Journal of Neurolinguistics 45 (2018) 1–12

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Gender constraint in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals

Lijuan Liang [a, c], Yun Wen [b, 1], Yanping Dong [b, c, *]

[a] School of English and Education, Guangdong University of Foreign Studies, Guangzhou 510420, China
[b] Center of Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou 510420, China
[c] Bilingual Cognition and Development Lab, Guangdong University of Foreign Studies, Guangzhou 510420, China

ARTICLE INFO

Article history:
Received 2 August 2016
Received in revised form 1 August 2017
Accepted 8 August 2017

Keywords:
Reflexive pronoun resolution
Semantic gender constraint
Chinese-English bilingual
ERP
Antecedent-reflexive pronoun binding

ABSTRACT

Previous behavioral research indicates that native speakers of Chinese are not very sensitive to lexical gender information, and one question that remains is how the rule of gender constraint functions in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. To answer the question, the present study conducted two ERP experiments in which participants' ERP responses to L1 and L2 reflexive pronouns that were either congruent or incongruent with their antecedents were recorded and analyzed. Gender incongruity effect was observed in the L1 in the form of left anterior negativity effect (LAN) within 350–450 ms, but in the L2 in the form of midline enhanced positivity within 300–500 ms. Then, within a later time window of 500–800 ms, the P600 component was observed for gender incongruity in both the L1 and the L2. However, large variance was observed in the participants' behavioral performance in the non-native language, and the amplitude of the P600 effect decreased linearly with the participants' inability to detect gender incongruity. The above findings demonstrate that the rule of gender constraint functions similarly in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals, but nevertheless the exact mechanism employed for the L1 and the L2 was not the same, with an LAN-P600 pattern for the L1 but an Earlier Positivity-P600 pattern for the L2. Moreover, the large variance in L2 processing should be taken into consideration in future research on L2 reflexive pronoun resolution.

1. Introduction

The prerequisite for successful sentence and discourse comprehension involves establishing certain links among the constituent words. One of the links is the binding between anaphoric elements such as reflexive pronouns (i.e., *himself, herself*) and their antecedents during listening or reading. According to Binding Principle A, a reflexive pronoun "*himself*" as in the sentences "*Jack praises himself*" and "*Henry says Jack praises himself*" must be bound to the antecedent "*Jack*" in the local domain (i.e., within the immediate clause containing the reflexive, "c-commanding") (Chomsky, 1981). This is the locality constraint on reflexives. Sturt (2003) found that the locality constraint on binding was applied immediately upon the participants' first reading of the reflexive.

* Corresponding author. Center of Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou 510420, China.
E-mail address: ypdong@gdufs.edu.cn (Y. Dong).
[1] Yun Wen has contributed equally with the first author.

http://dx.doi.org/10.1016/j.jneuroling.2017.08.001
0911-6044/

Apart from the locality constraint mentioned above, the reflexive pronoun must also match the possible antecedent in gender, person and number (Badecker & Straub, 2002). The gender congruence between the reflexive pronoun and the antecedent, i.e., the gender constraint, is a factor that could be manipulated in experiments to investigate how different elements in a sentence are linked. According to Bock and Miller (1991), the congruity of gender information through anaphora is a way to constrain sentence elements and establish appropriate within-phrase and across-phrase links.

As for the cognitive processes underlying reflexive pronoun resolution, a two-stage model, early bonding and subsequent resolution processes, was proposed for all kinds of anaphoric interpretation (Garrod & Terras, 2000; Sanford, Garrod, Lucas, & Henderson, 1983). In the early bonding phase, appropriate links are established automatically between an anaphor and the possible antecedents when their semantic features match (Garrod & Terras, 2000; Sturt, 2003). In other words, when the reader encounters a pronoun, he/she will initiate a search in previous context for the antecedent guided by the semantic features (e.g., gender, number) of the anaphor (Callahan, 2008). In a similar sense, Van Berkum, Koornneef, Otten, and Nieuwland (2007) suggests that upon encountering a pronoun, readers immediately inspect their situation model for a suitable discourse entity by referring to the semantic features. Moreover, Dillon (2014) further proposed that the problem of selecting a reflexive's antecedent is a memory retrieval problem, such that antecedent retrieval is cue-based direct access by matching a set of features against the contents of working memory. In the subsequent resolution phase, the link is checked against the broader context and the antecedent is finalized (Garrod & Terras, 2000).

However, there is one prerequisite for all the above theoretical elaborations concerning the gender constraint between the reflexive pronoun and its antecedent during the bonding stage: the gender information lexically specified in the reflexive pronoun (e.g., *himself, herself*) and the antecedent (i.e., nouns with definitional or stereotypical gender) is supposed to be activated as part of lexical access. Although gender information, as a semantic feature, is assumed to be automatically retrieved during semantic retrieval (Kreiner, Sturt, & Garrod, 2008), some studies found that this assumption is not true for all populations, especially for native Chinese speakers (Chen & Su, 2011; Dong, Wen, Zeng, & Ji, 2014).

Previous behavioral studies indicate that native speakers of Chinese are not as sensitive to lexical gender information as corresponding native speakers of English. Chen and Su (2011) compared the performance of English and Chinese participants in two tasks presented in their own respective native language. They were asked in one task to listen to a story and answer gender and non-gender related questions, and then in another task, to read sentences and determine which of the two pictures presented immediately afterwards matches the sentence. The results showed that relative to the non-gender-related condition, the Chinese participants were much slower and less accurate than the English participants in answering the gender-related questions and in responding to the gender-related sentences. Furthermore, Dong et al. (2014) conducted self-paced reading experiments on Chinese-English bilinguals to compare their performance in two conditions: pronoun-antecedent semantic gender matched or not matched. The mismatch effect was observed only when the gender saliency of the antecedent was enhanced with a human picture presented before the antecedent. Without this manipulation, however, the gender mismatch effect disappeared not only for L2 English but also for L1 Chinese. The authors concluded that Chinese-English bilinguals probably do not automatically activate the gender information incorporated in the antecedent and the pronoun in the bonding phase in both the L1 and the L2. Moreover, Chinese-English bilinguals make lots of gender errors when speaking English in various language tasks (Dong & Jia, 2011).

The reason for the reduced sensitivity of native Chinese speakers to lexically specified gender information probably lies in the linguistic realization of semantic gender information in person pronouns in oral Chinese. To be specific, oral Chinese uses the same pronunciation,/*ta*/, to refer to both female and male singular third person pronoun, and does not distinguish between "*she/her/herself …*" and "*he/him/himself …*". Hence, in oral Chinese, speakers may not have to activate semantic gender information when using pronouns. What's more, despite the fact that "he" and "she" are distinguished in written Chinese through orthographic forms "*他/ta/(he)*" and "*她/ta/(she)*", these gender-encoding Chinese characters are often omitted in reflexive pronouns, leaving a gap in the sentence, e.g.,"*周先生对(他$_{gap}$)自己很有信心。Mr. Zhou has confidence in (him $_{gap}$)self.*" Again, readers or listeners do not have to activate semantic gender information since there is no gender constraint here. Understanding Chinese, therefore, may not rely heavily on the constraints among sentence elements established with semantic gender information.

