Bilingualism: Language and Cognition

cambridge.org/bil

Anticipatory processing of cataphora is constrained by binding principles in L2 English

Jun Lyu[1], Zuzanna Fuchs[2] and Elsi Kaiser[2]

[1]Peking University, Beijing, China and [2]University of Southern California, Los Angeles, United States

Research Article

Cite this article: Lyu, J., Fuchs, Z., & Kaiser, E. (2024). Anticipatory processing of cataphora is constrained by binding principles in L2 English. *Bilingualism: Language and Cognition* **27**, 755–768. https://doi.org/10.1017/S1366728924000208

Received: 9 January 2023
Revised: 15 February 2024
Accepted: 18 February 2024
First published online: 22 March 2024

Keywords:
second language processing; anticipatory processing; cataphora; L1 influence; Binding Theory

Corresponding author:
Jun Lyu;
Email: junlyu@pku.edu.cn

Abstract

Language processing studies show that native speakers anticipate linguistic elements before their occurrence. However, it is debated to what extent second language (L2) learners do the same. To address this question, this study examines the processing of cataphora by Chinese-speaking L2 English learners. Additionally, we query whether L2 learners' expectations of upcoming antecedents are modulated by first language (L1) influence and constrained by Principle B of the Binding Theory (Chomsky, 1981). Two self-paced reading studies show that L1 English speakers' anticipation of upcoming referents is active and strictly constrained by Principle B. Crucially, L2 English learners also actively predict upcoming referents and are sensitive to Principle B. However, L2 processing patterns suggest that Principle B competes with semantics at later processing stages. Together with data from L1 Chinese and English control participants, these results support the view that anticipatory processing in English is not fundamentally different between monolinguals and bilinguals.

1. Introduction

In language comprehension, native speakers anticipate upcoming information at various linguistic levels (e.g., Altmann & Kamide, 1999; DeLong et al., 2005; Grüter et al., 2017; Kuperberg & Jaeger, 2016). In second language (L2) research, however, evidence is mixed regarding whether L2 learners formulate expectations to the same extent as native speakers do.

Some real-time studies suggest that L2 learners differ from native speakers in underusing linguistic cues (e.g., grammatical gender) for generating expectations (e.g., Grüter et al., 2012; Lew-Williams & Fernald, 2010), although the difference could be quantitative rather than qualitative (e.g., Chun & Kaan, 2019; Kim & Grüter, 2021). However, other studies indicate that L2 learners demonstrate native-like anticipatory behaviors (e.g., Hopp, 2013; Mitsugi, 2018), modulated by L2 proficiency and L1 influence (e.g., Dussias et al., 2013; Hopp, 2010; Hopp & Lemmerth, 2016).

Anticipatory processing in L2 is a crucial topic because it is key to understanding some fundamental differences between L1 and L2 processing, especially at the syntactic level (e.g., Kaan, 2014; Kaan et al., 2010). These differences have sometimes been ascribed to inefficient use of syntactic information during processing (e.g., Clahsen & Felser, 2006). But an alternative expectation-based account holds that non-native processing patterns could be related to weaker expectations. Our study builds on Kush and Dillon's (2021) work on L1 English processing and uses backward anaphora or CATAPHORA, shown in (1), to explore anticipatory processing in Chinese learners' L2 English.

(1) While driving him/his friend to school, Mike casually told James an interesting story.

Prior work suggests that upon encountering a pronoun without an antecedent in prior discourse, native speakers anticipate a postcedent (i.e., 'antecedent' following the pronoun) matching in gender and number in the upcoming discourse (e.g., Drummer & Felser, 2018; Kazanina et al., 2007). This study investigates to what extent Chinese learners of L2 English similarly anticipate a postcedent in real-time processing. This study has three main questions.

First, when cataphora is syntactically licensed, do L2 learners actively anticipate a semantically congruent postcedent in the upcoming discourse?

Second, is anticipatory processing in the L2 subject to L1 influence? As we shall see, cataphora in Chinese is less expected (Sun & Dennison, 2015) and infrequent (Y. Xu & He, 2007; Yuan, 2005). Leveraging the differences between Chinese and English, we query whether properties of L1 Chinese influence anticipatory processing in L2 English.

Third, is real-time anticipatory cataphora resolution in the L2 constrained by Principle B of the Binding Theory (Chomsky, 1981)? This question connects to L2 sentence processing work

© The Author(s), 2024. Published by Cambridge University Press

This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

examining syntactic representations and the time-course of access to grammatical constraints in the L2 (for a review, see Clahsen & Felser, 2006, 2018; Cunnings, 2017).

2. Research background

2.1. Debate on reduced ability to generate expectations in L2

One of our main goals is to examine whether L2 speakers show signs of anticipation in cataphora resolution. An influential recent hypothesis posits that L2 learners show a "reduced ability to generate expectations" (RAGE) during real-time processing compared to native speakers (Grüter et al., 2017). Although the RAGE hypothesis argues for "reduced reliance on expectations at the discourse level" (Grüter et al., 2017, p. 224), it has been suggested that reduced expectations could also account for non-native processing patterns at the syntactic level (e.g., Kaan, 2014; Kaan et al., 2010).

However, the suggestion that RAGE can be extended to (morpho)syntactic processing is debated. In the case of morphosyntax involving local domains, researchers investigated whether L2 learners can use gender-marked determiners and adjectives to formulate expectations for upcoming nouns (e.g., Dussias et al., 2013; Hopp, 2010, 2013; Hopp & Lemmerth, 2016; E. Lau & Grüter, 2015; Lew-Williams & Fernald, 2010), with mixed results, including some evidence suggesting that advanced learners can show native-like anticipatory behaviors (e.g., Hopp, 2010, 2013; Hopp & Lemmerth, 2016).

Here, we focus on long-distance dependencies which demand more cognitive resources than local dependencies and could provide more insights into RAGE. The line of research we pursue concerns cataphora resolution. In L1 processing, when cataphora is syntactically licensed in sentences like *while she was taking classes, Kathryn/Russell…*, native speakers expect *she* to refer to the matrix subject *Kathryn*. Empirically, this expectation is revealed by the gender mismatch between *she* and *Russell* leading to reading slowdowns (Kazanina et al., 2007). This slowdown is known as the GENDER MISMATCH EFFECT and is typically used as a diagnostic for detecting attempts at cataphoric binding (e.g., Badecker & Straub, 2002; the same diagnostic will be used for our study.) In L2 processing, we see a similar pattern: Chinese, Spanish, and Japanese learners of L2 English (Bertenshaw, 2009; Rodriguez, 2008) and Russian learners of L2 German (Drummer & Felser, 2018) show gender mismatch effects in syntactically licensed cataphoric constructions, although the gender mismatch effects do not always emerge immediately.

One exception is Wu et al. (2019), who found that Chinese learners of English did not show any gender mismatch effect when reading English sentences like *As he got closer to the group, Bill/Mary shouted hello*, suggesting that Chinese learners might not anticipate the upcoming postcedent in L2 English. But Wu et al. did not provide a detailed explanation for this finding, other than suggesting that the absence of a gender mismatch effect stems from Chinese learners' quick adjustment to gender (in)congruency. Importantly, prior work (including Wu et al.) did not directly examine the role of L1 influence and L2 proficiency. We address these gaps by investigating the processing of L1 Chinese (Experiment 1) to identify potential L1 influence, in addition to the processing of English as an L1/L2 (Experiment 2), and by using L2 proficiency as a continuous predictor in statistical models.

2.2. Potential L1 influence in cataphora resolution

When seeking to explain differences between L1 and L2 processing, potential differences between languages need to be considered. For example, could L1 influence explain the absence of anticipation for postcedents in Wu et al.'s (2019) study? Thus, a natural extension of the research on anticipatory processing is to test whether cataphora resolution is subject to L1 influence (here, we use the term "L1 influence" in a broad sense, including effects of prior L1 experience on the L2; this experience could be grammar- or usage-based; we return to this in Sections 5 and 6.)

Prior comparative work shows that cataphora in Chinese is more restricted in its syntactic environments than in English (e.g., Huang, 1982; Huang & Lin, 2021; C. Wang, 2000). It is also less acceptable (Sun & Dennison, 2015; Z. Wang, 1994) and less frequent (Y. Xu & He, 2007; Yuan, 2005), which motivated the proposal of a linearity constraint for Chinese (Tai, 1973; C. Wang, 2000; Z. Wang, 1994). Notably, Sun and Dennison (2015) found in a self-paced reading study that, unlike anaphora, cataphora is not strongly anticipated by native Chinese speakers, as indicated by a weak and delayed gender mismatch effect, which contrasts with the online cataphora resolution patterns in L1 English speakers (Kennison et al., 2009). Intriguingly, despite the relatively weak and delayed anticipation during online processing, Chinese speakers strongly prefer cataphoric coreference in offline forced-choice judgment. Nevertheless, the weak anticipation for upcoming postcedents by Chinese speakers could potentially impact Chinese learners' real-time cataphora resolution in L2 English, which we examine in this study.

Despite a large body of work showing that L1 properties influence the acquisition of L2s (see e.g., White, 2000 for a review), it is less clear whether L1 influence guides REAL-TIME L2 processing (see Lago et al., 2021; Roberts, 2013 for reviews).

On the one hand, some researchers argue that the role of the L1 is quite limited in grammatical processing: L1 influence has been described as "often absent" (Clahsen & Felser, 2018, p. 697) based on overviews of prior studies (e.g., Cunnings et al., 2010; Felser et al., 2009; Roberts et al., 2008). On the other hand, there is empirical evidence for L1 influence on some aspects of syntactic processing, such as morphosyntactic processing (e.g., Dussias et al., 2013; Dussias & Sagarra, 2007; Hopp, 2017).

