, they demonstrated a preference for interpreting NU sentences as ISR (SSR: 38.46 % vs. ISR: 61.54%). The statistical analyses returned a main effect of language group for UN (b = 2.08, p < .001), UE (b = -1.76, p = .001), and EU (b = 1.98, p = .003) sentences, respectively, suggesting that L1 and L2 participants differed in in their preference for SSR over ISR in these instances. Specifically, L2 English learners exhibited a higher likelihood of interpreting EU and UN sentences as SSR compared to L1 English speakers. Interestingly, L2 learners displayed a higher tendency to interpret UE sentences as ISR compared to L1 speakers. No language group effect was found for NU sentences.

The influence of L2 English proficiency on participants' preference for SSR vs. ISR in SSR_ISR scenarios among L2 learners was examined. However, no significant influence was found for any of these structures, including UE, EU, UN, and NU (all *ps* > .05). Definiteness knowledge was also explored for its influence on L2 English learners' performance on their choices of pictures representing ISR with different types of quantifier scope in the SSR_ISR scenario. A medium negative correlation between one's definiteness knowledge scores and picture choices for ISR was found only in UN sentences (r = -.35, p = .02). No such correlation was observed in other cases, including UE, EU, and NU sentences.

6.2.2.2 Individual results for sentence-picture matching task with L1 and L2 English

Individual analyses were conducted to assess whether these results align with the patterns of results observed at the group level. Given the higher variability exhibited in the SSR_ISR scenario compared to the other two evaluated scenarios with false readings, the individual analysis concentrated on trials conducted within the SSR_ISR scenario. In this experiment, I utilized the sentence stimuli from Experiments 1 and 2, for which six items for each type of quantifier scope were created. These six items from the same type of quantifier scope were distributed into three different scenarios (SSR_FR, ISR_FR, and SSR_ISR), resulting in two items for each quantifier scope type within each scenario. To perform the individual analysis, I calculated the percentages of participants who consistently selected pictures representing a particular interpretation for all items within each quantifier scope type.

Table 6.1 Percentages of participants consistently selecting pictures representing a particular interpretation across all items in L1 and L2 English

	Universal-Existential		Existential-Universal		Universal-Negation		Negation-Universal	
	SSR	ISR	SSR	ISR	SSR	ISR	SSR	ISR
L1 English	50.85%	5.08%	44.07%	11.86%	27.87%	22.95%	20.97%	27.42%
L2 English	28.74%	20.69%	72%	1.33%	56.96%	7.59%	18.39%	31.03%

As shown in Table 6.1, the individual analysis results were largely in line with the group analysis results from the sentence-picture matching task. For UE sentences, L1 English speakers consistently preferred pictures representing the SSR over those depicting the ISR. By comparison, L2 English learners showed relatively similar preferences for both interpretations, as they consistently chose pictures for either interpretation to a similar degree. For EU sentences, although both groups of participants showed a strong preference for SSR as opposed to ISR, this preference was even more pronounced in the L2 English group (SSR: 72% vs. ISR: 1.33%) compared to the L1 English group (SSR: 44.07% vs. ISR:11.86%). For UN sentences, L1 English speakers demonstrated a similar preference for SSR and ISR, as they consistently selected pictures representing either interpretation to a comparable extent (SSR: 27.87 % vs. ISR: 22.95%). In contrast, the L2 English group demonstrated a strong preference for SSR, with 56.96% of the L2 participants consistently selecting pictures representing SSR, while 9.59% consistently chose pictures for ISR. These findings highlight the significant preference for SSR in the L2 English group. Regarding NU sentences, both participant groups displayed a preference for ISR, as higher percentages of participants consistently chose pictures representing ISR (L1: 27.42%; L2: 31.03%) compared to pictures for SSR (L1: 20.97; L2 18.39%).

6.2.2.3 Results of eye fixations

The mean proportions of eye fixations are illustrated in Figure 6.7 for UE sentences, Figure 6.8 for EU sentences, Figure 6.9 for UN sentences, and Figure 6.10 for NU sentences. The graph illustrates the average proportion of eye fixations towards both AOIs, calculated by aggregating data across times from individual trials and participants. Detailed mean proportions of eye fixations for various language groups (L1 and L2 English), quantifier scope types, presentation scenarios, and AOIs are provided in Appendix C.

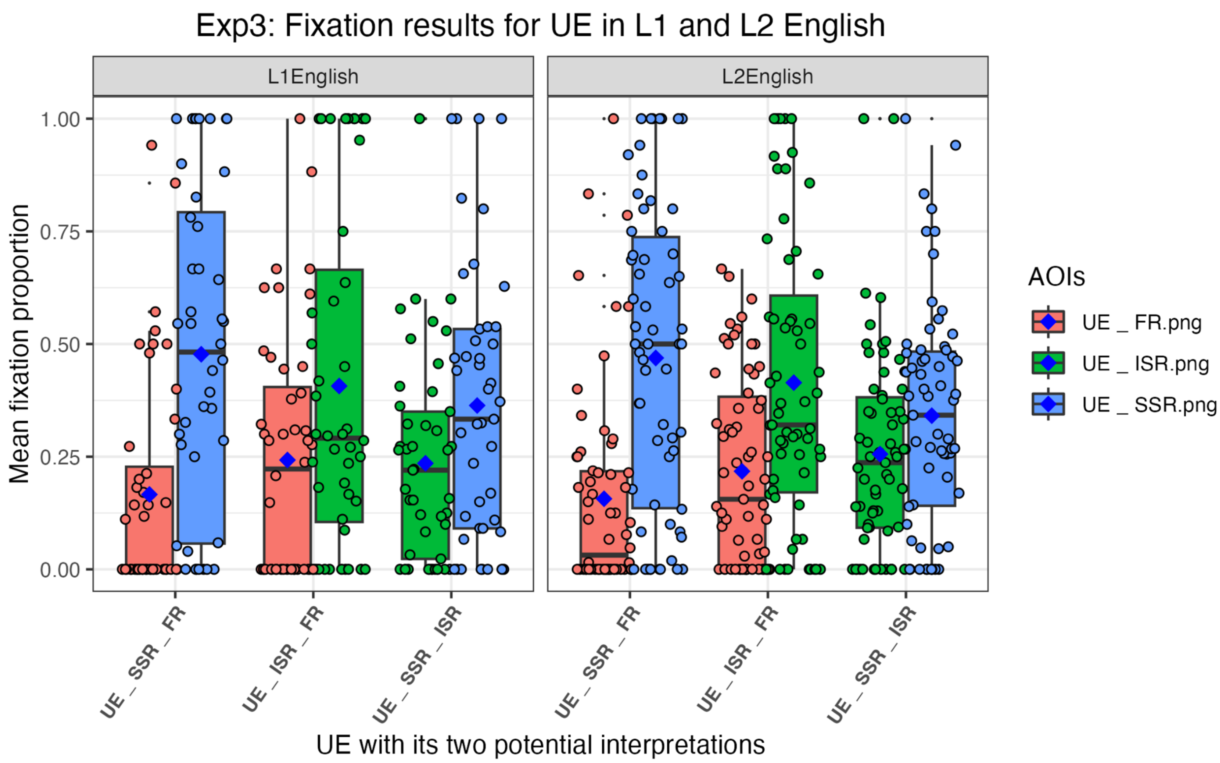

Figure 6.7 Exp3a: Mean proportions of eye fixations for UE sentences in L1 & L2 English

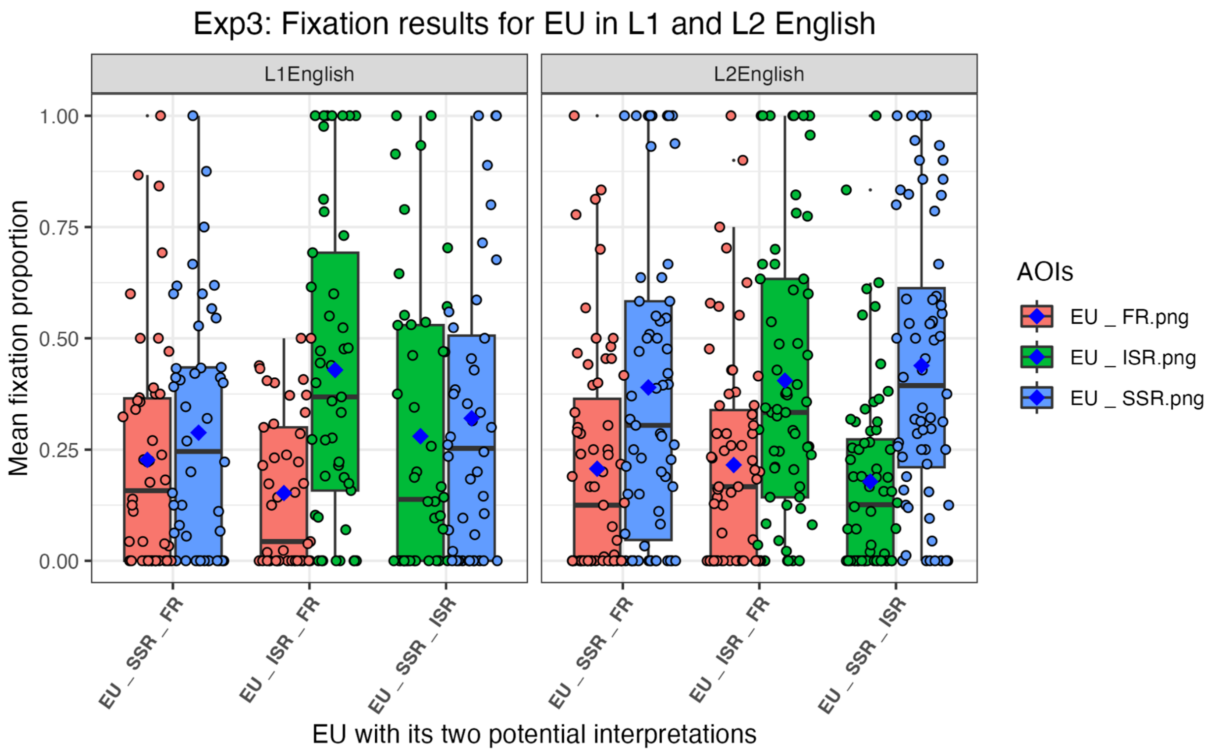

Figure 6.8 Exp3a: Mean proportions of eye fixations for EU sentences in L1 & L2 English

Descriptively, in the SSR_FR scenario, both L1 and L2 participants consistently directed their attention more towards pictures depicting SSR than those depicting FR, regardless of the type of quantifier scope. This observation was confirmed in the statistical analyses, for which AOIs and language groups were included as the fixed effects. Specifically, pictures depicting SSR received significantly higher proportions of looks compared to pictures representing FR. This difference can be attributed to the main effects of AOIs in UE (b = 0.36, p < .001), EU (b = 0.21, p < .001), UN (b = 0.26, p < .001), and NU sentences (b = 0.27, p < .001). No interaction existed between AOIs and language group in any of these cases. The observed pattern of results aligns with the findings from the sentence-picture matching task, indicating that both L1 and L2 participants were able to access the given sentence on its SSR when evaluated against an entirely impossible interpretation, i.e., FR. For the ISR_FR scenario, both participant groups showed a significantly higher proportion of looks towards pictures for ISR compared to SSR across various quantifier scope types. This was evidenced by the main effects of AOIs in UN (b = 0.23, p < .001), NU (b = 0.16, p < .001), UE (b = 0.22, p < .001), and EU (b = 0.24, p < .001). These findings suggest that participants were able to access ISR as a viable interpretation when contrasting it with FR as an impossible interpretation. This pattern holds consistently across various quantifier scope types and language groups. No meaningful interaction was found between AOIs and language group in the case of ISR_FR.

Now, let us shift our focus to the SSR_ISR scenario and examine each type of quantifier scope individually. For UE sentences, a main effect of AOIs was found (b = 0.08, p = .003), driven by higher proportions of looks to pictures for SSR than for ISR among both participant groups. For EU sentences, although a main effect of AOIs demonstrated higher proportions of looks towards SSR than ISR overall (b = 0.16, p < .001), an interaction between AOIs and language groups (b = 0.15, p = .01) suggests that this difference was specifically observed in the L2 group. For UN sentences, both a main effect of AOIs (b = 0.10, p < .001) and an interaction between AOIs and language group (b = 0.24, p < .001) were observed. This interaction specifically resulted in differences in the proportions of looks towards SSR and ISR being observed in the L2 group only. There were no significant differences in participants' proportions of looks towards SSR relative to ISR within each group in the case of NU sentences (all ps > .05).

The study examined how L2 English proficiency and learners' definiteness knowledge influenced interpretation preferences in the SSR_ISR scenario. The results did not show any influences of either L2 English proficiency or definiteness knowledge on the interpretation preferences for any type of quantifier scope.

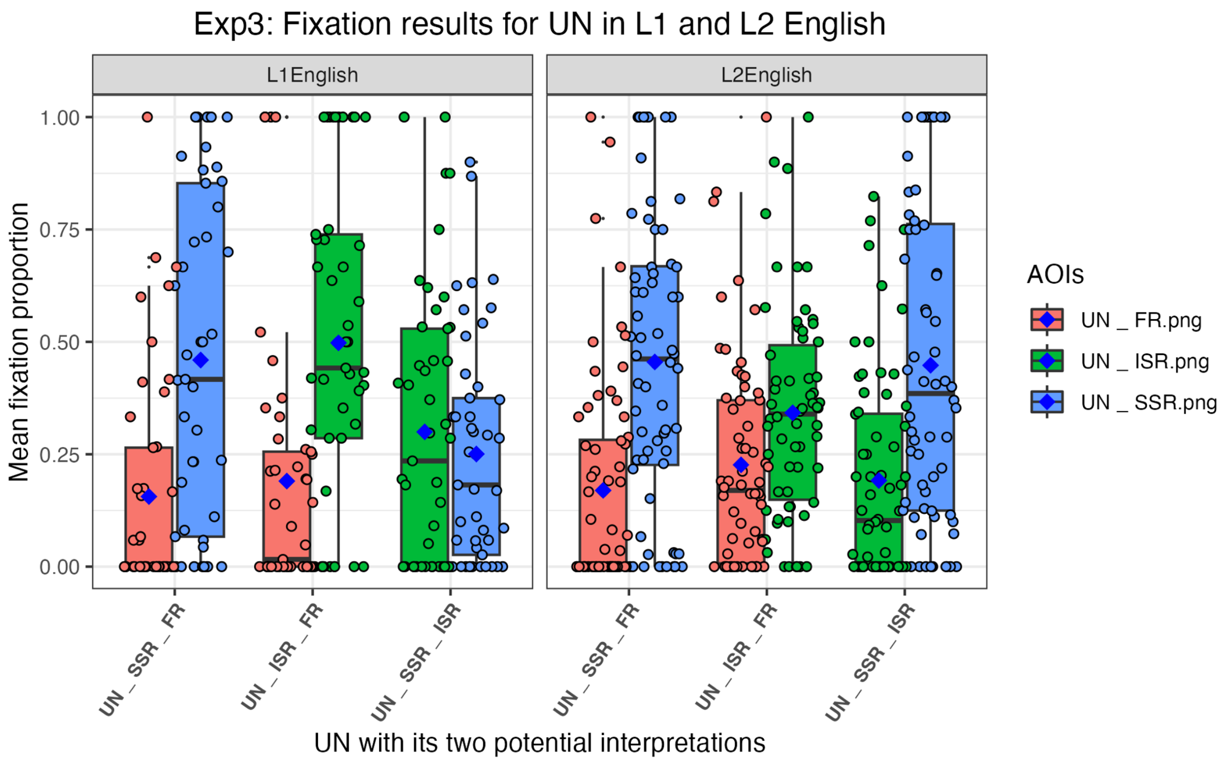

Figure 6.9 Exp3a: Mean proportions of eye fixations for UN sentences in L1 & L2 English

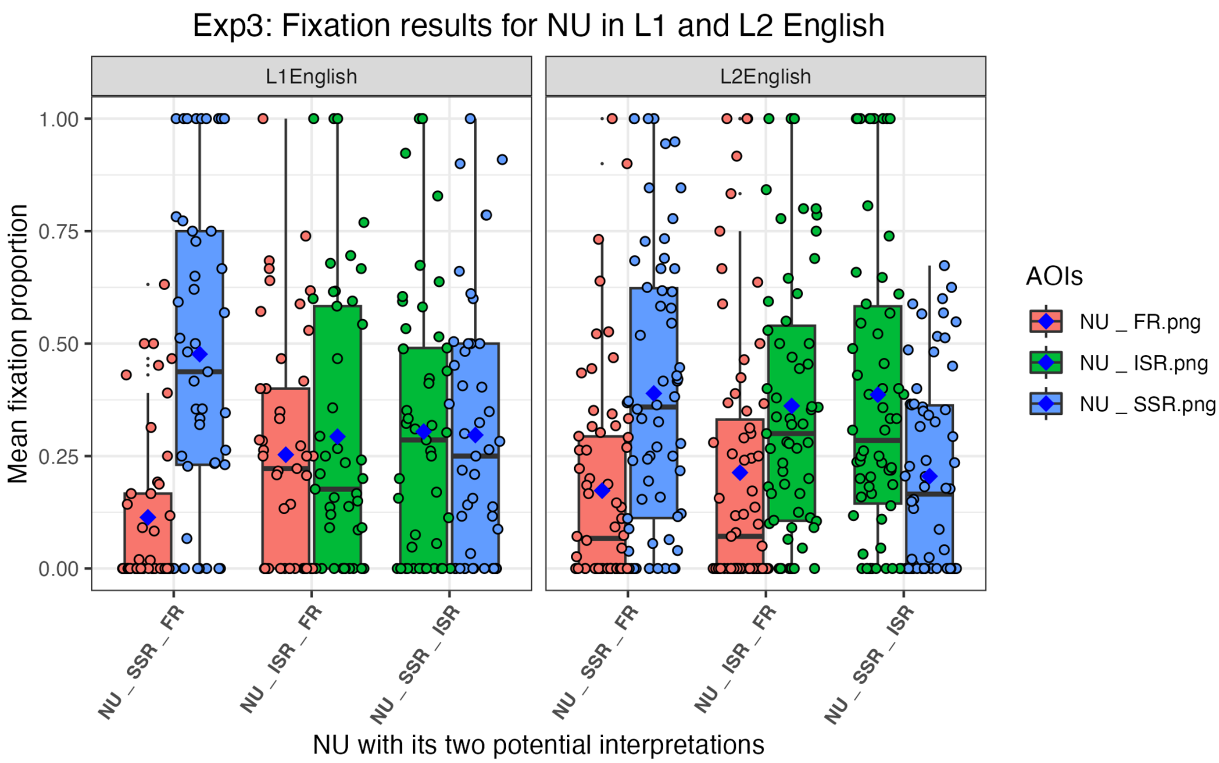

Figure 6.10 Exp3a: Mean proportions of eye fixations for NU sentences in L1 & L2 English

6.3 Experiment 3b: L2 Chinese visual-world eye-tracking experiment

6.3.1 Methods

6.3.1.1 Participants

The participants involved in the current experiment were also part of Experiment 2b. For information about the participants, refer to sections 5.3.1.1 of Experiment 2b.

6.3.1.2 Design and materials

The experimental design closely followed that of Experiment 3a with L1 and L2 English participants, with the exception that the auditory sentences were presented in Chinese for L1 and L2 Chinese participants in this experiment. The sentence stimuli were recorded by a female native speaker of Mandarin Chinese who was unaware of the experimental purpose. The total duration of the target Chinese sentences did not differ significantly between conditions, as confirmed by the results of one-way ANOVA tests (all *ps* > .05) (UE: 3.0s, EU: 3.07s, UN: 3.16s, NU: 3.17s).

6.3.1.3 Procedure

The experimental implementation procedures were identical to those outlined in section 6.2.1.3 of Experiment 3a.

6.3.1.4 Analysis

The analysis procedures followed those reported in section 6.2.1.4 of Experiment 3a.

6.3.2 Results

6.3.2.1 Results of picture selection from sentence-picture matching

The examination of fillers resulted in an accuracy rate of 92% for L1 Chinese speakers and 90% for L2 Chinese speakers, indicating a high level of understanding and engagement among the participants. The mean percentages for experimental items across conditions are shown in Figure 6.11 for UE sentences, Figure 6.12 for EU sentences, Figure 6.13 for UN sentences, and Figure 6.14 for NU sentences. Both L1 and L2 Chinese participants chose the pictures representing SSR for the SSR_FR scenario and the pictures representing ISR for the ISR_FR scenario most of the time (all above 90%, except for UN_ISR_FR in L1 (75.73 %) and L2 (73.08%) Chinese). The observed pattern holds consistently across different types of quantifier scope.

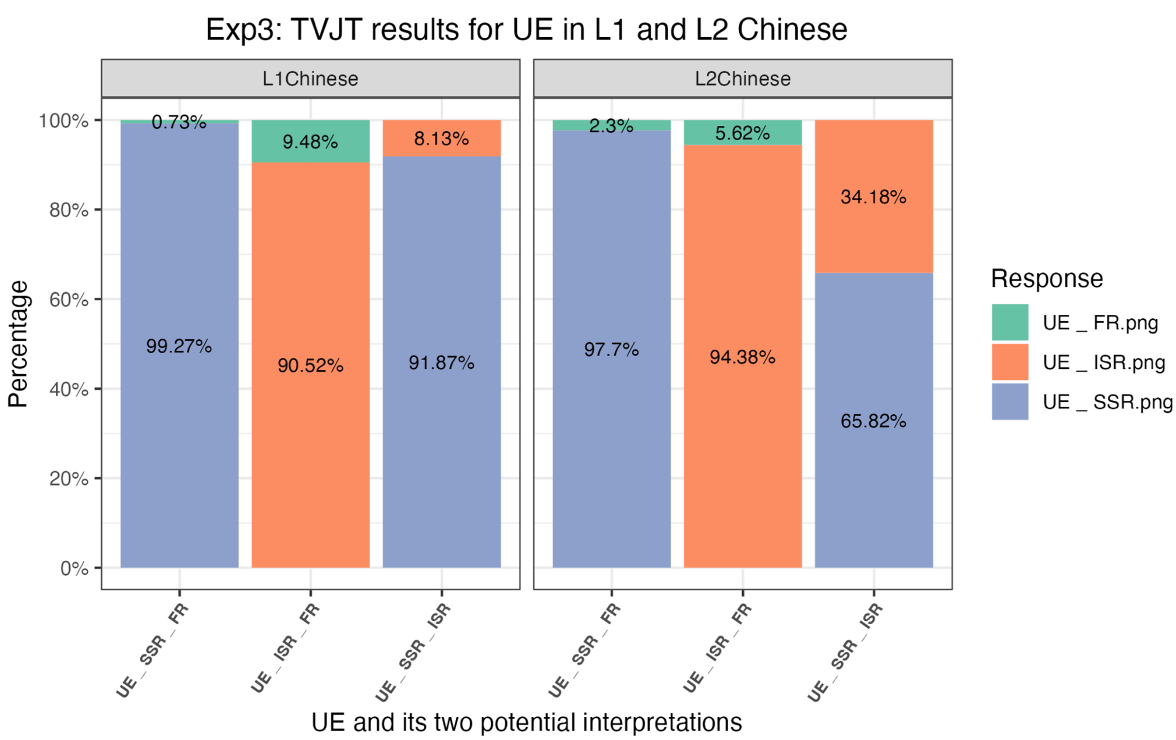

Figure 6.11 Exp3a: Mean percentages of UE sentences from TVJT in L1 & L2 Chinese

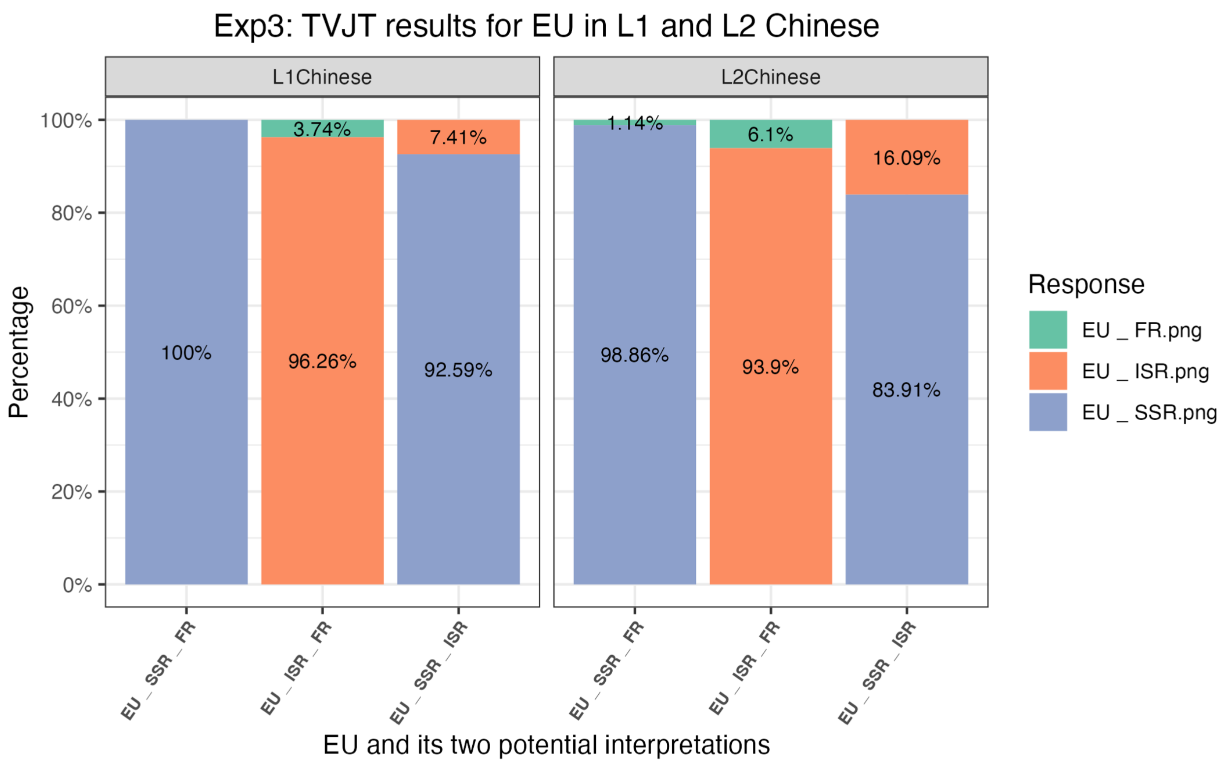

Figure 6.12 Exp3a: Mean percentages of EU sentences from TVJT in L1 & L2 Chinese

The other notable observation is that both groups of participants found the test sentences to be more felicitous when assessed in the scenario where SSR was the only possible interpretation, as opposed to the scenario where ISR was the sole correct interpretation (SSR_FR vs. ISR_FR). This trend is supported by consistently higher percentages favoring SSR over ISR in these two scenarios, as seen in the case of UE, EU, and UN sentences. A notable exception arises with NU sentences, as both L1 and L2 Chinese participants found the test sentences to be more felicitous when considering ISR as the only possible interpretation, compare to when SSR was considered as the sole correct interpretation (L1 Chinese: 100% for NU_ISR_FR vs. 99.28% for NU_SSR_FR; L2 Chinese: 97.47% for NU_ISR_FR vs. 85.71% for NU_SSR_FR). These results suggest that, for both L1 and L2 Chinese participants, NU sentences with ISR appeared to be more accessible than those with SSR.