Then, one question arises in the field of research on reflexive pronoun resolution: will the binding force of the gender constraint, as a coreferential tool, weaken for speakers who have reduced sensitivity to lexical gender information? To be more specific, the issue under investigation in the current study is whether people like native Chinese speakers apply gender constraint during reflexive pronoun resolution when processing sentences such as "*The man prepared himself for the interview*" where the locality constraint within the clause is good enough for the pronoun-antecedent binding. Based on the above reasoning, it could be hypothesized that since the locality constraint has already guaranteed the intended reading, the gender constraint may not play its full role for readers who have reduced sensitivity to lexical gender information.

The factor of gender congruity was manipulated in the current study in a way that the reflexive pronoun was either congruent or incongruent with the antecedent (i.e., Congruent vs. Incongruent). The gender incongruity effect would demonstrate whether gender constraint is applied or not. To avoid the confounding factor of linearity, the experimental materials (e.g., "*Mr./Miss Williams found himself/herself the center of attention.*") were designed in a way that there was no ambiguity in co-reference between the reflexive pronoun and its antecedent because these two items were successfully constrained by their position within the clause (within-phrase anaphora) (Binding Principle A, Chomsky, 1981). Antecedents (e.g., *Mr./Miss Williams*) with unambiguous and definitional semantic gender instead of stereotypical gender were used to

guarantee the immediate access of gender information. This ruled out the potential factor of gender retrievability so that the gender incongruity effect would be brought into the spotlight.

Because of its fine temporal resolution, the electrophysiological technique was used in the current study to investigate the real-time processing of gender congruity between the reflexive pronoun and the antecedent. Two major ERP components were expected to be elicited for gender incongruity: the P600 and the Anterior Negativity (AN). Evidence from L1 semantic gender agreement processing shows that a P600 component with centro-parietal or parietal distribution within 500–800 ms is elicited upon the onset of the pronoun when the gender feature of the pronoun is incompatible with the only possible antecedent (Hammer, Jansma, Lamers, & Münte, 2005; Osterhout & Mobley, 1995; Osterhout, Bersick, & McLaughlin, 1997; Xu, Jiang, & Zhou, 2013). P600 is an index of reanalysis and repair of an anomaly, syntactic anomaly in most cases (Friederici, Hahne, & Saddy, 2002; Hagoort & Brown, 2000). In addition, syntactic violations may also yield an increased ERP negativity for the violation stimuli in comparison to the correct stimuli around 300–500 ms across more anterior electrodes (Morgan-Short, 2014). When the distribution of this anterior negativity (AN) is maximal over the left hemisphere, it is called left anterior negativity (LAN) (Molinaro, Barber, & Carreiras, 2011; Molinaro, Barber, Caffarra, & Carreiras, 2015). The anterior negativity effect elicited upon gender-mismatching pronouns has been observed in the study by Nieuwland (2014). Therefore, the two time windows for the P600 and anterior negativity effects were analyzed in the current study.

Two experiments were conducted on Chinese-English bilinguals, with one focusing on the processing of reflexives in their L1 Chinese, and the other focusing on their L2 English. Both languages were tested, as it is assumed that lexical gender information is probably not well processed in the Chinese-English bilingual' conceptualizer, impacting both languages (Chen & Su, 2011; Dong et al., 2014). The two experiments used similar designs. The gender information specified in the reflexive pronoun is as clearly marked in the written Chinese characters "*他(him)自己(self)*" and "*她(her)自己(self)*" as in English. Besides, the gender congruity between the reflexive pronoun and the antecedent is obligatory both in Chinese and English. Hence, similar results were expected from both languages. However, as previous studies have revealed individual variability in L2 sentence processing (Roberts, 2012), probably more variance in behavioral and real-time processing would be observed from the L2.

To sum up, the present study aimed to investigate how the rule of gender constraint functions in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals. If the rule of gender constraint functions well, we would expect a gender incongruity effect in the form of (L)AN-P600 ERP pattern from the two languages, with probably more variance in the processing of L2 reflexive pronouns. If not, the gender incongruity effect would not appear.

2. Experiment 1

Experiment 1 focused on how the rule of gender constraint functions in L1 reflexive pronoun resolution by Chinese-English bilinguals.

2.1. Method

2.1.1. Participants

Sixteen Chinese-English bilinguals (5 males, mean age: 19.28 ranging from 19 to 24 years old) were recruited from Guangdong University of Foreign Studies, and were paid for their participation. They were undergraduate or postgraduate students majoring in English, and have been learning English for almost 11 years. All of them were right-handed assessed by the adapted version of Coren's (1992) handedness questionnaire, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric impairment. They signed a consent form before the experiment. This study was approved by the Ethics Committee of Guangdong University of Foreign Studies.

2.1.2. Design and materials

The experiment is a single factorial design, manipulating the gender congruity (i.e., Congruent, Incongruent) of the antecedent and the reflexive pronoun. Altogether 160 Chinese sentences were presented, 50% of which were anomalous.

Altogether 80 Chinese experimental sentences were constructed (see Table 1 for example), with an antecedent as the subject for each sentence. The antecedents were Chinese equivalents of Mr. or Miss plus common Chinese family names, e.g., "*赵先生/小姐 (Mr./Miss Zhao)*". Chinese reflexive pronouns "*她自己(herself)*" and "*他自己(himself)*" were the region of interest. For half of the 80 sentences, the antecedents were male and the other half were female. The length of these sentences ranged from five to seven word segments, with the reflexive pronoun being the third. Two versions of the sentences corresponding to the two conditions of gender Congruent/Incongruent were created and assigned to one of the two experimental lists. In each list, both the gender congruent sentences and incongruent sentences had an equal number of male and female antecedents as the subject. Experimental lists were counterbalanced across the 16 participants.

Another set of 80 sentences were constructed as fillers (see Table 1 for example). Among them, 40 fillers had "*我的(my)*" plus a noun indicating an occupation or title, e.g., "*老师(teacher)*", as the subject, and half of these fillers were anomalous. The other 40 fillers had pronoun "*我(I)*" as the subject, and half of these fillers were anomalous. These anomalous sentences contained a classifier-noun collocation violation appearing in different positions across sentences. The length of the fillers ranged from 5 to 8 word segments.

Table 1
Materials used in Experiment 1.