However, evidence for L1 influence involving long-distance dependencies is indeed limited, which makes cataphora resolution a good test case. In our case, as cataphora is less frequent in Chinese (Y. Xu & He, 2007; Yuan, 2005) and less anticipated in real-time processing (Sun & Dennison, 2015), Chinese learners might show a similarly weak (or even null) anticipatory processing pattern in their L2 English as in their L1 Chinese. In this work, we ask whether L2 learners' experience with cataphora in L1 Chinese impacts their expectations for postcedents in L2 English.

2.3. Cataphora resolution and Principle B

If L2 learners can actively anticipate the postcedent, we can ask whether their expectations are guided by syntactic constraints. This study probes how Principle B influences L2 learners' anticipatory behavior in sentences like (1) (*while PRO driving him to school, Mike…*). According to Principle B, a pronoun cannot be bound by a clausemate antecedent – the PRO in the adverbial

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

clause. This means that the matrix subject *Mike* that controls PRO cannot be coreferential with the pronoun *him*. Therefore, if L2 processing is syntactically guided in similar ways as L1 processing, L2 learners should not anticipate the matrix subject to be the postcedent.

However, it is debated whether L2 learners can employ binding principles in a native-like manner during real-time processing (e.g., Felser et al., 2009, 2012; Puebla & Felser, 2022). The Shallow Structure Hypothesis (SSH; Clahsen & Felser, 2006) posits that L2 speakers build shallower syntactic representations in real-time parsing and over-rely on lexical-semantic cues (e.g., Frenck-Mestre & Pynte, 1997; Marinis et al., 2005). But this view has been challenged by studies showing that L2 speakers' real-time processing is sensitive to syntactic constraints such as syntactic islands (e.g., Felser et al., 2012; Omaki & Schulz, 2011).

A more recent version of the SSH argues that L2 speakers can have fully specified syntactic representations but show reduced and delayed sensitivity to syntactic constraints in real-time parsing (Clahsen & Felser, 2018). For example, in an eye-tracking study on reflexive resolution in L2 English, Felser et al. (2012) found that despite native-like grammatical knowledge about Principle A in the off-line task (i.e., binding of reflexives must be local), during real-time processing German learners of English nonetheless showed an early bias for the discourse-prominent matrix subject as the antecedent, contrary to Principle A. By contrast, native English speakers only considered local antecedents at early processing stages. This suggests that L2 speakers are less sensitive to syntactic constraints than discourse-pragmatic constraints in anaphora resolution.

Regarding cataphora resolution, evidence is mixed. Rodriguez (2008) and Bertenshaw (2009) presented evidence of sensitivity to Principle C (i.e., a pronoun cannot c-command its coreferenced DP as in he_i *opened the window while the fireman*$_{*i/j}$…) with Chinese, Japanese, and Spanish learners of English. But Drummer and Felser (2018) observed that native German speakers and Russian learners of German both allowed interpretations that violate Principle C in German. These differences could be due to language-specific properties. Another reason could be that, as Kush and Dillon (2021) suggested, Principle C might not be a good test case for real-time application of syntactic constraints, given some exceptions to Principle C. In our work, we follow Kush and Dillon's paradigm by making use of Principle B to test whether Chinese learners of English can restrain their expectations for postcedents in syntactically unlicensed positions.

3. Overview of the aims

This study has three overarching goals.

First, we aim to explore whether Chinese learners of English can anticipate an upcoming postcedent when cataphora is syntactically licensed by the English grammar.

Second, based on the processing patterns of cataphora in L1 Chinese, we aim to assess to what extent L2 processing is impacted by L1 influence.

Third, we hope to shed light on whether Principle B strictly guides anticipatory processing in the L2 when cataphoric binding is ruled out syntactically.

In two experiments, we combine postcedent choice and self-paced reading tasks to investigate L2 learners' grammatical knowledge and their processing of cataphora to further our understanding of anticipatory processing involving long-distance dependencies.

4. Experiment 1: Cataphora in Chinese

Experiment 1 examines cataphora processing in Chinese by native Chinese speakers, using self-paced reading and sentence-final postcedent choices. The postcedent choice task tests to what extent cataphora is constrained by Principle B in Chinese, which to our knowledge has not been examined experimentally in prior work. The Chinese speakers' processing patterns in self-paced reading can illuminate how actively they anticipate the postcedent and what the time-course is of multiple constraints that are at play.

4.1. Participants

Eighty-five adult native Chinese speakers participated over the internet. Participants with comprehension accuracies below 75% in filler trials were excluded (5 participants). This screening criterion was applied to all experiments in this paper.

4.2. Materials

We crossed Constraint (Licensed/Unlicensed) and Gender (Match/Mismatch) in a factorial design, as exemplified in (2). The null subject in the adverbial clause is indicated by PRO/*pro*[1] (see Sun & Dennison, 2015, for a similar analysis). Constraint refers to whether cataphoric binding by the matrix subject is syntactically licensed. In (2a), cataphoric binding does not violate any syntactic constraints. By contrast, in (2b), the object pronoun *ta* cannot refer to the matrix subject (e.g., *Liqiang*): the null subject (PRO/*pro*) – coreferential with the matrix subject – c-commands *ta*, thereby precluding coreference between *ta* and the null subject/matrix subject. Gender refers to the semantic gender congruency between the pronoun and the matrix subject. The gender mismatch effect, as introduced in Section 2.1, is a diagnostic for detecting cataphoric binding in real-time processing.

(2) a. **Licensed:**
Zai$_1$ PRO/*pro* song$_2$ {ta$_{male}$/ta$_{female}$} pengyou$_3$ qu xuexiao$_4$ de
at PRO/*pro* take 3SG.M/3SG.F friend to school DE
shihou,$_5$ Liqiang$_{male6}$ suikou$_7$
time Liqiang casually
gaosu-le$_8$ {Wangfan$_{male}$/Wangli$_{female}$}$_9$ yi-ge$_{10}$ gushi.$_{11}$
tell-ASP Wangfan/Wangli one.CL story
'While taking {his/her} friend to school, Liqiang$_{male}$ casually told {Wangfan$_{male}$/Wangli$_{female}$} a story.'

b. **Unlicensed:**
Zai$_1$ PRO/*pro* song$_2$ {ta$_{male}$/ta$_{female}$}$_3$ qu xuexiao$_4$ de shihou,$_5$
at PRO/*pro* take 3SG.M/3SG.F to school DE time
Liqiang$_{male6}$ suikou$_7$ gaosu-le$_8$
Liqiang casually tell-ASP
{Wangfan$_{male}$/Wangli$_{female}$}$_9$ yi-ge$_{10}$ gushi.$_{11}$
Wangfan/Wangli one-CL story
'While taking him/her to school, Liqiang$_{male}$ casually told Wangfan$_{male}$/Wangli$_{female}$ a story.'

In the target sentences, presented region-by-region as indicated by subscripts in (2), there are two potential postcedents: the matrix subject (e.g., *Liqiang*) and the matrix object (e.g., *Wangfan*/*Wangli*). The matrix subject is the first available postcedent and matches/mismatches the gender of the pronoun.

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

Following Kush and Dillon (2021), the matrix object always matches the gender of the pronoun (i.e., the pronoun has an available referent even if it mismatches the gender of the matrix subject). Genders of the postcedents are indicated by stereotypically male/female names.

These Chinese target sentences have the same structural design as the English sentences in Experiment 2. However, the sentences are not lexically identical, because Chinese and English verbs differ in subcategorization and semantic/pragmatic properties. Because the target sentences in both languages involve a variety of meanings, the results are not dependent on any specific meaning configuration. Crucially, the Chinese and English sentences closely match in structure and thus the semantic differences are not expected to impact our conclusions.

Twenty-four sets of target items were distributed into 4 lists using Latin square. These were mixed with 24 filler sentences including relative clauses and anaphora. Half of targets have female and half have male matrix subjects. The critical region is the matrix subject (e.g., *Liqiang*). The two spillover regions are the adverb ('casually') and the matrix verb ('told').

4.3. Procedure

The self-paced reading experiment was implemented on *Ibex Farm* (Drummond, 2013). Participants read each sentence region-by-region. After each sentence, they answered a two-alternative question. On target trials, this question probes the interpretation of the cataphor. The answer choices were the matrix subject (e.g., *Liqiang*) and object (e.g., *Wangfan/Wangli*). The order of the choices was randomized. See (3) for an example:

(3) Example question for a target trial (presented in Chinese):
'{_____ /_____ 's friend} was taken to school.'
A. Liqiang B. Wangfan/Wangli

4.4. Predictions

We first consider the predictions for offline postcedent choices. First, a main effect of Constraint is expected as native Chinese speakers should prefer matrix objects in the Unlicensed conditions due to Principle B, relative to the Licensed conditions. Second, we anticipate a Constraint x Gender interaction. In the Licensed conditions, when the genders of the pronoun and the matrix subject mismatch, Chinese speakers are expected to prefer the matrix object (*Wangfan/Wangli*); but when the genders of the pronoun and the matrix subject match, they are expected to prefer the matrix subject (*Liqiang*). In the Unlicensed conditions, theoretically we could anticipate a null gender effect because the matrix subject is unlikely to be considered (i.e., ruled out by Principle B), regardless of its gender. However, given Kush and Dillon's finding regarding postcedent choices in English, we may see a weak 'intrusion' effect from the gender-congruent matrix subject in Unlicensed conditions in Chinese. This would not be empirically surprising because in offline tasks gender information often impacts referent choices at some very late processing stage (e.g., Cunnings & Sturt, 2014; Sturt, 2003).