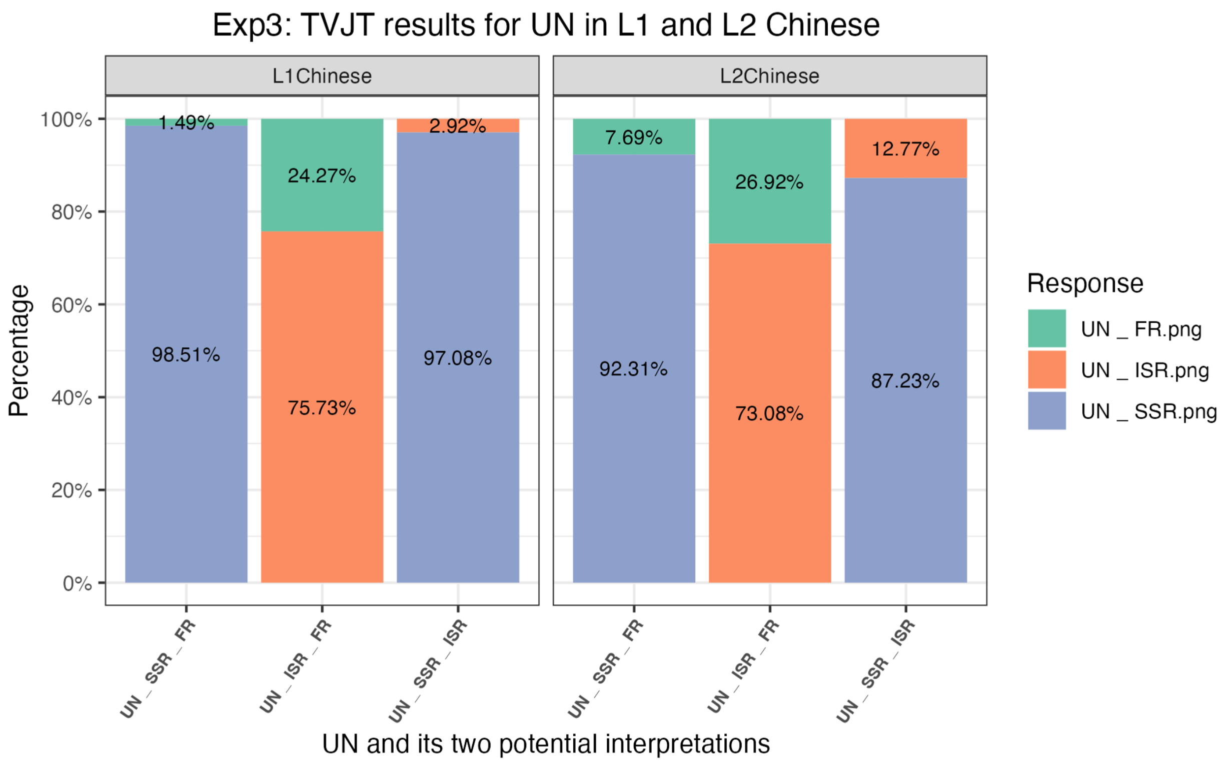

Figure 6.13 Exp3a: Mean percentages of UN sentences from TVJT in L1 & L2 Chinese

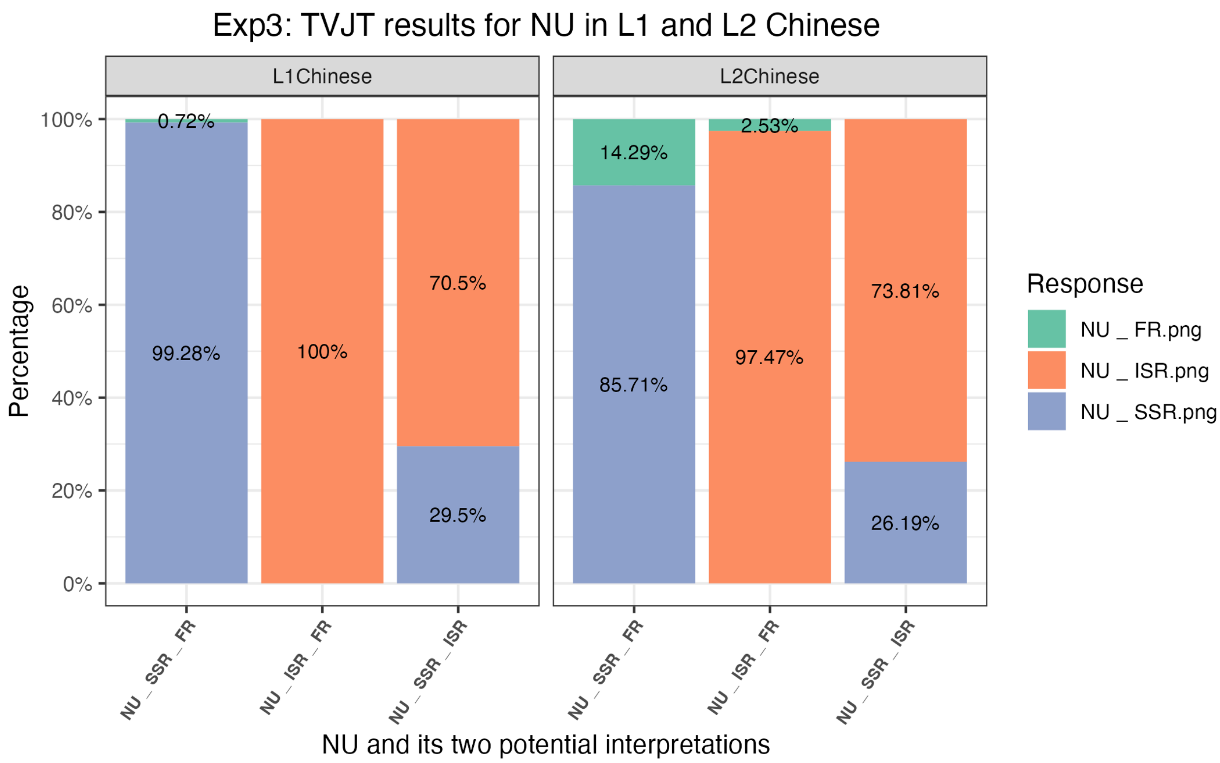

Figure 6.14 Exp3a: Mean percentages of NU sentences from TVJT in L1 & L2 Chinese

The scenario of SSR_ISR was examined to determine the relative preferences for SSR and ISR. For UE, EU, and UN sentences, both L1 and L2 Chinese participants consistently demonstrated a preference for pictures representing SSR as opposed to ISR, as evidenced by higher percentages of choices made for pictures depicting SSR than for pictures depicting ISR. Furthermore, the preference for SSR was more pronounced in L1 Chinese speakers. Statistical analyses returned a main effect of language group for UN sentences (b = -2.17, p < .001), suggesting that L1 Chinese speakers significantly exhibited a preference for SSR than L2 Chinese learners in this case. A marginal effect of language group arose for UE sentences (b = -3.21, p = .0899), which was further examined using a Bayesian analysis that did not provide conclusive support for this effect (The Bayes factor being 0.07, smaller than 1/3). No significant effect of language group was found for EU sentences, suggesting a lack of statistical support for the observed differential preference for SSR vs. ISR. In the case of NU sentences where the ISR was preferred, there were no significant differences between L1 Chinese speakers and L2 Chinese learners in their preference patterns (*b* = -0.84, *p* = .34).

In addition, I explored the influence of L2 proficiency and one's DOU knowledge on L2 Chinese learners' performance in the sentence-picture matching task. The study found no significant impact of L2 proficiency on individuals' preference for SSR vs. ISR in the SSR_ISR scenario, regardless of the quantifier scope type (all *ps* > .05)[68]. The correlation was examined between individual learners' knowledge about DOU and their preferences for SSR vs. ISR in the SSR_ISR scenario, in the case of the UE and UN sentences where the DOU particle was present. The analysis showed no significant correlation for UE or UN sentences.

6.3.2.2 Individual results for sentence-picture matching task with L1 and L2 Chinese

Table 6.2 Percentages of participants consistently selecting pictures representing a particular interpretation across all items in L1 and L2 Chinese

	Universal-Existential		Existential-Universal		Universal-Negation		Negation-Universal	
	SSR	ISR	SSR	ISR	SSR	ISR	SSR	ISR
L1 Chinese	81.82%	3.64%	75.44%	3.51%	88.89%	0%	18.75%	40.63%
L2 Chinese	42.31%	23.08%	61.54%	3.85%	60.38%	1.89%	18%	54%

[68] A marginal effect of L2 Chinese proficiency was observed in L2 Chinese learners' picture choices for SSR vs. ISR in the SSR_ISR scenario (*b* = 0.12, *p* = .063). However, a Bayesian analysis with a Bayes factor of 0.946 did not provide conclusive support for the existence of this influence.

Individual analysis results for L1 and L2 Chinese were provided in Table 6.2. The pattern of results was consistently observed across UE, EU, and UN sentences, with both participant groups displaying a preference for SSR over ISR. Notably, this preference was more pronounced in the L1 group. For NU sentences, both participant groups demonstrated a preference for ISR, as evidenced by a higher percentage of participants choosing pictures representing ISR (L1: 40.63%; L2: 54%) compared to pictures representing SSR (L1: 18.75%; L2: 18%).

6.3.2.3 Results of eye fixations

The mean proportions of eye fixations are visualized in Figure 6.15 for UE sentences, Figure 6.16 for EU sentences, Figure 6.17 for UN sentences, and Figure 6.18 for NU sentences. Appendix D provides a table of eye fixations for their mean proportions calculated across various language groups (L1 and L2 Chinese), quantifier scope types, presentation scenarios, and AOIs.

Similar to the findings in L1 and L2 English fixation data, both L1 and L2 Chinese participants looked at pictures representing SSR/ISR more than pictures representing FR, regardless of the quantifier scope type. This observation received statistical support. Specifically, for UE sentences, a main effect of AOIs arose for the SSR_FR scenario ($b = 0.35$, $p < .001$), driven by the fact that participants looked at pictures depicting SSR more than pictures depicting ISR, regardless of language group. In addition, a main effect of AOIs also arose of the ISR_FR scenario ($b = 0.28$, $p < .001$), due to higher proportions of looks towards pictures depicting ISR compared to pictures depicting FR among both L1 and L2 Chinese participants. As for EU sentences, there was a main effect of AOIs observed in the SSR_FR scenario ($b = 0.34$, $p < .001$), indicating higher proportions of looks to pictures for SSR than to pictures for ISR. A main effect of AOIs emerged in the ISR_FR scenario ($b = 0.23$, $p < .001$), which was induced by higher proportions of looks to pictures for ISR than to pictures for FR in the case of EU sentences. As for UN sentences, both groups of participants also directed more attention towards pictures for SSR than pictures for FR in the SSR_FR scenario, manifested in the main effect of AOIs ($b = 0.41$, $p < .001$). In the ISR_FR scenario, a main effect of AOIs also arose for UN sentences ($b = 0.08$, $p = .004$), driven by higher proportions of looks to pictures for ISR than to pictures for FR. Finally, analyses on NU sentences yielded a main effect of AOIs for the SSR_FR scenario ($b = 0.37$, $p < .001$) and also for the ISR_FR scenario ($b = 0.29$, $p < .001$), suggesting that both L1 and L2 Chinese participants looked at pictures for either SSR or ISR more than pictures for FR.

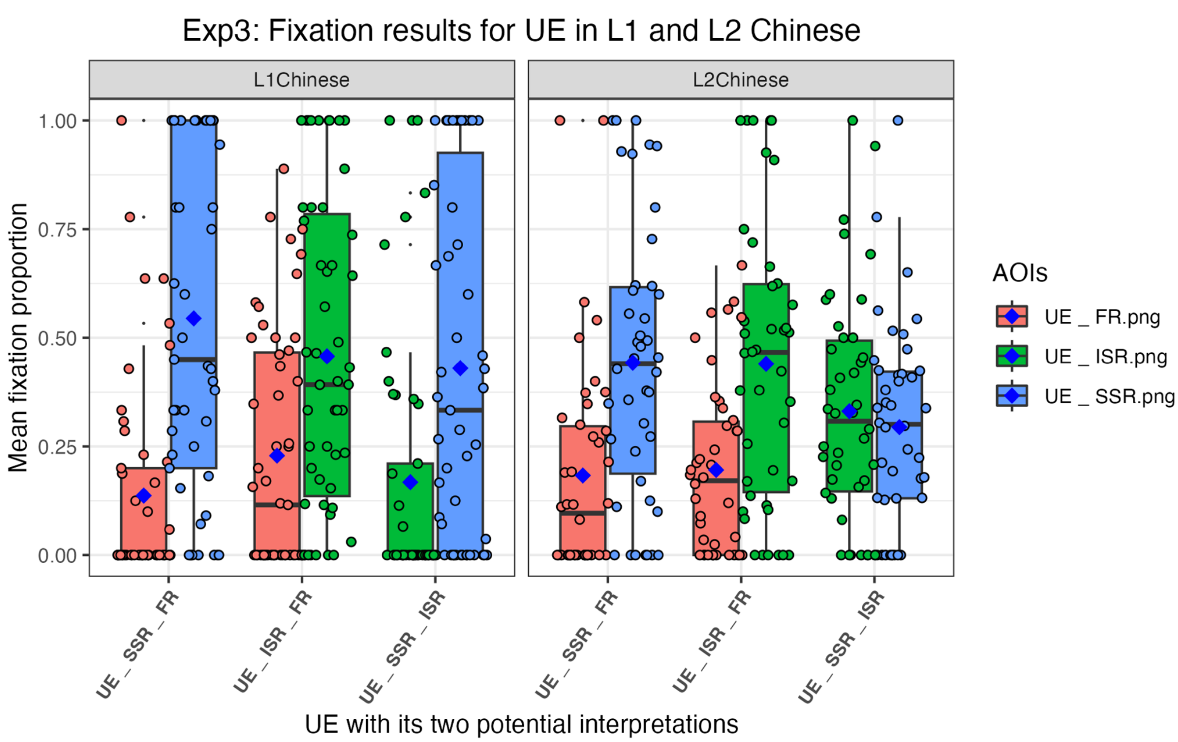

Figure 6.15 Exp3a: Mean proportions of eye fixations for UE sentences in L1 & L2 Chinese

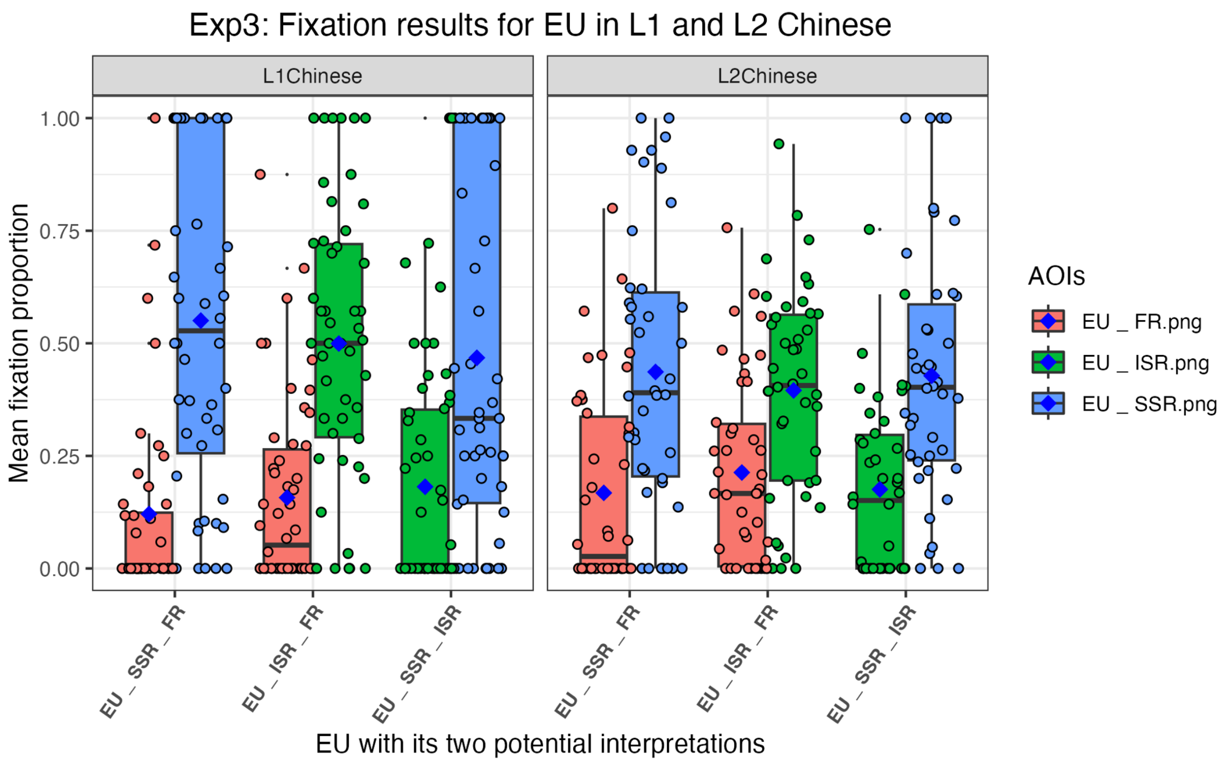

Figure 6.16 Exp3a: Mean proportions of eye fixations for EU sentences in L1 & L2 Chinese

When it comes to the SSR_ISR scenario, descriptively, L1 Chinese speakers seemed to predominantly favor SSR over ISR when a scenario provided both options in the case of UE, EU and UN sentences. NU sentences represent an exception, in that ISR was preferred over SSR by L1 Chinese speakers. Regarding L2 Chinese learners, their results followed a similar pattern to that of L1 Chinese participants, but their preference for one interpretation over the other was not as pronounced as that of the L1 Chinese speakers. Statistical analyses were conducted for each type of quantifier scope including AOIs and language group as the fixed effects. As for UE sentences, a main effect of AOIs was found ($b = 0.12$, $p < .001$), indicating that participants fixated on pictures for SSR more than pictures for ISR. More importantly, an interaction between AOIs and language group was observed ($b = -0.17$, $p = .011$), due to the fact that significant higher proportions of looks to pictures of SSR as opposed to ISR were only found for L1 Chinese speakers but for L2 Chinese learners. As for EU sentences, only a main effect of AOIs emerged ($b = 0.27$, $p < .001$), driven by higher proportions of looks to SSR than to ISR in both groups of participants. As for UN sentences, both groups of participants were found to look more at pictures for SSR than at pictures for ISR, manifested in the main effect of AOIs ($b = 0.32$, $p < .001$). Different from the other three cases, NU sentences led to higher proportions of looks to pictures for SSR than to pictures for ISR from both L1 Chinese speakers and L2 Chinese learners, reflected in the main effect of AOIs ($b = -0.17$, $p < .001$).

This experiment investigated the impact of L2 Chinese proficiency and learners' DOU knowledge on their interpretation preferences in the SSR_ISR scenario. However, the findings revealed no significant effects of either Chinese proficiency or DOU knowledge on interpretation preferences across different types of quantifier scope.