	Conditions	Examples
Experimental	Congruent	赵小姐/发现/她自己/陷入了/两难的/境地。 (Miss Zhao found herself in the dilemma.)
	Incongruent	赵小姐/发现/他自己/陷入了/两难的/境地。 (Miss Zhao found himself in the dilemma.)
Filler	Correct	我/喜欢/和/朋友/一起/打/篮球。 (I like playing basketball with friends.) 我的/医生/建议/我/坚持/运动。 (My doctor advised me to keep a habit of doing sports.)
	Incorrect	我/给/自己/买了/一部/鞋子。 (I bought myself a group of shoes.) 我的/爸爸/给/我/买了/一篇/兔子。 (My father brought me a piece of rabbit.)

2.1.3. Procedure

All the trials were presented in a pseudo-randomized order to avoid three correct or incorrect sentences appearing in succession. Each trial began with a fixation ("+") in the center of the screen for 500 ms, followed by a 500 ms blank interval. Then each word segment appeared for 400 ms with a 250 ms inter-stimulus interval. At the end of the sentence, the last word appeared with the period. After a 500 ms blank screen, a question mark was presented and remained there up to a maximum of 2000 ms. The question mark was a sign for participants to respond and judge whether the sentence was correct (i.e., syntactically well-formed and semantically coherent) or incorrect (i.e., syntactically ungrammatical or semantically incoherent) by pressing one of two buttons on the keyboard. The assignment of response type to the left and right hands was counterbalanced across participants. The whole experiment (with a break every 20 trials) lasted for about 1.5 h. Before the formal experiment, participants practiced for 12 trials to get familiar with the procedure.

2.1.4. EEG recording and data analysis

Electroencephalogram (EEG) was recorded at a 1000 Hz sampling rate using a 64-channel NeuroScan net. Eye movements were measured using vertical (electrodes below and above the left eye) and horizontal (two electrodes placed lateral to the outer canthi of the two eyes) electrooculogram. The common EEG and electrooculogram (EOG) reference was attached to the left mastoid, and re-referenced off-line to the mean of the activity at the left and right mastoids. Electrode impedances were kept <10 kΩ. The electrophysiological signals were filtered on-line with a bandpass of 0.05–100 Hz and later low-pass (30 Hz) filtered off-line.

EEG data analysis was performed using Scan 4.3. Eye movements were corrected by means of correlation. Epochs time-locked to the onset of the reflexive pronoun were extracted from −200 to 1000 ms, and were averaged off-line for each group of stimuli (gender congruent, gender incongruent) separately for each participant. Baseline correction was performed in reference to prestimulus activity (−200 to 0 ms). Epochs with EEG exceeding either ±100 μV at any channel within intervals of 200 ms were automatically rejected off-line.

ERP components of interest were quantified using mean amplitude measure. Two time windows, 300–500 ms for the AN and 500–800 ms for the P600, were selected in accordance with previous reports and visual inspection of the data. In order to investigate the topographic distribution of the relevant effects, nine ROIs were computed. Data from midline and lateral electrodes were treated separately. Specifically, the midline electrode groups include: midline anterior (FZ, FCZ), midline medial (CZ, CPZ), and midline posterior (PZ, POZ). The lateral electrode groups include: left anterior (F1, F3, F5, FC1, FC3, FC5), left medial (C1, C3, C5, CP1, CP3, CP5), left posterior (P1, P3, P5, PO3, PO5, PO7), right anterior (F2, F4, F6, FC2, FC4, FC6), right medial (C2, C4, C6, CP2, CP4, CP6), right posterior (P2, P4, P6, PO4, PO6, PO8). Repeated-measures ANOVAs over midline sites were conducted with gender congruity (congruent, incongruent) and electrode region (anterior, medial, and posterior) as the within-subjects factors. Similar analysis over lateral sites was conducted with gender congruity (congruent, incongruent), hemisphere (left, right), and electrode region (anterior, medial, and posterior) as the within-subjects factors. Since the main concern of the present study was the presence of gender incongruity effect, only when reliable interactions involving gender congruity were found, further analysis was performed. Significance levels of the *F* ratios were adjusted with the Greenhouse-Geisser correction where appropriate and the corrected *p* values are reported.

2.2. Results

2.2.1. Behavioral results

The mean accuracy rate for gender congruent trials was 95.47% (SD = 4.3%), for gender incongruent trials 95% (SD = 4.91%), and for fillers 96.41% (SD = 2.77%). Then d-prime score was calculated by taking gender congruent and incongruent trials into a whole based on the algorithm d-prime = Z (hit rate) − Z (false alarm rate), where function Z(p), p ∈ [0,1], is the inverse of the cumulative distribution function of the Gaussian distribution. A d-prime of 4 indicates near-perfect processing in signal detection. The mean d-prime score for gender congruity processing was 3.32 (SD = 0.56). On average, all participants

performed very well in the sentence acceptability judgment task. Reaction times were not analyzed because delayed responses may not be a good indicator of on-line processing.

2.2.2. Electrophysiological results

The grand average ERPs time-locked to the onset of the reflexive pronoun and the topographic maps depicting the gender incongruity effect in the incongruent minus congruent condition during 300–500 ms and 500–800 ms are shown in Fig. 1.

300-500 ms: In this time window, no significant main effect of gender congruity was found [lateral: $F(1, 15) = 0.61$, $p = 0.44$, $\eta^2_P = 0.04$; midline: $F(1, 15) = 0.01$, $p = 0.92$, $\eta^2_P = 0.001$], and no gender congruity effect was found in any brain regions either [lateral: gender congruity × hemisphere × region, $F(2, 30) = 0.25$, $p = 0.68$, $\eta^2_P = 0.02$; midline: gender congruity × region, $F(2, 30) = 2.6$, $p = 0.12$, $\eta^2_P = 0.15$].

However, a visual inspection of the ERP waveforms and the topographic map seemed to indicate that the gender incongruent condition elicited a larger negativity than the congruent condition specifically in the anterior region. We therefore adopted a more sensitive measure, i.e., consecutive analyses of 50 ms time windows from 300 ms to 500 ms (which means four time windows 300–350 ms, 350–400 ms, 400–450 ms, and 450–500 ms). Repeated-measures ANOVAs over left anterior electrode sites (F7, F5, F3, F1) and right anterior electrode sites (F8, F6, F4, F2) were conducted with gender congruity (congruent, incongruent) and electrode sites as the within-subjects factors. Analyses over left anterior electrode sites revealed a significant main effect of gender congruity within 350–400 ms, $F(1, 15) = 5.56$, $p < 0.05$, $\eta^2_P = 0.27$, and marginal significant main effect of gender congruity within 400–450 ms, $F(1, 15) = 4.49$, $p = 0.051$, $\eta^2_P = 0.23$. No effect of gender congruity was found within the other two time windows over left anterior electrode sites (all $ps > 0.05$), and within all the four time windows over right anterior electrode sites (all $ps > 0.05$). Taken together, the gender incongruent condition evoked a larger negativity than the congruent condition within 350–450 ms over the left anterior electrode sites F7/F5/F3/F1. This negativity elicited over the left anterior region could be considered as the Left Anterior Negativity component (LAN) (Molinaro et al., 2011), because the LAN is a well-documented ERP component for L1 syntactic processing falling into a time range of 300–450 ms across more anterior electrodes over the left hemisphere (Dowens, Vergara, Barber, & Carreiras, 2009), and it suggests an early detection of a syntactic violation (Molinaro et al., 2015; Münte, Matzke, & Johannes, 1997).