As regards reading times, a Constraint x Gender interaction is expected at and/or after the matrix subject. Following prior work, we use the gender mismatch effect as an indicator of binding between cataphoric pronouns and postcedents. In the Licensed conditions, if Chinese speakers actively anticipate a postcedent, we expect an immediate gender mismatch effect at the critical region. However, given Sun and Dennison's (2015) findings, we might alternatively see a reduced or delayed gender mismatch effect. For the Unlicensed conditions, following Kush and Dillon (2021), our default hypothesis is that Chinese speakers observe Principle B when the matrix subject is first encountered and do not show any gender effect.

4.5. Data analysis

For the analysis of postcedent choices, we fitted mixed-effect logistic models in R (R Core Team, 2020) using the *glmer* function (*lme4* package, Bates et al., 2015a). Contrasts were set for fixed effects Constraint (Unlicensed: +05, Licensed: −0.5) and Gender (Match: +0.5, Mismatch: −0.5). Participants and items were treated as grouping factors for random effects. Statistical models were fitted parsimoniously (Bates et al., 2015b): a simpler model is preferred if it does not differ significantly from a more complex one in model comparisons. A p-value of 0.2 was selected for model comparisons (Matuschek et al., 2017).

For RT analysis, extreme RTs outside the 100–3000 ms range were first trimmed. Then, RTs 2.5 standard deviations above the mean by region and by condition were also removed (e.g., Ratcliff, 1993; Kush et al., 2017). This affected 2.67% of the original data. We then fitted mixed-effect linear models (*lmer* function) and used the contrast coding described above. Analyses were run over both raw RTs and log-transformed RTs.[2] For brevity, we report results based on raw RTs; when crucial discrepancies exist, we report both. T-values were transformed to p-values using the Satterthwaite approximation implemented by the *lmerTest* package (Kuznetsova et al., 2015).

4.6. Results

The 80 participants included in the analysis had an average accuracy of 96% on filler-trial comprehension questions. We report postcedent choice results first.

Postcedent choices

Participants' mean proportions of matrix object choices are in **Table 1**. See supplementary materials (Table A1) for statistics.[3] The statistical model reveals main effects of Gender and Constraint. The main effect of Constraint indicates that Chinese speakers have a stronger preference for matrix objects in the Unlicensed than in the Licensed conditions, suggesting that Principle B constrains postcedent choices.

The main effect of Gender indicates that Chinese speakers are overall sensitive to the gender of the matrix subject. This gender effect is especially revealing in the Licensed conditions: when the genders of the pronoun and the matrix subject match, matrix subjects are strongly preferred (72%), suggesting that participants prefer the first available referent as the postcedent (also see Giskes & Kush, 2021). This FIRST-REFERENT BIAS even characterizes the Licensed/Mismatch condition, in which 46% of postcedent choices went to the gender-incongruent matrix subject. As will be discussed later, a similar pattern characterizes English native speakers' antecedent choices as well (also see Kush & Dillon, 2021) and suggests that native Chinese speakers are sensitive to the orthographical gender of *ta* (also see e.g., Liang et al., 2018; Qiu et al., 2012; X. Xu et al., 2013). In addition, as expected, we see a small gender effect – a weak 'intrusion' effect – in the Unlicensed conditions.

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

Table 1. Mean proportions of matrix object choices by L1 Chinese speakers in Experiment 1. Numbers in parentheses indicate standard errors in percentages.

Licensed		Unlicensed	
Match	Mismatch	Match	Mismatch
28% (2%)	54% (2%)	78% (2%)	89% (1%)

Finally, gender congruency plays a more prominent role in the Licensed than in the Unlicensed conditions as indicated by a marginal Constraint x Gender interaction, a pattern which is similar to what we shall see in native English speakers. Overall, these data suggest that Principle B constrains Chinese speakers' postcedent choices.

Reading times

Mean RTs across regions are in **Figure 1**. See supplementary materials (Table A2) for statistics. Before the critical region, we find main effects of Gender at regions 1 ('At') and 4 ('to school') which are probably spurious because the gender-(in)congruent postcedent has not been encountered yet.

There are main effects of Constraint from region 3 ('pronoun (NP)') to region 6 (*Liqiang*): the Licensed conditions elicited longer RTs than the Unlicensed conditions. This is expected because the possessive phrase ('his friend') is longer and referentially more complex than a bare pronoun ('him'). No other significant effects were found at the pre-critical regions.

At the critical region (the matrix subject), the main effect of Gender and the Constraint x Gender interaction are not significant. At the first spillover region (region 7), the main effect of Constraint is marginal in the raw RT analysis but significant in the log-transformed analysis ($\beta = -0.04$, SE = 0.02, $t = -2.18$, $p = 0.03$). We find no main effect of Gender. Crucially, the Constraint x Gender interaction is significant. Gender mismatch caused reading slowdowns in the Licensed conditions ($\beta = -33.61$, SE = 15.61, $t = -2.15$, $p < 0.05$). However, in the Unlicensed conditions, gender match caused slowdowns ($\beta = 46.86$, SE = 14.26, $t = 3.29$, $p = 0.001$), exhibiting an intriguing 'gender match' effect. We return to this finding in Section 4.7.

Another interaction characterizes the second spillover region (region 8), due to a gender mismatch effect in the Licensed conditions ($\beta = -32.80$, SE = 9.11, $t = -3.60$, $p < 0.001$) and a gender match effect in the Unlicensed conditions ($\beta = 21.15$, SE = 8.71, $t = 2.43$, $p < 0.05$). No other effects are significant at this region.

Together, these results suggest that gender modulates cataphora resolution by native Chinese speakers in their first language in both Licensed and Unlicensed conditions.

4.7. Discussion

Experiment 1 tested how native Chinese speakers process cataphora in Chinese, in syntactically licensed and unlicensed contexts. In combination with Experiment 2, this experiment provides an L1 Chinese control group that allows us to compare the processing of cataphora in L1 Chinese and L1/L2 English.

The postcedent choice results show that native Chinese speakers strongly prefer a gender-congruent matrix subject in a syntactically licensed position due to a first-referent bias. Crucially, we found that Principle B guides Chinese speakers' postcedent choices as predicted. Interestingly, we find that in the syntactically unlicensed conditions, gender congruency plays a minor role: gender match between the pronoun and the matrix subject led to more matrix subject choices. Although not crucial to our aims and expected given prior work in English (Kush & Dillon, 2021), it may be caused by misretrieval of the gender-congruent matrix subject under similarity-based interference (Gordon et al., 2001, 2004) – as the matrix subject and object have the same gender – or because the effect of Principle B weakens over time. We will not disentangle these two possibilities as they are not critical to our research goals.

For the self-paced reading results, there are two major findings. First, using the gender mismatch effect as a diagnostic, we find

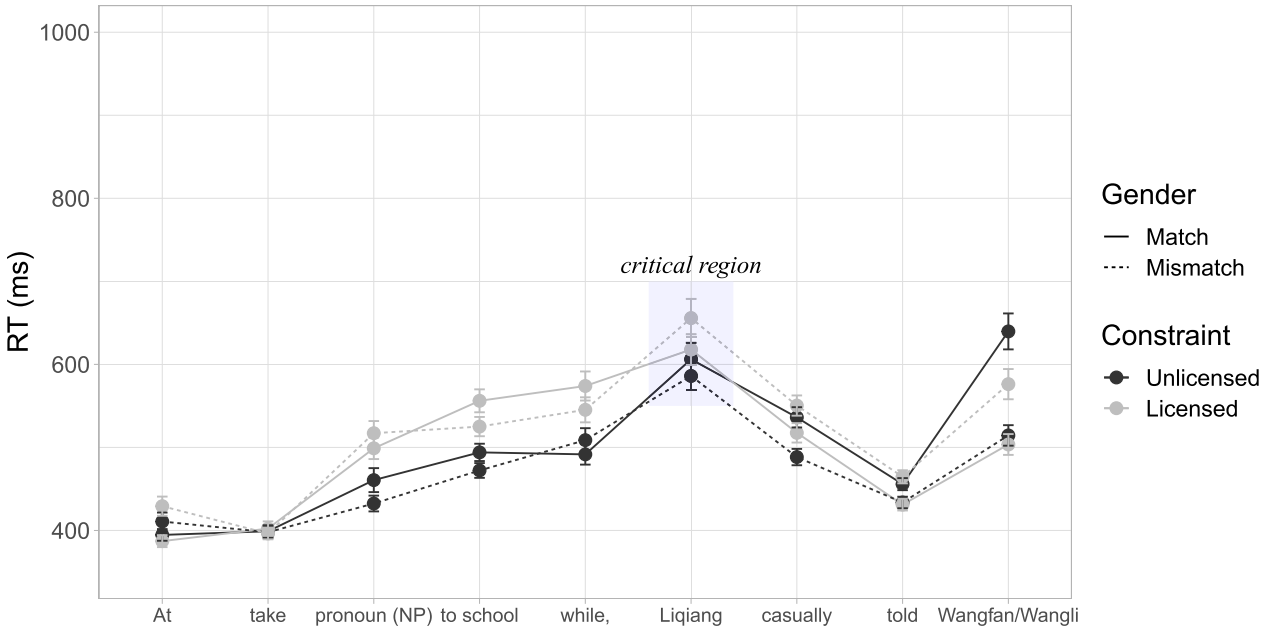


Figure 1. Mean RTs (ms) across regions and conditions for L1 Chinese speakers in Experiment 1.

