Proceedings of CLS 56 (2020), 295-307

Processing of anaphora and cataphora in weak crossover in English (non-)native speakers

Jun Lyu[1], Jiwon Yun[2], Daniel Finer[2]
[1]University of Southern California and [2]Stony Brook University

1 Introduction

Pronoun resolution has received a lot of attention in psycholinguistic research as the interpretation of a pronoun depends on various factors. Among the factors known to impact pronoun resolution, syntax has been argued to play a pivotal role. We know this from experimental work which found that the Binding Principles (Chomsky 1981) constrain the real-time processing of anaphors (Badecker & Straub 2002; Sturt 2003; Felser & Cunnings 2012) and cataphors (Kazanina *et al.* 2007; Kazanina & Phillips 2010; Drummer & Felser 2018). Besides syntax, the position in which they appear relative to their binder (i.e. linear precedence) may also impact the interpretation process and/or outcome (Bresnan 1994;Bresnan 1998; Kazanina & Phillips 2010; Drummer & Felser 2018). This is because while anaphora resolution involves retrospective search where the language parser needs to identify and retrieve the optimal antecedent from the working memory, cataphora resolution involves prospective search. In the former, parsing is more prone to errors as resolution happens only once. The intervening nominal elements might thus all interfere and compete with the legitimate binder (Lewis & Vasishth 2005). In the case of cataphora resolution, each potential binder is assessed once at a time and can be discarded if deemed inappropriate. Despite the difference in search mechanisms, anaphora and cataphora resolution have been both shown to obey syntactic constraints in several studies. The modular account argues that this is because syntax filters out potential antecedents in syntactically inaccessible positions at the early processing stage (Sturt 2003; Chow *et al.* 2014; Kush *et al.* 2015), although this was challenged by the multiple-constraint account which posits that non-syntactic factors also modulate real-time processing (Runner *et al.* 2006; Kaiser *et al.* 2009; Jäger *et al.* 2015; Patil *et al.* 2016).

Previous research on pronoun resolution has produced interesting insights and meaningful debates. However, we can identify some limitations. First, the scope of research was largely restricted to the Binding Principles although exceptions exist (e.g., Kush *et al.* 2017; Felser & Drummer 2017). Second, there was not enough attention on anaphora and cataphora processing underlied by the same syntactic constraint. Third, real-time pronoun resolution by non-native speakers has not been well understood, especially for cataphora resolution. The last point is especially interesting as crosslinguistic variations exist in licensing cataphoric dependencies. For example, Wang (1994) and Wang (2006) claimed that, based on native intuition, cataphoric coreference between the pronoun ta ('s/he') in Chinese and its antecedent is not allowed. In this case, one might wonder whether negative transfer prevents Chinese learners of English from acquiring the cataphoric dependencies in

a second language (L2). In this study, we will address these gaps in our knowledge by examining the native and non-native processing of anaphors and cataphors in English weak crossover (WCO).

In a WCO configuration, an operator or binder moves across a co-indexed pronoun. In the theoretical literature, a bound reading of the pronoun is assumed to be unavailable (Koopman & Sportiche 1982; Safir 1984; Bresnan 1994; Bresnan 1998; Ruys 2000; Barker 2012). For an example of WCO, see (1b):

(1) a. I knew the waitress$_i$ who$_i$ t$_i$ said that her$_{i/j}$ friend had divorced.
b. I knew the waitress$_i$ who$_i$ it seemed that her$_{*i/j}$ friend had divorced t$_i$.

Given space limitation, we will briefly mention a subset of WCO theories although the findings of this paper will be relevant to the others as well. Koopman and Sportiche (1982) proposed that the WCO effect was due to the Bijection Principle which says that an operator cannot simultaneously bind two variables. This principle was later modified by Safir (1984) to hold only between variables of different [αpronominal] classes as in (1b). Another influential school of thought based on syntactic notions (i.e. c-command) is attributed to Reinhart (1983) who argues that a pronominal variable must be bound from an A-position. In (1a), the trace in an A-position binds the pronoun *her*. The trace is then licitly bound by the *wh*-operator. Thus, pronominal binding can be established via transitivity. In (1b), however, there is no intermediate landing site that can license pronominal binding, hence its degradedness. Aside from the syntactic accounts, scope (Sauerland 1998; Ruys 2000; Barker 2012) and linearity (Bresnan 1994; Bresnan 1998; Jäger 2005; Shan & Barker 2006) have also been argued to contribute to the anti-binding property of WCO. Among these theories, Barker's scope-plus-reconstructed-linearity approach is a general constraint on pronominal binding in all syntactic structures, including that of WCO. Barker argues that the operator must scope over and precede the pronoun after reconstruction. Therefore, the degradedness of binding in (1b) is because after reconstruction the binder follows the pronoun.