500-800 ms: In this time window, no significant main effect of gender congruity was found [lateral: $F(1, 15) = 0.90$, $p = 0.35$, $\eta^2_P = 0.06$; midline: $F(1, 15) = 2.29$, $p = 0.15$, $\eta^2_P = 0.13$]. However, the lateral analysis revealed a significant interaction of gender congruity × hemisphere × region, $F(2, 30) = 8.18$, $p < 0.001$, $\eta^2_P = 0.35$. Further analysis showed a significant gender incongruity effect in the left posterior region, $F(1, 15) = 5.89$, $p < 0.05$, $\eta^2_P = 0.28$, with the gender incongruent condition eliciting more positive responses than the congruent condition, indicating a P600 effect to reflexive pronouns with gender anomaly. In the midline analysis, the interaction of gender congruity × region did not reach significance, $F(2, 30) = 3.21$, $p = 0.089$, $\eta^2_P = 0.17$. Nevertheless, visual inspection of the ERP waveforms found the midline posterior region showing pronounced positivity in the gender incongruent condition compared to the congruent condition. Further

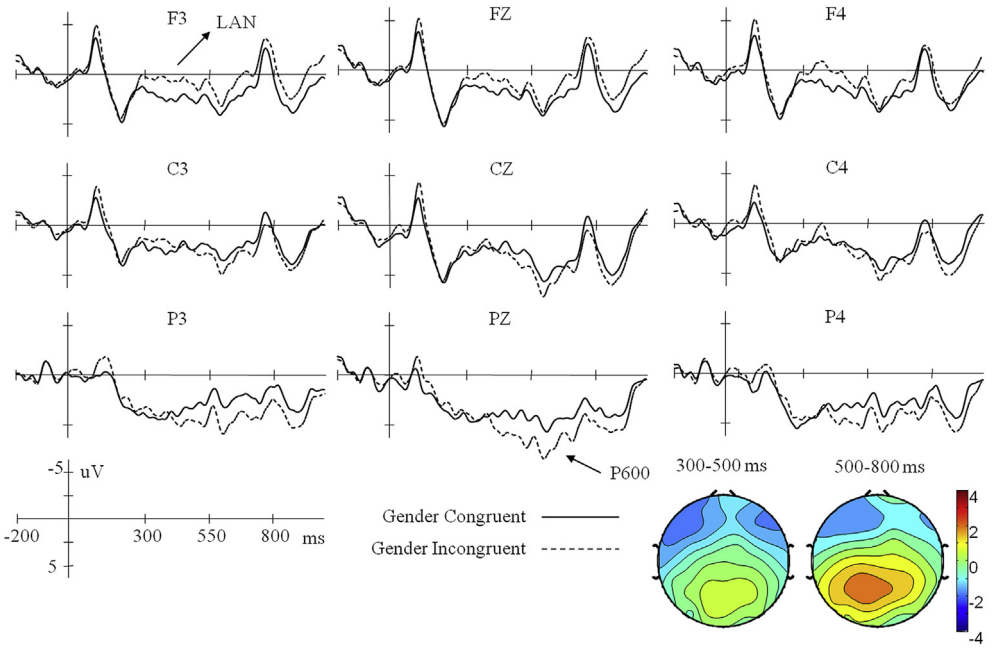

Fig. 1. Grand average ERPs time-locked to the reflexive pronoun for the gender congruent and incongruent conditions, and the topographic maps for difference waves (gender incongruent minus congruent) within 300–500 ms and 500–800 ms in Experiment 1.

analysis over electrode PZ and POZ found that this positivity was significant at the PZ site, $F(1, 15) = 10.51$, $p < 0.01$, $\eta^2_P = 0.41$, but not at the POZ site, $F(1, 15) = 2.70$, $p = 0.12$, $\eta^2_P = 0.15$.

2.3. Discussion

The results in Experiment 1 showed gender incongruity effect within 350–450 ms over left anterior electrode sites in the form of LAN. The LAN was then followed by the P600 component within 500–800 ms in the left posterior area and the midline posterior electrode PZ.

The gender incongruity effect elicited upon processing the reflexives in the native language indicates that gender constraint was applied during sentence comprehension, regardless of the strong binding force of the locality constraint over the antecedent and the reflexive pronoun. Also, it means the semantic gender information was successfully retrieved, which is the prerequisite for gender constraint, despite participants' reduced sensitivity to lexical gender information as reported in the literature (e.g. Chen & Su, 2011; Dong et al., 2014).

3. Experiment 2

Using similar designs as used in Experiment 1, Experiment 2 focused on how the rule of gender constraint functions in L2 reflexive pronoun resolution by Chinese-English bilinguals.

3.1. Method

3.1.1. Participants

Originally, 16 Chinese-English bilinguals (3 males, mean age: 21.62 ranging from 21 to 23 years old), who did not take part in Experiment 1, were recruited from the same population as those in Experiment 1, and were paid for their participation. However, unlike Experiment 1, the results from these 16 participants in Experiment 2 showed huge variance in their behavioral accuracy in gender congruity processing (gender congruent: 89.69%, SD = 8.23%, range: 67.5%–97.5%; gender incongruent: 64.22%, SD = 35.69%, range: 5%–97.5%), but their accuracy was quite high in the filler condition (93.75%, SD = 2.92%, range: 88.13%–96.25%). In order to guarantee the validity of the ERP data and take the huge individual differences into consideration, more participants were recruited.