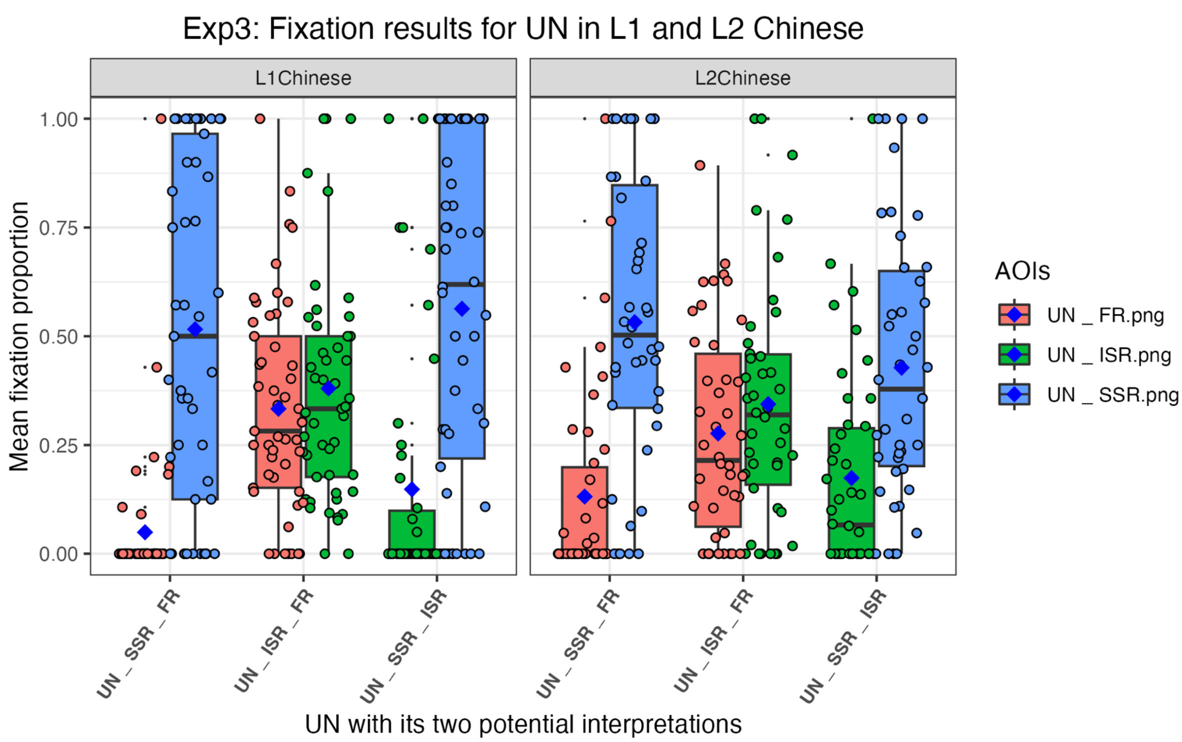

Figure 6.17 Exp3a: Mean proportions of eye fixations for UN sentences in L1 & L2 Chinese

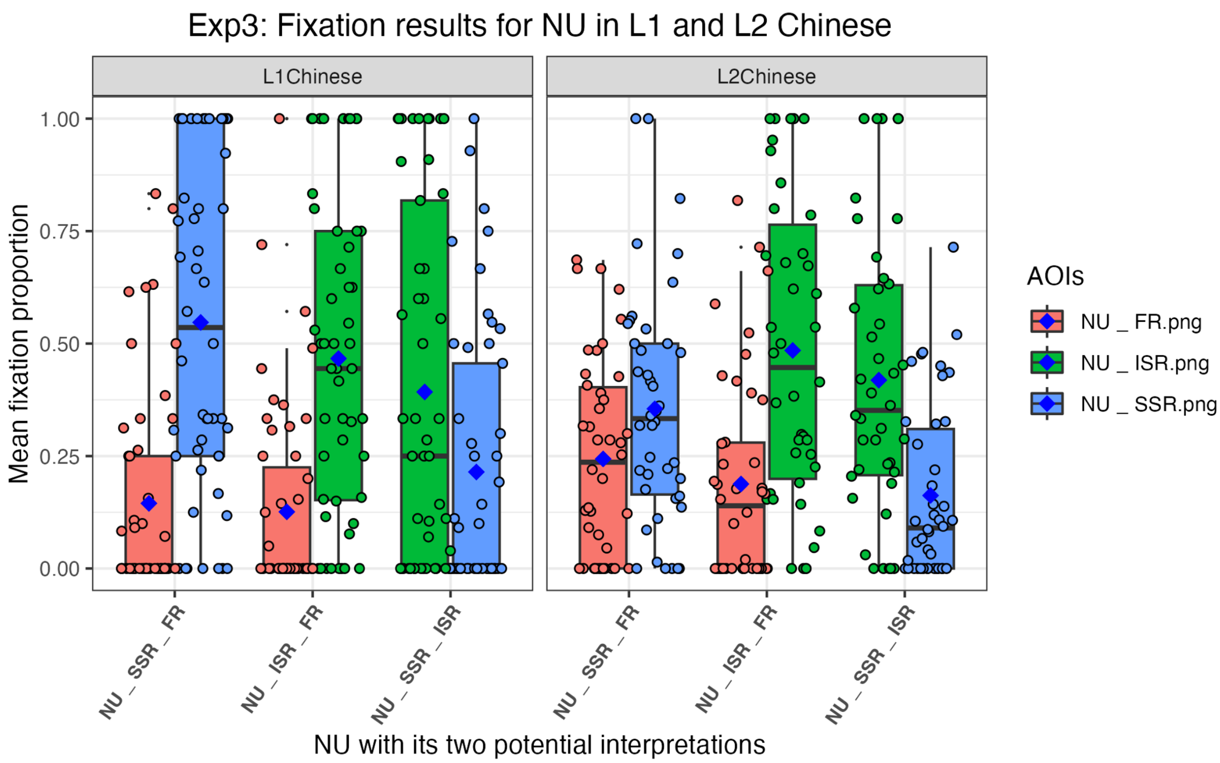

Figure 6.18 Exp3a: Mean proportions of eye fixations for NU sentences in L1 & L2 Chinese

6.4 Discussion for Experiment 3

The primary goal of Experiment 3 was to examine how L2 learners of English and Chinese process quantifier scope in real time[69]. Additionally, it was to determine if the patterns observed in offline settings through written truth-value judgment tasks can extend to online settings. To this end, I used eye-tracking in the visual-world paradigm to reveal L2 learners' processing of quantifier scope. The study collected two types of data: 1) picture selection representing various interpretations, and 2) eye fixations from sentence offset until picture selection, both of which provided evidence for learners' quantifier interpretation in online settings. In addition to its theoretical contributions in investigating quantifier scope processing in Chinese-speaking learners of English and English-speaking learners of Chinese, this study makes two significant methodological contributions. Firstly, it utilizes webcam-based eye-tracking as a more ecologically valid means to conduct this SLA investigation, and secondly, it incorporates a false reading for comparison with possible readings (i.e., SSR and ISR) to directly assess their accessibility.

First, I summarize the results obtained from the sentence-picture matching task and the eye-tracking task. Subsequently, I compare the findings from both tasks. Finally, I compare the results obtained through the online methods used in Experiment 3 with those obtained from the offline method used in Experiment 2. Patterns of results were explained with respect to relevant theoretical accounts. In the sentence-picture matching task with either the SSR_FR or the ISR_FR

---

[69] If *processing* is narrowly defined to include time-locked processes, it may be more appropriate to consider eye-tracking measures in this study as implicit measures of quantifier scope processing, while the sentence-picture matching task serves as explicit measures of preference.

scenario, participants were presented with two visual scenes, including one for false readings. This presentation allowed us to assess participants' access to the surface and inverse scope readings for each quantifier scope, respectively. Across all quantifier scope types, participants consistently selected the possible interpretation of each heard sentence rather than the completely impossible interpretation. This indicates that their performance exceeded the chance level, suggesting their ability to access the possible readings provided in the visual contexts. These findings held true across different language groups, including L1 English, L2 English, L1 Chinese, and L2 Chinese.

Comparing scenarios of SSR_FR and ISR_FR, it was notable that all participants in most cases found the test sentences to be more felicitous when assessed in the scenario where SSR was the only possible interpretation, as opposed to the scenario where ISR was the sole correct interpretation (SSR_FR vs. ISR_FR), as evidenced by higher percentages of participants favoring SSR over ISR in these two scenarios. This finding indirectly suggests that accessing inverse scope readings may be more challenging compared to surface scope readings, although both types of readings were accessible. One plausible explanation is that processing inverse scope is more challenging due to the additional syntactic operation it requires at the level of logical form (LF) (Anderson, 2004). One exception arose with NU sentences (e.g., The horse didn't jump over every fence) in L1 and L2 Chinese. These sentences were found to be more felicitous when assessed in the scenario where ISR was the only possible interpretation, as opposed to the scenario where SSR was the sole correct interpretation. This finding aligns with the results of Experiment 2, which examined the performance of L1 and L2 Chinese participants in written truth-value judgment tasks. In that experiment, it was found that both L1 and L2 Chinese speakers significantly favored the inverse scope reading of Chinese NU sentences, as evidenced by their high mean ratings of ISR (above 5 out of 7).

Regarding the results of the sentence-picture task involving L1 and L2 participants in the SSR_ISR scenario, which specifically investigated their interpretation preferences, the following findings emerged. Although both L1 and L2 English speakers preferred SSR over ISR in the cases of EU and UN sentences, L1 English speakers were found to accept ISR to a larger extent than L2 English learners. Notably, L1 English speakers exhibited a comparable degree of acceptance for UN sentences on both the ISR (47.42%) and SSR (52.58%). It was somewhat surprising to find that L2 English learners were even more likely to choose pictures representing ISR than L1 English native speakers in UE sentences. However, this fact alone suggests that Chinese learners of English may have acquired the ability to interpret UE sentences with inverse scope readings, which aligns with the performance of the same group of learners in the written TVJT conducted in Experiment 2. Regarding NU sentences, L2 English learners were not different from L1 English speakers in their ability to acquire the availability of ISR, which was in line with these learners' performance in the written TVJT.

L2 Chinese participants demonstrated a pattern similar to native Chinese speakers, exhibiting a strong preference for SSR in the case of UE, EU, and UN sentences. However, L2 Chinese participants exhibited a slightly lower preference for SSR compared to L1 Chinese speakers specifically for UN sentences. Nevertheless, both the L1 and L2 groups converged in their preference patterns for UE and EU sentences. Both the L1 and L2 Chinese groups exhibited a convergent preference pattern for NU sentences, albeit with a shift towards ISR. Taken together, English-speaking learners of Chinese demonstrated an overall ability to acquire the native speakers' way of processing quantifier scope, as measured by the sentence-picture matching task.

Regarding eye fixation patterns, it was observed that all participant groups, across all kinds of quantifier scope, were able to access both the SSR and ISR when presented along with a false reading. This was evidenced by significantly higher proportions of fixations towards SSR in the SSR_FR scenario and towards ISR in the ISR_FR scenario. Chinese-speaking learners of English failed to perform like native speakers of English in EU and UN sentences. Unlike native English speakers, who exhibited no preference difference between SSR and ISR, Chinese-speaking learners of English exhibited a preference for SSR over ISR in these two cases. Regarding UE sentences, both participant groups exhibited a preference for SSR over ISR, and there was no significant difference between the two groups in this tendency. Additionally, each group of participants showed no discernible difference in paying attention to SSR vs. ISR for NU sentences. The eye tracking data presented a mixed picture of L1 Chinese L2 English learners' ability to process quantifier scope similarly to L1 English speakers, with L2 proficiency playing no significant role.

When it comes to English-speaking learner of Chinese in the eye-tracking task, they converged with L1 Chinese speakers in their processing of quantifier scope in EU, UN, and NU sentences. For EU and UN sentences, both participant groups exhibited a preference for SSR. By contrast, both participant groups exhibited a preference for ISR for NU sentences. Only L1 Chinese speakers exhibited a distinct preference for SSR in the case of UE sentences, indicating that English-speaking learners of Chinese did not demonstrate native-like performance.

When comparing the performance of L1 Chinese L2 English learners in the offline settings (i.e., written TVJT) and online settings (i.e., sentence-picture matching and eye-tracking), the results were overall comparable and consistent. Specifically, L2 English learners' ability to acquire inverse scope interpretations exhibited a structure-dependent pattern. Specifically, they demonstrated a relatively easy acquisition of inverse scope interpretations for UE, NU and UN sentences, which are typically absent in their L1. It is noteworthy that Chinese learners of English exhibited a distinct preference for the ISR in NU sentences. This preference for ISR over the SSR remained consistent across both sentence-picture matching and eye-tracking tasks. This observation suggests a potential transfer of properties from the learners' L1. As previously argued, the availability of the ISR for NU sentences in L1 Chinese may account for this phenomenon. When appropriate contexts are provided, the prominence of ISR for NU sentences becomes even more apparent. This can be observed through the performance of L1 Chinese speakers in both the written TVJT and sentence-picture matching tasks using Chinese stimuli.

Regarding the challenges faced by Chinese-speaking learners of English in acquiring the ISR for EU sentences in online settings, the two explanations that accounted for their performance in the written TVJT also remained applicable in the online context. These explanations are as follows: 1) The positive evidence in the L2 input was not sufficiently robust for learners to incorporate the ISR into their target grammar; 2) the rejection of the ISR for EU sentences can be attributed to the influence of the Single Reference Principle (Kurtzman & MacDonald, 1993). Similar to the findings in the TVJT, the ISR for UN sentences did not pose a significant challenge for L2 learners of English specially measured by eye fixations. This may be attributed to the fact that L2 input offers opportunities for learners to encounter exemplars that demonstrate the possibility of the ISR for UN sentences. This is supported by the results from L1 English speakers in both eye-tracking and sentence-picture tasks.

English-speaking learners of Chinese exhibited consistent patterns of results in both the sentence-picture matching and eye tracking tasks. They successfully acquired the absence of inverse scope interpretations for EU and UE sentences, as well as the presence of inverse scope interpretations for NU sentences. Although English-speaking learners of Chinese did not exhibit a significant difference from native speakers of Chinese in their dispreference for the ISR for UN sentences as measured by eye fixations, L2 Chinese learners showed a stronger preference for the ISR compared to native Chinese speakers as measured in the sentence-picture matching task. This preference might stem from the influence of L2 Chinese learners' native language, which allows for the ISR in UN sentences. Based on the observations in the written TVJT, English-speaking learners of Chinese appeared to face challenges in fully acquiring the absence of inverse scope interpretations for UE sentences (UE on ISR: 3.97 for L1 Chinese vs. 4.76 for L2 Chinese), at least to the extent observed in native Chinese speakers. Although these learners did not differ significantly from L1 Chinese speakers in their dispreference for inverse scope interpretations in UE sentences both in the sentence-picture and eye tracking tasks, there were numerical differences in the choices and fixations of L2 Chinese participants towards pictures representing the ISR. This suggests that the ISR for UE sentences might be somewhat more acceptable to L2 Chinese speakers than to L1 Chinese speakers. Another significant discovery is that Chinese NU sentences allow for inverse scope interpretations. Interestingly, both native Chinese speakers (L1) and non-native Chinese speakers (L2) showed a preference for the inverse scope interpretation. Therefore, the extent to which Chinese permits inverse scope interpretations varies depending on the sentence structure. This flexibility is influenced not only by syntax and semantics but also by additional factors, including discourse contexts and the linear ordering of logical operators in the surface syntax.

7.0 General Discussion

The goal of this dissertation was to investigate the interpretation and processing of quantifier scope in L2 learners of English and Chinese. In the literature, a distinction has been made between English and Chinese regarding their capacity to accommodate quantifier scope ambiguity. It has been observed that English permits both surface scope and inverse scope interpretations when logical operators, such as universal quantifiers, existential quantifiers, and negation, interact in sentences. However, English tends to favor one interpretation over the other due to factors such as processing cost and frequency in the linguistic input. In contrast, Chinese demonstrates a greater rigidity in allowing only the surface scope interpretation. Nevertheless, a few empirical studies present mixed evidence regarding the dominant interpretation in English and the extent to which Chinese permits the inverse scope interpretation (cf. Chung, 2009; Lee, 2010; Musolino & Lidz, 2006; Scontras & Polinsky, 2017; Wu & Ionin, 2019; Zhou & Gao, 2009). The inclusion of data from native speakers as control groups provides novel evidence in this regard.

The cross-linguistic variation in quantifier scope interpretation can potentially result in divergence between L1 and L2 grammars due to the influence of L1 transfer. Furthermore, this cross-linguistic difference points towards the fact that English in general is a wider grammar than Chinese in the case of quantifier scope interpretation. Thus, based on the positive evidence from L2 input, Chinese-speaking learners of English are likely to achieve a target-like interpretation, particularly when it comes to the acquisition of the presence of the inverse scope interpretation, which is not present in their L1. Conversely, English-speaking learners of Chinese may encounter greater challenges in acquiring a target-like interpretation. The learning scenario they face presents a serious learnability problem, namely, how to acquire the absence of the inverse scope interpretation in Chinese, as there is a lack of positive evidence to indicate that inverse scope interpretations are disallowed. As such, the bidirectional nature of this study allowed for examining how L2 input influenced the extent to which L2 learners arrived at target-like quantifier scope interpretation by overcoming these crosslinguistic differences. The literature review suggests that limited research with Chinese-speaking learners of English used untimed offline tasks. Additionally, there is scarce research investigating the interpretation of quantifier scope by English-speaking learners of Chinese. To this end, I conducted offline sentence-interpretation matching, TVJT, and online eye-tracking tasks across three experiments.

In what follows, I will present a summary of major findings from each experiment conducted and discuss them in relation to relevant theoretical accounts, highlighting their implications for second language acquisition research. This chapter concludes with suggestions for future directions in this line of inquiry.

7.1 Major findings and discussion

In Experiment 1, quantifier scope interpretation was examined across various types of quantifier scope (UE (e.g., Every child climbed a tree), EU (e.g., A child climbed every tree), UN (e.g., Every horse didn't jump over the fence), and NU (e.g., The horse didn't jump over every fence)) in the absence of supportive contexts. Although interpretations are not forced by discourse contexts as in Experiment 2, it is crucial to compare interpretation patterns between English and Chinese across conditions. The testing paradigm in which sentences were presented out of contexts indeed is not uncommon for the examination of linguistic domains with a focus on interpretation, such as scalar implicature (e.g., Guasti et al., 2005; Noveck, 2001). It might be the case that a particular interpretation may arise either through the creation of a context on one's own or by considering a broader context based on our world knowledge regarding the likelihood of different interpretations. Results of the sentence-interpretation matching tasks suggested that both English and Chinese speakers exhibited a dominant preference for the Surface Scope Reading (SSR) in their interpretations. However, it is worth noting that this pattern diverged when it came to English UN sentences, where the Inverse Scope Reading (ISR) emerged as the dominant interpretation.

These patterns of results align with two prominent models proposed to explain the mechanism of quantifier scope interpretation. The preference for the SSR (except English Universal-Negation sentences) observed in my study is consistent with the proposal put forth by Anderson (2004), which underscores the role of abstract linguistic structure in computing the scope relationship between logical operators. Native speakers of English were found to endorse UN sentences for the inverse scope interpretation, due to the interplay of syntactic structural computation and scalar implicature calculation. This result is consistent with a recent study by Attali et al. (2021), which also presented UN sentences for consideration without context by asking participants to rate paraphrases of 'every-not' sentences using a sliding scale. In their study, the English L1 participants' ratings indicated that, for most of the target sentences, the paraphrases supporting the ISR were much more likely to be endorsed than the SSR paraphrases, although some sentences showed varying levels of ambiguity. The interpretation of UN sentences in English, which involves the interplay between syntax and pragmatics (Musolino & Lidz, 2006), aligns with the model positing that the processing of quantifier scope relies on the integration of various sources of information, such as syntax and semantics (Filik et al. 2004; Kurtzman & MacDonald 1993).

An important observation is that English and Chinese exhibit distinct patterns in permitting the ISR, and these differences vary depending on the sentence structure. In doubly quantified sentences, English speakers preferred SSR more than Chinese speakers. As previously suggested in this study, when examining negatively quantified sentences with UN word order, the preference for the SSR was found to be more pronounced in English compared to Chinese. A noteworthy observation pertains to Negation-Universal (NU) quantifier sentences, where a higher preference for the ISR was observed in Chinese compared to English. The consistent pattern observed in Chinese NU sentences, as evidenced across experiments, can be attributed to the linguistic properties inherent to these sentence structures in Chinese. Later in the discussion, I will delve into these linguistic properties, highlighting their influence on L2 performance.

Experiments 2 and 3 were set up to investigate the L2 interpretation and processing of quantifier scope in the offline and online settings. Both experiments were conducted with L1 Chinese L2 English and L1 English L2 Chinese learners in a bidirectional design. Native speakers of English and Chinese were respectively included as controls. L2 proficiency and knowledge about English definiteness and Chinese DOU were explored for their potential influences in the L2 patterns of results. Experiments 2 and 3 revealed a significant finding: L2 learners of English and Chinese demonstrated the ability to acquire surface scope interpretations across various structures. This was evident from their high ratings, preference for pictures representing SSR, and increased eye fixations to SSR for pictures. L1 transfer facilitated the acquisition of these interpretations, in line with Schwartz and Sprouse's (1994, 1996) full transfer/full access model of L2 acquisition, which suggests that interlanguage initially involves transferring abstract properties from the L1 grammar. The experiment also found that L1 English L2 Chinese learners increased their acceptance of DQ and UN sentences for the surface scope interpretations as their proficiency increased. The findings suggest that the synergy between L1 (via transfer) and L2 input can aid learners in developing a better understanding of surface scope interpretations.

Chinese-speaking learners of English tended to perform in a more native-like manner in offline interpretation, as measured by TVJT, compared to English-speaking learners of Chinese. Specifically, they are expected to excel in achieving target-like interpretations, particularly in cases of inverse scope interpretation, which often present differences between L1 and L2. The acquisition of a new interpretation, such as in the case of Chinese-speaking learners of English, is generally considered easier than the unlearning of an interpretation that can be transferred from the L1, as in the case of English-speaking learners of Chinese. This conclusion is supported by a substantial body of evidence from studies addressing questions of preemption in SLA (e.g., Grüter et al., 2010; Inagaki, 2001; Juffs, 1996). In this study, Chinese-speaking learners of English indeed demonstrated significant success in acquiring inverse scope interpretations across various types of quantifier scope. Chinese-speaking learners of English successfully acquired inverse scope interpretations, which were generally not present in their L2, primarily based on positive evidence exemplified in the L2 input, even though the L2 input may not provide abundant unambiguous evidence in this regard.

Unambiguous evidence supporting inverse scope interpretations is often insufficient because learners typically encounter contexts that yield surface scope interpretations for quantified sentences. Unless explicitly provided, contexts in favor of inverse scope interpretations are likely limited for L2 learners in the input. In experimental settings, test sentences were assessed within discourse contexts, leading to increased accessibility of inverse scope interpretations, particularly in English due to its grammar allowing for such interpretations, compared to when test sentences were evaluated out of context as shown in Experiment 1. For the same reason, L2 learners may permit/accept inverse scope interpretations more readily when they were exposed to discourse contexts that facilitated accessibility. When examining UN sentences and the ISR in Chinese-speaking learners of English, there appeared to be a trend where L1 speakers exhibited a stronger preference for the ISR ($p = .071$) compared to L2 learners. However, the Bayesian analysis did not provide sufficient evidence to establish this difference as reliable.

For Chinese-speaking learners of English, EU sentences are the only ones that present a distinct challenge, as they deviate significantly from the pattern of obtaining inverse scope interpretations. One plausible explanation for the L2 learners' strong preference against the inverse scope interpretation in EU sentences is Kurtzman and MacDonald's (1993) Single Reference principle. According to this principle, the parser tends to associate an indefinite NP with a single referent, which creates a preference for the non-inverse scope interpretation of EU sentences. The control group, consisting of L1 English speakers, also demonstrated the influence of this principle. They consistently rated surface scope readings higher than inverse scope readings, regardless of whether the sentences were presented with or without contexts, when compared to UE sentences, where the Single Reference principle does not come into play. Likewise, Scontras et al. (2017) observed a similar contrast between ISR and SSR in native English speakers when it came to EU sentences. The negative correlation between definiteness test scores and ISR ratings for EU sentences provides further evidence of the role of this principle in relation to the definiteness property of NPs in the L2 group. As L2 learners became more sensitive to the definiteness property of NPs in the context of EU sentences, the principle gained prominence, resulting in a stronger reluctance to accept inverse scope readings for EU sentences. This principle may have further complicated the influence of L2 input. As L1 English speakers were similarly affected by this principle, it is likely that instances of ISR in the ambient language for EU sentences were infrequent, providing limited exposure for L2 learners compared to the case of ISR in UE sentences.

This study also discovered that English-speaking learners of Chinese generally attained native-like proficiency in judging the lack of acceptability of inverse scope interpretations, despite the presence of L1 influence on their acquisition process. For EU and UN sentences which were rejected for their inverse scope interpretations among native Chinese speakers, English-speaking learners of Chinese had been able to successfully acquire the absence of inverse scope interpretations in these two cases. In both cases, it was found that as the language proficiency of L2 learners increased, they exhibited a greater reluctance to accept inverse scope interpretations. As predicted by Schwartz and Sprouse's (1994, 1996) full transfer/full access model, the findings suggested that English-speaking learners exhibit L1 transfer at low levels of proficiency. It is also possible that learners made more random and indeterminate choices at lower proficiency levels. Furthermore, at higher levels of proficiency, they demonstrate increasing native-like abilities, overcoming L2 poverty of stimulus.

The question arises as to how these learners managed to overcome the learnability problem. Explicit instruction should never be overlooked as a valuable source of evidence in the context of L2 acquisition. I conducted a search through several textbooks, including Wu's (2020) guidebook for L2 learners of Chinese at different proficiency levels, but found no explicit instruction regarding quantifier scope interpretation. However, it is not impossible that instructors may draw students' attention to certain properties when introducing DOU and discussing its interaction with negation. While explicit instruction cannot be entirely discounted, I believe that the success of these learners cannot be solely attributed to this factor. In a similar vein, the acquisition of inverse scope interpretations cannot occur if L2 learners rely solely on positive evidence from L2 input because the L2 lacks exemplars that would indicate the unavailability of such interpretations.

Alternatively, these learners might have accessed an internal mechanism like UG, limiting the necessary interpretive space for Chinese. Access to UG may be guided by specific triggering input that facilitates the acquisition of underlying properties. The positive impact of triggering input on inverse scope interpretations in UN sentences has been observed, wherein learners with higher sensitivity to DOU knowledge are found to be less inclined to accept such interpretations. An additional view posits that L2 learners may employ a Bayesian engine for statistical inference, enabling them to compute relative probabilities of various scope interpretations based on observed data (Hsu et al., 2017; Perfors et al., 2011). This domain-general capacity is not reliant on language-specific innate biases. In Hsu et al. (2017), it was found that the power of this Bayesian engine is constrained by the probability of inferred events' absence, where less probable absences are more salient in guiding inferences from absent data. In other words, when an event is more likely to be inferred as less probable, language learning is more likely to occur through these Bayesian inference capacities. The data patterns for UN and EU sentences are consistent with this approach. For UN sentences, the L2 input provides minimal evidence, evident from the low acceptance of ISR (1.72). Similarly, compared to UE sentences, EU sentences exhibit very low ISRs (1.76). Learners make inferences from absence based on their observation of the L2 input, leading them to generalize that inverse scope interpretations for these structures are unlikely. As a result, they acquire the target interpretations in these learning scenarios.