What is particularly relevant to our concern is whether comprehenders can foresee the WCO configuration before reaching the trace. To identify the WCO configuration, the parser must predict the presence of an unfilled gap by the time the pronoun is revealed. Whether L2 speakers can predict the presence of a gap before it is encountered and whether they can represent such complex structures as in (1b) are currently under debate. Probing L2 learners' processing patterns can also inform us whether their real-time interpretation is sensitive to the WCO constraint, a question that is as important in L1 processing as it is in L2.

2 L2 processing Theories

In this paper, we will consider two processing theories relevant to the questions we aim to address. One can notice that the WCO configuration in (1b) is fairly complex as embedding and extraction are involved. According to the *Shallow Structure Hypothesis* or SSH (Clahsen & Felser 2006), L2 learners have difficulty in representing discontinuous dependencies with hierarchical structures. This claim was built on early work on filler-gap dependencies (Marinis *et al.* 2005; Felser & Roberts

2007) and parasitic gaps in subject islands (Boxell & Felser 2017). But, over the years, counterevidence has prompted Felser (2015) and Clahsen & Felser (2018) to reassess and refine their earlier claims about SSH. The most recent version of SSH revolves around two new claims.

The first novel claim is that L2 learners have difficulty with retrospective dependencies but not with prospective dependencies. The crucial difference between the two is that in the former, "there is usually not any need for the processing system to verify whether the required c-command relationship does indeed obtain" (Felser 2015: 13). In other words, in a prospective dependency (e.g., "The nurse who$_i$ the doctor argued t$_i$ that the rude patient had angered t$_i$ is ..."), the *wh*-operator sits in a higher position and the c-command relationship comes for free. So, when L2 speakers do not need to mobilize syntax, the processing of non-local dependencies would be native-like. This contrasts with retrospective dependencies where looking backward involves identifying syntactic relationships (Felser *et al.* 2009; Felser & Cunnings 2012). The second novel claim about SSH is that L2 learners rely more on non-syntactic cues (e.g., topic prominence, semantic cues) than on syntactic cues. This argument ties in nicely with the first claim. The reason for the non-native performance of L2 speakers with retrospective dependencies is exactly because L2 speakers under-use syntax. For example, Felser & Cunnings (2012) in an eye-tracking study on the processing of English reflexives suggests that German native speakers were more sensitive to topic/discourse prominence than Condition A, although German obeys Condition A. Therefore, SSH predicts that L2 learners will be unable to rule out binding in retrospective WCO sentences like (1b) because plausibility and semantic feature match have more weights than the WCO constraint. Furthermore, SSH predicts that cataphoric dependencies as in (2) are also susceptible to erroneous interpretation where the quantifier phrase (QP) binds the cataphor due to "shallow" interpretation based on feature match.

(2) The toy that his$_i$ parents bought made every boy$_{*i/j}$ in the room happy.

Contrary to SSH, the *Reduced Ability to Generate Expectations* model or RAGE (Grüter *et al.* 2014) posits that there is no deficit in L2 speakers' ability to represent complex structures. While SSH claims that the non-native-like processing of filler-gap dependencies is due to reduced sensitivity to syntactic cues, the RAGE model argues that this is because L2 learners under-generate expectations for the upcoming syntactic structures, and that their impoverished ability to generate expectations is due to limited cognitive resources. Although RAGE was originally used to account for the recency bias in L2 learners' pronoun coreference in offline sentence completion tasks (e.g., "Patrick$_i$ gave/was giving a towel to Ron$_j$. He$_{i/j}$..."), Kaan *et al.* (2010) and Kaan (2014) extended the RAGE model to real-time processing. The evidence comes from ERP studies which showed that the early left anterior negativity (ELAN) was absent in L2 processing (Hahne & Friederici 2001) and from self-paced reading studies which suggest filler-gap effects for advanced L2 learners appeared with some delay (Dallas, 2008). Interestingly, studies on the positioning of classifier phrases in Chinese relative clauses also show that even advanced L2 learners cannot predict the positioning of classifiers depending on the type of the relative clauses (Wu & Sheng 2014; Wu & Lyu 2016), lending further evidence to

RAGE. To revisit the WCO structures in (1b) and (2), RAGE predicts that, depending on their proficiency level or cognitive resources, L2 learners might not be able to predict the WCO structure by the time they reach the pronoun in (1b) and thus insensitive to the WCO violation (or sensitive to the violation with some lag), but they will not attempt to bind the cataphora in (2) as their representation of syntax is non-defected.