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

that Chinese speakers attempted to bind the cataphor to the first available postcedent in the Licensed conditions. However, this effect does not surface until the spillover regions and appears small, echoing previous findings of a delayed and reduced effect from Sun and Dennison (2015). These results suggest that Chinese speakers' expectations for syntactically licensed postcedents are less active and weaker, relative to what has been found for native English speakers in prior work (Kush & Dillon, 2021) and what we will report for our group of native English speakers in Experiment 2.[4]

Second, focusing on the Unlicensed conditions, we did not find a gender mismatch effect at the critical region, in line with our predictions. However, we found an unpredicted gender match effect in the two spillover regions. In fact, a prior self-paced reading study on cataphora resolution in Russian found similar gender match effects (Kazanina & Phillips, 2010), which was taken to suggest that the syntactically unlicensed gender-congruent postcedent is temporarily considered but later rescinded or reanalyzed. Here, we additionally speculate that the unpredicted gender match effect may reflect a clash between the gender cue and Principle B during incremental processing (e.g., Badecker & Straub, 2002; also see Kush & Dillon, 2021, for discussion of a gender match effect in their Experiment 2): gender congruency favors the matrix subject as the postcedent, only to be contradicted by syntax which penalizes cataphora, hence the slowdown in the Unlicensed/Match condition.

Overall, Experiment 1 shows that native Chinese speakers can anticipate syntactically licensed postcedents in real-time processing, but the gender mismatch effect is delayed and weak. Importantly, the processing differences in Licensed vs. Unlicensed conditions show that Chinese speakers are sensitive to Principle B, but Principle B may not be inviolable in Chinese in real-time processing, as shown by the gender match effects seen in later-stage (spillover) RTs.

These results raise two questions related to L2 processing. First, do Chinese learners of English show a similar delayed gender mismatch effect in the Licensed conditions in L2 cataphora resolution, especially given the low frequency of cataphora in their L1? Second, will they show a gender match effect in the Unlicensed conditions? We turn to these questions in Experiment 2.

5. Experiment 2

In this experiment, we query to what extent L2 learners' processing patterns resemble those of native English speakers and whether influence from L1 Chinese modulates cataphora resolution in L2 English. We report data from Chinese learners of English and a control group of native English speakers.

5.1. Participants

The L2 learners were 93 adult Chinese learners of English recruited from several universities in China. One was removed due to low comprehension accuracy. The L2 learners were born and raised in China, spoke Chinese as their L1, and were living in China at the time of the experiment. Among the remaining 92 L2 participants, eight reported having lived outside China but had spent no more than 6 months in an English-speaking country, except for three (two spent 7 and 20 months in the USA, one spent 11 months in the UK). All L2 participants learned English in China. The mean age for initial contact with English is 8.09 years (SD = 2.4).

To obtain a measure of L2 proficiency, the L2 participants completed a cloze test (Oshita, 1997) after the self-paced reading experiment. The cloze test consisted of three passages with 75 blanks. Six participants did not finish the cloze test; their data were included for postcedent choice and RT analyses but not in the analysis with proficiency as a predictor. The 86 L2 participants who finished the cloze had an average score of 65.9/75 (SD = 6.2).

The L1 control group consisted of 102 adult native English speakers recruited on Amazon Mechanical Turk. Ten were removed due to low accuracy on filler trials, leaving 92 for the final analysis.

5.2. Materials

The target stimuli in Experiment 2 are structurally similar to those in Experiment 1. The same factors were manipulated. See (4) for an example. (Subscripts indicate presentation regions in the self-paced reading experiment.) Twenty-four sets of target items were mixed with 24 fillers and distributed into 4 lists using Latin square.

(4) a. **Licensed:**
While$_1$ driving$_2$ {his/her} friend$_3$ to$_4$ school,$_5$ Mike$_6$ casually$_7$ told$_8$ {James/Jane}$_9$ an$_{10}$ interesting$_{11}$ story.$_{12}$

b. **Unlicensed:**
While$_1$ driving$_2$ {him/her}$_3$ to$_4$ school,$_5$ Mike$_6$ casually$_7$ told$_8$ {James/Jane}$_9$ an$_{10}$ interesting$_{11}$ story.$_{12}$

5.3. Procedure

For native English participants, the procedure was the same as in Experiment 1. L2 participants additionally completed a cloze test and a language background questionnaire after the self-paced reading task.

5.4. Predictions

We first make predictions for postcedent choices for the two groups. For native English speakers, we predict a main effect of Constraint as they should prefer matrix objects in the Unlicensed conditions. We also predict a Constraint x Gender interaction: the gender effect should be stronger in the Licensed than in the Unlicensed conditions.

As regards L2 learners, if they are sensitive to pronoun gender in English in comprehension (e.g., Rodriguez, 2008; Yu & Dong, 2018) and have knowledge of Principle B, we predict the same choice patterns as native English speakers. Alternatively, the interaction effect may not appear because in the Unlicensed conditions, L2 learners might be less constrained by Principle B in English, compared to native English speakers.

Next, we turn to predictions for RTs. Given prior findings in Kush and Dillon (2021), native English speakers are expected to show an early gender mismatch effect at the postcedent region in the Licensed conditions but no gender effect in the Unlicensed conditions. For L2 learners, RT predictions differ based on: (i) whether L2 learners can anticipate an upcoming syntactically licensed postcedent; (ii) if they can, whether they will be subject to L1 influence; and (iii) whether they will strictly obey Principle B in cataphora resolution.

We hypothesize that if RAGE can be extended to cataphora resolution and if experience with cataphora in L1 Chinese impacts L2 processing, L2 learners should not show an early mismatch

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

effect in the Licensed conditions. In addition, if L2 learners adopt shallower parsing strategies by overusing semantic information and underusing syntactic information (Clahsen & Felser, 2006, 2018), we should see gender mismatch effects in both Licensed and Unlicensed conditions.

5.5. Data analysis

Data analysis was the same as in Experiment 1, except that we included Group (L1/L2) as a between-participant factor (sum-coded, L1: +0.5, L2: −0.5) in the omnibus data analysis. Trimming impacted 1.96% and 2.96% of the raw RTs in L1 and L2 English data, respectively.

5.6. Results

The 92 L1 and 92 L2 participants included in the final analysis answered comprehension questions with 96% and 95% accuracy, respectively. Below, we first report between-group analyses on postcedent choices and RTs to identify potential by-group differences. When differences of interest are found, we then conduct within-group analyses.

5.6.1. Between-group analysis

Postcedent choices

The omnibus analysis reveals main effects of Constraint, Gender, and Group, qualified by Constraint x Gender and Constraint x Group interactions. See supplementary materials (Table A3) for statistics. The Constraint x Gender interaction suggests that the gender congruency effect is weaker in the Unlicensed conditions due to Principle B. Furthermore, the lack of a significant Gender x Group interaction in postcedent choices indicates that L2 English participants utilized gender cues similarly to native speakers. The main effect of Group and the Constraint x Group interaction are due to the fact that, within the Licensed conditions, L2 English speakers preferred matrix subject choices more strongly compared to native speakers ($\beta = 0.51$, $SE = 0.14$, $t = 3.67$, $p < 0.001$); but within the Unlicensed conditions, there is no group-related difference ($\beta = 0.12$, $SE = 0.21$, $t = 0.56$, $p = 0.58$). In other words, when cataphoric binding is syntactically licensed, L2 speakers tend to resolve the dependency earlier than native speakers by choosing the referent linearly closer to the pronoun, presumably due to limited cognitive resources (e.g., Borghini & Hazan, 2018; Declerck & Kormos, 2012).

Reading times

Here, we focus on the critical matrix subject region and the two spillover regions. See supplementary materials (Table A4) for statistics.

At the critical region, we found main effects of Constraint and Group. Crucially, there is a significant Constraint x Gender interaction, which aligns with Kush and Dillon's (2021) findings. (However, the log-transformed analysis did not reveal a significant interaction.) There is an additional Constraint x Group interaction, not crucial to our purposes. The Group x Gender interaction and the three-way interaction are not significant, suggesting that L1 and L2 English participants were similarly sensitive to gender cues in Licensed and Unlicensed conditions at the early processing stages (critical region).

At the first spillover region (e.g., *casually*), in addition to a main effect of Group, the Constraint x Gender interaction, the Constraint x Group interaction, and the Gender x Group interaction all reach significance. The crucial Gender x Group interaction suggests that L1 and L2 English groups reacted differently to gender (mis)match between the pronoun and the matrix subject at this stage. Given this, we analyze the two groups' RT patterns separately in Sections 5.6.2 and 5.6.3. The second spillover region (e.g., *told*) shows similar results, except that there is also a main effect of Constraint.

Table 2. Mean proportions of matrix object choices by L1 English speakers in Experiment 2. Numbers in parentheses indicate standard errors in percentages.

Licensed		Unlicensed	
Match	Mismatch	Match	Mismatch
22% (2%)	51% (2%)	73% (2%)	81% (2%)

5.6.2. L1 English results

Postcedent choices

Table 2 shows the native English speakers' mean proportions of matrix object choices. See supplementary materials (Table A5) for statistics. The statistical model reveals expected main effects of Constraint and Gender, qualified by an interaction. As discussed in our between group-analysis in Section 5.6.1, these results suggest that native English speakers are sensitive to Principle B, similar to L1 Chinese results in Experiment 1.

Reading times

Native English speakers' mean RTs are plotted in **Figure 2**. See supplementary materials (Table A6) for statistics. Before the postcedent region, the main effect of Constraint reaches significance at regions 3 (pronoun (NP)) and 4 (e.g., to), which is not relevant for our predictions: the longer RTs in Licensed conditions are attributable to the possessive phrase (e.g., his/her friend) being phonologically and referentially more complex. The Constraint x Gender interaction at region 3 is probably spurious because gender (mis)match cannot matter before the postcedent appears. No other effects at the pre-critical regions are significant.