Altogether, 32 Chinese-English bilinguals (3 males, mean age: 21.94 ranging from 20 to 24 years old) participated in Experiment 2. They were undergraduate or postgraduate students majoring in English, and have been learning English for an average of 11.19 years (SD = 1.63). At the time of the experiment, they had passed TEM4 (Test for English Majors Band 4), which is administered annually to tens of thousands of intermediate English majors by the official National Advisory Commission on Foreign Language Teaching in Higher Education in China and is recognized nationwide as a proof of English proficiency. Their self-rating score in reading was 5.18 (SD = 0.69), in writing 5.03 (SD = 0.54), in listening 4.78 (SD = 0.79), and in speaking 4.62 (SD = 0.66) on a 7-point scale (1 = very poor, 7 = native-like) language proficiency questionnaire developed by Li, Sepanski, and Zhao (2006). They can be considered intermediate to advanced English learners. All of them were right-handed assessed by the adapted version of Coren's (1992) handedness questionnaire, had normal or corrected-to-normal vision, and had no history of neurological or psychiatric impairment. They signed a consent form before the experiment.

3.1.2. Design and materials

Adapted from Osterhout et al. (1997), 80 experimental sentences were constructed (see Table 2 for example). Each of them began with a gender-specific antecedent as the subject (e.g., *Mr.* or *Miss* followed by a common English family name) and had a reflexive pronoun as the target. In half of the experimental sentences, the gender of the reflexive pronoun was congruent with the gender of the antecedent, and in the other half, they were incongruent. This forms the critical contrast between *Gender Congruent condition* and *Gender Incongruent condition* in the experiment. In 40 of the experimental sentences, the antecedent was male and in the other 40 it was female. Two counterbalanced lists were created. In each list, the two conditions had an equal number of male and female antecedents as the subject of the sentence. The length of the experimental sentences ranged from 6 to 9 words, with the reflexive pronoun being the third. Altogether one hundred and sixty filler sentences ranging from 4 to 10 words were included, with half of them being correct sentences and the other containing various linguistic violations at different sentence positions (see Table 2 for examples).

3.1.3. Procedure

Sentences were presented in a pseudo-randomized order ensuring that no more than three correct or incorrect sentences appeared in succession. Each sentence began with a fixation ("+") in the center of the screen for 500 ms, followed by a 500 ms blank interval. Each word appeared on the screen for 500 ms with a 300 ms inter-stimulus interval. At the end of the sentence, the last word appeared together with a period to the right bottom of the word. After a 500 ms blank screen, a question mark appeared in the center indicating that the participant should respond within 2000 ms. The inter-trial interval was 1000 ms. Participants were instructed to press one of two keys on a keyboard to indicate whether the sentence was correct, and the

Table 2
Materials used in Experiment 2.

	Conditions	Examples
Experimental	Congruent	Miss Wright lost herself in thought.
	Incongruent	Miss Wright lost himself in thought.
Filler	Correct	These reports are very encouraging.
		I moved to a new apartment last month.
	Incorrect	These reports is very encouraging.
		I had a dream of become famous.

assignment of response type to the left and right hands was counterbalanced across participants. Prior to the experiment, a practice session of 12 trials was given to familiarize the participants with the procedure. Participants were prompted to have a rest after every 24 trials.

3.1.4. EEG recording and data analysis

Collecting and analyzing the EEG data were conducted in the same way as in Experiment 1.

3.2. Results

3.2.1. Behavioral results

Data from five participants showed that their accuracy in gender congruity processing was quite low, as only 11% of gender incongruent sentences were correctly answered (SD = 5.75%), although the mean accuracy rate for gender congruent sentences was 87.5% (SD = 6.61%). However, they did quite well in the fillers which involves various language errors (mean accuracy for fillers: 92.25%, SD = 2.52%), indicating that these participants were serious enough in the experiment. Still, for the validity of the ERP results, these data of chance level performance was not included in the subsequent statistical analyses, resulting in 27 participants in the final dataset.

For the remaining 27 participants, the mean accuracy rate for gender congruent trials was 86.78% (SD = 8.82%), for gender incongruent trials 84.07% (SD = 14.5%), and for fillers 94.28% (SD = 3.52%). The mean d-prime score for gender congruity processing was 2.27 (SD = 0.81), which was calculated using the same measure as in Experiment One. Overall, the accuracy rate was high above the chance level of performance in gender congruity processing. Also, all participants performed very well for the fillers, suggesting that they read the stimuli carefully during the experiment. Reaction times were not analyzed because delayed responses may not be a good indicator of on-line processing.

3.2.2. Electrophysiological results

The grand average ERPs time-locked to the onset of the reflexive pronoun and the topographic maps depicting the gender incongruity effect in the incongruent minus congruent condition during 300–500 ms and 500–800 ms are shown in Fig. 2.

Grand mean analyses for 300–500 ms. In this time window, a significant main effect of gender congruity was found, but only in the midline analysis [lateral: $F(1, 26) = 2.37$, $p = 0.13$, $\eta^2_P = 0.08$; midline: $F(1, 26) = 7.96$, $p < 0.01$, $\eta^2_P = 0.23$], with the gender incongruent condition eliciting more positive responses than the congruent condition, suggesting the onset of a P600 effect at the midline electrodes. No effect of gender congruity was found in any other brain regions [lateral: gender congruity × hemisphere × region, $F(2, 52) = 0.89$, $p = 0.39$, $\eta^2_P = 0.03$].

Grand mean analyses for 500–800 ms. In this time window, a significant main effect of gender congruity was found at the midline electrodes, $F(1, 26) = 26.43$, $p < 0.001$, $\eta^2_P = 0.50$, and the bilateral medial, $F(1, 26) = 17.20$, $p < 0.001$, $\eta^2_P = 0.39$, and posterior regions, $F(1, 26) = 26.72$, $p < 0.001$, $\eta^2_P = 0.51$, but not in the bilateral anterior regions, $F(1, 26) = 2.79$, $p = 0.10$, $\eta^2_P = 0.09$ [lateral: gender congruity × region, $F(2, 52) = 14.62$, $p < 0.001$, $\eta^2_P = 0.36$]. Over the above brain regions where gender incongruity effect was found, the gender incongruent condition elicited more positive responses than the congruent condition, indicating a P600 effect to reflexive pronouns with gender anomaly.

Analyses of individuals' ERP responses. The behavioral data showed large variance in the participants' d-prime scores for gender congruity processing (mean = 2.27, SD = 0.81, range: 0.84–3.92). In order to investigate the relationship between participants' performance in the sentence acceptability judgment task and the magnitude of their online brain responses, a correlation analysis was conducted using participants' d-prime scores and their P600 effect magnitudes over the electrode PZ, where the largest ERP effect was found. The results showed the P600 effect magnitudes were significantly correlated with d-prime scores, $r = 0.39$, $p < 0.05$ (see Fig. 3), such that the participants' P600 effect magnitudes decreased linearly with their inability to detect gender incongruity between the reflexive pronoun and the antecedent. To further validate the above finding, visual inspection of the ERP waveforms and topographic maps[2] of the five participants who were eliminated from

[2] Here the larger negativity visually detected in the incongruent condition within 650–850 ms over left and midline centro-parietal electrodes is probably the Error-Related Negativity (ERN) (Gehring, Liu, Orr, & Carp, 2012) because so many trials were incorrectly answered. The ERN component may be elicited even when the participant is not explicitly aware of making the error (Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). The latency of the ERN in the current study is not the same as previous studies probably because it is time-locked to the onset of the pronoun rather than the button-press.