3 Experiments

The aim of the experiments is to investigate the real-time interpretation of pronouns in WCO in native/non-native speakers of English. A self-paced reading paradigm was adopted. Although there were two types of dependencies, anaphoric and cataphoric, these two experiments were merged into one. Cataphoric WCO sentences acted as fillers to anaphoric sentences, and vice versa. Nevertheless, the results for these two experiments will be reported separately.

3.1 Participants

Twenty-seven English native speakers (mean age = 20, SD = 1.12) in the Stony Brook University community participated as the L1 group. Thirty Chinese natives (mean age = 22, SD = 1.37) of English majors recruited online from different Chinese universities participated as the L2 group. Table 1 below shows the English proficiency backgrounds of these L2 participants.

Proficiency Test	Mean score/Max score	SD	Number
CET-6	590.18/710	44.20	11
TEM-4	78.56/100	5.14	9
TEM-8	69.60/100	7.83	10

Table 1: Language proficiency backgrounds of the L2 participants.

3.2 Experiment 1: Anaphoric dependency

3.2.1 Materials

Gender Congruency (Match/Mismatch) and *Clause Type* (Complement/WCO) were crossed in a 2 x 2 factorial design. When necessary, the third factor *Group* (L1/L2) was included to identify L1/L2 differences. See (3) for a set of examples of anaphoric WCO. Gender mismatch effect (GMME) was used as a diagnostic for coreference. For example, in the licensing complement clauses, if pronoun *her* caused longer reaction time (RT) in the mismatch condition (e.g., "...waiter...her..."), this would suggest that binding is attempted. Otherwise, there should be no difference in RTs related to the processing of the pronoun in a minimal pair. Crucially, if the WCO constraint blocks binding, no GMME should occur. In other words, we would expect an interaction of *Gender Congruency* and *Clause Type*. It should be noted that the WCO structure in the current experiment is situated in an RC, different from previous experiments which used indirect questions (Kush *et al.* 2017; Felser & Drummer 2017; Lyu 2020).

(3) a. *Complement/(Mis)match*: Jason$_1$/ talked to$_2$/ the {waitress/waiter}$_3$/ who$_4$/ heard$_5$/ that$_6$/ her$_7$/ friend$_8$/ had$_9$/ divorced$_{10}$/ last week.$_{11}$

b. *WCO/(Mis)match*: Jason$_1$/ talked to$_2$/ the {waitress/waiter}$_3$/ who$_4$/ it seemed$_5$/ that$_6$/ her$_7$/ friend$_8$/ had$_9$/ divorced$_{10}$/ last week.$_{11}$

Twenty-four sets of stimuli as in (3) as well as 24 sets of cataphoric sentences were created (see Section 3.3.1). Whenever possible, nouns with unambiguous gender roles (e.g., waiter/waitress) were used as local antecedents. The matrix subject (i.e. "Jason") always disagrees in gender with the pronoun to force participants to only consider the local antecedent.

3.2.2 Procedure

The experiment was hosted on Ibex Farm (Drummond 2013). The stimuli were presented pseudo-randomly in a Latin square design to the participants. Anaphoric and cataphoric sentences were shuffled and interleaved for presentation. Each participant saw one condition per item. After each sentence, a comprehension question appeared to make sure participants were paying attention.

3.2.3 Results

Before data analyses, participants whose comprehension accuracies below 70% were excluded. This procedure left out 1 native and 2 non-native participants. For the rest of the data, RTs smaller than 80ms and 2.5 SDs above the mean for each region were excluded, which lead to the deletion of 3.14% and 7.04% of the L1 and L2 data. Data analyses were carried out in R (R Core Team 2018) using mixed-effect linear models provided by the lme4 package. Figure 1 shows the mean RTs across all conditions starting from the pre-critical region is (i.e. "that"). Before the critical pronoun region, two regions showed significant effects for the native group only. At the region 4 (i.e. "who"), the main effect of *Gender Congruence* was significant (β = -49.29, SE = 17.24, t = -2.86, $p < 0.005$), reflecting the processing effort of the preceding words ("waiter/waitress") which differed in word length and frequency. The main effect of *Clause Type* was significant at region 5 (e.g., "heard/it seemed") because "it seemed" took more time to read than verbs like "heard" (β = 61.62, SE = 20.07, t = 3.07, $p < 0.005$).