At the critical region, we found main effects of Constraint and Gender but no significant interaction. However, planned comparisons show that the main effect of Gender is mainly driven by a gender mismatch effect in the Licensed conditions: gender mismatch led to significant reading slowdowns in the Licensed conditions ($\beta = -43.68$, $SE = 16.77$, $t = -2.61$, $p < 0.01$) but not in the Unlicensed conditions ($\beta = -6.43$, $SE = 15.60$, $t = -0.41$, $p = 0.68$). These results replicate Kush and Dillon's (2021) findings for the critical region. The two spillover regions show the same main effects as well as a significant interaction: gender mismatch modulates cataphora resolution in the Licensed conditions (region 7: $\beta = -50.38$, $SE = 10.69$, $t = -4.71$, $p < 0.001$; region 8: $\beta = -47.39$, $SE = 9.32$, $t = -5.09$, $p < 0.001$). Note that native English speakers did not show any gender match effect in the Unlicensed conditions, unlike L1 Chinese speakers.

5.6.3. L2 English results

Postcedent choices

Table 3 shows the L2 learners' mean preferences for matrix objects. See supplementary materials (Table A7) for statistics. The main effects of Constraint and Gender as well as the interaction are significant, similar to the results for native English speakers. This suggests that L2 learners are similarly sensitive to Principle B as L1 English speakers.

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

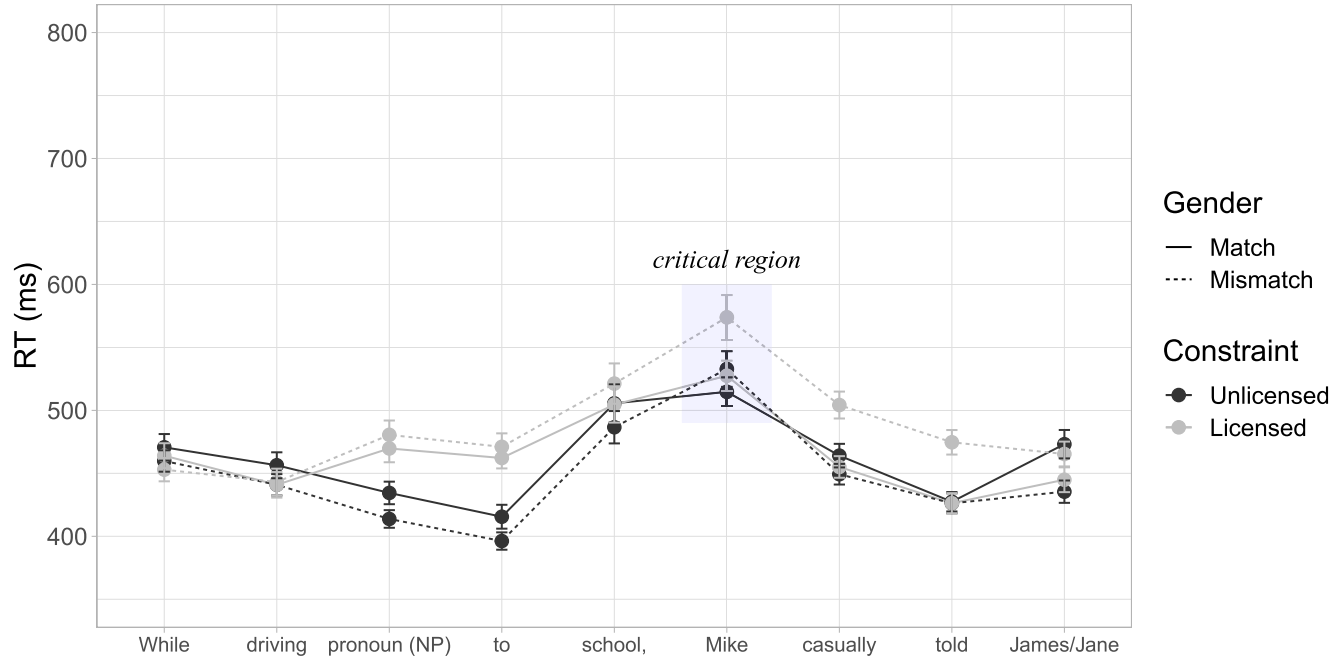


Figure 2. Mean RTs (ms) across regions and conditions for L1 English speakers in Experiment 2.

Reading times

The L2 learners' RTs are plotted in **Figure 3**. See supplementary materials (Table A8) for statistics. Before the critical region, as expected, the main effects of Constraint are significant for regions 3 (*pronoun (NP)*) and 4 (e.g., *to*), due to possessive phrases eliciting longer RTs than bare pronouns. Another main effect of Constraint in the opposite direction characterizes region 2 (e.g., *driving*), probably spurious as this region is identical across conditions. No other effects are significant prior to the critical region.

At the critical postcedent region, there is only a main effect of Constraint: the Unlicensed conditions induced longer RTs than the Licensed conditions. Although not relevant for our purposes, we speculate that the involvement of Principle B in the Unlicensed conditions may lead to extra processing costs for L2 speakers. The Gender x Constraint interaction is only marginal, driven by a numerically trending mismatch effect ($\beta = -43.85$, $SE = 22.94$, $t = -1.91$, $p = 0.06$) in the Licensed conditions and no gender effect in the Unlicensed conditions ($\beta = 17.70$, $SE = 26.48$, $t = 0.67$, $p = 0.50$). We come back to this trending gender mismatch effect in Section 5.6.4.

At the first spillover region, the main effect of Constraint and the interaction are significant. But unlike native English speakers whose RTs show a significant gender mismatch effect in the Licensed conditions, with L2 English learners, we only find a marginal gender mismatch effect in the same conditions ($\beta = -29.64$, $SE = 15.31$, $t = -1.94$, $p = 0.053$). Furthermore, L2 learners differ from native English speakers by exhibiting a striking gender match effect at the first spillover region, echoing what we saw in Experiment 1. Indeed, planned comparisons indicate slowdowns due to gender match in the Unlicensed conditions ($\beta = 53.90$, $SE = 18.49$, $t = 2.92$, $p = 0.004$). As alluded to in our discussion of RTs in Experiment 1, a gender match effect may reflect a clash between Principle B and gender cues, which means that Principle B is not inviolable in online cataphora resolution in Chinese. Observing a similar effect in L2 English processing suggests that Principle B may be similarly violable for Chinese speakers in their L2, at least during the later processing stages.

There are still hints of this gender match effect at the second spillover region, as the Constraint x Gender interaction is marginal. But planned comparisons show that the trending gender match effect in the Unlicensed conditions is not significant ($\beta = 25.12$, $SE = 16.88$, $t = 1.49$, $p = 0.13$).

Table 3. Mean proportions of matrix object choices by L2 English speakers in Experiment 2. Numbers in parentheses indicate standard errors in percentages.

Licensed		Unlicensed	
Match	Mismatch	Match	Mismatch
18% (2%)	36% (2%)	72% (2%)	81% (2%)

5.6.4. Session and proficiency analyses for L2 speakers

L2 learners' RTs show no clear sign of a gender mismatch effect (despite a numerical trend) in contrast to L1 English speakers. In this section, we consider two possible reasons for this and provide two further analyses that explore these possibilities.

First, the possessive pronoun *his/her* in the Licensed conditions is structurally less prominent (more embedded) compared to the object *him/her* in the Unlicensed conditions. Thus, L2 learners might be less likely to attend to gender mismatch in the Licensed conditions, especially at the beginning of the experiment when they are still getting used to the task. In addition, it has been shown that L2 learners' ability to make predictive use of gender cues increases with more trials (Hopp, 2016). To see if L2 learners become more sensitive to the gender mismatch over the course of the experiment, we ran a session analysis comparing the first and second halves of the experiment.

Second, proficiency could also play a role: it could be that the more proficient L2 learners anticipate the upcoming postcedent and exhibit a gender mismatch effect, while the less proficient

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

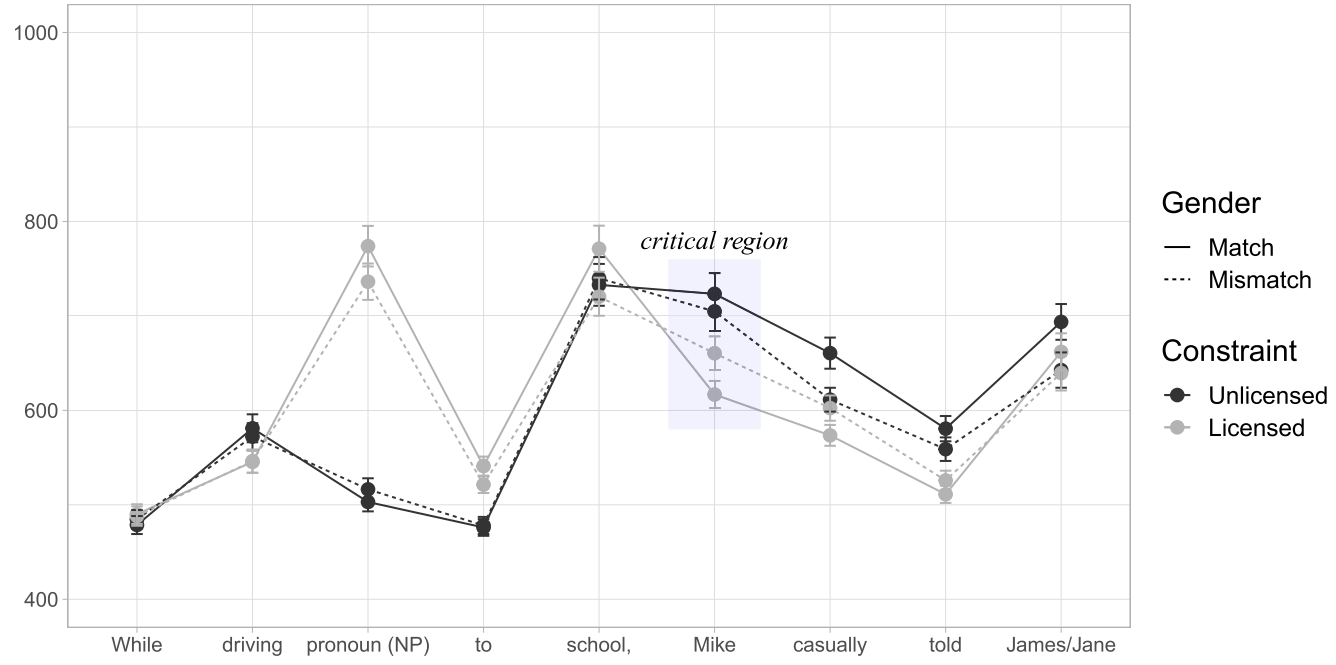


Figure 3. Mean RTs (ms) across regions and conditions for L2 English speakers in Experiment 2.

