J Psycholinguist Res
DOI 10.1007/s10936-014-9314-6

Exploring the Cause of English Pronoun Gender Errors by Chinese Learners of English: Evidence from the Self-paced Reading Paradigm

Yanping Dong · Yun Wen · Xiaomeng Zeng · Yifei Ji

Abstract To locate the underlying cause of biological gender errors of oral English pronouns by proficient Chinese-English learners, two self-paced reading experiments were conducted to explore whether the reading time for each 'he' or 'she' that matched its antecedent was shorter than that in the corresponding mismatch situation, as with native speakers of English. The critical manipulation was to see whether highlighting the gender information of an antecedent with a human picture would make a difference. The results indicate that such manipulation did make a difference. Since oral Chinese does not distinguish 'he' and 'she', the findings suggest that Chinese speakers probably do not usually process biological gender for linguistic purposes and the mixed use of 'he' and 'she' is probably a result of deficient processing of gender information in the conceptualizer. Theoretical and pedagogical implications are discussed.

Keywords Biological gender · Gender errors · Bilingualism · Chinese learners of English · L1 thinking for L2 speaking

Introduction

Assessing the biological gender of a human person is a simple piece of common sense, but gender errors (i.e., 'errors of biological gender' in the present paper) of oral English pronouns are prevalent among Chinese learners of English. No matter how proficient they are in English, they may produce 'she' when they should have used 'he' to refer to a male person. Dong and Jia (2011) found an average error rate of 6.47 % in a corpus survey (data from second-year English majors in college in China), and a much higher rate of 15 % in a task of Chinese-English interpreting by proficient Chinese learners of English (i.e., 10 years learning English in primary and middle schools and 2.5 years as English majors in a university). According

Y. Dong (✉) · Y. Wen · X. Zeng · Y. Ji
National Center of Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Baiyun Dadao North 2#, Guangzhou 510420, China
e-mail: ypdong@gdufs.edu.cn

Published online: 02 September 2014

Table 1 English-Chinese oral forms for 'he' and 'she' and other corresponding grammatical cases

	Nominative	Accusative	Possessive	Reflexive
Male	he–ta	him–ta	his–tade	himself–ta ziji
Female	she–ta	her–ta	her–tade	herself–ta ziji

to James' (1998) criterion of comprehensibility and Khalil's (1985) criteria of intelligibility, acceptability and irritation, this mixed use of 'he' and 'she' (and corresponding pairs of other grammatical cases) is a serious problem.

The problem of gender errors with Chinese learners of English seems different from that of other English learners reported in the literature. Felix and Hahn (1985), for example, reported that German EFL beginning learners made more gender errors than personal, number or case errors for pronouns. However, gender errors ceased to occur as learners' language proficiency increased, which is not true with Chinese-English bilinguals. Besides, the error rates found in Dong and Jia (2011) are much higher than those reported in Antón-Méndez (2010a, 2010b) with Dutch-English, Spanish-English and Italian-English bilinguals. Using data in the LINDSEI corpus (Louvain International Database of Spoken English Interlanguage corpus), Chen (2004) found that Chinese EFL learners' gender error rate was 17.65 %, much higher than that of Japanese and French EFL learners whose native language distinguishes third-personal pronouns for male and female antecedents (error rate: 4.2 %, 0.92 %, respectively). The influence of the particular L1 background in the case of Chinese-English bilinguals, therefore, seems stronger and more persistent. Chen and Su (2011) even found that Chinese speakers were less accurate than English speakers when answering questions later about the gender of a protagonist in a story they had heard or read.

This gender problem with Chinese learners of English is probably a result of L1 transfer. First of all, in its oral forms, the Chinese language does not distinguish between the male, the female or the 'neutral' 3rd person singular pronouns and uses the same sound 'ta' for each of the corresponding English forms of 'he', 'she' and 'i', and their corresponding forms in other grammatical cases (see Table 1). In other words, the biological gender is unmarked in the 3rd person pronouns in Chinese while it is marked in English. Second, there are a lot of 'φ pronouns' in Chinese and to use a pronoun to refer to an antecedent is therefore less frequent than in English (Xu 2000).[1] But the question remains what L1 transfer means exactly or how it happens exactly.