At the critical pronoun region, the native group showed no effects. But at the first spillover region, gender mismatch lead to significant slowdowns (β = 49.17, SE = 19.98, t = 2.46, $p < 0.05$). The absence of significant interaction indicates that binding was attempted for both types of structures at first. However, at the second spillover region, the interaction reached significance (β = -41.38, SE = 20.19, t = -2.05, $p < 0.05$) in addition to the main effects of *Gender Congruence* and *Clause Type* (ps < .05). Post-hoc pairwise comparisons revealed that GMME only occurred in the licensing conditions, but not in the WCO conditions.

The L2 group displayed a different processing pattern. At the critical region, the interaction already reached significance (β = 128.02, SE = 60.69, t = 2.12, $p < 0.05$), it was the WCO structures that showed GMME in the pairwise comparison (β = 128.93, SE = 45.33, t = 2.84, $p < 0.005$). The following two spillover regions were characterized by significant main effects of *Gender Congruence* (ps < .05) while

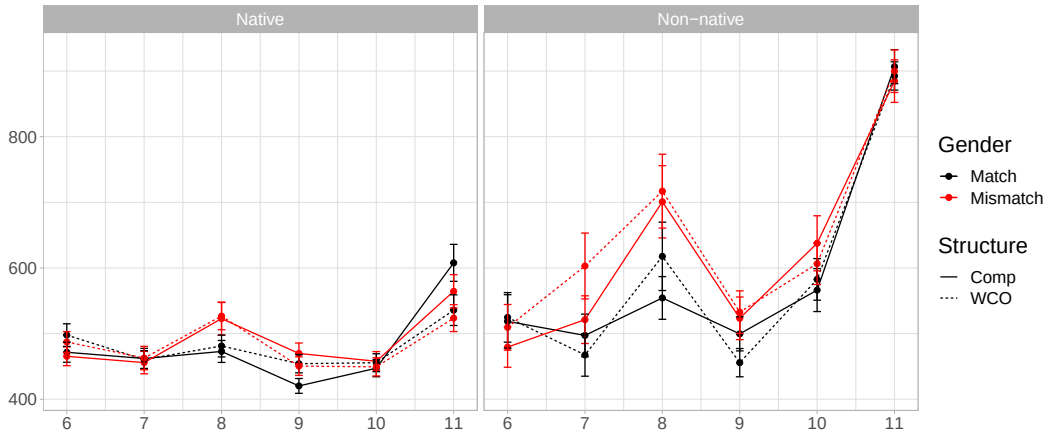

Figure 1: Mean RTs across conditions for L1/L2 participants in Exp. 1. Region 7 is the pronoun.

the main effects of *Clause Type* and the interactions were not significant. No other significant results were found at the next two regions although, numerically, the GMME at the third spillover seems to be weaker in the WCO conditions compared to the complement controls (24ms vs. 72ms).

To summarize, English native speakers displayed a general GMME at the target pronoun region followed by a significant interaction, suggesting that the WCO constraint constrained online parsing, but at a delay. This contrasts with the L2 processing pattern where WCO constraint was absent, except a numerical GMME trend at the third spillover region.

3.3 Experiment 2: Cataphoric dependency

3.3.1 Materials

Gender Congruency (Match/Mismatch) and *Antecedent Type* (DP/QP) were similarly crossed in a Latin square. Gender (mis)match was likewise used to diagnose coreference. The DP conditions were the licensing conditions while the QP conditions represented WCO with covert quantifier raising. *Group* (L1/L2) will be included as a between-group factor to assess L1 and L2 differences when needed. Twenty-four sets of stimuli were created. See (4) for an example set:

(4) a. *DP/(Mis)match*: The toy$_1$/ that$_2$/his$_3$/ parents$_4$/ bought$_5$/ made$_6$/ the {boy/girl}$_7$/ in$_8$/ the room$_9$/ happy.$_{10}$

b. *QP/(Mis)match*: The toy$_1$/ that$_2$/ his$_3$/ parents$_4$/ bought$_5$/ made$_6$/ every {boy/girl}$_7$/ in$_8$/ the room$_9$/ happy.$_{10}$

3.3.2 Procedure

The procedure was the same as in Experiment 1.

3.3.3 Results

Similar to the previous experiment, RTs below 80ms and 2.5SDs above the mean by region were excluded, which lead to the deletion of 2.96% and 5.57% of the L1

and L2 RT data. Figure 2 dispalys the mean RTs across all conditions by region for L1 and L2 participants.

