Research Article

Shaohua Fang* and Zhiyi Wu

Syntactic prediction in L2 learners: evidence from English disjunction processing

https://doi.org/10.1515/iral-2021-0223
Received December 23, 2021; accepted August 12, 2022; published online August 26, 2022

Abstract: This study examined the extent to which second language (L2) learners' syntactic prediction resembles or differs from native speakers' and the role of L2 proficiency in this process. An experimental group of 135 Chinese learners of English and a control group of 58 English native speakers participated in a phrase-by-phrase self-paced reading task where the experimental sentences were contrastive in whether the DP disjunction was preceded by *either* or not. To ensure that the participants had sufficient knowledge about the target stimuli, they were asked to additionally judge the acceptability of the experimental sentences. The results showed that like native speakers, L2 learners read the critical region consisting of "or" and a DP disjunct faster when it was preceded by *either* compared to when *either* was absent. In addition, this effect of the presence versus absence of *either* spilled over to the post-critical region. Moreover, L2 proficiency was not found to robustly modulate this process. It is concluded that L2 learners, like native speakers, can make predictions at the level of syntax.

Keywords: disjunction; L2 learners; processing; proficiency; syntactic prediction

1 Introduction

The human brain has been claimed to operate as an intricate predictive device (Clark 2013). In language use and learning, people often predict what comes next before the actual input becomes available. For example, when one encounters the verb *drink* in a sentence (e.g., *The boy will drink …*), they would expect something drinkable to appear in the upcoming speech or text. Meanwhile, a

***Corresponding author: Shaohua Fang**, Department of Linguistics, University of Pittsburgh, 4200 Fifth Avenue, Pittsburgh, PA, 15260, USA, E-mail: shf64@pitt.edu
Zhiyi Wu, Graduate Program in Second Language Acquisition, School of Languages, Literatures, and Cultures, 3215 Jiménez Hall, University of Maryland, College Park, MD, 20742, USA, E-mail: zhiyiw1@umd.edu

noun (e.g., coffee, soda), following *drink*, would be the most likely grammatical category, among others, to be anticipated. Ample psycholinguistic evidence has shown that language listeners and readers make predictions based on already processed linguistic information from different sources such as lexical semantics (e.g., Altmann and Kamide 1999), morphosyntax (e.g., Van Berkum et al. 2005), phonology (e.g., DeLong et al. 2005), and discourse (Kaiser and Trueswell 2004). Moreover, it is assumed that the representations for the expected materials are more highly activated than those for the less expected materials, thus requiring less cognitive effort, signaled by the reduced reading times that one spends on the expected materials relative to the unexpected ones (Roland et al. 2012).

In second language (L2) studies, however, to what extent non-native speakers (NNSs) resemble or differ from native speakers (NSs) in making predictions remains a topic of recent debate (Grüter et al. 2017; Kaan 2014). Grüter et al. (2017) proposed the *Reduced Ability to Generate Expectations* (RAGE) hypothesis, according to which L2 learners, relative to NSs, essentially generate expectations to a lesser extent compared to NSs during language comprehension. A considerable body of research on L2 prediction has focused on gender agreement in Romance languages such as Spanish and French (Dussias et al. 2013; Grüter et al. 2012; Lew-Williams and Fernald 2010). In Spanish, for example, a noun and its preceding modifiers, such as determiners and adjectives, are gender-marked. Grüter et al. (2012) showed that even highly proficient Spanish NNSs, unlike Spanish NSs, failed to use gender information to anticipate upcoming nouns. Hopp (2015) also found that NNSs across different proficiency levels used information from lexical semantics only, rather than from morphosyntax, to make predictions. Both representation-based and processing-based accounts could provide some explanation for the discrepancy between native and non-native predictive processing. According to the representation-based account, syntactic representations built up by late L2 learners tend to be shallower and less detailed than those by NSs (Clahsen and Felser 2006), and hence may not be used by learners to make predictions, especially in the domain of syntax. On the processing-based account, L2 prediction has been argued to be not fully automatic (Ito and Pickering 2021), because NNSs may be slow in accessing lexical information during processing to construct representations necessary for prediction (Hopp 2018; Ivanova and Costa 2008).

Although NNSs may predict in a non-native manner, Kaan (2014) argued that L1 and L2 predictive processing are underlyingly the same and that potential differences in predictive processing can be accounted for by mediating factors that influence sentence processing in general. Among these factors, L2 proficiency is perhaps the most often pinpointed factor that could potentially modulate the extent to which NNSs engage in predictive processing, as reflected by the proposal of Kaan and her associates that NNSs with low proficiency may struggle with prediction but may eventually become indistinguishable from NSs in their ability to make predictions, as their language proficiency reaches a certain level (Kaan et al. 2010). In line with this proposal, a few studies have shown that L2 proficiency modulated predictive processing such that more proficient learners are more likely to generate predictions compared to those with lower proficiency (e.g., Chambers and Cooke 2009; Dussias et al. 2013; Henry et al. 2022; Hopp 2013; Leal et al. 2017).

However, another recent body of research targeting various linguistic phenomena yielded a rather different picture where the modulatory effect of L2 proficiency has not been detected (Dijkgraaf et al. 2017; Hopp 2015; Ito et al. 2018; Kim and Grüter 2021; Mitsugi 2021). There seem to be two possible scenarios for the absence of the L2 proficiency effect – 'prediction does not automatically come with global proficiency and, vice versa, that high overall proficiency does not imply L1-like prediction', as claimed by Kaan and Grüter (2021). Specifically, one scenario would be that NNSs engage in making predictions regardless of proficiency (e.g., Kim and Grüter 2021; Mitsugi 2021). For example, Mitsugi (2021) took advantage of the structural dependency between negative polarity items (e.g., *zenzen* 'at all') and negation in Japanese to examine whether Japanese NNSs could use information from negative polarity adverbs to predict negation. In a visual-world eye-tracking experiment, all L2 participants, like NSs, were found to make more eye movements towards the picture depicting a negative reading when they heard the adverb compared to when they did not, suggesting they were able to employ such a predictive cue during processing. Interestingly, such a native-like prediction was not modulated by their L2 proficiency. The other scenario would be that NNSs do not necessarily engage in prediction like NSs, regardless of proficiency (e.g., Grüter and Rohde 2021; Hopp 2015). For example, Hopp (2015) found that even advanced English learners of German failed to use the highly reliable morphosyntactic cues represented by case markers to facilitate prediction during L2 sentence processing.

There could be many reasons for the mixed findings regarding the effects of proficiency on L2 prediction. One reason for the lack of L2 proficiency effect is that variance in proficiency among the investigated sample was too limited to detect the relationship between L2 prediction and proficiency that would otherwise be captured by a sample with a wider proficiency range if the effect were present. To fix this problem, a larger sample targeting a wider proficiency range could be adopted such that the chance of detecting the effect of proficiency, if present, could increase. It would also allow us to take a closer look at the L2 development of processing skills and ultimately language competence. Another possible reason stems from the practice of artificially dividing participants into different proficiency groups (e.g., 'beginner', 'intermediate', 'advanced') based on a set of cutoff points, which may misrepresent learner proficiency, thus leading to reduced statistical power and sometimes spurious effects (see Leal 2018 for a critical review). Moreover, complex linguistic computation as a result of combining different cues for prediction may complicate the proficiency effect on L2 predictive processing. For example, unlike Japanese NSs (Kamide et al. 2003), Japanese NNSs were found to have failed to predict upcoming noun phrases (NPs) based on case markers (Mitsugi and MacWhinney 2016). The authors argued that the experimental task might be too challenging for the participants, as it required them to integrate both the syntactic and semantic information to make the intended prediction. By contrast, the role of L2 proficiency, which has been well documented in learners during integrative processing, appears to be rather consistent (see Roberts 2012 for a comprehensive review). For example, only highly proficient learners were found to be able to process case markers for them to be integrated into the constructed L2 representations (e.g., Hopp 2006, 2010) and recover from structural misanalysis due to garden-path effects during L2 sentence processing (Jackson 2008).