L2 learners do not. Thus, in the second analysis we use L2 learners' cloze scores as a continuous predictor to see if L2 proficiency modulates the gender mismatch effect.

Session analysis

We analyzed the L2 learners' RTs for the first and second half of the trials separately (**Figure 4**; see supplementary materials Set B for statistics). Here, we focus on gender effects at the critical and spillover regions.

For the first half, only the two spillover regions show significant Constraint x Gender interactions. As shown in Figure 4, gender match elicited longer RTs in the Unlicensed conditions (first spillover: $\beta = 51.79$, $SE = 27.46$, $t = 1.89$, $p = 0.06$; second spillover: $\beta = 48.74$, $SE = 24.08$, $t = 2.02$, $p = 0.04$) while gender mismatch led to longer RTs in the Licensed conditions, which reach significance at the first spillover region ($\beta = -47.95$, $SE = 21.81$, $t = -2.20$, $p = 0.03$).

In the second half, at the critical region, the Constraint x Gender interaction is significant. Planned comparisons indicate that gender mismatch led to reading slowdowns in the Licensed conditions ($\beta = -92.61$, $SE = 30.17$, $t = -3.07$, $p = 0.002$). No gender effect was observed in the Unlicensed conditions. The first spillover region also shows a significant interaction. However, this is caused by the gender match effect in the Unlicensed conditions ($\beta = 56.69$, $SE = 25.95$, $t = 2.19$, $p = 0.03$). No gender effect was observed for the second spillover region. Thus, the RT data suggest that L2 English speakers can promptly anticipate an upcoming postcedent, but their ability to do so (reflected by the smaller gender mismatch effect in the whole session analysis) is overall weaker compared to native English speakers.

Proficiency analysis

To analyze the effect of proficiency, L2 learners' cloze scores were included as a continuous predictor. At the critical and spillover regions, L2 proficiency was not found to correlate significantly with Gender in the RT analyses (see supplementary materials, Table A9). Thus, the L2 learners' anticipatory processing seems unimpacted by their L2 proficiency as measured by the cloze test in our study.

5.7. Discussion

Experiment 2 investigates to what extent L2 learners' processing patterns resemble those of native English speakers and whether influence from L1 Chinese modulates cataphora processing in L2 English.

Postcedent choice patterns of L1 and L2 English speakers are similar. Both groups preferred the first available postcedent in the Licensed conditions when the genders of the pronoun and the matrix subject match. Furthermore, both groups' postcedent choices were guided by Principle B as they strongly preferred matrix objects in Unlicensed conditions.

However, reading times from L1 and L2 English speakers differ. First, while native English speakers showed an early gender mismatch effect in the Licensed conditions, the same gender mismatch effect is weaker in L2 learners. However, this does not mean that L2 learners are insensitive to semantic gender in English. Indeed, as we saw in our session analysis, in the second half of the trials, L2 learners displayed an early gender mismatch effect. In this sense, L2 learners can indeed anticipate postcedents by making use of gender cues. Thus, although L2 speakers show weaker gender mismatch effects, there seem to be no fundamental differences between L2 and L1 English speakers relating to the nature and time-course of cataphora resolution in syntactically licensed environments. These results do not favor a strong interpretation of RAGE (Grüter et al., 2017).

Another difference between L1 and L2 anticipatory processing is that while native English speakers were insensitive to gender cues in Unlicensed conditions, L2 learners showed a gender match effect, similar to L1 Chinese patterns in Experiment 1. Building on our discussion in Sections 4.7 and 5.6.3 as well

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

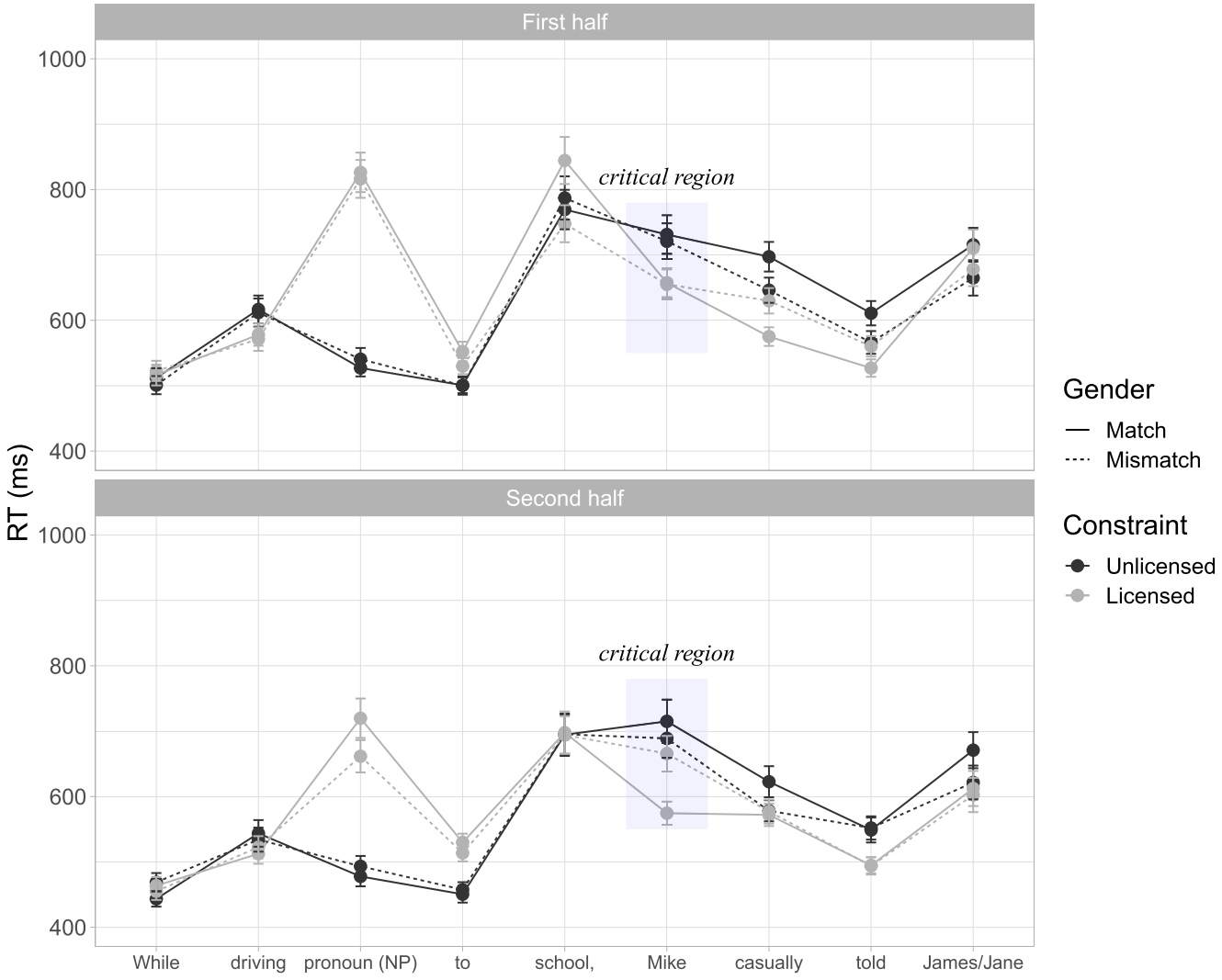


Figure 4. L2 English speakers' mean RTs (ms) across regions and conditions in the first and second half of trials.

as discussion in prior work (Badecker & Straub, 2002; Kush & Dillon, 2021), this gender match effect could suggest that Principle B competes with semantic gender at later processing stages in the L2. Furthermore, we argue that the gender match effect in L2 processing could be due to the influence of L1 Chinese where Principle B cannot completely suppress gender congruency. This could suggest that grammar-based L1 influence – Principle B is less weighted in L2 English due to how it is weighted in L1 Chinese – is harder to overcome than usage-based L1 influence in L2 processing.

However, as aptly pointed out by an anonymous reviewer, the grammar-based L1 influence account ought to be tested by comparison with another L2 group that does not show the gender match effect in the Unlicensed conditions. Until this possibility can be tested, it is important to consider that factors other than L1 influence may cause the gender match effect. For example, in the presence of two same-gendered but non-coreferential referents, L2 speakers might be particularly susceptible to similarity-based interference, which causes reading slowdowns. We leave this to future research.