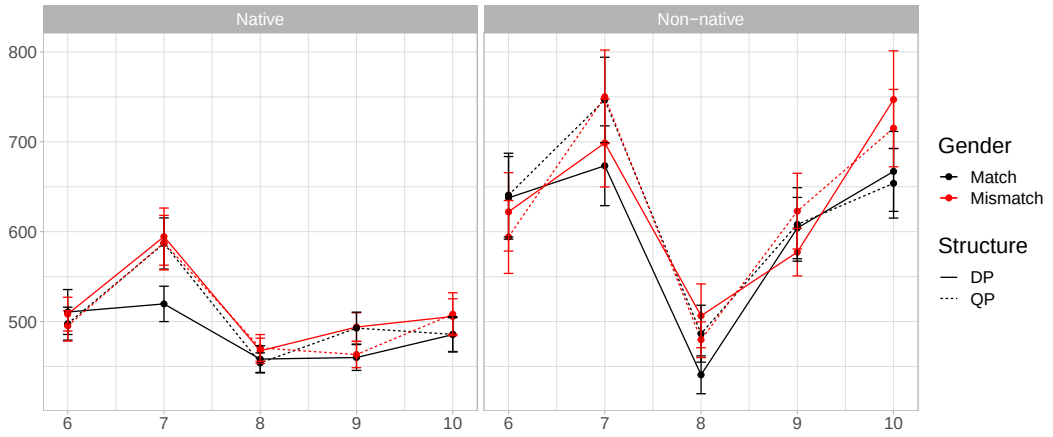

Figure 2: Mean RTs across conditions for L1/L2 participants in Exp. 2. Region 7 is the antecedent.

No region prior to the critical antecedent region showed significant effects. For the native group, at the critical region, the main effects of *Gender Congruence* and *Clause Type* reached significance level (ps < 0.05). The interaction was marginally significant (β = -83.39, SE = 40.30, t = -1.93, p = 0.05). Pairwise comparisons suggest that the GMME was significant in the DP conditions but not in the QP conditions. Thus, native speakers were sensitive to the WCO constraint right at the antecedent position. At the second spillover region (i.e. "the room"), the GMME effect was marginally significant (β = 30.00, SE = 15.69, t = 1.91, p = 0.05), in addition to a marginally significant main effect of *Clause Type* (β = 25.89, SE = 15.20, t = 1.70, p = 0.08) and a significant interaction (β = -50.53, SE = 21.54, t = -2.35, p < 0.05). In Figure 2, it seems that the slowdowns were caused by gender mismatch in the DP conditions but by gender match in the QP conditions. Pairwise comparisons showed that only the GMME in the DP structures was statistically meaningful (β = 29.76, SE = 15.66, t = 1.90, p = 0.05). The 28ms gender match effect (GME) did not reached significance.

Next, we turn to the L2 results. At the critical region, the QPs took more time to read than the DPs although this slowdown was not significant. At the first spillover region, the *Gender Congruence* (β = 66.08, SE = 34.34, t = 1.92, p = 0.05) and the interaction (β = -62.73, SE = 38.65, t = -1.73, p = 0.09) were marginally significant, because GMME lead to slowdowns only the in DP conditions (β = 67.85, SE = 37.24, t = 1.82, p = 0.07), suggesting Chinese natives were sensitive to gender mismatch when binding was licensed. No other statistically meaningful results were found at the following regions.

In summary, cataphoric dependencies in WCO were processed differently in L1 and L2 speakers. While L1 speakers were immediately sensitive to GMME in DP conditions at the critical region, L2 speakers only showed delayed weak sensitivity to GMME. However, both groups did not attempt to bind the cataphor to the quantified antecedent, in contrast anaphora processing. Furthermore, L1 speakers

displayed a transient GME with some delay. We will discuss the implications of these findings in the context of linguistic and sentence processing theories in the next section.

4 Discussion

This study used psycholinguistic methods to investigate pronominal binding in English WCO. The scope of this study covers anaphoric and cataphoric A'-dependencies. Despite copious theorizing and some inspiring progress achieved by experimental research, the present study has produced new findings that carry implications to linguistic and psycholinguistic theories. Crucially, it also contributes to our understanding of L2 processing of WCO, a topic seldom explored in the literature. In the following subsections, we will first discuss the L1 and L2 results separately.