The interpretation and processing of quantifier scope also provide a unique opportunity to test the Interface Hypothesis (e.g., Tsimpli & Sorace 2006; Sorace 2011, 2012). Based on the reasoning presented by Özçelik (2018), the argument previously was that the acquisition task for

Chinese-speaking learners of English would be more challenging than for English-speaking learners of English due to the additional need to consider pragmatics in interpretation. Unlike L1 English learners of L2 Chinese, when acquiring inverse scope interpretations in English, L1 Chinese learners of L2 English demonstrate an additional involvement of pragmatics, due to the role of the Principle of Charity in determining which of the two possible interpretations to choose. As revealed from the above experimental findings, not different from L2 leaners of Chinese in terms of the level of success achieved in the acquisition task, Chinese-speaking learners of English had no particular difficulty acquiring the presence of inverse scope interpretations in most sentences, with the exception of EU sentences. In this regard, L2 learners do not encounter significant challenges with the external interface, providing no support for the Interface Hypothesis.

The reason behind the failure of English-speaking learners of Chinese to completely reject the inverse scope interpretation for UE sentences remains unclear. This is particularly puzzling since the same group of learners was able to successfully acquire the absence of inverse scope interpretations for EU and UN sentences, despite the lack of relevant evidence in the L2 input. One possibility is that while L1 Chinese speakers rated UE sentences significantly lower on the ISR compared to L2 Chinese speakers, the mean rating given by L1 Chinese speakers for this interpretation was 3.97, which is not considered at the floor level, accounting for more than half of the scale's maximum rating of 7. As a result, this may not constitute a classical learnability problem since the L2 input could potentially include exemplars that indicate the possible presence of ISR for UE sentences. Similarly, Chinese NU sentences for L2 learners should not be considered a classical learnability problem since inverse scope interpretations were entirely possible in the target language (L1 Chinese: ISR for NU being 5.74 out of 7), and learners were able to acquire these interpretations because of the positive evidence available through the L2 input. One implication that arises from UE and NU sentences is the need to empirically establish the L1 baseline, as the existing evidence for it has been scarce in L1 Chinese. This empirical investigation is crucial as it can also provide further substantiation regarding whether an L2 learning scenario truly qualifies as a classical learnability problem. The data presented in this dissertation, particularly from NU sentences, suggests that Chinese, despite being claimed as scope rigid, can in fact permit inverse scope interpretations to some extent, especially when evaluated within supportive discourse contexts. I observed this phenomenon in Experiment 1 (mean rating: 3.36) vs. Experiment 3 (mean rating: 5.74), where the acceptance of ISR in NU sentences among L1 Chinese speakers increased significantly when tested under discourse contexts.

Experiment 3 aimed to provide evidence for how quantifier scope was computed and accessed online by L2 learners of English and Chinese. In contrast to Experiment 2, which assessed quantifier scope interpretation offline, the investigation was conducted online using the VWP, thus minimizing reliance on explicit knowledge. The findings indicated that both L1 native speakers and L2 learners were able to access both surface scope readings and inverse scope readings across different sentence structures. This was evident from the higher likelihood of selecting pictures corresponding to SSRs and ISRs in the sentence-picture matching task, as well as the increased proportion of eye fixations towards pictures representing SSRs and ISRs in the eye fixation data, when false readings were presented visually.

The results of picture selections and eye fixations on visual scenes, featuring pictures representing various possible readings, showed mixed outcomes regarding the convergence of L2 English learners with L1 English native speakers, while L2 Chinese learners with English as their L1 exhibited more consistent convergence with L1 Chinese native speakers. To summarize, the results for Chinese-speaking learners of English in sentence-picture matching and eye-tracking tasks are as follows: In the sentence-picture matching task, except for NU sentences in which L1 and L2 English participants were not different in their preference for the ISR (preferred by both groups), L2 learners of English accessed the ISR to smaller extent than L1 English native speakers in other structures. In the eye-tracking task, while both groups of participants showed similar levels of attention towards the ISR which was the target interpretation for L2 learners to acquire, L2 learners were slightly less inclined to direct their gaze towards pictures representing ISRs compared to L1 English speakers. However, this difference did not reach statistical significance.

The results for English-speaking learners of Chinese in the sentence-picture matching and eye-tracking tasks can be summarized as follows: In the sentence-picture matching task, both L1 Chinese speakers and L2 Chinese learners generally converged in their strong preference for surface scope interpretations in UE, EU, and UN sentences. Additionally, both groups exhibited a bias towards inverse scope interpretations for NU sentences. In the eye-tracking task, English-speaking learners of Chinese exhibited a tendency to align with the interpretation preferences of native Chinese speakers. In general, L2 Chinese participants showed similar patterns to L1 Chinese speakers, displaying a lack of preference for ISR in UN, UE, and EU sentences, while demonstrating a preference for ISR in NU sentences. Notably, L1 Chinese speakers displayed a stronger preference for SSR in UN sentences compared to L2 Chinese learners, and SSR was significantly more favored over ISR by L1 Chinese speakers in UE sentences. Nevertheless, the results from both the sentence-picture matching and eye-tracking tasks demonstrated convergence among L2 learners of Chinese with data from L1 speakers of Chinese.

In both offline and online experiments, both L1 native speakers and L2 learners consistently exhibited a preference for surface scope interpretations for most of the structures in question. This preference is attributed to the lower processing demands associated with such interpretations (Anderson, 2004; O'Grady & Lee, 2006). While a direct comparison between Experiments 2 and 3 is not feasible due to different setups, a notable pattern emerges: L1 English speakers accessed inverse scope interpretations to a lesser extent in Experiment 3 compared to Experiment 2. This pattern specifically applies to English quantifier scope interpretation, where inverse scope interpretations should be accessible, albeit generally less preferred than surface scope interpretations. O'Grady's et al. (2009) processing-based account provides a direct explanation, suggesting that inverse scope readings, which could impose processing burden, are generally less preferred, particularly in time-constrained online tasks. Nonetheless, Chinese-speaking learners of English generally produced results consistent with both Experiment 2 and Experiment 3, particularly when comparing the findings between the Truth Value Judgment Task (TVJT) and the eye-tracking task. Specifically, L2 learners of English obtained the ISR in a manner comparable to that of L1 English speakers.

The findings from Chinese-speaking learners of English suggest that, despite being tested online under time pressure, they exhibited performance similar to that of native English speakers in the case of eye-tracking, particularly regarding inverse scope interpretations. This ability can be attributed to the availability of positive evidence in the input, which informs the accessibility of the less preferred interpretation in L2 English. Only a limited number of studies, such as Chung and Shin (2022), have examined the online processing of quantifier scope interpretation in L2 learners of English with Korean as their L1. These studies revealed a clear inclination among L2 learners to prefer the most economical interpretation when tested online using the self-paced reading paradigm. In contrast, L1 English speakers, who arguably possessed the ability to consult detailed syntactic and semantic information, were capable of accessing both interpretations.

The simultaneous presentation of two pictures in a single visual scene, as discussed in Lohiniva and Panizza (2016), effectively highlighted both surface and inverse scope interpretations as valid responses to the Question Under Discussion. This setup potentially helped L2 learners access inverse scope interpretations, which may have been difficult due to processing factors when using alternative online measures such as self-paced reading, as noted by Chung and Shin (2022). The interpretation of quantifier scope by Chinese-speaking learners of English involves the incorporation of pragmatics, particularly the Principle of Charity. This acquisition task highlights the relevance of external interfaces, as discussed by Sorace (2011). The question of whether specific challenges are encountered by L2 learners of English in acquiring inverse scope interpretations involving the task of handling external interfaces was addressed through the use of eye-tracking data, which directly measures processing (Leal & Hoot, 2022). The eye fixation findings from Chinese-speaking learners of English suggest that inverse scope interpretations can be acquired by these learners through the consideration of the Principle of Charity. This observation is consistent with Özçelik's (2018) argument that L2 learners are capable of incorporating new interpretations into their L2 grammar based on positive evidence present in the L2 input.

A noteworthy pattern that deserves attention is observed in NU sentences (e.g., 'The horse didn't jump over every fence'). It has been consistently found across multiple experiments that the inverse scope interpretation of NU sentences can be accessed and, in some cases, even preferred by both L1 and L2 English participants. The availability of contextual information in the form of written stories or pictures may have facilitated the accessibility of this interpretation, particularly among native speakers of English. L2 learners, who have been observed to access the inverse scope interpretation for such sentences, may have been aided by the presence of evidence in the L2 input that occasionally supports this interpretation. The preference of Chinese-speaking learners of English for the inverse scope interpretation in such sentences may also indicate the influence of L1 transfer in scope interpretation (Lee, 2017; O'Grady, 2013), as L1 Chinese speakers in the TVJT, sentence-picture data, and eye-tracking studies have been observed to access and even prefer the inverse scope interpretation for such sentences.

English-speaking learners of Chinese consistently demonstrated their ability to acquire the interpretation patterns observed in L1 Chinese speakers. They showed a preference for surface scope interpretations in UN, EU, and EU sentences, and for inverse scope interpretations in NU sentences. Despite English allowing both interpretations, these learners were able to acquire the knowledge that Chinese primarily favors the surface scope interpretation for UN, UE, and EU sentences, even in the absence of explicit positive or negative evidence in the instructional contexts. As such, it constitutes evidence that L2 learners can overcome learnability problems in these cases. The role of the triggering input from 'DOU' in revealing the underlying abstract knowledge has been sporadically observed, primarily in offline settings. The acquisition of the inverse scope interpretation for NU sentences by L2 Chinese learners is not surprising, as it does not pose a significant learnability problem. This acquisition can be facilitated by both L1 transfer and exposure to L2 input. The influence of L1 transfer on the acquisition of the absence of interpretation in L2, while not dominant, can be observed. For instance, L1 Chinese speakers exhibited a stronger preference for SSR in UN sentences compared to L2 Chinese learners. This can be attributed to the fact that the L1 of these learners allows for a more extensive acceptance of the inverse scope interpretation for UN sentences, as can be seen in the sentence-picture matching task by L1 English speakers on UN sentences.

A task effect was primarily observed in Chinese-speaking learners of English, specifically in their contrasting performance between eye-fixation and sentence-picture matching tasks. Consequently, the eye-tracking data, rather than the sentence-picture matching data, closely aligns with the TVJT data.

7.2 Limitations and future directions

When drawing conclusions based on a comparison between findings from two learner groups regarding their success in acquiring target-like interpretations, it is crucial to establish the legitimacy of this comparison in terms of tasks and learners. The tasks administered to one group were also used for the other, and the English materials were nearly directly translated into Chinese to accommodate different target learner groups. Thus, the findings are comparable in terms of the similarity of tasks for both groups. However, caution is necessary when comparing findings concerning factors specific to L2 learners, such as proficiency, age of first exposure, and length of instruction. Admittedly, both groups differed significantly in these aspects: L2 English learners were exposed to the language at an average age of 6.4 years, whereas L2 Chinese learners began at 15.4 years old. Additionally, L2 English learners received 15.1 years of instruction, while L2 Chinese learners had only 2.9 years. Moreover, it is likely that Chinese learners of English were more proficient in English than English learners of Chinese were in Chinese. Future research should aim to include learner participants from both groups with comparable learner factors, ensuring that findings from both groups can be reliably compared. Prosody has been identified as an influential factor in scope interpretation among native English speakers (Syrett et al., 2014). However, limited research has been conducted to investigate the independent role of prosody and its interaction with syntax and pragmatics in shaping quantifier scope interpretations among L2 learners. Future studies could examine whether and how prosody might influence L2 quantifier scope interpretation.

Changes to the task setup have been observed to elicit more adult-like behavior from children, resulting in increased endorsement of the utterance (Musolino & Lidz, 2006; Viau et al., 2010). One significant contextual factor hypothesized to influence participants' expectations regarding pragmatics variables related to quantifier scope interpretation is the Question Under Discussion (QUD), which determines the conversational goals (Roberts, 2012). For an utterance to be considered felicitous, it needs to address the QUD, either partially or fully. The role of the QUD has garnered attention in the computation of scalar implicatures in L2 (Starr & Cho, 2022), but its influence on L2 quantifier scope interpretation remains largely unexplored. Future inquiries in this area could investigate the potential impact of the QUD on L2 learners' interpretation of quantifier scope, providing valuable insights into the interplay between pragmatics and semantic interpretation in L2.

Future research could also systematically examine the availability of inverse scope readings in Chinese, shedding light on whether Chinese lacks inverse scope readings (as a scope-rigid language) or if their availability depends on the structural properties of quantifier scope. This issue arises due to the current finding that Chinese speakers can access the inverse scope interpretation for NU sentences, which structurally differ from those used in Zhou and Crain (2009). Additionally, exploring how L2 learners of Chinese perform on different types of NU sentences using various methods remains an intriguing avenue for further investigation.

Furthermore, deriving complex pragmatic reasoning is essential for interpreting sentences involving scalar implicature, such as 'The horse didn't jump over the fence.' In future research, it would be valuable to investigate whether individual differences in the ability to engage in such reasoning could modulate the preference of L2 learners for one interpretation over the other. Previous studies in L1 sentence processing have shown correlations between these individual differences and abilities related to scalar implicature derivation (Nieuwland et al., 2010; Xiang et al., 2013). These studies measured individual differences in implicature derivation abilities using cognitive tasks like the Autism Spectrum Quotient (Baron-Cohen et al., 2001). For future research, it would be valuable to investigate the impact of individual differences on such reasoning among L2 learners.

This study holds the potential to inform language teaching by highlighting the significance of pragmatic contexts in enhancing learning outcomes. The findings demonstrate that L2 learners benefited from pragmatic contexts to better grasp the less accessible scope interpretation. Consequently, language teaching can incorporate pragmatic contexts alongside sentences to support and facilitate accurate semantic interpretations. Processing Instruction (PI) holds promise for teaching quantifier scope, an area that has received little attention in classroom instruction. This method focuses on manipulating input in specific ways to enhance processing strategies and facilitate better language acquisition (VanPatten, 2020). For instance, PI activities may involve presenting students with two pictures—one representing a possible interpretation and the other displaying a false or impossible reading. During this training stage, students receive feedback on their performance. After the training, the learners can be assessed using a written truth-value judgment task to evaluate their understanding of quantifier scope interpretation. To assess retention in long-term memory, a post-delayed test can be conducted one month later to determine if the learners' knowledge of quantifier scope interpretation endures. To establish a direct link between linguistic theory and language teaching, future intervention studies could also investigate the impact of various input and instruction methods on linguistic knowledge representations, specifically regarding semantic interpretation of quantifier scope (Ionin & Montrul, 2023).

7.3 Conclusion

This dissertation used various research methods to explore how L2 learners, whose L1s are Chinese and English, interpret, process, and acquire different types of quantifier scope. Results from Experiment 1 showed that both English and Chinese native speakers generally preferred surface scope interpretations across structures when tested without supportive discourse contexts. However, English speakers accepted inverse scope interpretations to a greater extent than Chinese speakers. This finding is in line with the claim in the literature that Chinese grammar restricts the potential for inverse scope interpretations and avoids scope ambiguity. Experiments 2 and 3 utilized a bidirectional design to investigate how L2 learners of English and Chinese interpret quantifier scope. Experiment 2 employed an offline truth-value judgment task, while Experiment 3 utilized an online visual-world eye-tracking task.

The results of Experiment 2 consistently showed that both English and Chinese L2 learners acquired the surface scope interpretations of the L2. Chinese-speaking learners of English successfully acquired inverse scope readings, influenced by positive L2 input. English-speaking learners of Chinese faced learnability challenges but occasionally acquired ISRs, guided by triggering input. L2 proficiency mainly affected the acquisition of the less preferred ISR in Chinese L2 learners. This dissertation contributes novel evidence on L2 learners scope interpretation (with Chinese and English as L1) of quantified sentences, particularly UE and NU sentences. It explores the influences of triggering input and L2 proficiency on scope interpretation. Furthermore, it sheds new light on the scope interpretation of quantified sentences, specifically DQ and NU sentences, among native Chinese speakers, addressing the limited and conflicting existing evidence.

The results of Experiment 3 suggest that both Chinese-speaking learners of English and English-speaking learners of Chinese could acquire the target interpretation to an extent comparable to L1 native speakers, except for Chinese-speaking learners of English on the sentence-picture matching task. L2 learners of English, despite negative transfer from their L1, can acquire the target interpretation due to positive evidence provided by L2 input. L2 learners of Chinese, while facing learnability challenges, can also acquire the target interpretation, potentially guided by constraints proposed in formal grammar. Notably, not all learning scenarios pose difficulties for L2 learners of Chinese. For instance, in the case of NU sentences, both L1 Chinese speakers and L2 learners can obtain the inverse scope interpretation. This new evidence suggests that scope rigidity in Chinese does not universally apply, and interpretations not available in UG may be facilitated for access by factors such as discourse contexts or experimental setups. Regarding the role of experimental setups, the visual-world paradigm differed from truth-value judgments by simultaneously presenting two pictures for two possible interpretations, thus making them more explicit and salient. This setup may increase the likelihood of participants acquiring the less accessible interpretation, effectively mitigating the potential negative impact of time pressure during testing. This contrasts with Experiment 2, where no time pressure was imposed on the participants.

In summary, this dissertation expands the empirical scope of research on quantifier scope interpretation among L2 learners by adopting a cross-linguistic perspective. Moreover, this project sheds light on the impact of multiple factors, such as L1 transfer, L2 (triggering) input, constraints proposed in formal grammar, and language proficiency, on the development of L2 grammar, specifically in relation to quantifier scope interpretation. Thus, any assertions regarding the nature of second language interpretation and processing must consider the intricate interplay between these linguistic and non-linguistic factors. This project also underscores the significance of employing diverse methodologies for data triangulation, which allows us to capture the linguistic competence of L2 learners more fully.

Appendix A： Experimental Items for Experiment 1

Experiment 1 used the following English items in tests conducted with native English speakers. These English items were translated into Chinese and then modified as needed before being presented to native Chinese speakers. It should be noted that the abbreviations UE (Universal-Existential), EU (Existential-Universal), UN (Universal-Negation), NU (Negation-Universal), and FIL (Filler) were employed.

Appendix Table 1 Items for Experiment 1 with native English speakers

Type	Item	Sentence	Surface	Inverse
		Doubly Quantified		
UE	1a	Every child climbed a tree.	Every child climbed a different tree.	Every child climbed the same tree.
EU	1b	A child climbed every tree.	The same child climbed every tree.	A different child climbed every tree.

UE	2a	Every dog chased a cat.	Every dog chased a different cat.	Every dog chased the same cat.
EU	2b	A dog chased every cat.	The same dog chased every cat.	A different dog chased every cat.
UE	3a	Every shark attacked a sailor.	Every shark attacked a different sailor.	Every shark attacked the same sailor.
EU	3b	A shark attacked every sailor.	The same shark attacked every sailor.	A different shark attacked every sailor.
UE	4a	Every student read a book.	Every student read a different book.	Every student read the same book.
EU	4b	A student read every book.	The same student read every book.	A different student read every book.
UE	5a	Every boy fed a dog.	Every boy fed a different dog.	Every boy fed the same dog.

EU	5b	A boy fed every dog.	The same boy fed every dog.	A different boy fed every dog.
UE	6a	Every girl washed a window.	Every girl washed a different window.	Every girl washed the same window.
EU	6b	A girl washed every window.	The same girl washed every window.	A different girl washed every window.
UE	7a	Every boy climbed a hill.	Every boy climbed a different hill.	Every boy climbed the same hill.
EU	7b	A boy climbed every hill.	The same boy climbed every hill.	A different boy climbed every hill.
UE	8a	Every child opened a present.	Every child opened a different present.	Every child opened the same present.
EU	8b	A child opened every present.	The same child opened every present.	A different child opened every present.

UE	9a	Every nurse examined a patient.	Every nurse examined a different patient.	Every nurse examined the same patient.
EU	9b	A nurse examined every patient.	The same nurse examined every patient.	A different nurse examined every patient.
UE	10a	Every student carried a suitcase.	Every student carried a different suitcase.	Every student carried the same suitcase.
EU	10b	A student carried every suitcase.	The same student carried every suitcase.	A different student carried every suitcase.
UE	11a	Every guest tried a dish.	Every guest tried a different dish.	Every guest tried the same dish.
EU	11b	A guest tried every dish.	The same guest tried every dish.	A different guest tried every dish.

UE	12a	Every boy crossed a bridge.	Every boy crossed a different bridge.	Every boy crossed the same bridge.
EU	12b	A boy crossed every bridge.	The same boy crossed every bridge.	A different boy crossed every bridge.

Negatively Quantified

UN	13a	Every horse didn't jump over the fence.	None of the horses jumped over the fence.	Some of the horses jumped over the fence.
NU	13b	The horse didn't jump over every fence.	The horse jumped over none of the fences.	The horse jumped over some of the fences.
UN	14a	Every child didn't climb up the tree.	None of the children climbed up the tree.	Some of the children climbed up the tree.
NU	14b	The child didn't smash up every cup.	The child smashed up none of the cups.	The child smashed up

				some of the cups.
UN	15a	Every bird didn't fly over the mountain.	None of the birds flew over the mountain.	Some of the birds flew over the mountain.
NU	15b	The bird didn't fly over every mountain.	The bird flew over none of the mountains.	The bird flew over some of the mountains.
UN	16a	Every kid didn't jump into the pool.	None of the kids jumped into the pool.	Some of the kids jumped into the pool.
NU	16b	The man didn't cut down every tree.	The man cut down none of the trees.	The man cut down some of the trees.
UN	17a	Every dog didn't sleep on the bed.	None of the dogs slept on the bed.	Some of the dogs slept on the bed.
NU	17b	The woman didn't take out every box.	The woman took out none of the boxes.	The woman took out some of the boxes.

UN	18a	Every person didn't go across the street.	None of the persons went across the street.	Some of the persons went across the street.
NU	18b	The girl didn't light up every candle.	The girl lit up none of the candles.	The girl lit up some of the candles.
UN	19a	Every woman didn't clean up the table.	None of the women cleaned up the table.	Some of the women cleaned up the table.
NU	19b	The woman didn't clean up every table.	The woman cleaned up none of the tables.	The woman cleaned up some of the tables.
UN	20a	Every student didn't get on the bus.	None of the students got on the bus.	Some of the students got on the bus.
NU	20b	The girl didn't put up every umbrella.	The girl put up none of the umbrellas.	The girl put up some of the umbrellas.

UN	21a	Every boy didn't climb up the ladder.	None of the boys climbed up the ladder.	Some of the boys climbed up the ladder.
NU	21b	The person didn't knock down every desk.	The person knocked down none of the desks.	The person knocked down some of the desks.
UN	22a	Every child didn't swim in the pool.	None of the children swam in the pool.	Some of the children swam in the pool.
NU	22b	The student didn't draw on every blackboard.	The student drew none of the blackboards.	The student drew on some of the blackboards.
UN	23a	Every girl didn't hide behind the door.	None of the girls hid behind the door.	Some of the girls hid behind the door.
NU	23b	The boy didn't pick up every apple.	The boy picked up none of the apples.	The boy picked up some of the apples.

UN	24a	Every child didn't ride on the elephant.	None of the children rode on the elephant.	Some of the children rode on the elephant.
NU	24b	The man didn't open up every book.	The man opened up none of the books.	The man opened up some of the books.