It is important to note that, with a few exceptions (e.g., Kim and Grüter 2021; Mitsugi 2021), studies on L2 prediction have been typically conducted with learners learning or experiencing an L2 in a naturalistic context (e.g., Grüter and Rohde 2021; Henry et al. 2022; Hopp 2013). Recent research has shown that the *type* of language exposure, i.e., naturalistic versus classroom, also plays some role in shaping L2 incremental and predictive processing (Leal et al. 2017; Pliatsikas and Marinis 2013). Therefore, it would be informative to see whether predictive patterns observed among learners with naturalistic exposure can be extended to those among learners with classroom exposure. In addition, L2 prediction research has primarily focused on how learners predict in the context of morphosyntax (e.g., grammatical gender, case marking) (e.g., Dussias et al. 2013; Grüter and Rohde 2021; Mitsugi and Macwhinney 2016) and lexical semantics (e.g., Chambers and Cooke 2009; Ito et al. 2018; Trenkic et al. 2014). To our knowledge, there have been very few studies directly investigating the prediction of syntactic structures in L2 learners. Among the very few studies, one was conducted by Kaan et al. (2016), who examined the extent to which NNSs may differ from NSs in their use of the preceding syntactic context to anticipate an elided noun. In this study, researchers recorded event-related potentials (ERPs) from advanced NNSs and NSs when they were reading sentences that allow and do not allow noun-ellipsis, as in (1a) and (1b) respectively.

(1) a. Ellipsis: Although Peter met John's surgeon, he did not meet Max's *of the operation.
    b. Non-ellipsis: Although the surgeon met John, he did not meet Max's *of the operation.

The results showed a greater posterior positivity for the Non-ellipsis condition compared to the Ellipsis condition at possessives only in the native group, suggesting that NNSs did not predict upcoming syntactic information to the same extent as NSs. However, this study only included advanced learners and did not examine the role of language proficiency in L2 prediction.

Research on L2 syntactic prediction is of particular importance, especially because syntactic information has a special status in framing some influential L2 sentence processing models, such as the Shallow Structure Hypothesis (Clahsen and Felser 2006), in that whether NNSs can build up structural representations to the same extent as NSs constitutes the hallmark for successful L2 competence development. It is, therefore, crucial to expand the scope of research on L2 prediction by examining a wider variety of linguistic properties. The present study contributes to the literature by investigating the L2 processing of another syntactic structure which involves determiner phrase (DP) disjunction. In the following section, we review some linguistic facts about English disjunction and previous empirical research pertaining to this phenomenon.

2 Disjunction and its processing

The syntactic phenomenon under investigation is disjunction in English. Consider the examples in (2a) and (2b), both of which contain disjunction. In (2a), it is assumed that 'the big house' and 'the old car', the two independent disjuncts, are coordinated with the disjunctor *or*. The two disjuncts are of the same linguistic category, namely DP (e.g., Chomsky 2009; Partee and Rooth 2012). Moreover, *either* appears adjacent to the disjunction phrase 'the big house or the old car' and thus marks the scope of this phrase (see Quine 1967; Sag et al. 1985; Stockwell et al. 1973).

(2) a. Jay painted either the big house or the old car for his family over the summer.
b. Jay painted the big house or the old car for his family over the summer.

Several studies used the *either-or* construction as a tool to investigate the extent to which the language processor is predictive at the level of pure syntax (e.g., Staub and Clifton 2006; Staub 2007; Yoshida et al. 2014; Warren et al. 2016). For example, building upon Frazier and Clifton (2001), Staub and Clifton (2006) manipulated the presence of *either* before a DP disjunction in a sentence (e.g., 'The team took the train or the subway to get to the game.') for an eye-tracking experiment testing English NSs. They found that participants spent less time on the DP region (e.g., 'or the subway') if *either* appeared previously, compared to when *either* was absent in the preceding text. They interpreted it as evidence of syntactic prediction where readers predicted that a DP disjunction would arrive later given the presence of *either*. Such predictions have facilitated sentence comprehension, as reflected in the RT differences in regions of interest between sentences with *either* and sentences without *either*. Similar processing behaviors have been observed in Mandarin NSs for their processing of Chinese constructions involving DP-disjunction (Chen and Yan 2012).

Similarly, facilitation associated with the processing of other parallel structures, such as conjunction (e.g., 'John saw a strange man and a tall woman …'), has also been argued to result from such structures being highly predictable (Frazier et al. 2000; Sturt et al. 2010). As such, prediction seems to be crucial for language comprehension, particularly when it comes to explaining why NSs process language input so fast and successfully in a highly constraining linguistic context. However, no consensus has been reached regarding whether prediction holds to the same extent across populations. As stated by Federmeier (2007: 495), 'predictive processing may not be the best—or even a viable—strategy for all individuals at all phases of the lifespan and/or in all processing situations'. Using a self-paced reading task (SPR), Warren et al. (2016), who also manipulated the DP disjunction for the presence/absence of *either* as in Staub and Clifton (2006), examined if syntactic prediction observed in young neurotypical adults would extend to people with aphasia and older neurotypical individuals. The main finding is that both participants with aphasia and older neurotypical participants performed similarly to those neurotypical participants in the control group. They all read the second DP disjunct beginning with *or* faster when it was preceded by *either*, compared to when *either* was absent in the preceding context. This finding, therefore, suggests that like neurotypical adults, people with aphasia and older neurotypical individuals can also engage in structural prediction cued by the lexical item *either*.

Although NSs and NNSs differ in many aspects, such as language proficiency and language experience, we may find that NNSs could predict syntax in a way similar to NSs if the cue used for prediction is reliable and salient, as in the case of *either-or* construction. If so, the current study could enrich our understanding of the relationship between L1 and L2 sentence processing. To our knowledge, no investigation has been conducted to examine the predictive processing of disjunction patterned as V either ([DP1] or [DP2]) among NNSs. The present study, therefore, aims to fill this gap and explore whether NNSs engage in syntactic prediction—and if so, how—by using this construction as a test window.

3 The present study

The present study specifically addresses the following research questions:

1. Do Chinese learners of English predict a disjunction consisting of two DPs based on the lexical item *either* to the same extent as English NSs?
2. What is the role of language proficiency in modulating the predictive processing behavior, if present, among NNSs?

4 Method

4.1 Participants

The study included two groups of participants consisting of 135 Chinese learners of English (107 female, *M* = 20.7, *SD* = 2.6) and 58 English NSs (42 female, *M* = 19.9 years of age, *SD* = 2.6). All NS participants were undergraduate students at a university in northeast U.S.A. and were given extra course credit for their participation. All NNSs were undergraduate or graduate students from several universities in Mainland China at the time of testing.[1] These NNS participants reported to have been monolingually raised and only had exposure to formal instruction in English in classroom settings. None had the experience of living in any English-speaking countries or was fluent in a language other than Chinese or English. They were paid for their participation. All participants had normal or corrected-to-normal vision and reported no history of language deficits. Table 1 summarizes NNSs' demographic information. Proficiency was measured using the LexTALE English Test (Lemhöfer and Broersma 2012). LexTALE has been shown to be highly correlated with some standard proficiency measures such as the Quick Placement Test (QPT), a test for measuring learners' general English proficiency. With reference to the Common European Framework (CEF) for language levels, learners are distinguished between intermediate (or lower) (score equal to 59 or below), upper intermediate (score of 60–80), and advanced (score of 80–00) levels depending on their performance in the LexTALE task. NNSs additionally self-rated on four of the English skills (speaking, listening, reading, and writing) on a 10-point Likert Scale. As shown in Figure 1, the proficiency of the sampled NNSs reflects a broad range and exhibits sufficient variability spanning from the low proficiency level to the advanced proficiency level.

1 One NNS had been resident in the US for two years at the time of testing. To avoid the potential confound of exposure type, this participant was excluded from the statistical modeling.

Table 1: NNSs' demographic information.