4.1 Asymmetry in anaphora and cataphora processing in L1

Exp. 1 on anaphoric dependencies showed that pronominal binding was licensed across complement and WCO conditions at the first spillover region. This was then followed by an interaction such that GMME only occurred in WCOs. In terms of the time course of this intricate pattern, the results are quite similar to an earlier study by Kush *et al.* (2017) where WCOs were situated in indirect questions. The only difference is that GMME and the interaction occurred at the critical region and the first spillover region respectively in Kush *et al.* (2017) but was delayed by one region in this study. This was unexpected although not surprising. The A'-dependency in Exp. 1 involves relativization which is presumably more complex than *w*-movement alone. It is conceivable that processing RC structures consumed more cognitive resources so that the parser to could not react to the gender mismatch immediately. In contrast, Exp. 2 on cataphoric dependencies presented a different picture. GMME showed up immediately at the critical region in the licensing DP conditions but was totally absent in the WCO conditions. A delayed albeit non-significant GME in the WCO conditions, however, could suggest that native participants briefly considered binding but quickly retracted such a possibility (Kazanina & Phillips 2010).

These results present a challenge to some linguistic theories on WCO. This is because any theory on WCO must account for the different processing profiles in anaphoric and cataphoric dependencies. The Bijection Principle (Koopman & Sportiche 1982) and the Parallelism Constraint on Operator Binding (Safir 1984) would make incorrect predictions as they would rule out the attested pronominal binding in anaphoric dependencies which we have seen in Exp. 1. Likewise, Barker's scope-plus-reconstructed-linearity constraint would incorrectly predict uniform processing patterns in these two types of dependencies as well. This is because although the binder is able to scope over the pronoun[1], the binder would follow the pronoun after reconstruction. Therefore, some alternative approaches must be pursued. One solution is to invoke linearity as a constraint on WCO. This

[1]Following Barker (2012), we can replace the pronoun his with someone's to check whether the operator can scope over the existential. This is indeed the case for the sentences in both experiments.

idea is attributed to Bresnan (1994, 1998) who argued, based on crosslinguistic evidence, that Linear Order and Syntactic Rank are violable constraints in WCO. The Syntactic Rank constraint says that "a pronominal P cannot outrank a constituent on which P is referentially dependent". The "rank" refers to the grammatical function hierarchy (Subject > Object > Oblique, etc.). These two constraints are independent and weighed differently across languages. For example, in Korean, Linearity Order is non-violable because "everyone, his mother likes" is allowed but "His mother likes everyone" is not, whilst for other languages such as Chichewa (Bresnan & McHombo 1987; Bresnan & Kanerva 1989) Linear Order is violable and does not have to be satisfied. Although English requires both constraints to be respected to license binding (Bresnan 1994), anaphora should still be easier to process, this is because cataphoric binding incurs two violations, but anaphoric binding only incurs one violation, hence the asymmetry in processing anaphors and cataphors. Furthermore, since both anaphoric and cataphoric binding in WCO violate the Syntactic Rank constraint, binding was ruled out ultimately in both experiments.

The findings in this paper also bear on the psycholinguistic theories. The multiple-constraint approach predicts that all linguistic cues, including syntax and semantics, can compete at any stage, early or late, of parsing. In this case, checking for A-binding and gender match is predicted to proceed in a parallel fashion. Full match of syntactic and semantic cues is more optimal than partial match of any one of the cues which is in turn better than no match. This argument is applicable to the WCO conditions where gender match facilitated processing compared to the gender mismatch condition at the first spillover region in Exp. 1. The absence of GMME in Exp. 2, as we suggested above, was due to the influence of Linear Order which participated in the multiple-constraint competition. Because neither Syntactic Rank nor Linear Order was satisfied in cataphoric dependencies, binding was prohibited. Note that the delayed weak GME effect is also compatible with the multiple-constraint approach. One possibility is that the parser tried to assess the possibility for binding based on gender match but retracted or gave up the binding later. The modular account, on the other hand, seems to fall short in accounting for the results in Exp. 1 because the interaction or WCO effect occurred after the GMME main effect, suggesting that syntax was applied too late to block the initial ungrammatical interpretation, a fact contrary to the claims of the modular view. However, more time-sensitive research methods are needed to examine this debate further because, in an eye-tracking study on indirect *wh*-questions, Lyu (2020) found that the WCO constraint successfully blocked binding at the early stage of processing. Only in the later stage was there a main effect of GMME. As suggested by Sturt (2003), self-paced reading might only reflect the late processing stage. Therefore, we suggest caution before one jumps to any quick conclusion.

4.2 "Shallow" parsing or poor prediction in L2?

In Exp. 1, the L2 participants showed a different processing pattern than the native speakers. Right at the pronoun region, an early GMME emerged in WCO conditions only, but not in the licensing complement conditions, indicating that the L2 speakers ignored the WCO constraint in anaphoric dependencies. The next two regions did not show any interaction either and only showed significant GMMEs in

both licensing and WCO conditions. Therefore, L2 speakers behaved differently from L1 speakers in that they were insensitive to the syntactic constraint in the anaphoric dependencies.