Appendix Table 2 Items for Experiment 1 with native Chinese speakers

Type	Item	Sentence	Surface	Inverse
		Doubly Quantified		
UE	1a	每一个孩子都爬了一棵树。	每一个小孩各自爬了一棵树。	每一个小孩都爬了同一棵树。
EU	1b	有一个孩子爬了每一棵树。	同一个小孩爬了每一棵树。	不同的小孩各自爬了一棵树
UE	2a	每一只狗都追了一只猫。	每一只狗各自追了一只猫。	每一只狗都追了同一只猫。
EU	2b	有一只狗追了每一只猫。	同一只狗追了每一只猫。	不同的狗各自追了一只猫。
UE	3a	每一只鲨鱼都攻击了一个船员。	每一只鲨鱼各自攻击了一个船员。	每一只鲨鱼都攻击了同一个船员。
EU	3b	有一只鲨鱼攻击了每一个船员。	同一只鲨鱼攻击了每一个船员。	不同的鲨鱼各自攻击了一个船员。

UE	4a	每一个学生都读了一本书。	每一个学生各自读了一本书。	每一个学生都读了同一本书。
EU	4b	有一个学生读了每一本书。	同一个学生读了每一本书。	不同的学生各自读了一本书。
UE	5a	每一个男孩都喂了一只狗	每一个男孩各自喂了一只狗。	每一个男孩都喂了同一只狗。
EU	5b	有一个男孩喂了每一只狗。	同一个男孩喂了每一只狗。	不同的男孩各自喂了一只狗。
UE	6a	每一个女孩都擦了一扇窗户。	每一个女孩各自擦了一扇窗户。	每一个女孩都擦了同一扇窗户。
EU	6b	有一个女孩擦了每一扇窗户。	同一个女孩擦了每一扇窗户。	不同的女孩各自擦了一扇窗户。
UE	7a	每一个男孩都爬了一座山。	每一个男孩各自爬了一座山。	每一个男孩都爬了同一座山。

EU	7b	有一个男孩爬了每一座山。	同一个男孩爬了每一座山。	不同的男孩各自爬了一座山。
UE	8a	每一个孩子都打开了一个礼物。	每一个孩子各自打开了一个礼物。	每一个孩子都打开了同一个礼物。
EU	8b	有一个孩子打开了每一个礼物。	同一个孩子打开了每一个礼物。	不同的孩子各自打开了一个礼物。
UE	9a	每一个护士都检查了一个病人。	每一个护士各自检查了一个病人。	每一个护士都检查了同一个病人。
EU	9b	有一个护士检查了每一个病人。	同一个护士检查了每一个病人。	不同的护士各自检查了一个病人。
UE	10a	每一个学生都提了一个箱子。	每一个学生各自提了一个箱子。	每一个学生都提了同一个箱子。
EU	10b	有一个学生提了每一个箱子。	同一个学生提了每一个箱子。	不同的学生各自提了一个箱子。

UE	11a	每一个客人都尝了一道菜	每一个客人各自尝了一道菜。	每一个客人都尝了同一道菜。
EU	11b	有一个客人尝了每一道菜。	同一个客人尝了每一道菜。	不同的客人各自尝了一道菜。
UE	12a	每一个男孩都过了一座桥。	每一个男孩各自过了一座桥。	每一个男孩都过了同一座桥。
EU	12b	有一个男孩过了每一座桥。	同一个男孩过了每一座桥。	不同的男孩各自过了一座桥。

Negatively Quantified

UN	13a	每一匹马都没有跳过篱笆。	没有任何一匹马跳过篱笆。	有一些马跳过了篱笆。
NU	13b	这匹马没有跳过每扇篱笆。	这匹马没有跳过任何一扇篱笆。	这匹马跳过了一些篱笆。
UN	14a	每一个孩子都没有爬上树。	没有任何一个孩子爬上树。	有一些孩子爬上了树。

NU	14b	这个孩子没有打碎每个杯子。	这个孩子没有打碎任何一个杯子。	这个孩子打碎了一些杯子。
UN	15a	每一只鸟都没有飞过山。	没有任何一只鸟飞过山。	有一些鸟飞过了山。
NU	15b	这只鸟没有飞过每座山。	这只鸟没有飞过任何一座山。	这只鸟飞过了一些山。
UN	16a	每一个孩子都没有跳进泳池。	没有任何一个孩子跳进泳池。	有一些孩子跳进了泳池。
NU	16b	这个男人没有砍倒每棵树。	这个男人没有砍倒任何一颗树。	这个男人砍倒了一些树。
UN	17a	每一只狗都没有睡在床上。	没有任何一只狗睡在床上。	有一些狗睡在床上。
NU	17b	这个女人没有拿出每个箱子。	这个女人没有拿出任何一个箱子。	这个女人拿出了一些箱子。

UN	18a	每一个人都没有穿过马路。	没有任何一个人穿过马路。	有一些人穿过了马路。
NU	18b	这个女孩没有点燃每根蜡烛。	这个女孩没有点燃任何一根蜡烛。	这个女孩点燃了一些蜡烛。
UN	19a	每一个女人都没有擦桌子。	没有任何一个女人在擦桌子。	有一些女人在擦桌子。
NU	19b	这个女人没有擦每张桌子。	这个女人没有擦任何一张桌子。	这个女人擦了一些桌子。
UN	20a	每一个学生都没有坐上公交车。	没有任何一个学生坐上公交车。	有一些学生坐上了公交车。
NU	20b	这个女孩没有打起每把伞。	这个女孩没有打起任何一把伞。	这个女孩打起了一些伞。
UN	21a	每一个男孩都没有爬上梯子。	没有任何一个男孩爬上梯子。	有一些男孩爬上了梯子。

NU	21b	这个人没有推倒每张桌子。	这个人没有推倒任何一张桌子。	这个人推倒了一些桌子。
UN	22a	每一个孩子都没有在泳池里游泳。	没有任何一个孩子在泳池里游泳。	有一些孩子在泳池里游泳。
NU	22b	这个学生没有画在每块黑板上。	这个学生没有画在任何一块黑板上。	这个学生画在了一些黑板上。
UN	23a	每一个女孩都没有躲在门后面。	没有任何一个女孩躲在门后面。	有一些女孩躲在门后面。
NU	23b	这个男孩没有捡起每个苹果。	这个男孩没有捡起任何一个苹果。	这个男孩捡起了一些苹果。
UN	24a	每一个孩子都没有坐在大象上。	没有任何一个孩子坐在大象上。	有一些孩子坐在大象上。

NU	24b	这个男人没有打开每本书。	这个男人没有打开任何一本书。	这个男人打开了一些书。

Appendix B： Experimental Items for Experiments 2 & 3

For the written truth-value judgments with native English speakers in Experiment 2, the following English items were employed. These items were directly translated into Chinese and modified wherever necessary. They were then assessed by native Chinese speakers. It is important to note that when presenting the Chinese materials to English-speaking learners of Chinese, *pinyin* was provided above each Chinese character. The abbreviations used are as follows: UE (Universal-Existential), EU (Existential-Universal), UN (Universal-Negation), NU (Negation-Universal), SSR (Surface), ISR (Inverse), and FIL (Filler). The sentence stimuli in Experiment 3 were identical to those used in Experiment 2. However, the presentation format differed: Experiment 2 used written stories as contexts, while Experiment 3 employed pictures as contexts.

Appendix Table 3 Items for Experiments 2 & 3 with L2 English learners

Item	Sentence	Condition	Context
		Doubly Quantified	
1a	Every child climbed a tree.	UE_SSR	At break one day, three kids decided to have a race and see who could climb to the top of a tree the fastest. There were three trees in the playground

			that were the same height and so they decided to each climb a different tree so it would be fair. They counted down and then each kid raced to the top of their tree.
1a	Every child climbed a tree.	UE_ISR	At break one day, three kids decided to have a race and see who could climb to the top of a tree the fastest. There was only one tree in the playground and so they decided they would take turns climbing it and see how long it took for each of them. So, one by one each of the kids took a turn climbing the same tree.
1b	A child climbed every tree.	EU_SSR	In the second grade, there was a boy who loved to climb trees. There were three trees in the playground in a row and one day he decided to see if he could climb them all in one day. So, he climbed them all, one after the other, and he succeeded in climbing all three before break ended.

1b	A child climbed every tree.	EU_ISR	In the second grade, there were three kids who loved to climb trees and there were three trees in the playground. One day they decided to have a race and see who could climb the fastest. They each started at the bottom of a tree, counted down and then each kid raced to the top of their tree.
2a	Every dog chased a cat.	UE_SSR	One day, a group of four dogs found a group of four cats sleeping in the sun. They were very excited because they loved chasing cats. They barked to wake the cats up and then since there were the same number of cats and dogs, each dog was able to chase a different cat.
2a	Every dog chased a cat.	UE_ISR	One day, a group of four dogs came across a cat exploring on his own. They were excited and all of them wanted to chase the cat. But since there was only one cat and four dogs, they decided that they would all chase him together and so they started running, all four dogs chasing a single cat.

2b	A dog chased every cat.	EU_SSR	One day a dog came across three cats playing in the street. She was very excited because she loved chasing cats. She began barking and chased after the first cat. When the first cat hid, she chased the second until it hid and then chased the third. She was very proud because she had chased every cat.
2b	A dog chased every cat.	EU_ISR	One day a pair of dogs came across a pair of cats playing in the street. The dogs were very excited because they loved chasing cats. Since there were two dogs and two cats, each dog decided to chase a different cat and they raced off in opposite directions, each chasing a different cat.
3a	Every shark attacked a sailor.	UE_SSR	One very windy day at sea, two sailors fell overboard from their boat. There were two sharks nearby who were very excited about this because they liked eating sailors. Each shark decided to attack a different sailor and that day they went home very full and happy.

3a	Every shark attacked a sailor.	UE_ISR	One very windy day at sea, a sailor fell overboard from a boat. There were three sharks nearby who were very excited about this because they liked eating sailors. Since there were three sharks and only one sailor, the sharks decided to share. So, all of them attacked the sailor.
3b	A shark attacked every sailor.	EU_SSR	One very windy day at sea, three sailors fell overboard from their boat. There was a shark nearby who was very excited about this because he liked eating sailors. He swam over to the sailors and attacked each of them, one after the other. He went home that day very full and happy.
3b	A shark attacked every sailor.	EU_ISR	One very windy day at sea three, sailors fell overboard from their boat. There were three sharks nearby who were very excited about this because they liked eating sailors. They swam over as fast as they could, and since there were the same

			number of sharks as sailors, a different shark attacked every sailor.
4a	Every student read a book.	UE_SSR	A reading group has three students: Tom, John, and Sally. They gathered yesterday doing some serious reading. Tom read a biology book; John read a geography book; Sally read a physics book. They had a discussion after one-hour reading.
4a	Every student read a book.	UE_ISR	A reading group has three students: Tom, John, and Sally. They gathered yesterday doing some serious reading. They decided to read together their physics textbook. After one-hour reading, they had a discussion.
4b	A student read every book.	EU_SSR	Kris is a college student and likes reading textbooks in her spare time. She is a big fan of science. She has books on biology, chemistry, and physics. This summer, Kris finished reading these books.

4b	A student read every book.	EU_ISR	A reading group has three students: Tom, John, and Sally. They gathered yesterday doing some reading. Tom read a biology book; John read a geography book; Sally read a physics book. They had a discussion after one-hour reading.
5a	Every boy fed a dog.	UE_SSR	Every lunchtime, the mother asked her three sons to feed the dogs: Rocky, Milo, and Gus. The eldest boy fed Rocky; the second eldest boy fed Milo; the youngest boy fed Gus. All of them enjoyed the responsibility of feeding the dogs.
5a	Every boy fed a dog.	UE_ISR	Every lunchtime, the mother asked her three sons to feed the dog named Rocky who has been in the family for 10 years. The three boys work together to feed the dog and they really enjoyed the responsibility of taking care of the dog.
5b	A boy fed every dog.	EU_SSR	Every lunchtime, the mother asked her son Lucas to feed the dogs:

			Rocky, Milo, and Gus. Lucas really enjoyed filling the dogs' bowls, because after then he can play with these dogs for a while.
5b	A boy fed every dog.	EU_ISR	Every lunchtime, the mother asked her three sons to feed the dogs: Rocky, Milo, and Gus. The eldest boy fed Rocky; the second eldest boy fed Milo; the youngest boy fed Gus. All of them enjoyed filling the dogs' bowls.
6a	Every girl washed a window.	UE_SSR	Three girls were hired to wash the display windows in a store. When they arrived, they saw that there were three big windows, one for fruit, one for meat and one for cakes. Each girl decided to clean a window so that they would all have the same amount of work. They got their buckets, and each girl washed a different window.
6a	Every girl washed a window.	UE_ISR	Three girls were hired to wash the display window in a store. When they arrived, they saw that there was one very large window that displayed a bit of

			everything that was in the store. Since it was such a big window, they each took a section, and all worked together to wash the one big window.
6b	A girl washed every window.	EU_SSR	One day, a toy store hired three girls from a cleaning company to clean all their front windows, so that the kids could see the toys better. But that day there was only one girl working because everyone else was out sick. So, she went by herself and washed every window until it was spotless.
6b	A girl washed every window.	EU_ISR	One day, a toy store hired three girls from a cleaning company to clean all their front windows, so that the kids could see the toys better. The store asked for three cleaners because they had three big windows, and they wanted a different cleaner for every window so the work could be done more efficiently.
7a	Every boy climbed a hill.	UE_SSR	In the park, four boys decided to compete climbing hills. To be fair, they chose four different hills of the same

			height. They started to climb at the same time and whoever first reached the top of each hill won.
7a	Every boy climbed a hill.	UE_ISR	In the park, four boys decided to compete climbing hills. To be fair, they chose to climb the same hill. They started to climb at the same time but on different sides of the hill and whoever first reached the top of the hill won.
7b	A boy climbed every hill.	7b_EU_SSR	Climbing hills is one of Oliver's favorite exercises. There are four different hills in the town he lives in. He made up his mind to climb each of the hills this summer. At the end of the summer, he had climbed them all.
7b	A boy climbed every hill.	7b_EU_ISR	In the park, four boys decided to compete climbing hills. To be fair, they chose four different hills at the same height. They started to climb at the same time and whoever first reached the top of each hill won.
8a	Every child opened a present.	UE_SSR	The father bought different Christmas presents for his three children:

			a basketball for Oliver, a toy car for William, and a dress for Mary. Each present was wrapped in paper. They happily opened their own present after dinner.
8a	Every child opened a present.	UE_ISR	The father decided to buy a football for his three boys to share as a Christmas present, because their favorite sport is football. After dinner, the three boys happily opened the present.
8b	A child opened every present.	EU_SSR	The father bought several presents for his daughter's birthday: a dress, a writing tablet, and a selfie camera. Each present was wrapped in paper. After the birthday party, the girl opened each present.
8b	A child opened every present.	EU_ISR	The father bought different Christmas presents for his three children: a basketball for Oliver, a toy car for William, and a dress for Mary. Each present was wrapped in paper. They each happily opened their own present after dinner.

9a	Every nurse examined a patient.	UE_SSR	In one room, there are three patients. Each patient needs to be examined every day. Because each examination takes some time, it takes too long for one nurse to examine them all. Therefore, one nurse is responsible for one patient. This morning, each nurse took about 30 minutes to examine their assigned patient.
9a	Every nurse examined a patient.	UE_ISR	There was one patient who was seriously injured in a car accident. To make sure this patient was thoroughly examined before the operation, three nurses were assigned to examine the patient. That morning, it took them about two hours to finish the examination.
9b	A nurse examined every patient.	EU_SSR	In one room, there are three patients. Each patient needs to be examined every day. One nurse is responsible for this room. This morning, each nurse took about 30 minutes to finish examining all patients.

9b	A nurse examined every patient.	EU_ISR	In one room, there are three patients. Each patient needs to be examined every day. Because each examination takes some time, it takes too long for one nurse to examine all of them. Therefore, one nurse is responsible for one patient. This morning, each nurse took about 30 minutes to examine their assigned patient.
10a	Every student carried a suitcase.	UE_SSR	A group of five students went on vacation together. Because the vacation would take around two weeks, each decided to take a suitcase. When they arrived at the hotel, they carried their own suitcase up the stairs.
10a	Every student carried a suitcase.	UE_ISR	A group of five students went on a short vacation together. They decided to share a big suitcase. When they arrived at the hotel, they carried this big suitcase up the stairs together, because it was quite heavy.

10b	A student carried every suitcase.	EU_SSR	A group of three students went on a short vacation together. Because the vacation would take around two weeks, each decided to take a small suitcase. When they arrived at the hotel, one strong student volunteered to carry all the suitcases up the stairs.
10b	A student carried every suitcase.	EU_ISR	A group of five students went on vacation together. Because the vacation would take around two weeks, each decided to take a suitcase. When they arrived at the hotel, they carried their own suitcase up the stairs.
11a	Every guest tried a dish.	UE_SSR	Every year my family gathers on New Year's Eve with lots of dishes. Usually, based on their own taste, every guest chooses a different dish. Last year, for example, my uncle tried beef with mashed potatoes, my cousin tried macaroni, my grandpa tried broccoli with chicken.
11a	Every guest tried a dish.	UE_ISR	Every year my family gathers on New Year's Eve. Last year, after having

			some drinks, all the guests shared a dish: chicken-noodle soup. Everyone agreed that it tasted really good.
11b	A guest tried every dish.	EU_SSR	Every year my family gathers on New Year's Eve with lots of dishes. One guest came very late and was quite hungry. He tried all the dishes available. According to him, every dish tasted really good.
11b	A guest tried every dish.	EU_ISR	Every year my family gathers on New Year's Eve with lots of dishes. Usually, based on their own taste, every guest chooses a different dish. Last year, for example, my uncle tried beef with mashed potatoes, my cousin tried macaroni, my grandpa tried broccoli with chicken.
12a	Every boy crossed a bridge.	UE_SSR	In the park, there are three bridges over the river. Three boys decided to cross the bridges. Considering that each bridge is too narrow for three all to cross together, each boy chose one bridge and

			reached the other side of the river almost at the same time.
12a	Every boy crossed a bridge.	UE_ISR	In the park, there is a wide bridge over the river. Three boys decided to cross this bridge to reach the other side of the river. They crossed the wide bridge together.
12b	A boy crossed every bridge.	EU_SSR	In the park, there are three bridges over the river. The three bridges are of different types. The boy decided to try these different bridges back and forth. He first crossed the stone bridge. Then he tried the wooden bridge and finally the metal bridge.
12b	A boy crossed every bridge.	EU_ISR	In the park, there are three bridges over the river. Three boys decided to cross the bridges. Consider that each bridge is too narrow for three all to cross together, each boy chose one bridge and reached the other side of the river almost at the same time.

Negatively Quantified

13a	Every horse didn't jump over the fence.	UN_SSR	Three horses looked at a fence in their field. One horse suggested jumping over the fence, but the other horses said the fence was too tall to jump over, so the plan was abandoned.
13a	Every horse didn't jump over the fence.	UN_ ISR	Three horses looked at a fence in their field. One horse suggested jumping over the fence. The first two horses succeeded. The third one, who had hurt its leg the day before, decided not to jump.
13b	The horse didn't jump over every fence.	NU_SSR	One day, a horse decided to test his jumping skills. He found a row of fences and then decided to challenge himself by jumping over all of them. But some of the fences were too high and he only succeeded in jumping over some of the fences.

13b	The horse didn't jump over every fence.	NU_ ISR	One day, a horse decided to test his jumping skills. He found a row of fences and then decided to challenge himself by jumping over all of them. But the fences were too high and so he failed in each attempt.
14a	Every child didn't climb up the trees.	UN_SSR	Every day at break, all the children love to climb trees. But then one day they got to school and found a new sandbox. They were very excited and that day everyone played in the sandbox and no kids climbed the trees.
14a	Every child didn't climb up the trees.	UN_ ISR	Every day at break, all the children love to climb trees. On the first day of school all the kids in the first grade climbed a tree. The next day half the kids wanted to play in the sandbox so only the other half climbed the trees.
14b	The child didn't smash every cup.	NU_SSR	There were three cups for the child to clean. However, the child was very careless. Of the three cups, he smashed two of them when cleaning them.

14b	The child didn’t smash every cup.	NU_ ISR	There were three cups for the child to clean. Usually, the child was very careless. Surprisingly, he didn’t smash one single cup when cleaning them.
15a	Every bird didn’t fly over the mountain.	UN_SSR	Five birds decided to have a contest to see who could fly the highest. They looked all over for a mountain that was tall enough and finally they found one. Then all the birds tried to fly over the mountain, but it was too tall and none of the birds succeeded in flying over it.
15a	Every bird didn’t fly over the mountain.	UN_ ISR	Five birds decided to have a contest to see who could fly the highest. They looked all over for a mountain that was tall enough and finally they found one. Of the five birds, only three of them succeeded in flying over it. The other two failed.
15b	The bird didn’t fly over every mountain.	NU_SSR	One day, a bird decided to see how high she could fly. She found a mountain range and tried to fly over all the mountains. But the mountains were

			very tall, and she only succeeded in flying over some of the mountains.
15b	The bird didn’t fly over every mountain.	NU_ ISR	One day, a bird decided to see how high she could fly. She found a mountain range and tried to fly over all the mountains. But the mountains were too tall to fly over. Unfortunately, she failed at flying over any single mountain.
16a	Every kid didn’t jump into the pool.	UN_SSR	A group of kids had a plan to play with water in the pool. Last weekend, they went to the pool. As the weather turned cooler quickly, they gave up their plan, and no one jumped into the pool.
16a	Every kid didn’t jump into the pool.	UN_ ISR	A group of kids had a plan to play with water in the pool. Last weekend, they went to the pool. As the weather turned cooler quickly, some of them gave up but the rest of them still wanted to have a try so they jumped into the pool.
16b	The man didn’t cut down every tree.	NU_SSR	Today, the man made up his mind to cut down three trees in the yard. He cut down the first tree and it was very easy. Then, he cut down the second tree,

			although it was harder this time. When it came to the third tree, he felt so tired that he gave up this tree.
16b	The man didn’t cut down every tree.	NU_ ISR	Today, the man made up his mind to cut down three trees in the yard. He tried to cut down the first tree and it was too tall to cut down. Then, he tried to cut down the next tree, but it was also too tall to cut down. After, the man tried to cut down the third one. Again, it was too tall so that he gave up.
17a	Every dog didn’t sleep on the bed.	UN_SSR	There are four baby dogs in the family. They are Rocky, Milo, Gus, and Max. Usually, they liked sleeping on the warm bed. But the blanket was being cleaned, so instead they slept on the floor.
17a	Every dog didn’t sleep on the bed.	UN_ ISR	There are four baby dogs in the family: Rocky, Milo, Gus, and Max. Usually, they liked sleeping on the bed, because the bed is warm. But the bed was not big enough for them all to sleep on, so only Rocky and Milo slept on the bed

			this week. Next week, it would be Gus and Max's turn.
17b	The woman didn't take out every box.	NU_SSR	There were five boxes that needed to be taken out from the storeroom for delivery. The woman succeeded taking out three of them. For the other two boxes, she failed because they were too heavy.
17b	The woman didn't take out every box	NU_ ISR	There were five boxes that needed to be taken out of the storeroom for delivery. The woman tried to take out them one by one. But they all were too heavy, so she had no choice but to give up.
18a	Every person didn't go across the street.	UN_SSR	Three people were waiting to go across the street. One of them noticed that the traffic light was not working properly. They decided not to cross the street at that place before the light was fixed.
18a	Every person didn't go across the street.	UN_ ISR	Three people were waiting to go across the street. Two of them crossed the street when the traffic light turned

			green. At the same time, the other person's phone was ringing. So she decided not to cross the street until she finished the phone call.
18b	The girl didn't light every candle.	NU_SSR	Last night Sophia came back home very late. Right after she took a shower, the lights suddenly went out. She found three candles on the table. She took out one candle and lit it. Then she started reading a magazine until she fell asleep.
18b	The girl didn't light every candle.	NU_ ISR	Last night Sophia came back home very late. Right after she took a shower, the lights suddenly went out. She found three candles on the table. However, since she was so tired, she didn't light the candles but went to sleep right away in the dark.
19a	Every woman didn't clean the table.	UN_SSR	Three women were setting up for an event. Before they set up the tables, they were asked to clean them all. But then someone spilled soup and they had

			to clean the floor. They were so busy that they forgot about the tables.
19a	Every woman didn't clean the table.	UN_ ISR	Three women were setting up for an event. They would need to clean the table first. But because the kitchen was so busy, one woman was asked to help in the kitchen. So only two women were left for cleaning the table. Luckily, they were able to handle it well.
19b	The woman didn't clean every table.	NU_SSR	The woman was asked to clean three tables in the room. She spent half an hour cleaning two tables. She was then asked to help in the kitchen. She was so busy that she forgot about the last table left for cleaning.
19b	The woman didn't clean every table.	NU_ ISR	The woman was asked to clean three tables in the room. Before she started cleaning these tables, she was asked to help in the kitchen first. She was so busy that she forgot about the tables and no single table was cleaned.