	NNS (*n* = 135)	
	M (*SD*)	Range
Age (years)	20.7 (2.6)	17–32
Age of first exposure	7.9 (2.2)	3–13
Instruction length	12.8 (3.1)	7–22
Self-ratings (1–10)		
Speaking	5.5 (1.7)	1–9
Listening	7.6 (1.9)	1–10
Reading	6.4 (1.5)	1–9
Writing	5.4 (1.5)	1–9
LexTALE (0–100)	69.4 (13.6)	43.8–100
Gender (male/female)	28/107	

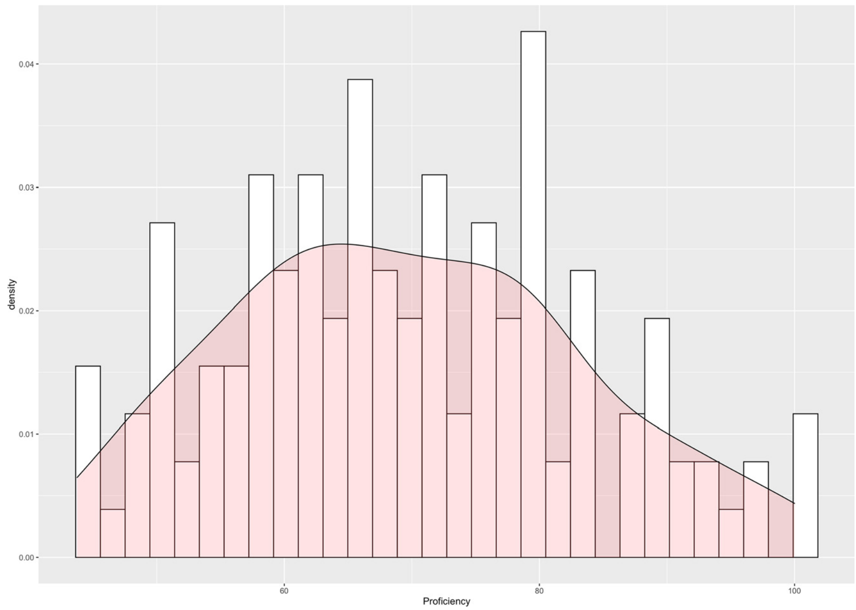

Figure 1: Distribution of LexTALE scores (range: 43.8–100) within the NNS group.

4.2 Experimental design and stimuli

Target stimuli for the SPR experiment were 20 pairs of sentences like those in (1a) and (1b). Most stimuli were adapted from Staub and Clifton (2006) and Warren et al. (2016) in a way that words presumably unfamiliar to NNSs were replaced with words that should be more readily accessible to NNSs. Each sentence was divided into four different regions: factor manipulating region, pre-critical region, critical region, and post-critical region. For demonstration, example sentences in (2) are repeated here as in (3).

(3) a. Jay painted either | the big house | or the old car
[factor manipulating region] [pre-critical region] [critical region]
| for his family over the summer.
[post-critical region]
b. Jay painted | the big house | or the old car
[factor manipulating region] [pre-critical region] [critical region]
| for his family over the summer.
[post-critical region]

(4) a. The Australian woman | saw the famous doctor | had been drinking | quite a lot.
b. The tall scientists | revealed that the valuable coins | had gone missing | from the site.

Twenty sets of experimental sentences in two conditions were used in this experiment. Each set contained two items (one item per condition) following the format as in (3). (3a) and (3b) in pair were identical except that (3a) involved *either* and (3b) did not. Each sentence had a transitive verb followed by two NPs joined by *or*. The critical region included a three-word DP or PP, together with *or*. Experimental effects, if present, were expected to emerge in this region. The experimental items were distributed across two lists in a Latin Square design such that each list contained only one item from each pair and each participant saw 10 items from each of the two conditions. Another 57 distractor items of unrelated syntactic structures from other experiments but in comparable sentence length and structural complexity were included in each list. (4a) and (4b) are example filler items.[2] Each participant thus read a total of 77 sentences in the SPR task. The experimental sentences used for the acceptability judgment task were the same as those in the SPR task. Items within each task were completely randomized, and lists from tasks were counterbalanced across participants to prevent participants from being exposed to identical items across tasks. We included six additional grammatically unacceptable catch trials in the acceptability judgment task to detect how well they understood and did the task.

2 Fillers items, adapted from Sturt et al. (1999), were for other unrelated experiments.

4.3 Procedure

We administered the following tasks sequentially: (a) Language background questionnaire; (b) Self-paced reading; (c) Independent English proficiency test; (d) Acceptability judgment task.

4.3.1 Language background questionnaire

Participants were asked to fill in a language background questionnaire, providing their age, gender, onset age of acquisition, length of learning English in the instructional context, and self-rating in English. They were instructed to rate their language ability in speaking, listening, reading, and writing, on a 10-point scale ranging from 1 (very poor) to 10 (very good).

4.3.2 Self-paced reading

The experiment was administered online through IbexFarm (Drummond 2013). Participants completed the experiment via a link distributed to them remotely. Upon clicking the link, participants were introduced to the experiment and the informed consent, after which they were instructed to read sentences at a natural pace and respond to comprehension questions as accurately as possible. To ensure they understood the instructions well, participants had to correctly answer questions on instructions before proceeding to the main experimental session. Sentences were presented phrase by phrase in a non-cumulative moving window paradigm (Just et al. 1982). Each trial began with a fixation cross ("+") at the center of the screen for 1,000 ms followed by a series of dashes (e.g., "——"), matching the length of the corresponding word. At the press of the space bar, the dashes were replaced by the first segment, and after each following press, one subsequent segment would appear, replacing the previous one. This process was self-paced. To engage participation, participants were asked to answer a yes/no comprehension question following each sentence. For example, "Did Jay paint something for his family?" was the comprehension question for (3a) and (3b). Participants answered each question by pressing the F or J key on the keyboard and received feedback if they answered incorrectly. The positions of the correct answers were counterbalanced. Participants were allowed to take a break after every 30 trials. Six practice trials were presented prior to the main experimental session. The whole task took about 30 min to complete.

4.3.3 English proficiency test

Participants' overall proficiency in English was tested with LexTALE. LexTALE is an untimed lexical decision task where participants had to decide whether a given string of letters was a real English word or not by responding Yes or No. They were required to respond to a string with Yes even if they did not know its meaning but believed it to be a real word. The task consisted of 63 trials with 42 real words. It took about 10 min to complete. 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):

$$((\text{number of words correct}/40*100)+(\text{number of nonwords correct}/20*100))/2$$

4.3.4 Acceptability judgment task

This task aimed to establish if the participants had sufficient knowledge about the target structures used in the SPR task. Participants were instructed to rate each sentence on a 7-point Likert Scale with 1 being highly unacceptable and 7 being highly acceptable. They were not allowed to change what had been decided on the ratings. This task took around 15 min.

5 Analysis

5.1 Data treatment

All data were analyzed in R (v. 4. 1. 1; R Core Team 2021). Data cleaning included the following procedures. First, participants who scored below 80% on comprehension questions to all test sentences were excluded, resulting in the removal of data from one NS participant and 13 NNS participants. The mean accuracy for the retained NS and NNS participant was 94.4% (SD = 0.2%) and 90.7% (SD = 0.2%) respectively. Such a difference reached significance as indicated by a one-way ANOVA test ($F(1, 162) = 17.43$, $p < 0.001$).[3] Next, reading times (RTs) beyond 2.5 SDs

3 This reflects that NNSs may engage in meaning inspection to a lesser extent relative to NSs. The focus of the current study is to explore the extent to which both groups of participants engage in generating syntactic prediction during online sentence processing. Therefore, the difference due to interpretation would impact little on our interpretation of the time-course data. One anonymous reviewer also pointed out that lower accuracy rates on comprehension questions by NNSs may also reflect that NNSs with a wide range of proficiency levels had more difficulty understanding the

from an individual's mean were excluded, affecting 1.7% of the NS data and 2.2% of the NNS data.

The remaining RT data were then log-transformed for the statistical analysis (Nicklin and Plonsky 2020). Plots of model residuals against fitted values and *Q–Q* plots based on log-transformed RTs revealed no apparent deviations of normality and homoscedasticity. To adjust for the variability in region length and individual reading speed, the log-transformed RTs were residualized. Residual RTs were obtained based on linear mixed-effects models constructed separately for NSs and NNSs. In each model, log RTs were modeled as a function of the scaled length of the regions in the number of characters with by-participant intercepts and by-participant slopes for scaled region length included as random effects structures. Subsequent models for the main statistical analyses used length-adjusted residual RTs as the dependent variable. To minimize potential differences in the way individuals within or across language groups use the 7-point scale (Spinner and Gass 2019), we converted raw ratings from all items into *z*-scores by participant and language group.