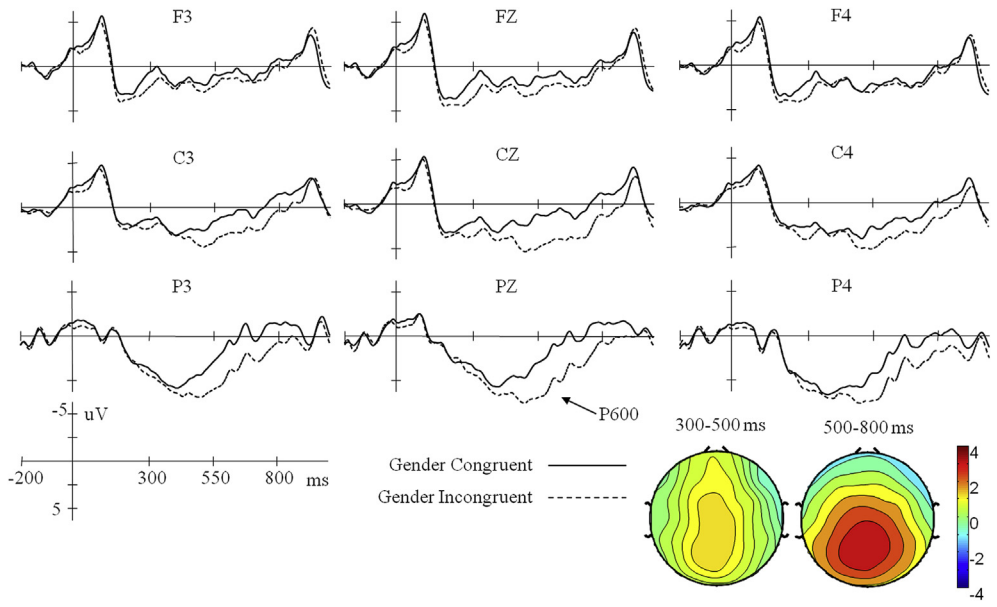

Fig. 2. Grand average ERPs time-locked to the reflexive pronoun for the gender congruent and incongruent conditions, and the topographic maps for difference waves (gender incongruent minus congruent) within 300–500 ms and 500–800 ms for the 27 participants in Experiment 2.

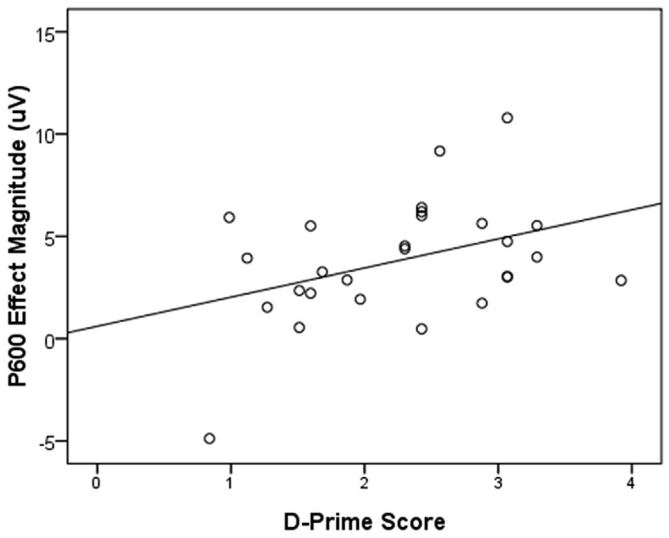

Fig. 3. Correlation between the P600 effect magnitude and the d-prime scores for gender congruity processing for the 27 participants in Experiment 2. The P600 effect magnitude is quantified as the mean amplitude in the 500–800 ms time window over the PZ electrode in the gender incongruent minus congruent condition. The solid line shows the best-fit line for the data from regression analysis. More positive values on the y-axis reflect larger P600 effects.

statistical data analyses due to their low accuracy in gender congruity processing found them showing no tendency for any P600 effect (see Fig. 4).

3.3. Discussion

The results in Experiment 2 showed reliable gender incongruity effect in the form of positivity at the midline electrode sites within 300–500 ms, suggesting the onset of a P600 effect. Then, this positivity (i.e., P600) continued among 500–800 ms at the midline electrodes and bilateral medial and posterior brain regions. However, there was large variance in the participants' behavioral and brain responses, and a correlation analysis showed that the magnitude of the P600 effect to gender incongruity was significantly correlated with the behavioral accuracy, in a way that the P600 magnitude decreased linearly with the participants' inability to detect gender anomaly. Moreover, the data of five participants, who were from the

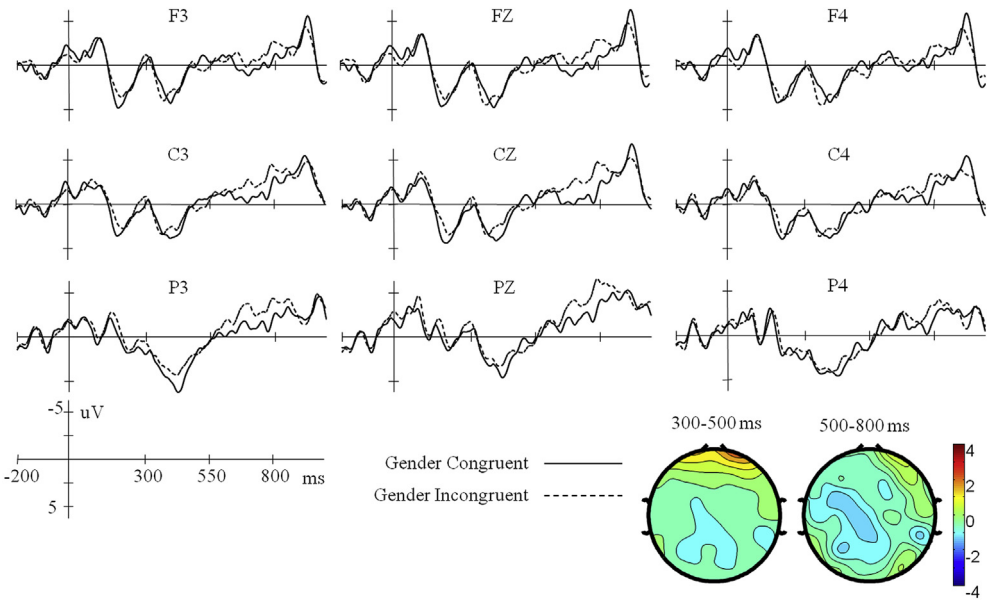

Fig. 4. Grand average ERPs time-locked to the reflexive pronoun for the gender congruent and incongruent conditions, and the topographic maps for difference waves (gender incongruent minus congruent) within 300–500 ms and 500–800 ms for the 5 excluded participants whose accuracy in gender congruity processing was quite low in Experiment 2.

same population as the rest of the participants, even showed that they were simply unaware of gender incongruity (mean accuracy rate in gender incongruent trials: 11%, SD = 5.75%), even though they were quite attentive and did very well in the filler condition (mean accuracy rate for fillers: 92.25%, SD = 2.52%).