The early GMME only in the WCO conditions was unexpected. We argue that this is presumably because L2 participants did not analyze the "it seemed" phrase. Recall that although L1 participants spent more time on "it seemed" than the embedded verb in the complement clauses (e.g., "heard"), L2 participants did not show any slowdown. We interpret this as evidence that L2 participants ignored the cumbersome "it seemed" phrase. This has lead to the consequence that "it seemed" was not represented on the syntactic tree all together. The reason for this purposeful omission could be due to structural "bizarreness" as the *wh*-phrase in an RC is either followed by a regular verb (e.g., "heard") in a subject relative or a noun in an object relative. The L2 participants might have some difficulty in representing this phrase on the syntactic tree given the infrequent bigram/trigram collocation (i.e. "who it seemed…"). More importantly, "it seemed" did not contribute much to the meaning of the whole sentence. To save processing effort, L2 participants might resort to a strategy where they simply raced through this phrase without integrating it. This would leave them with a simpler and less embedded structure in the case of the anaphoric WCO. Therefore, less burdened with a less complicated structure, L2 participants could react to GMME in the WCO more quickly than in the licensing complement clause where "she heard" is not semantically vacuous.

It should be pointed out that the L2 results in Exp. 1 could be predicted by both SSH and RAGE because the absence of the interaction (or WCO effect) could be due to more reliance on semantic cues in L2 processing or, alternatively, due to the poor ability to generate predictions. However, results from cataphoric binding in Exp. 2 seem to favor RAGE over SSH. First, the L2 speakers were sensitive to the WCO constraint in Exp. 2, suggesting that there was no deficit in L2 speakers' ability to represent the WCO structure or use syntactic constraints to rule out illicit readings. Crucially, they did not establish binding based on semantic feature match. Second, the GMME in the licensing DP conditions appeared at the spillover region for the L2 speakers, slower than the native speakers. The slower GMME compared to the native speakers suggest that the L2 speakers did not expect a gender-matched antecedent as quickly when the antecedent was encountered. These two findings can be unified under the RAGE model, but not under SSH.

5 Conclusion

In order to probe whether (non-)native speakers can promptly utilize syntactic constraints in real-time processing, the present study investigated the processing of anaphors and cataphors in weak crossover in English native and non-native populations. The study showed that native speakers processed anaphoric and cataphoric binding differently. The delayed WCO effect in anaphoric dependencies suggests that linear precedence relationship between the operator and the pronoun might play an important role in ruling in/out binding in WCO. The non-native speakers, on the other hand, did not show signs of sensitivity to the WCO constraint in anaphoric dependencies but did avoid binding in cataphoric dependencies. Furthermore, the L2 speakers showed a delayed GMME in the licensing DP conditions. Overall, these

results are consistent with the predictions of the RAGE model but not SSH.

References

Badecker, W., & K. Straub. 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.

Barker, C. 2012. Quantificational binding does not require c-command. *Linguistic Inquiry* 43.614–633.

Boxell, O., & C. Felser. 2017. Sensitivity to parasitic gaps inside subject islands in native and non-native sentence processing. *Bilingualism: Language and Cognition* 20.494–511.

Bresnan, J. 1994. Morphology competes with syntax: Explaining typological variation in weak crossover effects. In *The 13th Proceedings of Chicago Linguistics Society*, 57–89.

Bresnan, J. 1998. Morphology competes with syntax: Explaining typological variation in weak crossover effects. In *Is the Best Good Enough?*, ed. by P. B. et al., 59–92. Cambridge, MA: MIT Press.

Bresnan, J., & S. A. Kanerva. 1989. Locative inversion in Chichewxa: A case study of factorization in grammar. *Linguistic Inquiry* 20.1–50.

Bresnan, J., & S. A. McHombo. 1987. Topic, pronoun, and agreement in chicheŵa. *Language* 63.741–782.

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

Chow, Y. W., S. Lewis, & C. Phillips. 2014. Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology* 5.1–16.

Clahsen, H., & C. Felser. 2006. Continuity and shallow structures in language processing. *Applied Psycholinguistics* 27.107–126.

Clahsen, H., & C. Felser. 2018. Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition* 40.693–706.

Drummer, J.-D., & C. Felser. 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/`.

Felser, C. 2015. Native vs. non-native processing of discontinuous dependencies. *Second Language* 14.5–19.

Felser, C., & I. Cunnings. 2012. Processing reflexives in english as a second language: The role of structural and discourse-level constraints. *Applied Psycholinguistics* 33.571–603.