5.2 Statistical modeling

Statistical analyses were performed using linear mixed-effects models with the lme4 package (Bates et al. 2015). Following Warren et al. (2016), separate models were fit for the pre-critical, critical, and post-critical regions. In line with the recommendation of Schad et al. (2020), independent variables included in the fixed effects structures were contrast-coded and centered around the grand mean: Condition (–0.5 with-either, 0.5 without-either) and Group (–0.5 NS, 0.5 NNS) as the fixed effects, and by-participant and by-item in the random effects structures. The random effects structures were always kept maximal for the initial model allowed by the experimental design (Barr et al. 2013), as represented by the by-participant random slopes for Condition and by-item random slopes for Group. In cases where models failed to converge, we simplified the random effects structures by iteratively removing the correlation between random effects and the random effect contributing to the least variance until the convergence was achieved. To independently assess the potential role of proficiency in modulating predictive processing, models were constructed separately for the NNS group across regions. Following the procedures in Kim and Grüter (2021), participants' LexTALE scores and self-ratings and the composite scores from both measures either as a

sentences. This speculation was confirmed in a Pearson correlation test, which revealed a positive correlation between learners' LexTALE scores and comprehension accuracy ($r = 0.24$, $p < 0.001$).

continuous or a categorical variable were added to the models of the RT data in a separate step. Continuous variables included in the models were centered and standardized (*z*-score transformation). All *p*-values were estimated using the lmerTest package (Kuznetsova et al. 2017).

6 Results

To set up whether NNSs possess relevant knowledge that they can deploy online, we first reported the results from the acceptability judgment task. We then presented the results from the self-paced reading task. The full set of data, code, analyses, and experimental materials are publicly available at the Open Science Framework (OSF) website: https://osf.io/abhjv/.

6.1 Acceptability judgments

The mean rating of the ungrammatical catch trials was 2.87 ($SD = 1.95$) for NSs and 3.98 ($SD = 2.29$) for NNSs, suggesting that they were not simply accepting all items. In other words, participants were able to distinguish between grammatical and ungrammatical structures in the judgment task, and therefore the task should be sensitive enough to detect participants' grammatical knowledge. Figure 2 visualizes the distribution of the transformed data. For both NSs and NNSs, sentences with *either* tended to be more acceptable than those without *either*. The linear mixed-effects model on data combined from both groups did not show a reliable effect of Condition ($\beta = -0.09$, $p = 0.20$) and Group ($\beta = -0.06$, $p = 0.50$). Separate models for each group obtained a main effect of Condition only in the NNS group ($\beta = 0.20$, $p = 0.001$), but not in the NS group ($\beta = -0.09$, $p = 0.19$). A significant interaction between Condition and Proficiency was observed in the L2 group ($\beta = -0.22$, $p < 0.001$) such that learners of higher proficiency were more sensitive to the differences between sentences across conditions than learners of lower proficiency. We speculate that the overall low acceptance for sentences without *either* for learners may be because sentences without *either* tend to be scopally ambiguous on the assumption that *either* marks the scope of the disjunct (Schwarz 1999), hence much less acceptable compared to sentences with *either*. Albeit to different degrees, NNSs generally well accepted sentences from both conditions (with-either: $M = 5.97$, $SD = 1.34$; without-either: $M = 5.61$, $SD = 1.63$) and thus had sufficient knowledge about the experimental sentences, to an extent comparable to NSs (with-either: $M = 5.77$, $SD = 1.51$; without-either: $M = 5.61$, $SD = 1.54$). As such, participants' performance associated with predictive processing, if contrasting

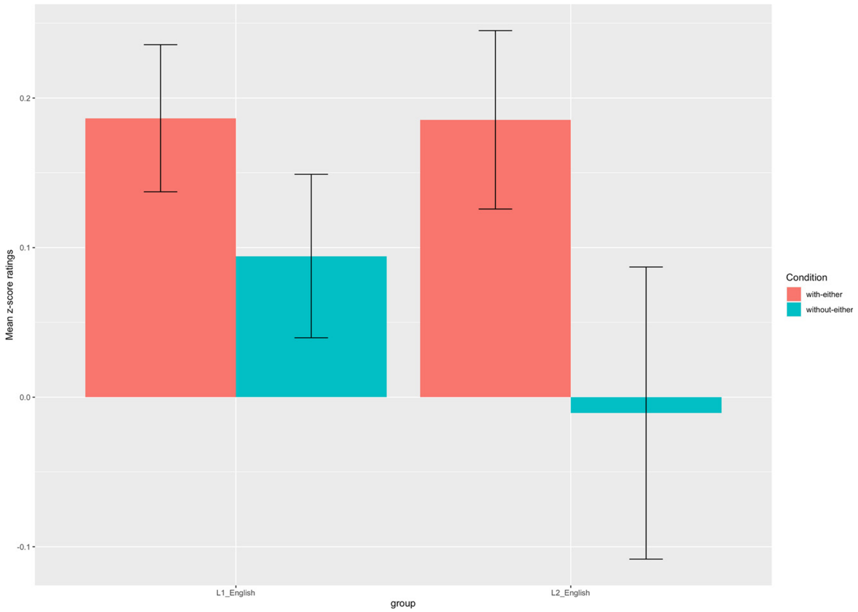

Figure 2: AJT task: Mean *z*-ratings by Condition across groups. Error bars indicate the standard error of the mean.

across conditions and language groups, should not be attributed to the possibility that they may lack the grammatical knowledge of the structures in question.

6.2 Reading times

Descriptive results for mean RTs by condition among NSs and NNSs were summarized in Table 2 and visualized in Figure 3. These results were statistically confirmed in a series of linear mixed-effects models. Table 3 presents a summary of modeling results. For the critical region, the modeling results revealed a reliable effect of Condition ($\beta = 0.127$, $p < 0.001$), indicating that participants spent less time reading this region when it was preceded by *either* compared to when it was not. A reliable effect of Group was also detected for this region ($\beta = -0.041$, $p = 0.0225$), driven by the longer RTs for the NS group than for the NNS group. To closely inspect each group's RT patterns across conditions and regions, we nevertheless conducted by-group analyses for all regions regardless of whether the interaction was significant or not to fully probe the effects of the presence versus the absence of *either* on predictive processing in different regions among NSs and NNSs.

Table 2: Mean reading times (in milliseconds) and standard deviations for each sentence region across conditions for NS and NNS speakers.

		Pre-critical region		Critical region		Post-critical region	
		With-either	**Without-either**	**With-either**	**Without-either**	**With-either**	**Without-either**
NS speakers	Mean	888.0	852.0	883.2	985.8	882.4	932.0
	SD	403.6	400.1	379.8	444.9	428.7	467.6
NNS speaker	Mean	1871.6	1870.3	1711.3	2044.9	1736.3	1733.8
	SD	1,347.8	1,351.7	1,175.5	1,506.9	1831.5	1,248.7

The mean RTs by Condition for different regions were summarized in Figure 4 for the NS group and Figure 5 for the NNS group. Analyses within each group revealed an effect of Condition both for the NS group ($\beta = 0.098$, $p = 0.000131$) and NNS group ($\beta = 0.158$, $p < 0.001$), suggesting that both groups of participants, in

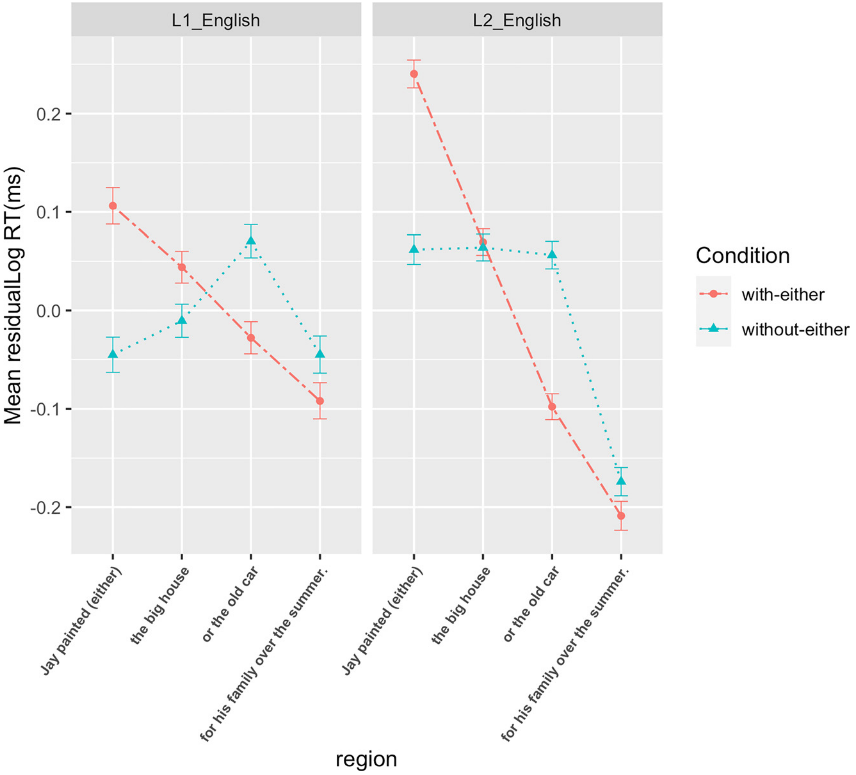

Figure 3: Mean residual reading times for NSs and NNSs by Condition across regions. Error bars indicate the standard error of the mean.