Together, the results from Experiment 2 indicate that Chinese-speaking learners of English can generate expectations for the upcoming postcedent in cataphora resolution in certain circumstances. But as the session analysis shows, these expectations do not 'kick in' right away. Moreover, L2 learners' expectations for postcedents are constrained by Principle B, like native speakers'. Last but not least, while usage-based L1 influence can be overcome in L2 cataphora resolution, grammar-based L1 influence seems more entrenched.

6. General discussion

We investigated anticipatory L2 processing by examining cataphora resolution in Chinese learners of English. The goals were to examine whether L2 learners can actively anticipate an upcoming postcedent and whether their expectations are constrained by L1 influence and Principle B. We tested cataphora processing in English as an L1 and an L2 as well as in Chinese as an L1.

6.1. Summary of findings

When asked to interpret cataphoric pronouns in postcedent choice tasks, all three groups behaved similarly in two aspects. First, all showed a first-referent bias in the Licensed conditions

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

when the pronoun and the matrix subject match in semantic gender. Second, all were sensitive to Principle B. These findings suggest that, in final-stage judgment, the global interpretation of cataphoric pronouns is similar in Chinese and in L1/L2 English, and that L2 learners have the grammatical knowledge of Principle B in L2 English. However, despite similar post-reading judgment preferences, the three groups exhibited different online processing patterns.

Using the gender mismatch effect as a diagnostic for cataphoric binding in self-paced reading, we observed that, in the Licensed conditions, native English speakers showed immediate sensitivity to gender mismatch between the pronoun and the matrix subject, suggesting that they actively anticipated the postcedent. This replicates Kush and Dillon (2021). In contrast, native Chinese speakers, when processing Chinese sentences, showed a delayed gender mismatch effect at the spillover regions, suggesting that Chinese speakers did not anticipate postcedents as strongly, echoing Sun and Dennison's (2015) findings. We conclude that cataphora resolution in Chinese is constrained by a usage-based linearity constraint (recall findings by Y. Xu & He, 2007; Yuan, 2005) even when cataphora is syntactically licensed.

L2 English learners displayed a more complex pattern, differing from both L1 Chinese and L1 English speakers. In the overall analysis, L2 learners' RTs show weaker (trending but non-significant) gender mismatch effects in the Licensed conditions compared to L1 English speakers; but when we look at the data from the first vs. second half of the trials, an early gender mismatch effect emerges in the second half, suggesting that L2 learners can overcome usage-based L1 influence. These findings suggest that anticipatory processing in L2 and L1 English are not fundamentally different. Therefore, the RAGE hypothesis should not be taken to mean that L2 learners cannot engage in native-like anticipatory processing at ALL linguistic levels.

Another key question for this study is whether L2 learners access syntactic constraints in a timely manner in online cataphora resolution. To this end, we examined the Unlicensed conditions where cataphoric binding by the matrix subject is prohibited by Principle B. Now, native English speakers differ from the other two groups. Consistent with Kush and Dillon (2021), native English speakers showed no gender effect in the Unlicensed conditions – as expected if Principle B strictly constrains processing. But L1 Chinese speakers reading Chinese or L2 English displayed gender match effects, pointing to a weaker role of Principle B in online parsing: semantic gender favoring coreference clashes with a non-absolute Principle B prohibiting coreference. However, these effects only emerged at the spillover regions, which may indicate that the clash only emerges at later stages. But overall, these RT results suggest that L2 learners are nonetheless sensitive to Principle B in early-stage processing.

6.2. Syntactic constraints and the L2 processing of cataphora

One crucial debate in the L2 processing literature is whether L2 learners prioritize syntactic constraints in real-time processing. It has been suggested that L2 learners are less sensitive to syntactic constraints than semantic/pragmatic constraints (e.g., Clahsen & Felser, 2018; Cunnings, 2017; Felser et al., 2009, 2012) – which constitutes the main proposal of the Shallow Structure Hypothesis (SSH) – even though they might have native-like grammatical knowledge. However, most evidence comes from anaphora resolution involving Principle A. There, L2 learners have been found to show a bias for non-local matrix subjects in the early processing stages, violating Principle A (Felser et al., 2009, 2012). Evidence from other syntactic constraints is more mixed, as in the case of processing of Principle C (e.g., Bertenshaw, 2009; Drummer & Felser, 2018; Rodriguez, 2008). In this study, we contributed to this debate by investigating how Principle B constrains cataphora resolution.

Contrary to the claims of the SSH, we found that Chinese learners of English are sensitive to Principle B: in the Unlicensed conditions they did not show gender mismatch effects. The intriguing gender match effect, as suggested earlier, is presumably due to the clash of semantics and syntax (e.g., Badecker & Straub, 2002; Kush & Dillon, 2021). This clash reflects that L2 learners are sensitive to Principle B. However, syntax does not have an absolute advantage in the later processing stages (spillover regions); but at the early processing stages (critical region), Principle B seems to be able to suppress semantics-based parsing in L2 learners. The role of Principle B during the early moments of processing is especially revealing in the second half of the trials for L2 learners: while the gender mismatch effect appeared early at the critical region in the Licensed conditions, no gender-related effect (neither gender mismatch nor match effect) was found in the Unlicensed conditions at this early region. This suggests that although L2 learners could make active use of gender cues, they did not do so when prohibited by syntax at the early stage. As the SSH predicts a gender mismatch effect for (Un)Licensed conditions at the early processing stages, it does not seem to fit well with our results.

6.3. Usage-based and grammar-based L1 influence

In this paper, we identify (at least) two types of L1 influence which seem to work differently in L2 cataphora resolution. USAGE-BASED L1 influence stems from L2 learners' experience with their L1 Chinese, where cataphora is less frequent (Y. Xu & He, 2007; Yuan, 2005) and less expected even when it is syntactically licensed (Sun & Dennison, 2015). Indeed, this is presumably what determines the processing patterns in L1 Chinese in Experiment 1. If this type of influence shapes L2 learners' expectations with cataphora in English, we would expect a delayed gender mismatch effect in L2 processing as well. However, as we saw in the second half of the trials in Experiment 2, L2 speakers of English show a gender mismatch effect at the postcedent region, despite cataphora being infrequent in their L1 Chinese. This suggests that anticipatory processing in the L2 English can be similar to that in L1 English when cataphora is syntactically licensed.

Such usage-based L1 influence is to be distinguished from GRAMMAR-BASED L1 influence. In our study, the source of grammar-based L1 influence is linked to the relative weighting of Principle B during online parsing in L1 Chinese. As suggested above, the weight of Principle B during incremental processing may be lower in L1 Chinese, relative to L1 English. This is reflected by the gender match effect where Principle B cannot completely suppress gender congruency during the later processing stages. Assuming that L2 learners make use of the same set of semantic and syntactic constraints in their L1 (Chinese) and L2 (English), one can imagine that the relative weighting of Principle B may be transferred from their L1 to L2. Indeed, this account is compatible with the proposal that properties of retrieval cues (here, structural and gender cues) used to identify potential referents can be transferred from learners' L1 (Cunnings, 2017). Thus, an intriguing finding from our study is that grammar-based influence seems harder to overcome

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

compared to usage-based influence. We regard this as a valuable direction for future research.

At any rate, it seems that Principle B outranks gender congruency in L1 English processing but not in L1 Chinese or L2 English processing in our study. However, a deeper question underlying the transfer of relative weights from L1 Chinese during processing of L2 English is: why would languages exhibit such differences? One possibility is that because Chinese has more flexible word order and less morphosyntactic features, semantic factors may be more relied upon for the interpretation of sentences than in English; thus, semantic information could carry more weight relative to structural information in Chinese than in English, especially during sentence processing. We will investigate this hypothesis further in our future research.

7. Conclusions

We investigated three open questions in L2 anticipatory processing by looking at cataphora resolution in Chinese learners of L2 English, as well as control groups of L1 Chinese and L1 English speakers. Reading time data show that L2 learners can anticipate upcoming postcedents in the syntactically licensed position, even though cataphora is infrequent in their L1. This suggests that RAGE (e.g., Grüter et al., 2017; Kaan, 2014; Kaan et al., 2010) should be assessed carefully before being extended to the syntactic level. Importantly, our results show that L2 learners are sensitive to Principle B in incremental processing, indicating that they can activate this constraint in a timely manner. The late-stage difference in L1 and L2 English cataphora resolution could reflect grammar-based L1 influence. Overall, this study supports the view that anticipatory processing in the L2 is not fundamentally different from that in the L1.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1366728924000208.

Abbreviations. ASP – aspect marker; CL – classifier; DE – a marker in noun phrases denoting a modification relation

Data availability. The data, analyses, and the supplementary materials are available at https://osf.io/b6gs7/.

Acknowledgements. We thank the anonymous reviewers and the handling editor for helpful questions and insightful comments. We also thank Audrey Li, Andrew Simpson, Heather Marsden, and the audience at the 28th annual meeting of *Architectures and Mechanisms of Language Processing* (AMLaP) for constructive feedback on earlier versions of this work. Finally, we are grateful to Shuo Feng for assistance in recruiting participants at Peking University.

Ethics and consent. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The experiments involving human subjects in this study were reviewed and approved by the Institutional Review Board at the authors' institutions.

Competing interests. The authors have no competing interests to declare.

Notes

1 In Chinese, the null subject could be a *pro* or PRO, but there are no interpretive differences between *pro* and PRO (Huang, 1982). We also highlight the findings of prior work showing that null subjects in fronted adverbial clauses strongly corefer with matrix subjects in cataphoric constructions (e.g., Sun & Dennison, 2015; C. Wang, 2006).