Felser, C., & J.-D. Drummer. 2017. Sensitivity to crossover constraints during native and non-native pronoun resolution. *Journal of Psycholinguist Research* 46.771–789.

Felser, C., & L. Roberts. 2007. Processing /wh-dependencies in a second language: A cross-modal priming study. *Second Language Research* 23.9–36.

Felser, C., M. Sato, & N. Bertenshaw. 2009. The online application of binding principle a in english as a second language. *Bilingualism: Language and Cognition* 12.485–502.

Grüter, T., H. Rohde, & A. Schafer. 2014. The role of discourse-level expectations in non-native speakers' referential choices. In *Proceedings of the 38th Annual Boston University Conference on Language Development*, 179–191.

Hahne, A., & A. D. Friederici. 2001. Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition* 4.123–141.

Jäger, G. 2005. *Anaphora and type logical grammar*. Dordrecht: Springer.

Jäger, L. A., F. Engelmann, & S. Vasishth. 2015. Retrieval interference in reflexive processing: Experimental evidence from mandarin, and computational modeling. *Frontiers in Psychology* 6.617.

Kaan, E. 2014. Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism* 4.257–282.

Kaan, E., A. C. Dallas, & F. Wijnen. 2010. Syntactic predictions in second-language sentence processing. In *Structure preserved. Festschrift in the honor of Jan Koster*, ed. by J.-W. Zwart & M. de Vries, 207–213. Amsterdam: John Benjamins.
Kaiser, E., J. T. Runner, R. S. Sussman, & M. K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112.55–80.
Kazanina, N., E. F. Lau, Lieberman, Y. M., M., & C. Phillips. 2007. The effect of syntactic constraints on the processing of backwards anaphora. *Journal of Memory and Language* 56.384–409.
Kazanina, N., & C. Phillips. 2010. Differential effects of constraints in the processing of russian cataphora. *Quarterly Journal of Experimental Psychology* 63.371–400.
Koopman, H., & D. Sportiche. 1982. Variables and the bijection principle. *The Linguistic Review* 2.139–160.
Kush, D., J. Lidz, & C. Phillips. 2015. Relation-sensitive retrieval: Evidence from bound variable pronouns. *Journal of Memory and Language* 82.18–40.
Kush, D., J. Lidz, & C. Phillips. 2017. Looking forward and backwards: The real-time processing of strong and weak crossover. *Glossa: a journal of general linguistics* 2.70. 1–29.
Lewis, R. L., & S. Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29.375–419.
Lyu, J. 2020. An eye-tracking study of pronominal binding in English weak crossover. *Modern Foreign Languages* 43.174–187.
Marinis, T., C. Roberts, L. qnd Felser, & H. Clahsen. 2005. Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27.53–78.
Patil, U., S. Vasishth, & R. L. Lewis. 2016. Retrieval interference in syntactic processing: The case of reflexive binding in english. *Frontiers in Psychology* 7.329.
R Core Team. 2018. R: A language and environment for statistical computing. `http://www.R-project.org/`.
Reinhart, T. 1983. Coreference and bound anaphora: A restatement of the anaphora questions. *Linguistics and Philosophy* 6.47–88.
Runner, J. T., R. S. Sussman, & M. K. Tanenhaus. 2006. Processing reflexives and pronouns in picture noun phrases. *Cognitive Science* 30.193–241.
Ruys, E. G. 2000. Weak crossover as a scope phenomenon. *Linguistic Inquiry* 31.513–539.
Safir, K. 1984. Multiple variable binding. *Linguistic Inquiry* 15.603–638.
Sauerland, U. 1998. *On the Making and Meaning of Chains*. MIT: Ph.D. dissertation.
Shan, C.-C., & C. Barker. 2006. Explaining crossover and superiority as left-to-right evaluation. *Linguistics and Philosophy* 29.91–134.
Sturt, P. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48.542–562.
Wang, C. 2006. Unidirectionality of the anaphora of third person pronouns in English and Chinese. *Foreign Language Teaching and Research* 38.17–24.
Wang, Y. 1994. Can "he/she" in English refer to the following nouns? *Foreign Language Teaching and Research* 4.36–38.
Wu, F., & J. Lyu. 2016. Ordering of Chinese relative clauses and determiner phrase: A second language perspective. *Chinese Language Learning* 4.84–94.
Wu, F., & Y. Sheng. 2014. Pre-RC determiner phrase bias and production preference for object relatives: Perspectives from L2-Chinese learners. *Foreign Language Teaching and Research* 46.401–411.