Table 3: Results of the linear mixed-effects models.

Region	Predictor	*β*	*SE*	*p*
Pre-critical	(Intercept)	0.044	0.023	0.06287
	Group	0.050	0.018	0.00731[b]
	Condition	−0.032	0.018	0.08108
	Group × Condition	0.050	0.036	0.17213
Critical	(Intercept)	0.002	0.024	0.9465
	Group	−0.041	0.018	0.0225[a]
	Condition	0.127	0.017	<0.001[c]
	Group × Condition	0.061	0.034	0.0701
Post-critical	(Intercept)	−0.128	0.030	<0.001[c]
	Group	−0.124	0.025	<0.001[c]
	Condition	0.043	0.018	0.0145[a]
	Group × Condition	−0.014	0.035	0.6912

Initial model formula for each region: lmer(corrected_log_rt ~ 1 + Condition*Group + (1 + Condition|Subject) + (1 + Group|Item)). [a]$p < 0.05$, [b]$p < 0.01$, [c]$p < 0.001$.

their responses to the critical region, were largely influenced by whether *either* appeared in the preceding context or not. As in Figures 4 and 5, the effect of Condition spilled over to the post-critical region. The difference in RTs between the *with-either* and *without-either* conditions reached significance at the post-critical region for both groups of participants ($\beta = 0.043$, $p = 0.0145$). NNSs overall read the post-critical region faster than NS speakers, reflected in the reliable effect of *Group* ($\beta = -0.124$, $p < 0.001$). Follow-up analyses by group revealed a main effect of Condition for the NS speakers ($\beta = 0.049$, $p = 0.0458$); such a significant effect did not surface for the NNS group ($\beta = 0.036$, $p = 0.0552$).

Analyses were also conducted on the pre-critical region which, relative to the critical and post-critical region, was of less theoretical interest though. The overall model demonstrated that both groups of participants showed no apparent effect of Condition ($\beta = -0.032$, $p = 0.081$).[4] Moreover, separate models conducted within each group confirmed the null effect of Condition for the NS group ($\beta = -0.054$, $p = 0.073$) and the NNS group ($\beta = -0.007$, $p = 0.700$).

Research has found that language users may show adaptation effects as their surrounding language context changes (Fine et al. 2013). Considering that such effects may be consequential in predictive processing, we conducted exploratory

4 Following the suggestion from an anonymous reviewer, in cases where the *p* values were slightly greater than 0.05 (traditionally taken to be marginally significant without additional justification), we computed the Bayes Factors (BFs) to assess the strength of the evidence for the alternative hypothesis (H_1) over the null hypothesis (H_0) (Dienes 2014). In these cases, the BFs were less than 1/3, indicating substantial evidence for H_0 (Jeffreys 1998).

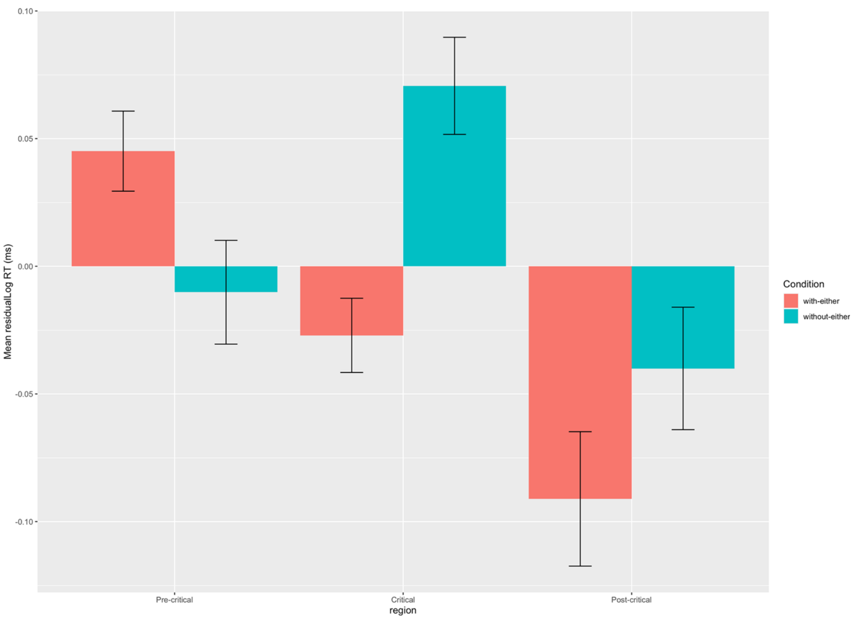

Figure 4: Mean reading times on each sentence region by Condition for NSs. Error bars indicate the standard error of the mean.

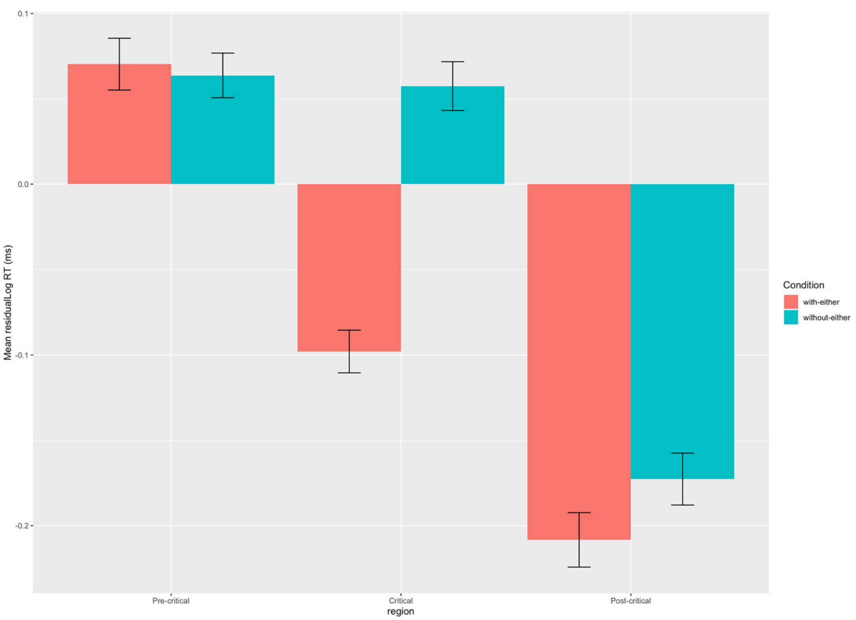

Figure 5: Mean reading times on each sentence region by Condition for NNSs. Error bars indicate the standard error of the participant mean.

analyses on the critical and post-critical regions where predictive effects are expected to arise. Models including Condition and the order of the experimental items for presentation as the fixed effects showed that Condition exhibited as a main effect but did not interact with the order of item presentation both for the NS and NNS groups, indicating that the effects of the presence versus absence of *either* on differences in RTs across conditions were held constant over the course of the experiment for the participants. These results, therefore, suggest that NSs and NNSs do not adapt their processing decisions within the experiment (See also Kaan et al. 2019).

6.3 Role of proficiency in the NNS group

The inclusion of a relatively large sample of participants with a broad range of L2 proficiency enables us to comprehensively examine the role of proficiency in modulating L2 prediction. In our case, we probed for the effects of proficiency on L2 predictive processing at the level of syntax. For the potential effects of L2 proficiency to be detected, we followed Kim and Grüter (2021) in their practice of submitting proficiency either as a continuous variable or a categorical variable to the statistical models. Specifically, each proficiency measure as a continuous variable was first centered and z-score transformed and then was individually included in the models for the NNS group: LexTALE scores, self-ratings averaged across the four language skills (speaking, listening, reading, and writing), and composite scores from the LexTALE test and self-ratings combined and averaged.[5] To categorically group participants, we adopted the median-split approach such that participants were divided into two groups based on the median of the proficiency as the cutoff point. Those who scored below the median of each pertained measure fell into the category of 'lower proficiency group' and otherwise 'higher proficiency group'. Results of each proficiency measure were summarized in Table 4.