Generally speaking, the significant gender incongruity effect (as reflected by the P600) elicited upon processing the reflexives in the non-native language indicates that gender constraint was applied during sentence comprehension. However, the observation that the gender incongruity effect decreased along with participants' inability to detect gender anomaly suggests that the rule of gender constraint may not function invariably in L2 reflexive pronoun resolution. In fact, some Chinese-English bilinguals did fail in applying gender constraint in several or more trials, and some even showed surprisingly low accuracy in gender incongruity detection (as revealed by their chance-level performance).

4. General discussion

The major goal of the present study was to explore how the rule of gender constraint functions in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals, who have reduced sensitivity to lexical gender information (Chen & Su, 2011; Dong et al., 2014). By analyzing participants' ERP responses to reflexive pronouns that are either congruent or incongruent with the antecedents in semantic gender, the two experiments revealed a gender incongruity effect in the L1 within 350–450 ms in the form of left anterior negativity (LAN), but in the L2 within 300–500 ms in the form of a larger positivity over midline electrode sites. Then the gender incongruity effect was also observed within 500–800 ms in the form of P600 component for both the L1 and the L2. However, in L2 reflexive pronoun resolution, the participants' P600 effect magnitudes decreased linearly with their inability to detect gender anomaly, indicating that individual differences exist among the L2 learners. Those who performed more poorly in detecting L2 gender incongruity tended to show a smaller or none P600 effect.

Taken together, Chinese-English bilinguals showed gender incongruity effect in both the L1 and the L2 as reflected by their real-time brain responses, but different mechanisms were employed between the L1 and the L2, with an LAN-P600 pattern for the L1 but an Earlier Positivity-P600 pattern for the L2. The Earlier Positivity observed in the gender incongruent condition in the L2 is probably an onset of a P600 effect. The LAN is considered to be an index of early detection of a syntactic violation (Molinaro et al., 2011; Münte et al., 1997). While the LAN-P600 pattern is a well-documented ERP experimental result for L1 syntactic processing, the LAN is not elicited as consistently as the P600 effect in L2 studies (Dowens et al., 2009; Morgan-Short, 2014; Tanner, 2015), probably because many participants were unbalanced bilinguals whose L2 proficiency level is not as high as the L1. As the focus of the present study is the function of gender constraint in reflective pronoun resolution, we would not discuss the different mechanisms between the L1 and the L2 in more details. The gender incongruity effect, either in the form of LAN-P600 pattern or Earlier Positivity-P600 pattern, could be considered as an indication that the rule of gender constraint was applied.

According to the two-stage model for anaphoric interpretation (Garrod & Terras, 2000; Sanford et al., 1983), semantic gender information specified in the lexical representation is activated automatically (Kreiner et al., 2008), and is used to facilitate the reflexive pronoun-antecedent binding, so that semantic gender coreference is established to guarantee correct language comprehension (Barber, Salillas, & Carreiras, 2004; Sturt, 2003). In the current study, the P600 component was evoked upon processing reflexive pronoun-antecedent gender incongruity in both L1 and L2. The P600, as an indication of reevaluation and reanalysis of grammatical anomaly (Friederici et al., 2002), suggests that the rule of gender constraint was enforced during reflexive pronoun resolution by Chinese-English bilinguals, despite their reduced sensitivity to lexical gender information.

However, while the rule of gender constraint functions consistently in L1 reflexive pronoun resolution where participants manifested homogenous performances, it may not function invariably in L2 processing where participants showed quite heterogeneous performances. To be specific, some Chinese-English bilinguals may fail to activate gender information automatically and efficiently when processing L2 English (Dong et al., 2014). Still, the coreference between the antecedent and the reflexive pronoun may have already been established because of the locality constraint before the retrieval of gender information. Thus, the violation of gender coreference between the reflexive pronoun and the antecedent was unattended to, and the sentence with gender incongruity was judged wrongly as correct in the sentence acceptability judgment task, and a smaller or even none P600 effect was elicited.

Similar to the results from the current study on L1 reflexive pronouns, some previous studies on L1 personal pronoun resolution by native Chinese speakers found that the rule of gender constraint also functions consistently during cross-phrase anaphoric processing. For example, Xu et al. (2013) compared sentences like "*这位**女**患者情绪低落，医生鼓励**她/他**振作起来。(This **woman** patient was in low spirits, doctors encouraged **her/him** to cheer up.)*", and found that the gender mismatch condition elicited a P600 effect, and the participants manifested homogenous behavioral performances. In addition, Qiu, Swaab, Chen, and Wang (2012) obtained similar results using long-distance cross-phrase anaphora like "*潘振获得了好评。在新一轮优秀个人评比活动中，**他/她**以绝对优势当选劳模。(**Panzhen**$_{\textbf{male name}}$ earned acclaim. In the new round of competition for the outstanding individual, **he/she** won the Model Worker award with an absolute advantage.)*".

However, up until now, there still lacks evidence for how the rule of gender constraint functions in L2 reflexive pronoun resolution by Chinese-English bilinguals. This is the first study to explore this issue. We observed, out of the large variance in the participants' performances, that the rule of gender constraint may not function invariably in L2 reflexive pronoun resolution. This finding seems inconsistent with previous evidence on processing reflexives in a second language (Felser & Cunnings, 2012; Felser, Sato, & Bertenshaw, 2009). For example, the study by Felser and Cunnings (2012) explored how German-speaking learners of English process reflexive pronouns, using materials like "*Helen has worked at the army hospital for years. **She** noticed that the soldier had wounded **himself/herself** while on duty in the Far East.* ". Their results showed that the non-native speakers of English unanimously exhibited a preference for gender constraint, and attempted to interpret reflexives through discourse-based coreference assignment rather than syntactic binding.