2 SD-based trimming was not applied for log-transformed analyses following Nicklin and Plonsky (2020).

3 Due to space, summaries of inferential statistics for this study are in supplementary materials.

4 An anonymous reviewer asks whether the reduced/delayed gender mismatch effect could be because gender cues are not relied upon in Chinese as much as they are in English. As mentioned in Section 2.2, Sun and Dennison (2015) found contrasting patterns for online anaphora and cataphora resolution, with the gender mismatch effect appearing earlier and stronger in anaphora resolution.

References

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

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*.748–769.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015b). Parsimonious mixed models. arXiv preprint arXiv, 1506.04967.

Bertenshaw, N. (2009). The Application of Binding Constraints by Japanese L2 Learners of English. Ph.D. dissertation. University of Essex.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*, 152.

Chomsky, N. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chun, E., & Kaan, E. (2019). L2 Prediction during complex sentence processing. *Journal of Cultural Cognitive Science*, *3*, 203–216.

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

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*(4), 569–678.

Cunnings, I., & Sturt, P. (2014). Coargumenthood and the processing of reflexives. *Journal of Memory and Language*, *75*, 117–139.

Cunnings, I., Batterham, C., Felser, C., & Clahsen, H. (2010). Constraints on L2 learners' processing of *wh*-dependencies. In B. VanPatten & J. Jegerski (eds.), *Research in Second Language Processing and Parsing* (pp. 95–118). Amsterdam: John Benjamins.

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language & Cognition*, *15*(4), 782–796.

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

Drummer, A., & Felser, C. (2018). Cataphoric pronoun resolution in native and non-native sentence comprehension. *Journal of Memory and Language*, *101*, 97–113.

Drummond, A. (2013). *Ibex Farm* (Version 0.3.9). http://www.spellout.net/

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.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*, 353–387.

Felser, C., Sato, M., & Bertenshaw, N. (2009). The online application of Binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, *12*(4), 485–502.

Felser, C., Cunnings, I., Batterham, C., & Clahsen, H. (2012). The timing of island effects in nonnative sentence processing. *Studies in Second Language Acquisition*, *34*, 67–98.

Frenck-Mestre, C., & Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *50A*(1), 119–148.

Giskes, A., & Kush, D. (2021). Processing cataphors: Active antecedent search is persistent. *Memory & Cognition*, *49*, 1370–1386.
Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(6), 1411–1423.
Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language*, *51*(1), 97–114.
Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215.
Grüter, T., Rhode, H., & Schafer, A. J. (2017). Coreference and discourse coherence in L2: The roles of grammatical aspect and referential form. *Linguistic Approaches to Bilingualism*, *7*(2), 199–229.
Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*, 901–931.
Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56.
Hopp, H. (2016). Learning (not) to predict: Grammatical gender processing in second language acquisition. *Second Language Research*, *32*(2), 277–307.
Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, *7*(1), 96–130.
Hopp, H., & Lemmerth, N. (2016). Lexical and syntactic congruency in L2 predictive gender processing. *Studies in Second Language Acquisition*, *40*(1), 171–199.
Huang, C.-T. J. (1982). Logical relations in Chinese and the Theory of Grammar. Ph.D. dissertation. MIT.
Huang, C.-T. J., & Lin, J.-W. (2021). Quantificational binding without surface c-command in Mandarin Chinese. In F. Si & L. Rizzi (eds.), *Current Issues in Syntactic Cartography: A Crosslinguistic Perspective* (pp. 184–215). Amsterdam: John Benjamins.
Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*(2): 257–282.
Kaan, E., Dallas, A., & Wijnen, F. (2010). Syntactic predictions in second-language sentence processing. In C. J. Zwart & M. de Vries (eds.), *Structure Preserved: Studies in syntax for Jan Koster* (pp. 207–214). Amsterdam: John Benjamins.
Kazanina, N., & Phillips, C. (2010). Differential effects of constraints in the processing of Russian cataphora. *Quarterly Journal of Experimental Psychology*, *63*, 371–400.
Kazanina, N., Lau, E. F., Lieberman, M., Masaya, Y., & Phillips, C. (2007). The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language*, *56*, 384–409.
Kennison, S. M., Fernandez, E. C., & Bowers, J. M. (2009). Processing differences for anaphoric and cataphoric pronouns: Implications for theories of referential processing. *Discourse Process 46*(1), 25–30.
Kim, H., & Grüter, T. (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, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31*(1), 32–59.
Kush, D., & Dillon, B. (2021). Principle B constrains the processing of cataphora: Evidence for syntactic and discourse predictions. *Journal of Memory and Language*, *120*, 104254.
Kush, D., Lidz, J., & Phillips, C. (2017). Looking forwards and backwards: The real-time processing of Strong and Weak Crossover. *Glossa: a journal of general linguistics*, *2*(1), 70. https://doi.org/10.5334/gjgl.280
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. R package (Version 2.0). http://CRAN.R-project.org/package=lmerTest
Lago, S., Mosca, M., & Garcia, A. S. (2021). The role of crosslinguistic influence in multilingual processing: Lexicon versus syntax. *Language Learning*, *71*(1), 163–192.
Lau, E., & Grüter, T. (2015). Real-time processing of classifier information by L2 speakers of Chinese. In E. Grillo & K. Jepson (eds.), *Proceedings of the 39th Annual Boston University Conference on Language Development* (311–323). Somerville, MA: Cascadilla Press.
Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.
Liang, L., Wen, Y., & Dong, Y. (2018). Gender constraint in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. *Journal of Neurolinguistics*, *45*, 1–12.
Marinis, T., Roberts, L., Felser, C., & Clahsen, H. (2005). Gaps in second language processing. *Studies in Second Language Acquisition*, *27*(1), 53–78.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. doi: 10.1016/j.jml.2017.01.001
Mitsugi, S. (2018). Generating predictions based on semantic categories in a second language: A case of numeral classifiers in Japanese. *International Review of Applied Linguistics in Language Teaching*, *58*(3), 323–349.
Nicklin, C., & Plonsky, L. (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-Analysis. *Annual Review of Applied Psycholinguistics*, *40*, 26–55.
Omaki, A., & Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Studies in Second Language Acquisition*, *33*(4), 563–588.
Oshita, H. (1997). "The Unaccusative Trap": L2 Acquisition of English Intransitive Verbs. Ph.D. dissertation. University of Southern California.
Puebla, C., & Felser, C. (2022). Discourse prominence and antecedent misretrieval during native and non-native pronoun resolution. *Discours*, *29*, 1–31.
Qiu, L., Swaab, T. Y., Chen, H. C., & Wang, S. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *Plos ONE*, *7*(5), e36156.
Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin 114*(3), 510–532.
R Core Team. (2020). *R: A language and environment for statistical computing*. http://www.R-project.org/
Roberts, L. (2013). Sentence processing in bilinguals. In R. P. G. van Gompel (ed.), *Sentence Processing* (pp. 234–253). Hove: Psychology Press.
Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*, 333–357.
Rodriguez, G. A. (2008). Second Language Processing: Is It Fundamentally Different? Ph.D. dissertation. University of Pittsburgh.
Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*(3), 542–562.
Sun, B., & Dennison, S. M. (2015). Comprehending pronouns in Chinese: Evidence for cross-language differences in referential processing. *Language Sciences*, *47*, 56–65.
Tai, J. H.-I. (1973). Chinese as a SOV language. In C. Corum, T. C. Smith-Stark, & A. Weiser (eds.), *Papers from the 9th Regional Meeting of the Chicago Linguistic Society* (pp. 659–671). Chicago: Chicago Linguistic Society.
Wang, Z. (1994). *Yingyu rencheng daici he/she neng yuzhi xiawen zhong de mingci ma?* [Can English pronoun he/she refer to the noun in the upcoming discourse?] *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, *4*, 36–39.
Wang, C. (2000). *Rencheng daici "ta" de zhaoying gongneng yanjiu* [On the anaphoric function of the personal pronoun *ta*]. *Zhongguo yuwen [Studies of the Chinese Language]*, *3*, 228–237.
Wang, C. (2006). *Yinghan di san rencheng daici zhaoying de danxiangxing ji xiangguan wenti* [Unidirectionality of the anaphora of the third person pronouns in English and Chinese]. *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, *38*(1), 17–24.
White, L. (2000). Second language acquisition from initial to final state. In J. Archibald (ed.), *Second Language Acquisition and Linguistic Theory* (pp. 130–155). Oxford: Blackwell.
Wu, D., Wu, M., Gao, X., & Zhang, J. L. (2019). *Eryu qianzhaoying he houzhaoying daici de zaixian jiagong yanjiu* [The online processing of anaphoric and cataphoric pronouns in L2]. *Waiyu yu waiyu jaoxue [Foreign Languages and Their Teaching]*, 90–100.
Xu, Y., & He, X. (2007). *Yinghan yu xiazhi de pianzhang gongneng he yuyong fenxi* [Cataphora in English and Chinese texts: A functional-pragmatic

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press

analysis]. *Waiyu jiaoxue yu yanjiu [Foreign Language Teaching and Research]*, *39*(6), 417–423.

Xu, X., Jiang, X., & Zhou, X. (2013). Processing Biological Gender and Number Information during Chinese Pronoun Resolution: ERP Evidence for Functional Differentiation. *Brain and Cognition*, *81*(2), p.223–236

Yu, Z., & Dong, Y. (2018). The neurocognitive mechanisms of semantic gender processing in L1 and L2 personal nouns by Chinese-English bilinguals. *Language, Cognition and Neuroscience*, *34*(7), 901–915.

Yuan, Y. (2005). A Contrastive Study of Backwards Anaphora in English and Chinese. MA Thesis. Shanghai International Studies University.

https://doi.org/10.1017/S1366728924000208 Published online by Cambridge University Press