20a	Every student didn't get on the bus.	UN_SSR	Every Monday is family day in this school. In case some parents may not be able to attend, the school bus last Monday was waiting for the students. In fact, all the students went home with their parents. So, no one left by bus.
20a	Every student didn't get on the bus.	UN_ ISR	After school, all the students were standing in line waiting for the school bus. Among them, one student was being picked up by her parents, so she didn't go home by bus.
20b	The girl didn't put up every umbrella.	NU_SSR	The girl told her friends that she was able to put up three umbrellas at the same time. She tried very hard, but she eventually could only put up two umbrellas at one time.
20b	The girl didn't put up every umbrella.	NU_ ISR	The girl told her friends that she was able to put up three umbrellas at the same time. She tried very hard, and eventually she failed in all three.
21a	Every boy didn't climb up the ladder.	UN_SSR	On a Saturday, one boy suggested climbing the ladder against the wall. His friends said the ladder was too tall when

			they saw it in the yard. The plan was therefore abandoned.
21a	Every boy didn’t climb up the ladder.	UN_ ISR	On a Saturday with three friends, one boy suggested climbing the ladder against the wall. The first two boys tried and succeed. The third one, who thought it was too tall, decided not to climb.
21b	The person didn’t knock down every desk.	NU_SSR	There were three desks in the room. The person was trying to test his strength by knocking down these desks. Only one desk was knocked down because it was not too heavy. The other two remained still although he tried every effort.
21b	The person didn’t knock down every desk.	NU_ ISR	There were three desks in the room. The person was trying to test his strength by knocking down these desks. Although he tried every effort, he couldn’t make knock down even one because these desks were too heavy.
22a	Every child didn’t swim in the pool.	UN_SSR	On a Saturday, four children planned to swim in the pool. When they arrived at the pool, they were surprised to

			find that the pool was open to adults only. They had to give up their plan.
22a	Every child didn’t swim in the pool.	UN_ ISR	On a Saturday, four children planned to swim in the pool. Three of them jumped into the pool and swam happily. The other one thought the water was too cold, so he decided to not swim.
22b	The student didn’t draw on every blackboard.	NU_SSR	The student volunteered to draw blackboard posters for the club. There were four blackboards that he could work on. The student decided to draw on two of them with each for a certain theme. He was praised for his great job.
22b	The student didn’t draw on every blackboard.	NU_ ISR	The student volunteered to draw blackboard posters for the club. There were four blackboards that he could work on. However, he didn’t find any chalk to draw with. So, he decided to draw with a pencil on the paper.
23a	Every girl didn’t hide behind the door.	UN_SSR	Five girls decided to play the hide-and-seek game. Among them, one girl played the role of seeker in the game. The rest of them hid under the bed,

			because they thought that they wouldn’t be easily found by doing so.
23a	Every girl didn’t hide behind the door.	UN_ ISR	Five girls decided to play the hide-and-seek game. Among them, one girl played the role of seeker in the game. Two of them hid under the bed and the other two hid behind the door. Those who hid behind the door were found first.
23b	The boy didn’t pick up every apple.	NU_SSR	In the yard, many fallen apples were scattered on the yard. On one Sunday morning, the boy decided to clear these apples in the yard. He spent one hour picking up most of the apples but decided to not finish, because he got tired of doing this work.
23b	The boy didn’t pick up every apple.	NU_ ISR	In the yard, many fallen apples were scattered on the yard ground. The boy had a plan to pick up all the apples in the yard later in the day. Unfortunately, he injured his leg in a football game, so he had to give up this plan.

24a	Every child didn't ride on the elephant.	UN_SSR	One elephant lived in the zoo and every weekend elephant rides were available to children. Early one morning, three children were waiting for a ride. They tried to climb on but found it was too tall for them. So eventually they gave up.
24a	Every child didn't ride on the elephant.	UN_ ISR	One elephant lived in the zoo and every weekend elephant rides were available to children. Early one morning, three children were waiting for a ride. They tried to climb on but one of them found it was too tall to ride on. So eventually only two of them had the ride.
24b	The man didn't open every book.	NU_SSR	On the table were three books on American history. The man decided to read these books one by one. This afternoon, he only opened up one of them about the civil war and started to read.
24b	The man didn't open every book.	NU_ ISR	On the table were three books on American history. The man decided to read these books next week. Instead, this

afternoon, he was taking a break doing

nothing but drinking a cup of coffee.

Appendix Table 4 Items for Experiments 2 & 3 with L2 Chinese learners

Item	Sentence	Condition	Context
		Doubly Quantified	
1a	每一个孩子都爬了一棵树。	UE_SSR	有一天，三个孩子决定进行一场比赛，看看谁能最快地爬到树顶。操场上有三棵树，它们的高度相同，所以他们决定每人爬一棵不同的树。最后，每个孩子都爬到了树顶。
1a	每一个孩子都爬了一棵树。	UE_ISR	有一天，三个孩子决定进行一场比赛，看看谁能最快地爬到树顶。操场上只有一棵树，所以他们决定轮流爬上它，看看每个人要花多少时间。于是，每个孩子一个接一个地爬同一棵树。

1b	有一个孩子爬了每一棵树。	EU_SSR	在这个学校里，有一个男孩喜欢爬树。操场上有三棵树，有一天他决定看看自己是否能爬上这些树。所以，他一个接一个地爬了所有的树。最后，他成功地爬上了所有三个。
1b	有一个孩子爬了每一棵树。	EU_ISR	在这个学校里， 有三个孩子喜欢爬树。操场上有三棵树。有一天，他们决定进行一场比赛，看看谁爬得最快。最后，每个孩子同时爬到了树顶。
2a	每 一 只 狗 都追了一只猫。	UE_SSR	有一天，四只狗发现四只猫睡在阳光下。他们非常开心，因为他们喜欢追猫。由于猫和狗一样多，一只狗分别追一只猫。

2a	每一只狗都追了一只猫。	UE_ISR	有一天，四只狗遇到了一只猫。他们很兴奋，都想追猫。因为只有一只猫和四只狗，所以他们决定一起追这只猫。于是，他们开始跑，四只狗都追一只猫。
2b	有一只狗追了每一只猫。	EU_SSR	有一天，一只狗在街上看到三只猫在玩。这只狗非常兴奋，因为她喜欢追猫。它开始追赶第一只猫。当第一只猫躲起来时，她开始追第二只猫，第二只猫也躲了起来，然后开始追第三只猫。
2b	有一只狗追了每一只猫。	EU_ISR	有一天，两只狗在街上看到两只猫在玩。这两只狗非常兴奋，因为它们喜欢追猫。因为有两只狗和两只猫，于是，小的狗追小的猫，大的狗追大的猫。

3a	每一只鲨鱼都攻击了一个船员。	UE_SSR	有一天，海上有两个水手从船上掉到水里。附近有两条鲨鱼非常兴奋，因为它们喜欢吃水手。每条鲨鱼决定攻击不同的水手。
3a	每一只鲨鱼都攻击了一个船员。	UE_ISR	有一天，在海上，一名水手从船上落水。附近有三条鲨鱼非常兴奋，因为它们喜欢吃水手。由于有三条鲨鱼，只有一个水手，鲨鱼决定一起攻击水手。
3b	有一只鲨鱼攻击了每一个船员。	EU_SSR	有一天，三名水手从船上落水。附近有一条鲨鱼非常兴奋，因为它喜欢

			吃水手。他游到水手旁边，攻击了他们每个人。
3b	有一只鲨鱼攻击了每一个船员。	EU_ISR	有一天，三名水手从船上落水。附近有三条鲨鱼非常兴奋，因为它们喜欢吃水手。由于鲨鱼的数量与水手相同，不同的鲨鱼袭击了每个水手。
4a	每一个学生都读了一本书。	UE_SSR	一个阅读小组有三个学生：小王、小李和小刘。他们昨天在一起阅读。小王读了一本生物学书; 小李读了一本地理书; 小刘读了一本物理书。他们在一小时的阅读后进行了讨论。
4a	每一个学生都读了一本书。	UE_ISR	一个阅读小组有三个学生：小王、小李和小刘。他们昨天在一起阅

			读。他们决定一起阅读他们的生物教科书。阅读了一个小时后，他们进行了讨论。
4b	有一个学生读了每一本书。	EU_SSR	小王是一名大学生，喜欢在空的时间阅读历史书。他还很喜欢科学。她有关于生物学、化学和物理学的书。今年夏天，小王读完了这些书。
4b	有一个学生读了每一本书。	EU_ISR	一个阅读小组有三个学生：小王、小李和小刘。他们昨天聚在一起读书。小王读 了一本生物学书; 小李读了一本地理书; 小刘读了一本物理书。阅读一小时后， 他们进行了讨论。

5a	每 一 个 男 孩 都喂了一只狗。	UE_SSR	每到午饭时间，妈妈都会让她的三个儿子喂狗：皮皮、球球和胖胖。大儿子喂皮皮; 二儿子喂球球; 最小的儿子喂胖胖。他们都很喜欢喂狗。
5a	每 一 个 男 孩 都喂了一只狗。	UE_ISR	每到午饭时间，妈妈都会让她的三个儿子喂一只名叫皮皮的狗，这只狗已经在这个家生活了 10 年。三个男孩一起喂狗，他们真的很享受喂狗。
5b	有 一 个 男 孩 喂了每一只狗。	EU_SSR	每到午饭时间，妈妈都会让儿子小李喂狗：皮皮、球球和胖胖。小李真的很喜欢把狗的碗装满，因为在那之后他可以和这些狗玩一会儿。
5b	有 一 个 男 孩 喂了每一只狗。	EU_ISR	每到午饭时间，妈妈都会让她的三个儿子喂狗：皮皮、球球和胖胖。大

			儿子喂皮皮; 二儿子喂球球; 最小的儿子喂胖胖。他们都喜欢把狗的碗装满。
6a	每一个女孩都擦了一扇窗户。	UE_SSR	三个女孩被雇来擦一家商店的橱窗。当他们到达时，他们看到有三个大橱窗，一个橱窗是卖水果，一个橱窗是卖肉，一个橱窗是卖蛋糕。每个女孩决定各自擦一扇窗户。他们拿到了水桶，每个清洁工洗了不同的窗户。
6a	每一个女孩都擦了一扇窗户。	UE_ISR	三个女孩被雇来擦商店的橱窗。当他们到达时，他们看到有一个非常大的橱窗，用来展示商店里的所有东西。由于橱窗很大，他们每人负责一部分，一起擦了这扇大橱窗。

6b	有一个女孩擦了每一扇窗户。	EU_SSR	有一天，一家玩具店请了三个女孩来擦他们所有的窗户，这样孩子们就可以更好地看到玩具。但那天有两个女孩生病了。所以，只有一个女孩去擦每一扇窗户。
6b	有一个女孩擦了每一扇窗户。	EU_ISR	三个女孩被雇来擦一家商店的橱窗。当他们到达时，他们看到有三个大橱窗，一个橱窗是卖水果，一个橱窗是卖肉，一个橱窗是卖蛋糕。每个女孩决定各自擦一扇窗户。他们拿到了水桶，每个清洁工洗了不同的窗户。
7a	每一个男孩都爬了一座山。	UE_SSR	在公园里，四个男孩决定参加爬山比赛。他们选择了四个高度一样的

			山。他们同时开始爬，谁先到达每座山顶，谁就赢了。
7a	每一个男孩都爬了一座山。	UE_ISR	在公园里，四个男孩决定参加爬山比赛。他们选择了爬同一座山。他们同时开始爬，但在山的不同侧面，谁先到达每座山顶，谁就赢了。
7b	有一个男孩爬了每一座山。	7b_EU_SSR	爬山是小刘最喜欢的运动之一。他在的小城市有四座不同的山。他下定决心今年夏天要爬上每一座山。夏天结束后，他把所有山都爬了。
7b	有一个男孩爬了每一座山。	7b_EU_ISR	在公园里，四个男孩决定参加爬山比赛。他们选择了四个高度一样的

		山。他们同时开始爬，谁先到达每座山顶，谁就赢了。
8a	UE_SSR	父亲为他的三个孩子买了不同的圣诞礼物，给大儿子买了篮球，给小儿子买了玩具车，给女儿买了衣服。每件礼物都用盒子包着。晚饭后，他们高兴地打开了自己的礼物。
每一个孩子都打开了一个礼物。		
8a	UE_ISR	父亲决定为他的三个儿子买一个足球作为 圣诞礼物分享，因为他们最喜欢的运动是足球。晚饭后，三个男孩高兴地打开了礼物。
每一个孩子都打开了一个礼物。		

8b	有一个孩子打开了每一个礼物。	EU_SSR	父亲为女儿的生日买了几件礼物：一件衣服、一个手写板和一个相机。每件礼物都用盒子包着。 生日聚会结束后，女孩打开了每一份礼物。
8b	有一个孩子打开了每一个礼物。	EU_ISR	父亲为他的三个孩子买了不同的圣诞礼物，给大儿子买了篮球，给小儿子买了玩具车，给女儿买了衣服。每件礼物都用盒子包着。晚饭后，他们高兴地打开了自己的礼物。
9a	每一个护士都检查了一个病人。	UE_SSR	在一个房间里，有三个病人。每个病人每天都需要接受检查。因为每次检查都需要一些时间，所以一个护士检查所有病人需要很长时间。因此，一名护士各自负责一名病人。今天早上，每

			位护士花了大约 30 分钟来检查他们负责的病人。
9a	每一个护士都检查了一个病人。	UE_ISR	有一个男人受了重伤。为了确保该患者在手术前得到彻底检查，指派了三名护士对这个男人进行了检查。 那天早上，他们花了大约两个小时才完成检查。
9b	有一个护士检查了每一个病人。	EU_SSR	在一个房间里，有三个病人。每个患者每天都需要接受检查。一名护士负责这个房间。今天早上，这个护士花了大约 30 分钟来完成对这个房间所有病人的检查。

9b	有一个护士检查了每一个病人。	EU_ISR	在一个房间里，有三个病人。每个病人每天都需要接受检查。因为每次检查都需要一些时间，所以一个护士检查所有病人需要很长时间。因此，一名护士各自负责一名患者。今天早上，每位护士花了大约 30 分钟来检查他们负责的病人。
10a	每一个学生都提了一个箱子。	UE_SSR	五个学生一起去度假。因为假期需要大约两周的时间，所以每个人都决定带一个行李箱。当他们到达酒店时，他们带着自己的行李箱上楼。
10a	每一个学生都提了一个箱子。	UE_ISR	五个学生一起度过了一个短的假期。他们决定共用一个大行李箱。当他

			们到达酒店时，他们一起把这个大行李箱抬上楼，因为它很重。
10b	有一个学生提了每一个箱子。	EU_SSR	三个学生一起度过了一个短暂的假期。因为假期需要大约两周的时间，所以每个人都决定带一个小行李箱。当他们到达酒店时，一名强壮的学生自愿将所有行李箱搬上楼梯。
10b	有一个学生提了每一个箱子。	EU_ISR	五个学生一起去度假。因为假期需要大约两周的时间，所以每个人都决定带一个行李箱。当他们到达酒店时，他们带着自己的行李箱上楼。
11a	每一个客人都尝了一道菜。	UE_SSR	每年过新年，我的家人都会聚在一起，吃很多菜。通常，根据自己的口味，每位客人选择不同的菜。例如，去

			年，我叔叔尝试了牛肉，我的表弟尝试了糖醋鱼，我爷爷尝试了豆腐。
11a	每一个客人都尝了一道菜。	UE_ISR	每年除夕夜，我的家人都会聚会。去年，喝了酒后，所有的客人都分享了一道菜：糖醋鱼。大家都觉得这个才味道真的很好。
11b	有一个客人尝了每一道菜。	EU_SSR	每年除夕夜，我的家人都会聚在一起，吃很多菜。一位客人来得很晚。他很饿。他尝了桌上所有的菜。他说，每道菜的味道都非常好。
11b	有一个客人尝了每一道菜。	EU_ISR	每年除夕夜，我的家人都会聚在一起，吃很多菜。通常，根据自己的口味，每位客人选择不同的菜。例如，去

			年，我叔叔尝试了牛肉，我的表弟尝试了糖醋鱼，我爷爷尝试了豆腐。
12a	每一个男孩都过了一座桥。	UE_SSR	在公园里，河上有三座桥。三个男孩决定过桥。考虑到每座桥都太小了，三个人一起过不去。于是，每个男孩选择一座桥，大家几乎同时到达桥对面。
12a	每一个男孩都过了一座桥。	UE_ISR	在公园里，有一座很大的桥。三个男孩决定穿过这座桥到达河的另一边。于是，他们一起过了这座桥。
12b	有一个男孩过了每一座桥。	EU_SSR	在公园里，河上有三座桥。三座大小不同的桥。男孩决定来回尝试这些

			不同的桥。他首先穿过最小的桥。然后他尝试了最大的桥，最后尝试了中等大小的桥。
12b	有一个男孩过了每一座桥。	EU_ISR	在公园里，河上有三座桥。三个男孩决定过桥。考虑到每座桥都太小了，三个人一起过不去。于是，每个男孩选择一座桥，大家几乎同时到达桥对面。
	Negatively Quantified		
13a	每一匹马都没有跳过篱笆。	UN_SSR	三匹马看着地上的篱笆。一匹马建议跳过篱笆，但其他马说篱笆太高，无法跳过，所以大家都不跳了。

13a	每一匹马都没有跳过篱笆。	UN_ ISR	三匹马看着地上的篱笆。一匹马建议跳过篱笆。前两匹马成功了。第三只，因为伤了腿，决定不跳了。
13b	这匹马没有跳过每扇篱笆。	NU_SSR	有一天，一匹马决定测试他的跳高能力。它找到了一排篱笆，然后决定跳过所有篱笆。但是，有些篱笆太高了，所以它只成功地跳过了一些篱笆。
13b	这匹马没有跳过每扇篱笆。	NU_ ISR	有一天，一匹马决定测试他的跳高能力。他找到了一排篱笆，然后决定跳过所有篱笆。但是篱笆太高了，所以它都没成功。
14a	每一个孩子都没有爬上树。	UN_SSR	在休息时，所有的孩子都喜欢爬树。但今天下雨，他们在教室内玩。没有孩子出去爬树。

14a	每一个孩子都没有爬上树。	UN_ ISR	在休息时，所有的孩子都喜欢爬树。开学第一天，一年级的所有孩子都爬上了一棵树。第二天，一半的孩子在教室里玩，所以只有另一半去爬了树。
14b	这个孩子没有打碎每个杯子。	NU_SSR	这个孩子在洗三个杯子。然而，孩子非常不小心。在洗的过程中，他打碎了其中的两个。
14b	这个孩子没有打碎每个杯子。	NU_ ISR	这个孩子在洗三个杯子。通常，这个孩子非常不小心。但令人感到奇怪的是，他在洗的过程中时没有打碎一个杯子。

15a	每一只鸟都没有飞过山。	UN_SSR	五只鸟决定进行了一场比赛，看谁能飞得最高。他们找到了一座很高的山。所有的鸟都试着飞过这座山，但它太高了，没有一只鸟能飞过它。
15a	每一只鸟都没有飞过山。	UN_ ISR	五只鸟决定举行一场比赛，看谁能飞得最高。他们找到了一座很高的山。在五只鸟中，只有三只成功飞过它。另外两个都失败了。
15b	这只鸟没有飞过每座山。	NU_SSR	有一天，一只鸟决定看看它自己能飞多高。于是，它找到了一些山，并试图飞过所有的山。但是有些山太高了，它只成功飞过了一些山。
15b	这只鸟没有飞过每座山。	NU_ ISR	有一天，一只鸟决定看看它自己能飞多高。于是，它找到了一些山，并试图飞过所有的山。但是这些山太高了，它连一座山都没能飞过去。

16a	每一个孩子都没有跳进泳池。	UN_SSR	一些孩子打算去游泳池里游泳。于是上周末，他们去了游泳池。但是，他们发现水太凉了，于是放弃了这个计划，没有人跳进游泳池。
16a	每一个孩子都没有跳进泳池。	UN_ ISR	一些孩子打算去游泳池里游泳。于是上周末，他们去了游泳池。但是，他们中的一些人感觉水太凉了，所以只有一部分人跳进了游泳池。
16b	这个男人没有砍倒每棵树。	NU_SSR	今天，这个男人下定决心，砍掉院子里的三棵树。他砍倒了第一棵树，这很容易。然后，他砍倒了第二棵树，

			尽管这次更难。 当砍第三棵树的时候，他感觉很累，于是放弃了。
16b	这个男人没有砍倒每棵树。	NU_ ISR	今天，这个男人下定决心，砍掉院子里的三棵树。他试图砍掉第一棵树，它太高了，无法砍倒。然后，他试图砍掉下一棵树，但它也太高了，无法砍倒。之后，该男子试图砍倒第三个。再一次，它太高了，最后他也放弃了。
17a	每一只狗都没有睡在床上。	UN_SSR	家里有四只小狗。他们是皮皮、球球、胖胖和毛毛。 通常，他们喜欢睡在温暖的床上。但是现在床上没有空的地方，于是所以他们睡在地板上。

17a	每一只狗都没有睡在床上。	UN_ ISR	家里有四只小狗。他们是皮皮、球球、胖胖和毛毛。 通常，他们喜欢睡在床上，因为床很暖和。但是床不够大，这周只有皮皮和球球睡在床上。胖胖和毛毛睡在地板上。
17b	这个女人没有拿出每个箱子。	NU_SSR	有五个盒子需要从房间里拿出来。这个女人成功地拿出来三个。另外两个盒子，她没有拿出来，因为它们太重了。
17b	这个女人没有拿出每个箱子。	NU_ ISR	有五个盒子需要从房间里拿出来。这个女人尝试一个接一个地把它们拿出来。但是它们都太重了，所以她只能放弃。

18a	每一个人都没有穿过马路。	UN_SSR	三个人正等着走到街对面。其中一个人注意到红绿灯工作不正常。他们决定在灯修好之前不在那个地方过马路。
18a	每一个人都没有穿过马路。	UN_ ISR	三个人正等着走到街对面。当红绿灯变绿时，其中两人过马路。这时，另外一个人的手机响了。所以她决定在打完电话之前不过马路。
18b	这个女孩没有点燃每根蜡烛。	NU_SSR	昨晚，小李很晚才回家。就在她洗完澡后，灯突然黑了。她在桌子上发现了三根蜡烛。她拿出一支蜡烛点燃。然后她开始看杂志，直到睡着。
18b	这个女孩没有点燃每根蜡烛。	NU_ ISR	昨晚，小李很晚才回家。就在她洗完澡后，灯突然黑了。她在桌子上发

			现了三根蜡烛。但是，由于太累了，她没有点蜡烛，而是在黑暗中入睡了。
19a	每一个女人都没有擦桌子。	UN_SSR	三个女人正在为一个活动做准备。在他们摆好桌子之前，他们被要求擦所有的桌子。但后来有人洒了汤，他们不得不打扫地板。他们太忙了，以至于忘记了擦桌子。
19a	每一个女人都没有擦桌子。	UN_ ISR	三个女人正在为一个活动做准备。他们需要先打扫桌子。但是因为厨房太忙了，一个女人留在了厨房帮忙。所以只剩下两个女人来擦桌子。
19b	这个女人没有擦每张桌子。	NU_SSR	这个女人被要求擦房间里的三张桌子。她花了半个小时擦了两张桌子。

			后来她去厨房帮了忙。最后，她忘了还有一张桌子要擦。
19b	这个女人没有擦每张桌子。	NU_ ISR	这个女人被要求擦房间里的三张桌子。在她开始擦这些桌子之前，她被叫去厨房帮忙。她太忙了，以至于忘记了擦桌子。最后，她一张桌子都没有擦。
20a	每一个学生都没有坐上公交车。	UN_SSR	每周一是这所学校的家庭日。上周一，学校的校车还是在等学生放学。但事实上， 所有的学生都和父母一起回家了。所以，没有人坐公交车离开。
20a	每一个学生都没有坐上公交车。	UN_ ISR	放学后，所有学生都排队等校车。其中，一名学生被父母接走，所以她没有坐公交车回家。

20b	这个女孩没有打起每把伞。	NU_SSR	这个女孩告诉她的朋友，她能够一次打起三把伞。她尝试了下，但最终只能一次打起两把伞。
20b	这个女孩没有打起每把伞。	NU_ ISR	女孩告诉她的朋友，她能够同时打起三把雨伞。他尝试了很多次，但最终都没有成功。
21a	每 一 个 男 孩都没有爬上梯子。	UN_SSR	在一个星期六，一个男孩建议爬上这个梯子。 其他几个男孩说，这个梯子太高了。因此，该计划被放弃。
21a	每 一 个 男 孩都没有爬上梯子。	UN_ ISR	在一个星期六，三个 朋友建议爬上这个梯子。前两个男孩尝试并成功