We reported the effect of each proficiency measure for the critical and the post-critical region respectively based on a set of statistical models. For the pre-critical and critical region, when proficiency was treated as a continuous variable, neither the main effect of L2 proficiency nor the interaction between Condition and Proficiency was significant in any case. Similarly, no effects of proficiency were obtained for either region when each proficiency measure was treated as a categorical variable. For the post-critical region, however, a somewhat different picture emerged with respect to the effects of proficiency. Each proficiency measure yielded, though not consistently, a significant main effect of L2 proficiency

5 Formula for this calculation is ((z-score of LexTALE) + (z-score of self-ratings))/2.

Table 4: Results of each proficiency measure based on the median-split.

Proficiency measure	High proficiency group		Low proficiency group	
	N	Mean (SD)	*N*	Mean (SD)
LexTALE (median = 0.059)	62	0.864 (0.618)	73	−0.766 (0.503)
Averaged self-ratings (median = −0.004)	57	0.918 (0.473)	78	−0.714 (0.562)
Composite score (median = 0.090)	59	0.748 (0.404)	76	−0.598 (0.428)

(LexTALE-continuous: $\beta = 0.041$, $p = 0.083$; self-ratings-continuous: $\beta = 0.036$, $p = 0.024$; composite-continuous: $\beta = 0.059$, $p = 0.003$; LexTALE-categorical: $\beta = 0.027$, $p = 0.013$; self-ratings-categorical: $\beta = -0.056$, $p = 0.077$; composite-categorical: $\beta = -0.063$, $p = 0.047$) such that NNS speakers with higher proficiency tended to spend longer time on the post-critical region, and the final region where the wrap-up effect was expected (Just and Carpenter 1980). However, no apparent interaction between Condition and any proficiency measure was observed, despite a trend towards the interaction when the continuous LexTALE score was included in the model ($\beta = -0.032$, $p = 0.093$), yet the effect did not reach significance. This indicates that the size of the RT difference between conditions (with-either vs. without-either) on the final region overall was not modulated by learner proficiency.

7 Discussion

This study investigated whether Chinese learners of English predicted at the level of syntax to the same extent as English native speakers during real-time sentence comprehension and probed the role of L2 proficiency as a potential factor to modulate this process. To this end, we employed the SPR task to measure the times that participants spent reading sentences with DP disjunction. The same group of participants rated the sentences that were used in SPR to establish how well NNSs understood the target sentences relative to the NSs. NNSs, like NSs, overall accepted the target stimuli (mean rating larger than 5 out of 7). Hence, any potential processing differences between NNSs and NSs would not be attributed to the possibility that NNSs lacked sufficient linguistic knowledge about the target structures.

Our results show that NNSs overall resemble NSs in their syntactic prediction by using the linguistic information encoded in the lexical item *either*. For the pre-critical region, no significant main effect of Condition was found for the RTs in both groups. This finding is crucial for two reasons. First, this suggests that the effects of Condition observed on the critical region were not contaminated by the inadvertent differences in the lexical items between conditions in the pre-critical region. Second, as pointed out by Staub and Clifton (2006), given the fact that there was no difference in RTs between conditions in this region, it would not be possible that *either* facilitated all subsequent regions following it in a general fashion. Thus, the effects of Condition culminating on the critical region could directly speak to whether there was structural prediction depending on the presence versus absence of *either*.

Turning to the results of the critical region where RT differences between conditions were anticipated, our analyses revealed a significant effect of Condition in both groups. Participants read the critical region faster when *either* was present compared to when *either* was absent. The pattern demonstrates that the presence of *either* allows both NSs and NNSs to predictively build up a structure consisting of two DP disjuncts and thus the processing of the constituent represented as *or* + DP 2 could be facilitated. This is in line with what has been formulated for the relationship between linguistic predictability and processing efforts in the expectation-based theory for sentence comprehension in that a high-constraining context where linguistic materials are assumed to be highly predictable would facilitate processing (Levy 2008; Roland et al. 2012). Such predictability effects, though particularly in the domain of semantics, have also been observed to exhibit for L2 sentence processing ((Fang and Enas 2021)Fang and Enas 2021). In the present study, the presence of *either* following the main verb makes the sentence highly constraining, and thus the DP-disjunction downstream is highly predictable. Consequently, the predicted structure should be easier to integrate when it is encountered by the parser, since a predictable structure relative to an unpredictable one is assumed to be more highly activated in memory and in turn less costly to process.

Before we firmly draw the conclusion that both groups of participants engaged in making syntactic predictions during sentence processing, two possible alternative accounts are worth evaluating and eventually should be rejected. One might wonder whether the reduced RTs on the second DP disjunct, when preceded by *either*, may simply be interpreted as an effect of integration rather than prediction. As discussed earlier, integration as a processing stage must be in operation during sentence comprehension in both NSs and NNSs (Altmann and Mirković 2009; Juffs and Rodríguez 2014), particularly when the parser encounters the structure that has been predictively built up. Although integration and prediction have been claimed to be very difficult to delineate with experimental techniques, including self-paced reading, EEG, and even visual world eye-tracking (Bovolenta and

Marsden 2021), the processing of DP disjunction, as argued in Ferreira and Qiu (2021), provides an ideal window for distinguishing between them. This is because when *either* is present in the context, the parser will predict that the upcoming DP is the first disjunct of the disjunction phrase and thereby does not experience any garden-path effects upon reading the second disjunct coupled with *or*. However, when *either* is absent, the parser will treat the first DP as the direct object headed by the main verb and thereby experience a garden-path effect upon reading the critical region where the first DP has to be reanalyzed as a constituent of the DP disjunction. The lack of garden-path effects in the condition of with-*either* relative to the condition of without-*either* indicates the presence of prediction that separates itself from integration.

One may also attribute the different RTs on the critical region between conditions to the contingency between *either* and *or*. As argued by Staub and Clifton (2006) and Warren et al. (2016), this account also seems unlikely. First, the facilitated processing of the short word *or* cued by the presence of *either* would not likely lead to a measurable RT difference (RT differences for NSs: 102.6 ms; for NNSs: 331.4 ms) across conditions in the critical region consisting of four words. Second, the Condition effect observed on the entire critical region continued to spill over to the post-critical region for both groups of participants, and furthermore, such an effect was measurable in size. Therefore, the observed effect should reflect the prediction for structures at levels beyond lexical items.

The overall picture emerging from the current study is that Chinese learners of English predicted to the same extent as English NSs despite some differences in the magnitude of the Condition effect between groups. This result is consistent with some recent studies which, though targeting other linguistic phenomena, suggested that NSs and NNSs behaved similarly in generating prediction during sentence processing (e.g., Dijkgraaf et al. 2017; Ito et al. 2018; Kim and Grüter 2021), but contrary to a few others which failed to find evidence for prediction among NNSs, particularly at the level of morphosyntax (e.g., Guillelmon and Grosjean 2001; Hopp 2015; Martin et al. 2013). As such, our findings seem to speak against the *RAGE* hypothesis which claims that NNSs have a reduced capacity for initiating predictions, but largely in line with the proposal by Kaan (2014) which claimed that NNSs are not qualitatively different from NSs in their capacity of generating predictions, albeit quantitatively modulated by factors such as L2 proficiency, lexical quality, etc.

In general, we found no evidence for the role of proficiency in the critical region. However, the effect of proficiency emerged later in the post-critical region, modulating only the overall speed at which participants processed this region, but not the size of the Condition effect. The fact that NNSs with higher proficiency spent longer in this region appeared somewhat surprising. One possible reason this might be the case is that the proficiency effect was driven by the wrap-up effect (in addition to the Condition effect) that was expected to occur in this region. Some research has shown that younger adults, as opposed to older adults, allocated extra processing resources at clause boundaries, thus exhibiting larger wrap-up effects (e.g., Stine 1990). In a similar vein, more proficient learners, who should be able to allocate more processing resources, are more likely to demonstrate larger wrap-up effects than less proficient learners. However, it should be noted that such interpretation remains speculative, and future research in this regard is of course needed. Despite the main effect of proficiency captured in the post-critical region, no significant interaction between Condition and Proficiency was observed such that proficiency did not influence predictive processing in the post-critical region to which the prediction effect may spill over from the critical region. Therefore, our findings collectively showed that proficiency did not modulate L2 predictive processes across regions.