Then, one question arises: why there was such variance within a biographically homogenous group of Chinese-English bilinguals[3] regarding the gender constraint in L2 reflexive pronoun resolution but not in the L1? One possible reason is that a stronger orthographic-semantic association of gender encodings may have been developed in their native language due to abundant language experience, which compensates for their reduced sensitivity to lexical gender information, so that the orthographic gender codes "*他(he/him)*" and "*她(she/her)*" activate more automatically semantic gender representation at the conceptual level, initiating the gender binding process between the reflexive pronoun and the antecedent. However, in their non-native language, there seems a continuum in the participants' accuracy in gender congruity processing from being low to high. Correspondingly, the orthographic-semantic association of gender encodings may be still developing along a continuum from being weak to strong, regardless of their overall language proficiency. Thus, for Chinese-English bilinguals who are at a lower scale along this continuum, the orthographic gender codes *"he/him"* and *"she/her"* probably could not activate the gender representation automatically at the conceptual level. Consequently, they could be simply unaware of the gender incongruity throughout the experiment.

Furthermore, the current findings suggest that caution should be taken in the field of L2 reflexive pronoun research regarding gender constraint (i.e., testing relevant theories: the Initial Filter approach, Nicol & Swinney, 1989; the Defeasible Filter approach, Sturt, 2003; the Multiple Constraints approach, Badecker & Straub, 2002) when using participants such as Chinese-English bilinguals or non-native speakers from other linguistic backgrounds (e.g., Spanish) who have reduced sensitivity to lexical gender. For example, Spanish is a pro-drop language and many sentences do not require gender to be encoded. The study by Antón-Méndez (2010) on Spanish speakers of L2 English found that they produced significantly more gender errors than any other type of pronoun error, and significantly more gender errors than French speakers of L2 English. Therefore, the variation of the function of the gender constraint among different groups of people should be taken into consideration in further research on the processing of reflexives in a second language.

Finally, the materials used in the current study contain within-phrase anaphora without the interference of inaccessible antecedent. A new set of materials with long-distance anaphora and inaccessible antecedents like the sentence "*Helen thinks*

[3] The Chinese-English bilinguals recruited in the current study were similar to each other in their proficiency level and in the years of time spent on English learning.

that David hates himself/herself" could be used in a future study to see whether consistent findings could be found, i.e., whether Chinese-English bilinguals prefer locality constraint to gender constraint in L2 reflexive pronoun resolution.

5. Conclusion

In summary, the present study demonstrates that the rule of gender constraint is applied in L1 and L2 reflexive pronoun resolution by Chinese-English bilinguals, as reflected by a gender incongruity effect in the form of LAN-P600 pattern in the L1 and Earlier Positivity-P600 pattern in the L2, which are elicited upon processing the reflexive pronoun that is incongruent with the antecedent in gender. However, the gender constraint in the L2 manifested large variance, such that the amplitude of the P600 effect decreased linearly with the participants' inability to detect gender incongruity. The variation of the gender constraint should be taken into consideration in future research on L2 reflexive pronoun resolution.

Acknowledgments

We thank Miss Yifei Ji and Mr. Zhibing Yu from Guangdong University of Foreign Studies for their helpful comments on the earlier version of the manuscript. The research is part of the project "Development of Language Competence" funded by the Ministry of Education in China.

References

Antón-Méndez, I. (2010). Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research, 39*, 119–139.
Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology, 28*(4), 748–769.
Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras, & C. Clifton (Eds.), *Online-study of sentence comprehension: Eye tracking, ERP and beyond* (pp. 309–327). Brighton, UK: Psychology Press.
Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology, 23*, 45–93.
Callahan, S. M. (2008). Processing anaphoric constructions: Insights from electrophysiological studies. *Journal of Neurolinguistics, 21*, 231–266.
Chen, J. Y., & Su, J. J. (2011). Differential sensitivity to the gender of a person by English and Chinese speakers. *Journal of Psycholinguistic Research, 40*, 195–203.
Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.
Coren, S. (1992). *The left-hander syndrome: The causes and consequences of left-handedness*. New York: Free Press.
Dillon, B. (2014). Syntactic memory in the comprehension of reflexive dependencies: An overview. *Language and Linguistics Compass, 8*(5), 171–187.
Dong, Y. P., & Jia, T. (2011). Zhongguo yingyu xuexizhe qianyanyu yuxin bianma: Laizi yingyu daici xingbie cuowude zhengjv [Composition of preverbal message: Gender errors of English pronouns by Chinese learners of English]. *Xiandai Waiyu (Modern Foreign Languages), 34*(3), 279–286.
Dong, Y. P., Wen, Y., Zeng, X. M., & Ji, Y. F. (2014). Exploring the cause of English pronoun gender errors by Chinese learners of English: Evidence from the self-paced reading paradigm. *Journal of Psycholinguistic Research*. http://dx.doi.org/10.1007/s10936-014-9314-6.
Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2009). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience, 22*(8), 1870–1887.
Felser, C., & Cunnings, I. (2012). Processing reflexives in a second language: The timing of structural and discourse-level constraints. *Applied Psycholinguistics, 33*, 571–603.
Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition, 12*(4), 485–502.
Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*, 45–63.
Garrod, S., & Terras, M. (2000). The contribution of lexical and situational knowledge to resolving discourse roles: Bonding and resolution. *Journal of Memory and Language, 42*, 526–544.
Gehring, W. J., Liu, Y., Orr, J. M., & Carp, J. (2012). The error-related negativity (ERN). In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 231–291). Oxford: Oxford University Press.
Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*, 1531–1549.
Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2005). Pronominal reference in sentences about person and things: An electrophysiological approach. *Journal of Cognitive Neuroscience, 17*(2), 227–239.
Kreiner, H., Sturt, P., & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language, 58*, 239–261.
Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods, 38*(2), 202–210.
Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement. *Cortex, 66*, 156–159.
Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*, 908–930.
Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics, 34*, 15–36.
Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and pseudo-words. *Journal of Cognitive Neuroscience, 9*(3), 318–329.
Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*(1), 5–19.
Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38*, 752–760.
Nieuwland, M. S. (2014). "*Who's he?*" Event-related brain potentials and unbound pronouns. *Journal of Memory and Language, 76*, 1–28.
Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition, 25*(3), 273–285.
Osterhout, L., & Mobley, L. A. (1995). Event-related potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.
Qiu, L. J., Swaab, T. Y., Chen, H. C., & Wang, S. P. (2012). The role of gender information in pronoun resolution: Evidence from Chinese. *PLoS One, 7*(5), e36156.
Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning, 62*(2), 172–188.
Sanford, A. J., Garrod, S. C., Lucas, A., & Henderson, R. (1983). Pronouns without explicit antecedents. *Journal of Semantics, 2*, 303–318.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542–562.
Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A commentary on "grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex, 66*, 149–155.
Van Berkum, J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.
Xu, X. D., Jiang, X. M., & Zhou, X. L. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain and Cognition, 81*, 223–236.