			了。第三个，认为它太高了，决定不爬了。
21b	这个人没有推倒每张桌子。	NU_SSR	房间里有三张桌子。这个人试图通过推倒这些桌子来测试他的力量。只有一张桌子被推倒，因为它不太重。另外两张桌子没有被推倒，虽然他试了很多次。
21b	这个人没有推倒每张桌子。	NU_ ISR	房间里有三张桌子。这个人试图通过推倒这些桌子来测试他的力量。尽管他尝试了很多次，但一张桌子都没被推倒，因为这些桌子太重了。

22a	每一个孩子都没有在泳池里游泳。	UN_SSR	星期六，四个孩子计划在游泳池里游泳。当他们到达游泳池时，他们发现游泳池只对成年人开放。他们不得不放弃他们的计划。
22a	每一个孩子都没有在泳池里游泳。	UN_ ISR	星期六，四个孩子计划在游泳池里游泳。其中三个人跳进水池，快乐地游泳。另一个人觉得水太冷了，所以他决定不游泳。
22b	这个学生没有画在每块黑板上。	NU_SSR	这个学生为俱乐部画黑板海报。一共有四块黑板可以画。学生决定在其中的两块黑板上画。他因其出色的工作而受到表扬。

22b	这个学生没有画在每块黑板上。	NU_ ISR	这个学生为俱乐部画黑板海报。一共有四块黑板可以画。但是，他没有找到任何可以在黑板上画画的笔。所以，他决定用笔在纸上画画。
23a	每一个女孩都没有躲在门后面。	UN_SSR	五个女孩决定玩捉迷藏游戏。其中，一个女孩在游戏中扮演找人的角色。其余的人躲在床底下，因为他们认为这样做不容易被发现 。
23a	每一个女孩都没有躲在门后面。	UN_ ISR	五个女孩决定玩捉迷藏游戏。其中，一个女孩在游戏中扮演找人的角色。其中两个躲在床底下，另外两个躲在门后。那些躲在门后的人先被发现了。

23b	这个男孩没有捡起每个苹果。	NU_SSR	院子里有许多落下的苹果。.在一个星期天的早上，男孩决定捡起院子里的这些苹果。他花了一个小时捡起大部分苹果，但没有全部捡完，因为他累了。
23b	这个男孩没有捡起每个苹果。	NU_ ISR	院子里有许多落下的苹果。男孩打算在当天晚些时候把院子里所有的苹果都捡起来。 但是，他的腿在一场足球比赛中受伤了，所以他不得不放弃这个计划。
24a	每一个孩子都没有坐在大象上。	UN_SSR	动物园里住着一头大象，每个周末孩子们　都可以去骑大象。一天早上，三个孩子在等着骑大象。他们试图

			爬上去，但发现大象太高了。 所以他们最后放弃了。
24a	每一个孩子都没有坐在大象上。	UN_ ISR	动物园里住着一头大象，每个周末孩子们 都可以去骑大象。一天早上，三个孩子在等着骑大象。他们试图爬上去，但其中一个孩子发现大象太高了，不能骑上去。所以最终只有另外两个孩子骑到了大象。
24b	这个男人没有打开每本书。	NU_SSR	桌子上放着三本关于美国历史的书。虽然男人决定把这些书全读完，但今天下午，他 只 打开了其中一本关于内战的书，开始读。

24b	这 个 男 人 没 有打开每本书。	NU_ ISR	桌子上放着三本关于美国历史的书。这个男人决定下周读这些书。 今天下午，他除了喝了一杯咖啡，其他什么都没做。

Appendix C : Mean eye fixations by L1 and L2 English speakers

Appendix Table 5 Mean eye fixations by L1 and L2 English speakers

		AOIs					
		L1 English			L2 English		
Sentence with interpretations		SSR	ISR	FR	SSR	ISR	FR
UE	SSR_FR	0.43 (0.50)	NA	0.18 (0.39)	0.40 (0.49)	NA	0.15 (0.36)
	ISR_FR	NA	0.32 (0.47)	0.20 (0.41)	NA	0.33 (0.48)	0.19 (0.40)
	SSR_ISR	0.36 (0.48)	0.31 (0.47)	NA	0.29 (0.46)	0.32 (0.47)	NA
EU	SSR_FR	0.24 (0.43)	NA	0.24 (0.43)	0.46 (0.50)	NA	0.24 (0.43)
	ISR_FR	NA	0.47 (0.50)	0.11 (0.32)	NA	0.48 (0.50)	0.13 (0.34)
	SSR_ISR	0.23 (0.42)	0.32 (0.47)	NA	0.36 (0.48)	0.23 (0.43)	NA
UN	SSR_FR	0.44 (0.50)	NA	0.20 (0.40)	0.39 (0.49)	NA	0.23 (0.43)

	ISR_FR	NA	0.40 (0.50)	0.27 (0.44)	NA	0.26 (0.44)	0.26 (0.44)
	SSR_ISR	0.20 (0.40)	0.40 (0.50)	NA	0.33 (0.47)	0.25 (0.44)	NA
	SSR_FR	0.49 (0.51)	NA	0.13 (0.34)	0.34 (0.48)	NA	0.19 (0.40)
NU	ISR_FR	NA	0.27 (0.44)	0.29 (0.46)	NA	0.43 (0.50)	0.22 (0.42)
	SSR_ISR	0.33 (0.48)	0.22 (0.42)	NA	0.19 (0.40)	0.42 (0.50)	NA

Appendix D : Mean eye fixations by L1 and L2 Chinese speakers

Appendix Table 6 Mean eye fixations by L1 and L2 Chinese speakers

		AOIs					
		L1 Chinese			L2 Chinese		
Sentence with interpretations		SSR	ISR	FR	SSR	ISR	FR
UE	SSR_FR	0.59 (0.50)	NA	0.14 (0.35)	0.36 (0.48)	NA	0.29 (0.46)
	ISR_FR	NA	0.55 (0.50)	0.10 (0.30)	NA	0.36 (0.48)	0.21 (0.42)
	SSR_ISR	0.35 (0.48)	0.35 (0.48)	NA	0.29 (0.46)	0.33 (0.48)	NA
EU	SSR_FR	0.52 (0.50)	NA	0.13 (0.33)	0.33 (0.48)	NA	0.14 (0.35)
	ISR_FR	NA	0.42 (0.50)	0.18 (0.39)	NA	0.19 (0.40)	0.24 (0.43)
	SSR_ISR	0.44 ()	0.26 (0.44)	NA	0.48 (0.51)	0.24 (0.43)	NA
UN	SSR_FR	0.57 (0.50)	NA	0.10 (0.31)	0.52 (0.51)	NA	0.12 (0.33)

	ISR_FR	NA	0.43 (0.50)	0.18 (0.39)	NA	0.33 (0.48)	0.36 (0.49)
	SSR_ISR	0.50 (0.51)	0.12 (0.33)	NA	0.38 (0.49)	0.10 (0.30)	NA
	SSR_FR	0.54 (0.50)	NA	0.20 (0.40)	0.33 (0.48)	NA	0.31 (0.47)
NU	ISR_FR	NA	0.43 (0.50)	0.16 (0.37)	NA	0.38 (0.49)	0.24 (0.43)
	SSR_ISR	0.20 (0.41)	0.43 (0.50)	NA	0.31 (0.47)	0.33 (0.48)	NA

Bibliography

Altman, D. G., & Bland, J. M. (1995). Statistics notes: Absence of evidence is not evidence of absence. *Bmj, 311*(7003), 485. https://doi.org/10.1136/bmj.311.7003.485

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of memory and language, 57*(4), 502-518. https://doi.org/10.1016/j.jml.2006.12.004

Ambridge, B., Barak, L., Wonnacott, E., Bannard, C., & Sala, G. (2018). Effects of both preemption and entrenchment in the retreat from verb overgeneralization errors: four reanalyses, an extended replication, and a meta-analytic synthesis. *Collabra: Psychology, 4*(1). https://doi.org/https://doi.org/10.1525/collabra.133

Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity*. Evanston, IL: Northwestern University dissertation.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods, 52*, 388-407.

Aoun, J., & Li, Y. H. A. (1989). Scope and constituency. *Linguistic inquiry, 20*(2), 141-172.

Aoun, J., & Li, Y. H. A. (1993). *Syntax of scope* (Vol. 21). Mit Press.

Apfelbaum, K. S., Klein-Packard, J., & McMurray, B. (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of memory and language, 121*, 104279. https://doi.org/10.1016/j.jml.2021.104279

Attali, N., Scontras, G., & Pearl, L. S. (2021). Pragmatic factors can explain variation in interpretation preferences for quantifier-negation utterances: A computational approach. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

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

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of autism and developmental disorders, 31*, 5-17. https://doi.org/10.1023/a:1005653411471

Bassetti, B., & Lu, M. (2016). Effects of interword spacing on native English readers of Chinese as a second language. *International Review of Applied Linguistics in Language Teaching, 54*(1), 1-22. http://dx.doi.org/10.1515/iral-2016-0014

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed- effects models using lme4. *J. Stat. Softw. 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In A. Szabolcsi (Ed.), *Ways of scope taking* (pp. 71– 107). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-011-5814-5_3

Berwick, R. C., Pietroski, P., Yankama, B., & Chomsky, N. (2011). Poverty of the stimulus revisited. *Cognitive Science, 35*(7), 1207-1242. https://doi.org/10.1111/j.1551-6709.2011.01189.x

Bice, K., & Kroll, J. F. (2021). Grammatical processing in two languages: How individual differences in language experience and cognitive abilities shape comprehension in heritage bilinguals. *Journal of Neurolinguistics, 58*, 100963. https://doi.org/10.1016/j.jneuroling.2020.100963

Bley-Vroman, R., & Masterson, D. (1989). Reaction time as a supplement to grammaticality judgements in the investigation of second language learners' competence. *University of Hawai'i Working Papers in ESL, 8*(2), 207-237.

Boersma, P., & Weenink, D. (2018). Praat: Doing phonetics by computer [Computer program]. Version 6.0.40, retrieved 11 May 2018 from *http://www.praat.org/*

Brooks, P. J., & Sekerina, I. (2006). Shortcuts to quantifier interpretation in children and adults. *Language Acquisition, 13*(3), 177-20. https://www.jstor.org/stable/20462462

Carnie, A. (2021). *Syntax: A generative introduction*. John Wiley & Sons.

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. Amsterdam, the Netherlands: John Benjamins.

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition, 20*(1), 3-16. https://doi.org/10.1017/S1366728915000863

Cheng, L. L. S. (1995). On dou-quantification. *Journal of East Asian Linguistics, 4*(3), 197-234. https://www.jstor.org/stable/20100676

Cheng, L. L. S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic inquiry, 30*(4), 509-542. https://www.jstor.org/stable/4179080

Cheng, L. L. S., Giannakidou, A., Gil, K., Harlow, S., & Tsoulas, G. (2013). The non-uniformity of wh-indeterminates with polarity and free choice in Chinese. *Oxford Studies in Theoretical Linguistics*, 123-151. https://doi.org/10.1093/acprof:oso/9780199692439.003.0007

Cheng, L. L.-S. (1994). Wh-words as polarity items. In *Chinese Language and Linguistics II*, Symposium Series of Institute of History and Philology (pp. 615– 640). Taiwan: Academia Sinica. https://hdl.handle.net/1887/2450

Cho, J., & Slabakova, R. (2014). Interpreting definiteness in a second language without articles: The case of L2 Russian. *Second language research, 30*(2), 159-190. https://www.jstor.org/stable/44280133

Chomsky, N. (1986). *Barriers* (Vol. 13). MIT Press (MA).

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Greenwood Publishing Group.

Chomsky, N. (1992). Some notes on economy of derivation and representation. *Anuario del Seminario de Filología Vasca" Julio de Urquijo"*, 53-82.

Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.

Chu, C. Y., Gabriele, A., & Minai, U. (2014). Acquisition of quantifier scope interpretation by Chinese-speaking learners of English. In *Selected proceedings of the 5th Conference on Generative Approaches to Language Acquisition North America* (pp. 157-168).

Chung, E. S. (2009). The interpretation of English sentences containing quantification and negation by L1-Korean L2-English adult learners. In *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 144-156).

Chung, E. S. (2013). Sources of difficulty in L2 scope judgments. *Second Language Research, 29*(3), 285-310. https://www.jstor.org/stable/43103926

Chung, E. S., & Shin, J. A. (2022). Native and second language processing of quantifier scope ambiguity. *Second Language Research*, 02676583221079741. https://doi.org/10.1177/0267658322107974

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics, 27*(1), 3-42. https://doi.org/10.1017/S0142716406060024

Core Team, R. (2021). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

Crain, S. (2012). *The Emergence of Meaning*. New York, NY: Cambridge University Press.

Crain, S. (2013). What's parsing got to do with it? *Linguistic Approaches to Bilingualism, 3*(3), 301-307. https://doi.org/10.1075/lab.3.3.05cra

Crain, S., & McKee, C. (1985). Acquisition of structural restrictions on anaphora. In *North East Linguistics Society* (Vol. 16, No. 1, p. 8).

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar*. Cambridge, MA: MIT Press.

Crain, S., & Thornton, R. (2000). *Investigations in universal grammar: A guide to experiments on the acquisition of syntax and semantics*. MIT Press.

Crain, S., Thornton, R., Boster, C., Conway, L., Lillo-Martin, D., & Woodams, E. (1996). Quantification without qualification. *Language Acquisition, 5*(2), 83-153. https://www.jstor.org/stable/20011437

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition, 20*(4), 659-678. https://doi.org/10.1017/S1366728916000675

Davidson, D., 1984. *Inquires Into Truth and Interpretation*. Clarendon Press, Oxford.

DeKeyser, R. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning, 55*(Suppl. 1): 1– 25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

Dekydtspotter, L., & Sprouse, R. A. (2001). Mental design and (second) language epistemology: Adjectival restrictions of wh-quantifiers and tense in English-French interlanguage. *Second Language Research, 17*(1), 1-35. https://www.jstor.org/stable/43104601

Dekydtspotter, L., Sprouse, R. A., & Swanson, K. A. (2001). Reflexes of mental architecture in second-language acquisition: The interpretation of combien extractions in English-French interlanguage. *Language Acquisition, 9*(3), 175-227. https://doi.org/10.1207/S15327817LA0903_1

Derrick, D., & Archambault, D. (2010). TreeForm: Explaining and exploring grammar through syntax trees. *Literary and linguistic computing, 25*(1), 53-66. https://doi.org/10.1093/llc/fqp031

Dink, J., & Ferguson, B. (2016). _eyetrackingR. R package version 0.1.6 http://www.eyetracking-R.com.

Dotlačil, J., & Brasoveanu, A. (2015). The manner and time course of updating quantifier scope representations in discourse. *Language, Cognition and Neuroscience, 30*(3), 305-323. https://doi.org/10.1080/23273798.2014.918631

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and cognition, 10*(1), 101-116. https://doi.org/10.1017/S1366728906002847

Dwivedi, V. D. (2013). Interpreting quantifier scope ambiguity: Evidence of heuristic first, algorithmic second processing. *PloS one, 8*(11), e81461. https://doi.org/10.1371/journal.pone.0081461

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition, 24*(2), 143-188. https://doi.org/10.1017/S0272263102002024

Ellis, N. C., & Wulff, S. (2020). Usage-based approaches to L2 acquisition. In *Theories in second language acquisition* (pp. 63-82). Routledge.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in second language acquisition, 27*(2), 141-172. https://doi.org/10.1017/S0272263105050096

Fan, L. (2017). No "Chinese-speaking phase" in Chinese Children's Early Grammar–A Study of the Scope between Negation and Universal Quantification in Mandarin Chinese. *Lingua, 185*, 42-66. https://doi.org/10.1016/j.lingua.2016.07.005

Fang, S., & Xu, Y. (2022). Processing and acquisition of temporality in L2 Mandarin Chinese: Effects of grammatical and lexical aspects. Frontiers in Psychology, 5137. https://doi.org/10.3389/fpsyg.2022.964861

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Feng, S. (2019). The acquisition of English definite noun phrases by Mandarin Chinese speakers. *Studies in Second Language Acquisition, 41*(4), 881-896. https://doi.org/10.1017/S0272263118000323

Filik, R., Paterson, K. B., & Liversedge, S. P. (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin & Review, 11*(5), 953-959. https://doi.org/10.3758/bf03196727

Fodor, J. D. (1982). The mental representation of quantifiers. *Processes, beliefs, and questions,* ed. by Stanley Peters and Esa Saarinen, 129– 64. Dordrecht: Reidel.

Fox, D. (2000). *Economy and semantic interpretation*. Cambridge, Massachusetts: MIT Press.

Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments, II: collective versus distributive readings. *Cognition, 70*(1), 87-104. https://doi.org/10.1016/s0010-0277(99)00002-5

Gabriele, A. (2009). Transfer and transition in the SLA of aspect: A bidirectional study of learners of English and Japanese. *Studies in Second Language Acquisition, 31*(3), 371-402. https://www.jstor.org/stable/44485862

Gan, T., & Tsai, C. Y. E. (2020). The syntax of mandarin dative alternation: An argument from quantifier scope interpretation. *International Journal of Chinese Linguistics, 7*(2), 187-222. https://doi.org/10.1075/ijchl.20006.gan

Gennari, S. P., & MacDonald, M. C. (2006). Acquisition of negation and quantification: Insights from adult production and comprehension. *Language Acquisition, 13*(2), 125-168. https://www.jstor.org/stable/20462458

Gil, K. H., & Marsden, H. (2010). Semantics before syntax: L2 knowledge of anyone by Korean speaking learners. In *Proceedings of the mind-context divide workshop* (pp. 40-51).

Gil, K. H., Marsden, H., & Whong, M. (2019). The meaning of negation in the second language classroom: Evidence from 'any'. *Language Teaching Research, 23*(2), 218-236. https://doi.org/10.1177/136216881774014

Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York, NY: Routledge.

Godfroid, A., Loewen, S., Jung, S., Park, J. H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality judgments measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition, 37*(2), 269-297. https://doi.org/10.1017/S0272263114000850

Grice, H. P. (1975). *Logic and conversation*. In Speech acts (pp. 41-58). Brill.

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grüter, T., Lieberman, M., & Gualmini, A. (2010). Acquiring the scope of disjunction and negation in L2: A bidirectional study of learners of Japanese and English. *Language Acquisition, 17*(3), 127-154. https://doi.org/10.1080/10489223.2010.497403

Gualmini, A., Hulsey, S., Hacquard, V., & Fox, D. (2008). The question–answer requirement for scope assignment. *Natural language semantics, 16*, 205-237. https://doi.org/10.1007/s11050-008-9029-z

Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes, 20*(5), 667-696. https://doi.org/10.1080/01690960444000250

Hara, M. (2007). Input, triggering and poverty of the stimulus in the second language acquisition of Japanese passives. *Second Language Research, 23*(4), 419-458. https://doi.org/10.1177/0267658307080

Heil, J., & López, L. (2020). Acquisition without evidence: English infinitives and poverty of stimulus in adult second language acquisition. *Second Language Research, 36*(4), 415-443. https://doi.org/10.1177/0267658319850

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second language research, 22*(3), 369-397. https://doi.org/10.1191/0267658306sr272

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua, 120*(4), 901-931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second language, 17*(1), 5-27.

Hopp, H., Bail, J., & Jackson, C. N. (2020). Frequency at the syntax–discourse interface: A bidirectional study on fronting options in L1/L2 German and L1/L2 English. *Second Language Research, 36*(1), 65-96. https://doi.org/10.1177/02676583188029

Horn, L. R. (1989). *A natural history of negation*. Stanford, CA: Center for the Study of Language & Information (CSLI), Stanford University.

Hornstein, N., & Lightfoot, D. (1985). Explanation in linguistics. The logical problem of language acquisition. *Tijdschrift Voor Filosofie, 47*(2).

Hoshino, N., Dussias, P. E., & Kroll, J. F. (2010). Processing subject–verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition, 13*(2), 87-98. https://doi.org/10.1017/S1366728909990034

Hsu, A. S., Horng, A., Griffiths, T. L., & Chater, N. (2017). When absence of evidence is evidence of absence: Rational inferences from absent data. *Cognitive science, 41*, 1155-1167. https://doi.org/10.1111/cogs.12356

Huang, C. T. J. (1987). Existential sentences in Chinese and (in) definiteness. *The representation of (in) definiteness*, 226, 253.

Huang, C.-T.J., 1982. *Logical relations in Chinese and the theory of grammar*. Ph.D. Dissertation. MIT.

Huang, J., White, R., & Buscher, G. (2012, May). User see, user point: gaze and cursor alignment in web search. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1341-1350). https://doi.org/10.1145/2207676.2208591

Huang, S. F. (1981). On the scope phenomena of Chinese quantifiers. *Journal of Chinese Linguistics, 9*(2), 226-243. https://www.jstor.org/stable/23753508

Inagaki, S. (1997). Japanese and Chinese learners' acquisition of the narrow‐range rules for the dative alternation in English. *Language Learning, 47*(4), 637-669. https://doi.org/10.1111/0023-8333.00024

Ionin, T. (2021). Acceptability Studies in L2 Populations. In G. Goodall (Ed.), The Cambridge Handbook of Experimental Syntax (Cambridge Handbooks in Language and Linguistics, pp. 373-393). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108569620.015

Ionin, T., & Luchkina, T. (2018). Focus on Russian scope: An experimental investigation of the relationship between quantifier scope, prosody, and information structure. *Linguistic Inquiry, 49*(4), 741-779. https://doi.org/10.1162/ling_a_00288

Ionin, T., & Montrul, S. (2023). *Second Language Acquisition: Introducing Intervention Research*. Cambridge University Press. https://doi.org/10.1017/9781009031585

Ionin, T., Zubizarreta, M. L., & Maldonado, S. B. (2008). Sources of linguistic knowledge in the second language acquisition of English articles. *Lingua, 118*(4), 554-576. https://doi.org/10.1016/j.lingua.2006.11.012

Ionin, T., & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics, 34*, 37-64. https://doi.org/10.1017/S0267190514000026

Ioup, G. L. (1975). *The Treatment of Quantifier Scope in a Transformational Grammar*. City University of New York.

Jeffreys, H. 1939/1961. *Theory of probability*. Oxford: Oxford University Press.

Juffs, A. (1996). Semantics-syntax correspondences in second language acquisition. *Second Language Research, 12*(2), 177-221. https://www.jstor.org/stable/43104511

Juffs, A. (2022). 2 Generative considerations of communicative competence. *Communicative Competence in a Second Language: Theory, Method, and Applications*, 21.

Juffs, A., & Fang, S. (2022). A Generative Approach to the Instructed Second Language Acquisition of Spanish 'se'. *Language Learning, 72*(1), 83-124. https://doi.org/10.1111/lang.12483

Juffs, A., & Fang, S. (2023). Experimental methods for studying second language learners. In P. Gygax & S. Zufferey (Eds.), *The Routledge Handbook in Experimental Linguistics*. New York: Taylor Francis.

Juffs, A., & Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Language learning, 46*(2), 283-323. https://doi.org/10.1111/j.1467-1770.1996.tb01237.x

Juffs, A., & Rodríguez, G. A. (2014). *Second language sentence processing*. New York: Routledge

Kim M (2010) Korean EFL learners' interpretation of quantifier-negation scope interaction in English. *English Language and Linguistics, 16*(1), 164–183.

Kiss, K. É., & Pafel, J. (2017). Quantifier scope ambiguities. *The Wiley Blackwell Companion to Syntax, Second Edition*, 1-36.

Kitagawa, Y. (1994). *Subjects in Japanese and English* (1st ed.). Routledge. https://doi.org/10.4324/9780429401381

Koopman, H., & Sportiche, D. (1991). The position of subjects. *Lingua, 85*(2-3), 211-258. https://doi.org/10.1016/0024-3841(91)90022-W

Kume, K., & Marsden, H. (2021). L2 acquisition of definiteness in Japanese floating numeral quantifiers: Can overt L1 morphology help?. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.20110.kum.

Kuno, S., Takami, K. I., & Wu, Y. (1999). Quantifier scope in English, Chinese, and Japanese. *Language, 75*(1), 63-111. https://www.jstor.org/stable/417474

Kuroda, S. Y. (1988). Whether we agree or not: A comparative syntax of English and Japanese. *Lingvisticae Investigationes, 12*(1), 1-47. https://doi.org/10.1075/li.12.1.02kur

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition, 48*(3), 243-279. https://doi.org/10.1016/0010-0277(93)90042-T

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software, 82*, 1-26. https://doi.org/10.18637/jss.v082.i13

Kwak, H. Y. (2010). Scope interpretation in first and second language acquisition: Numeral quantifiers and negation. *Unpublished doctoral dissertation. University of Hawai'i at Manoa, Honolulu, USA.*

Ladusaw, W. A. (1979). *Polarity Sensitivity as Inherent Scope Relations*. The University of Texas at Austin.