We argue that the lack of a robust proficiency effect cannot be attributed to low variability associated with a small sample size, since we have included a relatively large sample of NNS participants with L2 proficiency spanning a broad swath. In addition, the finding that RTs in the post-critical region vary as a function of proficiency may itself be an indication that the proficiency measures adopted were sensitive enough to capture the variance associated with learner proficiency. As such, it may simply be the case that proficiency does not drive the variation associated with predictive behaviors among L2 learners, contrary to the observation by Kaan (2014) that proficiency as a factor among others modulates the extent to which L2 learners engage in predictive processing.[6] This finding, in fact, echoes several recent studies on L2 predictive processing where the proficiency effect was not evident (Dijkgraaf et al. 2017; Ito et al. 2018; Kim and Grüter 2021; Mitsugi 2021, but see Dussias et al. 2013; Hopp 2013; Leal et al. 2017).

6 An anonymous reviewer raised the possibility that the absence of proficiency effect may reflect that learners from a broad range of proficiency performed at the ceiling, due to the *either … or* structure being pedagogically elementary. Unless further evidence is available, we would like to remain cautious here. First, while no apparent proficiency effect was observed on learners' predictive processing profiles, we detected some proficiency effects at the post-critical region. Aside from the nature of the processing behavior in this region (e.g., the spillover effect of predictive processing vs. integrative processing at the clause boundary), the proficiency effect itself was an indication that not all L2 participants performed equally well, and that the variance associated with L2 performance was captured by the proficiency measures. Second, previous L2 predictive processing studies that have examined structurally more complex structures also exhibited no effect of proficiency (e.g., Kim and Grüter 2021; Mitsugi 2021), which may suggest that learners, regardless of proficiency, tend to predictively process structures of varying linguistic complexity. However, further investigation in one single study regarding the interaction between linguistic complexity and L2 is needed.

The above discussion now allows us to answer the two research questions: (1) Similar to NSs, NNSs predict syntax based on the information encoded in the lexical item; (2) NNSs' predictive processing is not modulated by proficiency. One last question left for us to think about is why studies on L2 prediction, especially with respect to morphosyntax, yield mixed results. In other words, why is the effect of syntactic prediction among NNSs detectable and rather robust in the current study, as opposed to others (e.g., Grüter et al. 2012; Hopp 2015; Kaan et al. 2016) in which such an effect was much less evident? There are several possible reasons for this discrepancy.

First, the lexical item, as argued in Warren et al. (2016), is a highly reliable cue in English for prediction to initiate in the sense that the English parser should be fairly confident that a DP disjunction will come up upon reading *either*. Despite a higher restriction for the placement of the verb relative to *either* in Chinese than in English, DP disjunction in Chinese is quite similar to that in English, and *either* therefore is well used for prediction to be made in the processing of English disjunction structures by Chinese learners of English. A related but different account for why the L2 parser was a successful predictor in the current study is Kuperberg and Jaeger's (2016) utility-based model, according to which prediction involves the pre-activation of aspects of upcoming linguistic input even before it is encountered. The parser is committed to maximizing the utility of prediction for language comprehension, which is weighed for its benefits and costs and estimated as a function of the reliability of its bottom-up input as well as the processing goals. In this case, the cost of generating predictions for upcoming DP disjunction based on the reliable lexical item, *either*, should be low. Moreover, the use of prediction eases the processing difficulty, ultimately benefiting sentence comprehension. As such, the utility being high leads to a robust predictive parser in this case.

In addition, differences in the experimental methods used across experiments may also account for the divergent results. Compared to the phrase-by-phrase self-paced reading employed in the current study, methods for measuring L2 prediction of morphosyntax like eye-tracking (e.g., Grüter et al. 2012) and ERPs (e.g., Kaan et al. 2016) are more fine-grained, and therefore the differences between NSs and NNSs may more likely be revealed. This speculation stems from the assumption that NNSs tend to be more susceptible to time constraints, simply because they are typically slower in processing than NSs (Frenck-Mestre 2002). This being so, NNSs may not be qualitatively different from NSs in their ability to generate predictions. However, prediction among NNSs may not be observable in certain timing conditions.

Finally, the linguistic phenomena that have been examined differ across studies. For example, in Kaan et al. (2016), the target structures involve ellipsis (e.g., *Although the surgeon met John, he did not meet Max's…*) in which the expectation and interpretation of the elliptical structures depended on the integration of syntactic information across clauses. Compared to the *either … or* structure, elliptical structures being structurally more complex could enhance processing difficulties. These accounts, however, remain speculative unless various syntactic structures could be investigated with different experimental measures in one single study.

8 Conclusion

The present study used *either … or* construction as a test window to investigate differences in predictive processing at the level of syntax between NSs and NNSs and further explored the potential role of L2 proficiency in this process by testing participants with a wide proficiency range. Our results showed that NNSs were very similar to NSs in the use of the information encoded in the lexical item, *either*, to predict a DP disjunction structure consisting of two smaller DP disjuncts, reflected in the fact that both groups of participants read faster the critical region consisting of *or* and the second DP disjunct when this region was preceded by the prediction trigger, *either*. Such an effect spilled over to the post-critical region. This finding is in line with Kaan's (2014) claim that NNSs do not fundamentally differ from NSs in predictive processing, but is inconsistent with the *RAGE* hypothesis that NNSs do not predict to the same extent as NSs (Grüter et al. 2017). L2 proficiency does not seem to play a notable role in modulating NNSs' predictive processing, particularly in the critical region, even though a weak effect of proficiency was observed in the post-critical region such that proficiency overall modulated the speed at which this region was read. The null effect of proficiency is inconsistent with Kaan (2014), who identified L2 proficiency as a key factor driving L2 predictive processing. In the hope of drawing a complete picture of L2 predictive processing, future research can explore how other potential factors (e.g., cross-linguistic influence, cue reliability, Foucart 2021; van Bergen and Flecken 2017) and their interaction with proficiency constrain NNSs' engagement in predictive processing as a certain mechanism driving incremental sentence processing.

References

Altmann, Gerry T.M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3). 247–264.

Altmann, Gerry T.M. & Jelena Mirković. 2009. Incrementality and prediction in human sentence processing. *Cognitive Science* 33(4). 583–609.
Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
Bates, Douglas, Martin Mächler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.
Bovolenta, Giulia & Emma Marsden. 2021. Prediction and error-based learning in L2 processing and acquisition: A conceptual review. *Studies in Second Language Acquisition*. 1–26. https://doi.org/10.1017/s0272263121000723.
Chambers, Craig G. & Hilary Cooke. 2009. Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory & Cognition* 35(4). 1029.
Chen, Qingrong & Huang Yan. 2012. Processing coordinate structures in Chinese: Evidence from eye movements. *PLoS One* 7(4). e35517.
Chomsky, Noam. 2009. *Syntactic structures*. Berlin: De Gruyter Mouton.
Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27(1). 3–42.
Clark, Andy. 2013. Whatever next? Predictive brains, situated agents & the future of cognitive science. *Behavioral and Brain Sciences* 36(3). 181–204.
DeLong, Katherine A., Thomas P. Urbach & Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8). 1117–1121.
Dienes, Zoltan. 2014. Using Bayes to get the most out of non-significant results. *Frontiers in Psychology* 5. 781.
Dijkgraaf, Aster, Robert J. Hartsuiker & Wouter Duyck. 2017. Predicting upcoming information in native-language and non-native-language auditory word recognition. *Bilingualism: Language and Cognition* 20(5). 917–930.
Drummond, Alex. 2013. *Ibex Farm*. Available at: http://spellout.net/ibexfarm/.
Dussias, Paola E., Jorge R. Valdés Kroff, Rosa E. Guzzardo Tamargo & Chip Gerfen. 2013. When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition* 35(2). 353–387.
Fang, Shaohua & Enas Albasiri. 2021. Effects of syntactic and semantic predictability on sentence comprehension: A comparison between native and non-native speakers. In *Proceedings of the annual meeting of the cognitive science society*, Vol. 43. No. 43. Cognitive Science Society.
Federmeier, Kara D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4). 491–505.
Ferreira, Fernanda & Zhuang Qiu. 2021. Predicting syntactic structure. *Brain Research* 1770. 147632.
Fine, Alex B., T. Florian Jaeger, Thomas A. Farmer & Ting Qian. 2013. Rapid expectation adaptation during syntactic comprehension. *PLoS One* 8(10). e77661.
Foucart, Alice. 2021. *Language prediction in second language: Does language similarity matter?*, 12. Amsterdam: John Benjamins Publishing Company.
Frazier, Lyn & Charles Clifton, Jr. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4(1). 1–22.
Frazier, Lyn, Alan Munn & Charles Clifton. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29(4). 343–370.