Larson, R., & Wu, H. (2018). Quantifier scope and topicality in Mandarin. In *the 26th Annual Conference of International Association of Chinese Linguistics*.

Larson–Hall, J. (2017). Moving beyond the bar plot and the line graph to create informative and attractive graphics. *The Modern Language Journal, 101*(1), 244-270. https://doi.org/10.1111/modl.12386

Larson‐Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*(S1), 127-159. https://doi.org/10.1111/lang.12115

Lasnik, H. (1972). *Analyses of negation in English* (Doctoral dissertation, Massachusetts Institute of Technology).

Leal, T. (2018). Data analysis and sampling: Methodological issues concerning proficiency in SLA research. In A. Gudmestad, & A. Edmonds (Eds.), *Critical reflections on data in second language acquisition*. Philadelphia, PA: John Benjamins. pp. 63-68. https://doi.org/10.1075/lllt.51.04lea

Leal, T., & Hoot, B. (2022). L2 representation and processing of Spanish focus. *Language Acquisition, 29*(4), 410-440. https://doi.org/10.1080/10489223.2022.2049599

Lee, S (2010). Quantifier scope in English: The view from sentence processing. *English Language and Linguistics 16*: 163–86. https://doi.org/10.17960/ell.2010.16.3.008

Lee, S. (2009). *Interpreting scope ambiguity in first and second language processing: Universal quantifiers and negation*. University of Hawaii at Manoa.

Lee, S. (2010). Interpretation of scope by Korean L2 learners of English: A self-paced reading study. *영어교육, 65*(4), 59-78. https://doi.org/10.15858/engtea.65.4.201012.59

Lee, S. (2017) Adults' online processing of scopally ambiguous sentences. *The Journal of Studies in Language, 33*: 245–64.

Lee, S. (2018). Semantic processing of scope ambiguity by Korean L2 learners of English. *The Journal of Studies in Language 33*: 695–713. https://doi.org/10.1177/02676583221079741

Lee, T.H.-T., 1986. *Studies on quantification in Chinese*. Ph.D. Dissertation. University of California, Los Angeles.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods, 44*, 325-343. https://doi.org/ 10.3758/s13428-011-0146-0

Lenth, R. (2020). Emmeans: Estimated Marginal Means aka Least-Squares Means. *R package Version 1.7.0.* Available online at: https://cran.r-project.org/ package=emmeans (Accessed May 2022).

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. *MIT press*. https://doi.org/10.7551/mitpress/5526.001.0001

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

Li, Y. (2012). What English-speaking Learners of Chinese Don't Know about Dou:--A Study on the Acquisition of '都'. *Journal of Chinese Language Teachers Association, 47*(3): 115-149.

Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist*., 1(1), 333-353. https://doi.org/10.1146/annurev-linguist-030514-125236

Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition, 84*(2), 113-154. https://doi.org/10.1016/S0010-0277(02)00013-6

Linck, J. A., and Cunnings, I. (2015). *The utility and application of mixed-effects models in second language research*. Lang. Learn. 65, 185–207. https://doi:org/10.1111/ lang.12117

Liu, Y. H. (2011, June). *Topic-comment and the Chinese existential construction*. In Proceedings of the 23rd North American Conference on Chinese Linguistics (Vol. 2, pp. 48-65).

Lohiniva, K., & Panizza, D. (2016). *When pragmatics helps syntax: an eye tracking study on scope ambiguity resolution in 4-to 5-year-old children*. In BUCLD 40: Proceedings of the 40th Annual Boston University Conference on Language Development (pp. 216-28). Somerville, MA: Cascadilla Press.

Luque, A., & Morgan-Short, K. (2021). The relationship between cognitive control and second language proficiency. *Journal of Neurolinguistics, 57*, 100956. https://doi.org/10.1016/j.jneuroling.2020.100956

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological review, 101*(4), 676. https://doi.org/10.1037/0033-295X.101.4.676

Marsden, H. (2009). Distributive quantifier scope in English-Japanese and Korean-Japanese interlanguage. *Language Acquisition, 16*(3), 135-177. https://doi.org/ 10.1080/10489220902967135

Marty, P., Chemla, E., & Spector, B. (2015). Phantom readings: The case of modified numerals. Language, *Cognition and Neuroscience, 30*(4), 462-477. https://doi.org/10.1080/23273798.2014.931592

Marty, P., Chemla, E., & Sprouse, J. (2020). The effect of three basic task features on the sensitivity of acceptability judgment tasks. *Glossa: a journal of general linguistics (2016-2021), 5*(1), 72. https://doi.org/10.5334/gjgl.980

May, R. (1985). *Logical form: Its structure and derivation* (Vol. 12). MIT press.

May, R. 1977. *The grammar of quantification*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language, 55*(3), 381-401. https://doi.org/10.1016/j.jml.2006.06.006

McMahon, E. L., Lidz, J., & Pierrehumbert, J. (2004). *Suprasegmental cues to meaning in child-directed speech*. In 17th CUNY Sentence Processing Conference.

Minai, U., Jincho, N., Yamane, N., & Mazuka, R. (2012). What hinders child semantic computation: children's universal quantification and the development of cognitive control. *Journal of Child Language, 39*(5), 919-956. https://doi.org/10.1017/S0305000911000316

Montrul, S. (2001). Agentive verbs of manner of motion in Spanish and English as second languages. *Studies in Second Language Acquisition, 23*(2), 171-206. https://doi.org/10.1017/S0272263101002030

Montrul, S. (2010). Dominant language transfer in adult second language learners and heritage speakers. *Second Language Research, 26*(3), 293-327.

Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition, 13*(4), 449-473. https://doi.org/10.1017/S1366728910000040

Montrul, S., & Slabakova, R. (2003). *Competence similarities between native and near-native speakers: An investigation of the preterite-imperfect contrast in Spanish*. Studies in second language acquisition, 25(3), 351-398.

Morey, R. D., Rouder, J. N., & Jamil, T. (2018). *BayesFactor: Computation of Bayes Factors for common designs.* R package version 0.9. 12-4.2.

Musolino, J. (2006). On the semantics of the subset principle. *Language Learning and Development, 2*(3), 195-218. https://doi.org/10.1207/s15473341lld0203_3

Musolino, J. and Lidz, J. (2006) Why children are not universally successful with quantification. *Linguistics 44*, 4817–4852. https://doi.org/10.1515/LING.2006.026

Musolino, J., Crain, S., & Thornton, R. (2000). *Navigating negative quantificational space. Linguistics, 38* (1), 1–32. https://doi.org/10.1515/ling.38.1.1

Musolino, J., Laity d'Agostino, K., & Piantadosi, S. (2019). Why we should abandon the semantic subset principle. *Language Learning and Development, 15*(1), 32-46. https://doi.org/10.1080/15475441.2018.1499517

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of memory and language, 63*(3), 324-346. https://doi.org/10.1016/j.jml.2010.06.005

Nishigauchi T. 1990. *Quantification in the Theory of Grammar*. Dordrecht, Neth.: Springer.

Noveck, I. A. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition, 78*(2), 165-188. https://doi.org/10.1016/s0010-0277(00)00114-1

O'Grady, W., & Lee, S. (2006). The interpretation of inverse scope in Korean. *Inquiries into Korean Linguistics II*, 369-80.

O'Grady, W. (2013). The illusion of language acquisition. *Approaches to Bilingualism, 3*, 253-85. https://doi.org/10.1075/lab.3.3.01ogr

O'Grady, W. (2022). A calculus for L1 transfer. *In Second Language Acquisition Theory* (pp. 143-176). John Benjamins. https://doi.org/10.1075/bpa.14.08ogr

O'Grady, W., Lee, M., & Kwak, H. Y. (2009). Emergentism and second language acquisition. *The new handbook of second language acquisition*, 69-88.

O'grady, W. (2005). *Syntactic carpentry: An emergentist approach to syntax*. Routledge.

OpenAI, *personal communication*, April 4, 2023.

Orfitelli, R., & Polinsky, M. (2017). When performance masquerades as comprehension: grammaticality judgments in experiments with non-native speakers. *Quantitative approaches to the Russian language, 2017*, 197-214. https://doi.org/10.4324/9781315105048-10

Ouhalla, J. (1990). Sentential negation, relativised minimality and the aspectual status of auxiliaries. *The Linguistic Review, 7*, 183-231. https://doi.org/10.1515/tlir.1990.7.2.183

Ovans, Z., Novick, J., & Huang, Y. (2021). Virtual-world eye-tracking: The efficacy of replicating word processing effects remotely. In *Poster presented at the 34th annual CUNY conference on Human Sentence Processing. Philadelphia, PA. II. J* (Vol. 1).

Özçelik, Ö. (2016). Against Isomorphism and the Maxim of Charity in child language acquisition: Implications for the validity of the TVJT methodology. *Linguistics, 54*(2), 305-337. https://doi.org/10.1515/ling-2016-0003

Özçelik, Ö. (2018). Interface Hypothesis and the L2 acquisition of quantificational scope at the syntax-semantics-pragmatics interface. *Language Acquisition, 25*(2), 213-223. http://dx.doi.org/10.1080/10489223.2016.1273936

Pagliarini, E., Reyes, M. A., Guasti, M. T., Crain, S., & Gavarró, A. (2021). Negative sentences with disjunction in child Catalan. *Language Acquisition, 28*(2), 153-165. http://dx.doi.org/10.1080/10489223.2020.1860055

Pan, H. H. (2006). Focus, tripartite structure, and the semantic interpretation of Mandarin dou. *Research and exploration on grammar, 13*, 163-184.

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). Webgazer: Scalable webcam eye tracking using user interactions. *In Proceedings of the 25th international joint conference on artificial intelligence* (ijcai) (pp. 3839–3845). AAAI. https://par.nsf.gov/servlets/ purl/10024076. Peelle.

Park, S. H., & Kim, H. (2022). Cross-linguistic influence in the second language processing of Korean morphological and syntactic causative constructions. *Linguistic Approaches to Bilingualism, 12*(5), 687-713. http://dx.doi.org/10.1075/lab.20026.par

Paterson, K. B., Filik, R., & Liversedge, S. P. (2008). Competition during the processing of quantifier scope ambiguities: Evidence from eye movements during reading. *Quarterly Journal of Experimental Psychology, 61*(3), 459-473. https://doi.org/10.1080/17470210701255317

Patson, N. D., & Warren, T. (2010). Evidence for distributivity effects in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 782. https://doi.org/10.1037/a0018783

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*(3), 306-338.

Phillips, C., & Ehrenhofer, L. (2015). The role of language processing in language acquisition. *Linguistic approaches to bilingualism, 5*(4), 409-453. https://doi.org/10.1075/LAB.5.4.01PHI

Phillips, S. F. (2022). *Composing Mixed-Language Expressions in the Bilingual Mind* (Doctoral dissertation, New York University).

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning, 64*(4), 878-912. https://doi.org/10.1111/lang.12079

Polinsky, M., & Scontras, G. (2020). Understanding heritage languages. *Bilingualism: Language and Cognition, 23*(1), 4-20. https://doi.org/10.1017/S1366728919000245

Politzer-Ahles, S., Fiorentino, R., Jiang, X., & Zhou, X. (2013). Distinct neural correlates for pragmatic and semantic meaning processing: An event-related potential investigation of scalar implicature processing using picture-sentence verification. *Brain research, 1490*, 134-152. https://doi.org/10.1016/j.brainres.2012.10.042

Pollock, J. Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic inquiry, 20*(3), 365-424. https://www.jstor.org/stable/4178634

Portner, P. H. (2005). *What is meaning: Fundamentals of formal semantics*.

Rankin, T., & Unsworth, S. (2016). Beyond poverty: Engaging with input in generative SLA. *Second Language Research, 32*(4), 563-572. https://doi.org/10.1177/0267658316648732

Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and philosophy*, 335-397. https://doi.org/10.1023/A:1005349801431

Reinhart, T. M. (1976). *The syntactic domain of anaphora* (Doctoral dissertation, Massachusetts Institute of Technology).

Romano, F., & Guijarro-Fuentes, P. (2023). Task effects and the yes-bias in heritage language bilingualism. *International Journal of Bilingual Education and Bilingualism*, 1-21. https://doi.org//10.1080/13670050.2023.2206949

Rothman, J., & Iverson, M. (2008). Poverty-of-the-stimulus and SLA epistemology: Considering L2 knowledge of aspectual phrasal semantics. *Language Acquisition, 15*(4), 270-314. https://doi.org/10.1080/10489220802352206

Rothman, J., & Slabakova, R. (2011). The mind-context divide: On acquisition at the linguistic interfaces. *Lingua, 121*(4), 568-576. https://doi.org/10.1016/j.lingua.2011.01.003

Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in second language acquisition, 40*(2), 417-442. https://doi.org/10.1017/S0272263117000134

Rutherford, W. E. (1989). Preemption and the learning of L2 grammars. *Studies in Second Language Acquisition, 11*(4), 441-458.

Sæbø, K. J. (2001). The semantics of Scandinavian free choice items. *Linguistics and philosophy, 24*(6), 737-788. https://doi.org/10.1023/A:1012788916366

Schütze, C. T., & Sprouse, J. (2013). Judgment data. *Research methods in linguistics*, 27-50.

Schwartz, B. D., & Sprouse, R. (1994). Word order and nominative case in nonnative language acquisition: a longitudinal study of (L1 Turkish) German interlanguage. *Language acquisition studies in generative grammar, 31*(4), 71-89. https://doi.org/10.1075/lald.8.14sch

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second language research, 12*(1), 40-72. https://doi.org/10.1177/026765839601200103

Schwartz, B. D., & Sprouse, R. A. (2000). When syntactic theories evolve: Consequences for L2 acquisition research. *Second language acquisition and linguistic theory*, 156-186. https://doi.org/10.1017/S0267190510000097

Schwartz, B. D., & Sprouse, R. A. (2013). Generative approaches and the poverty of the stimulus. *The Cambridge handbook of second language acquisition*, 137-158. https://doi:10.1017/CBO9781139051729.010

Scontras, G., & Pearl, L. S. (2021). When pragmatics matters more for truth-value judgments: An investigation of quantifier scope ambiguity. *Glossa: a journal of general linguistics, 6*(1). https://doi.org/10.16995/glossa.5724

Scontras, G., Polinsky, M., Tsai, C. Y. E., & Mai, K. (2017). Cross-linguistic scope ambiguity: When two systems meet. *Glossa: A journal of general linguistics, 2*(1), 1-28. https://doi.org/10.5334/gjgl.198

Scontras, G., Tsai, C. Y. E., Mai, K., & Polinsky, M. (2014). Chinese scope: An experimental investigation. In *Proceedings of Sinn und Bedeutung* (Vol. 18, pp. 396-414).

Sekerina, I. A., & Sauermann, A. (2015). Visual attention and quantifier-spreading in heritage Russian bilinguals. *Second language research, 31*(1), 75-104. https://doi.org/10.1177/0267658314537292

Sekerina, I. A., & Sauermann, A. (2017). Quantifier spreading in child eye movements: A case of the Russian quantifier kazhdyj 'every'. *Glossa: a journal of general linguistics, 2*(1). https://doi.org/10.5334/gjgl.109

Sekerina, I. A., Brooks, P. J., Campanelli, L., & Schwartz, A. M. (2018). Quantifier Spreading in School-Age Children: An Eye-Tracking Study. *In Linguistic and Cognitive Aspects of Quantification* (pp. 171-192). Springer, Cham. https://doi.org/10.1007/978-3-319-91566-1_8

Shen, J., & Chen, Y. (2022). Simplified grammar in both languages? On scope assignment in Q-Neg sentences in English-dominant heritage Chinese speakers. *Proceedings of the Linguistic Society of America, 7*(1), 5228. https://doi.org/10.3765/plsa.v7i1.5228

Shimada, H., & Goro, T. (2021). On the source of children's conjunctive interpretation of disjunction: Scope, strengthening, or both?. *Language Acquisition, 28*(2), 98-130. https://doi.org/10.1080/10489223.2020.1844477

Singmann, H., and Kellen, D. (2019). An introduction to mixed models for experimental psychology. *New Methods Cogn. Psychol. 28*, 4–31. https://doi.org/10.4324/9780429318405-2

Slabakova, R. (2010). Scalar implicatures in second language acquisition. *Lingua, 120*(10), 2444-2462. https://doi.org/10.1016/j.lingua.2009.06.005

Slabakova, R. (2013). L2 semantics. In *The Routledge handbook of second language acquisition* (pp. 145-164). Routledge.

Slabakova, R. (2016). Second language acquisition. *Oxford University Press*.

Slabakova, R., Leal, T. L., & Liskin-Gasparro, J. (2014). We have moved on: Current concepts and positions in generative SLA. *Applied Linguistics, 35*(5), 601-606. https://doi.org/10.1093/applin/amu027

Song, H. S., & Schwartz, B. D. (2009). Testing the Fundamental Difference Hypothesis: L2 adult, L2 child, and L1 child comparisons in the acquisition of Korean wh-constructions with negative polarity items. *Studies in Second Language Acquisition, 31*(2), 323-361. https://doi.org/10.1017/S0272263109090329

Spinner, P., & Gass, S. M. (2019). *Using judgments in second language acquisition research*. London: Routledge.

Sprouse, J. (2018). Acceptability judgments and grammaticality, prospects and challenges. *Syntactic structures after, 60*, 195-224. https://doi.org/10.1515/9781501506925-199

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience, 31*(6), 817-840. https://doi.org/10.1080/23273798.2016.1161806

Starr, G., & Cho, J. (2022). QUD sensitivity in the computation of scalar implicatures in second language acquisition. *Language Acquisition, 29*(2), 182-197. https://doi.org/10.1080/10489223.2021.1990930

Stowell, T. (1985). Null antecedents and proper government. In *North East Linguistics Society* (Vol. 16, No. 1, p. 32).

Syrett, K., Simon, G., & Nisula, K. (2014). Prosodic disambiguation of scopally ambiguous quantificational sentences in a discourse context1. *Journal of Linguistics, 50*(2), 453-493. https://doi.org/10.1017/S0022226714000012

Szendroi, K. (2022). (Quantifier) *Scope Judgments*. Oxford University Press.

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye movements and spoken language comprehension. In *Handbook of psycholinguistics* (pp. 863-900). Academic Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634. https://doi.org/ 10.1126/science.7777863

Tang, S. W. (2001). On QU-features. *UCI working papers in linguistics, 7*, 191-204.

Thornton, R. 2017. The truth value judgment task: An update. In Nakayama, Mineharu & Su, Yi-Ching & Huang, Aijun (eds.), *Studies in Chinese and Japanese language acquisition: In honor of Stephen Crain, 13–40*. John Benjamins Publishing Company. https://doi.org/10.1075/ lald.60.02tho

Thornton, R., Nakayama, M., Su, Y., & Huang, A. (2017). The truth value judgment task. *Studies in Japanese and Chinese Language Acquisition: In honor of Stephen Crain*, 13-39.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in second language acquisition, 27*(2), 173-204. https://doi.org/10.1017/S0272263105050102

Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space, and time: A review of the role of cross-language similarity in L2 (morpho) syntactic processing as revealed by fMRI and ERP methods. *Studies in Second Language Acquisition, 33*(1), 91-125. https://doi.org/10.1017/S0272263110000549

Trahey, M., & White, L. (1993). Positive evidence and preemption in the second language classroom. *Studies in second language acquisition, 15*(2), 181-204. https://doi.org/10.1017/S0272263100011955

Truscott, J., & Sharwood Smith, M. (2011). Input, intake, and consciousness. The Quest for a Theoretical Foundation. *Studies in Second Language Acquisition, 33*(4), 497-528. https://doi.org/10.1017/S0272263111000295

Tsai, E., Scontras, G., Mai, K., & Polinsky, M. (2014). Prohibiting inverse scope: An experimental study of Chinese vs. English. *Empirical issues in syntax and semantics* 10:305-322.

Tsimpli, L., & Sorace, A. (2006). Differentiating interfaces: L2 performance in syntax-semantics and syntax-discourse phenomena. In *30th Annual Boston-University Conference on Language Development* (pp. 653-664). Cascadilla Press.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62*, 148-171. https://doi.org/10.1111/j.1467-9922.2012.00710.x

VanPatten, B. (2020). Input processing in adult L2 acquisition. In *Theories in second language acquisition* (pp. 105-127). Routledge. https://doi.org/10.1075/sibil.62.01lee

Viau, J., Lidz, J., & Musolino, J. (2010). Priming of abstract logical representations in 4-year-olds. *Language Acquisition, 17*(1-2), 26-50. https://doi.org/10.1080/10489221003620946

Vos, M., Minor, S., & Ramchand, G. C. (2022). Comparing infrared and webcam eye tracking in the Visual World Paradigm. *Glossa Psycholinguist*. 1, 131. https://doi.org/10.5070/G6011131

Wang, T. (2022). *Examining the Role of Subjecthood and Pronominalization in Reference Resolution in L1 and L2 Processing: An Eye-Tracking Study* (Doctoral dissertation, University of Kansas).

White, L. (2003). *Second language acquisition and universal grammar*. Cambridge University Press.

White, L. (2011). Second language acquisition at the interfaces. *Lingua, 121*(4), 577-590. https://doi.org/10.1016/j.lingua.2010.05.005

White, L. (2022). Should linguistics be applied and, if so, how?. *Language Teaching*, 1-13. https://doi.org/10.1017/S0261444822000313

White, L., Bruhn-Garavito, J., Kawasaki, T., Pater, J., & Prévost, P. (1997). The researcher gave the subject a test about himself: Problems of ambiguity and preference in the investigation of reflexive binding. *Language Learning, 47*(1), 145-172. https://doi.org/10.1191/026765898675690231

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wu, E. L. (2020). *Practical Chinese Usage: Common Words and Phrases*. Routledge.

Wu, M. J., & Ionin, T. (2020). INTONATIONAL EFFECTS ON ENGLISH SCOPALLYAMBIGUOUS SENTENCES. *Ilha do Desterro, 73*, 13-36.

Wu, M. J., & Ionin, T. (2022). Does explicit instruction affect L2 linguistic competence? An examination with L2 acquisition of English inverse scope. *Second Language Research, 38*(3), 607-637. https://doi.org/10.1177/0267658321992830

Wu, M. J., & Ionin, T. (2022, August). L1-Mandarin L2-English learners' acquisition of English double-quantifier scope. In *Generative SLA in the Age of Minimalism. Features, interfaces, and beyond. Selected proceedings of the 15th Generative Approaches to Second Language Acquisition Conference* (pp. 93-114). John Benjamins Publishing Company. https://doi.org/10.1075/lald.67.04wu

Wu, M. J., Ionin, T., Brown, M. M., & Dailey, B. (2019). L1-Mandarin L2-English speakers' acquisition of English universal quantifier-negation scope. In Proceedings of the 43rd Annual Boston University Conference on Language Development (pp. 716-729). https://doi.org/10.1075/lald.67.04wu

Xiang, Y. (2016). Mandarin particle dou: A pre-exhaustification exhaustifier. *Empirical issues in syntax and semantics*, 11.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in psychology, 4*, 708. https://doi.org/10.3389/fpsyg.2013.00708

Xu, L., & Yuan, B. (2022). Dependency resolutions of null and overt subjects in English speakers' L2 Chinese: Evidence for the cue-based model. *Second Language Research*. https://doi.org/10.1177/02676583221135189

Yang, X., & Wu, Y. (2020). On the scope of quantifier phrases in Chinese passive construction. *International Journal of Chinese Linguistics,* *7*(1), 71-89. https://doi.org/10.1075/ijchl.19010.yan

Yuan, B. (2010). Domain-wide or variable-dependent vulnerability of the semantics—syntax interface in L2 acquisition? Evidence from wh-words used as existential polarity words in L2 Chinese grammars. *Second Language Research,* *26*(2), 219-260. https://doi.org/10.1177/0267658309349421

Zhou, P., & Crain, S. (2009). Scope assignment in child language: Evidence from the acquisition of Chinese. *Lingua, 119*(7), 973-988. https://doi.org/10.1016/j.lingua.2009.01.001

Zhou, P., & Crain, S. (2011). Children's knowledge of the quantifier dou in Mandarin Chinese. *Journal of psycholinguistic research, 40*, 155-176. https://doi.org/10.1007/s10936-010-9161-z

Zhou, P., & Gao, L. (2009). Scope processing in Chinese. *Journal of Psycholinguistic Research,* *38*(1), 11-24. https://doi.org/10.1007/s10936-008-9079-x

Zyzik, E. (2009). The role of input revisited: Nativist versus usage-based models. L2 Journal, 1(1). https://doi.org/10.5070/l2.v1i1.9056