Frenck-Mestre, Cheryl. 2002. An on-line look at sentence processing in the second language *Advances in psychology*, Vol. 134, 217–236. North-Holland.
Grüter, Theres & Hannah Rohde. 2021. Limits on expectation-based processing: Use of grammatical aspect for co-reference in L2. *Applied Psycholinguistics* 42(1). 51–75.
Grüter, Theres, Casey Lew-Williams & Anne Fernald. 2012. Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research* 28(2). 191–215.
Grüter, Theres, Hannah Rohde & Amy J. Schafer. 2017. Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism* 7(2). 199–229.
Guillelmon, Delphine & François Grosjean. 2001. The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition* 29(3). 503–511.
Henry, Nick, Carrie N. Jackson & Holger Hopp. 2022. Cue coalitions and additivity in predictive processing: The interaction between case and prosody in L2 German. *Second Language Research* 38(3). 397–422.
Hopp, Holger. 2006. Syntactic features and reanalysis in near-native processing. *Second Language Research* 22(3). 369–397.
Hopp, Holger. 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua* 120(4). 901–931.
Hopp, Holger. 2013. Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research* 29(1). 33–56.
Hopp, Holger. 2015. Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching* 53(3). 277–306.
Hopp, Holger. 2018. The bilingual mental lexicon in L2 sentence processing. *Second language* 17. 5–27.
Ito, Aine & Martin J. Pickering. 2021. Automaticity and prediction in non-native language comprehension. In *Prediction in second language processing and learning*, 26–46. Amsterdam: John Benjamins Publishing Company.
Ito, Aine, Corley Martin & Martin J. Pickering. 2018. A cognitive load delays predictive eye movements similarly during L1 and L2 comprehension. *Bilingualism: Language and Cognition* 21(2). 251–264.
Ivanova, Iva & Albert Costa. 2008. Does bilingualism hamper lexical access in speech production? *Acta Psychologica* 127(2). 277–288.
Jackson, Carrie. 2008. Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning* 58(4). 875–909.
Jeffreys, Harold. 1998. *The theory of probability*. Oxford: OUP.
Juffs, Alan & Guillermo A. Rodríguez. 2014. *Second language sentence processing*. New York: Routledge.
Just, Marcel A. & Patricia A. Carpenter. 1980. A theory of reading: From eye fixations to comprehension. *Psychological Review* 87(4). 329.
Just, Marcel A., Patricia A. Carpenter & Jacqueline D. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111(2). 228.
Kaan, Edith. 2014. Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism* 4(2). 257–282.
Kaan, Edith, Andrea Dallas & Frank Wijnen. 2010. Syntactic predictions in second-language sentence processing. *Structure preserved. Festschrift in the honor of Jan Koster* 29. 207–213.

Kaan, Edith, Corinne Futch, Raquel Fernández Fuertes, Sonja Mujcinovic & Esther Álvarez de la Fuente. 2019. Adaptation to syntactic structures in native and nonnative sentence comprehension. *Applied Psycholinguistics* 40(1). 3–27.

Kaan, Edith & Theres Grüter. 2021. Prediction in second language processing and learning: Advances and directions. In T. Grüter & E. Kaan (eds.), *Prediction in second-language processing and learning*. Amsterdam: John Benjamins.

Kaan, Edith, Kirkham Joseph & Wijnen Frank. 2016. Prediction and integration in native and second-language processing of elliptical structures. *Bilingualism: Language and Cognition* 19(1). 1–18.

Kaiser, Elsi & John C. Trueswell. 2004. The role of discourse context in the processing of a flexible word-order language. *Cognition* 94(2). 113–147.

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

Kim, Hyunwoo & Theres Grüter. 2021. Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition* 43(1). 133–154.

Kuperberg, Gina R. & T. Florian Jaeger. 2016. What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience* 31(1). 32–59.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H.B. Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(1). 1–26.

Leal, Tania. 2018. Data analysis and sampling. *Critical Reflections on Data in Second Language Acquisition* 51. 63.

Leal, Tania, Roumyana Slabakova & Thomas A. Farmer. 2017. The fine-tuning of linguistic expectations over the course of L2 learning. *Studies in Second Language Acquisition* 39(3). 493–525.

Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods* 44(2). 325–343.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3). 1126–1177.

Lew-Williams, Casey & Anne Fernald. 2010. Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language* 63(4). 447–464.

Martin, Clara D., Guillaume Thierry, Jan-Rouke Kuipers, Bastien Boutonnet, Alice Foucart & Albert Costa. 2013. Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language* 69(4). 574–588.

Mitsugi, Sanako. 2021. Polarity adverbs facilitate predictive processing in L2 Japanese. *Second Language Research*. https://doi.org/10.1177/02676583211000837.

Mitsugi, Sanako & Brian Macwhinney. 2016. The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition* 19(1). 19–35.

Nicklin, Christopher & Luke Plonsky. 2020. Outliers in L2 research in applied linguistics: A synthesis and data Re-analysis. *Annual Review of Applied Linguistics* 40. 26–55.

Partee, Barbara & Mats Rooth. 2012. Generalized conjunction and type ambiguity. *Meaning, use & interpretation of language*, 361–383. Berlin: de Gruyter.

Pliatsikas, Christos & Theodoros Marinis. 2013. Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition* 16(1). 167–182.

Quine, Willard Van Orman. 1967. On a suggestion of Katz. *The Journal of Philosophy* 64(2). 52–54.

R Core Team. 2021. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.
Roberts, Leah. 2012. Individual differences in second language sentence processing. *Language Learning* 62. 172–188.
Roland, Douglas, Hongoak Yun, Jean-Pierre Koenig & Gail Mauner. 2012. Semantic similarity, predictability & models of sentence processing. *Cognition* 122(3). 267–279.
Sag, Ivan A., Gerald Gazdar, Thomas Wasow & Steven Weisler. 1985. Coordination and how to distinguish categories. *Natural Language & Linguistic Theory* 3(2). 117–171.
Schad, Daniel J., Shravan Vasishth, Sven Hohenstein & Reinhold Kliegl. 2020. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *Journal of Memory and Language* 110. 104038.
Schwarz, Bernhard. 1999. On the syntax of either... or. *Natural Language & Linguistic Theory* 17(2). 339–370.
Spinner, Patti & Susan M. Gass. 2019. *Using judgments in second language acquisition research*. New York: Routledge.
Staub, Adrian. 2007. The return of the repressed: Abandoned parses facilitate syntactic reanalysis. *Journal of Memory and Language* 57(2). 299–323.
Staub, Adrian & Charles Clifton, Jr. 2006. Syntactic prediction in language comprehension: Evidence from either. Or. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32(2). 425.
Stine, Elizabeth L. 1990. On-line processing of written text by younger and older adults. *Psychology and Aging* 5(1). 68.
Stockwell, Robert P., Schacter Paul, Barbara Hall Partee & ul Schachter. 1973. *The major syntactic structures of English*. Holt, Rinehart and Winston.
Sturt, Patrick, Frank Keller & Dubey Amit. 2010. Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language* 62(4). 333–351.
Sturt, Patrick, Martin Pickering & Matthew Crocker. 1999. Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language* 40(1). 136–150.
Trenkic, Danijela, Jelena Mirkovic & Altmann T.M. Gerry. 2014. Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition* 17(2). 237–257.
Van Bergen, Geertje & Monique Flecken. 2017. Putting things in new places: Linguistic experience modulates the predictive power of placement verb semantics. *Journal of Memory and Language* 92. 26–42.
Van Berkum, Jos J.A., Colin M. Brown, Pienie Zwitserlood, Valesca Kooijman & Hagoort Peter. 2005. Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory & Cognition* 31(3). 443.
Warren, Tessa, Michael Walsh Dickey & Chia-Ming Lei. 2016. Structural prediction in aphasia: Evidence from either. *Journal of Neurolinguistics* 39. 38–48.
Yoshida, Masaya, Michael Walsh Dickey & Patrick Sturt. 2014. *The effect of Either on the processing of disjunctive coordination*. Ms. Northwestern University, University of Pittsburgh & University of Edinburgh.