There is not much research about the problem of L2 gender errors in the literature up till now. Antón-Méndez (2010a) compared performances from Spanish-English and French-English bilinguals who were asked to answer questions in English in a task designed to elicit pronouns. Spanish-English bilinguals were found to not only make significantly more gender errors for the 3rd person singular nominative pronouns (i.e., mixed use of 'he' and 'she') than their French counterparts (4.30 % for Spanish, 0.68 % for French), but also make more gender errors than any other kinds of pronoun errors such as omissions. According to Antón-Méndez (2010a), these results are not surprising since, in Spanish which is a pro-

[1] For English, according to the data provided in Ariel (1990), pronouns make up 6 % of the total word count and 71 % of all the referring expressions (with 61 % of these pronouns referring to an antecedent in a preceding sentence and 21 % within the same sentence). For Chinese, according to the principles and data provided in Xu (2000), pronouns make up 1 % of the total word count and 14 % of all the referring expressions (with 39 % of these pronouns referring to an antecedent in a preceding sentence and 54 % within the same sentence). What makes up for the small percentage of the pronouns used in Chinese are the so-called 'φ pronouns'. They take up 3.8 % of the total word count, 22 % of which are used to refer to an antecedent in a preceding sentence, while 80 % are to an antecedent within the same sentence. The use of proper nouns as referring expressions is another remedy for the small percentage of pronouns in Chinese.

drop language, nominative person pronouns are often omitted except in cases of emphasis. The author suggests that this is not a result of L1 transfer because otherwise there would be more errors of nominative pronoun omission and that it is a failure of encoding gender information in the preverbal message (as described by Levelt 1989) when producing English. In a follow-up study, Antón-Méndez's (2010b) asked Dutch-English, Spanish-English and Italian-English bilinguals to turn an English description of a picture such as 'My mother accompanies the teacher to the school' into 'His/Her mother …'. By comparing the gender error rates of the three different groups of bilinguals, the author suggests that the problem of gender errors in possessive pronouns is a result of transfer of L1 syntactic procedures. Dong and Jia (2011) compared the error rates of different grammatical cases of the 3rd person singular pronouns both in a corpus survey and in an interpreting task by Chinese learners of English and found that there were more gender errors in places that were more easily neglected (the nominative and possessive cases instead of the accusative and reflexive cases), which the authors claimed is consistent with the hypothesis that gender information is encoded in the conceptualizer. In short, all the three relevant studies reviewed here suggest that the cause of English pronoun gender errors is still not clear and merits more research. The current study, employing the self-paced reading paradigm instead of production data, is an attempt in this direction.

Self-paced Reading Experiment

Like the study on co-reference processing by Kennison and Trofe (2003), participants read chunk by chunk in the self-paced reading task. A chunk may be a word or a phrase or even a clause with a maximum of five words. Readers move one step forward by clicking a key each time. Readers understand the present chunk in the context of preceding chunks and while clicking the key for the next chunk, predict what follows. The RT to a specific chunk is therefore the result of the prediction from preceding chunks and the processing of the present chunk. The RT to a pronoun like *he* or *she* is thus the result of the activation of relevant information in a preceding antecedent and the reading of the pronoun itself. If the pronoun is consistent with the activation of relevant information, the RT will be short; otherwise, it will be long. That is *the mismatch effect*.

As far as we know, *the mismatch effect* (i.e., longer reading time for the pronoun that mismatched its antecedent in gender than that for the pronoun that matched) has been found in all the relevant studies conducted with native speakers of languages that distinguish 'he' and 'she' (e.g., Cacciari et al. 1997; Carreiras et al. 1996; Foertsch and Gernsbacher 1997; Kennison and Trofe 2003; Duffy and Keir 2004; Sturt 2003). Some of these studies (Duffy and Keir 2004; Sturt 2003) recorded reading time with eye movements, a paradigm quite similar to self-paced reading. For example, Duffy and Keir (2004) found the mismatch effect in which native speakers of English read English sentences that contained a reflexive ('himself' or 'herself') mismatching/matching its role name antecedent (e.g., 'electrician') in a neutral context (without specifying the gender of the role name). According to Duffy and Keir (2004), the mismatch effect indicates that gender stereotypes are automatically activated in the absence of disambiguating information. In fact, the mismatch effect is a basic finding with languages that distinguish 'he' and 'she', which suggests that gender information of an antecedent is automatically activated in reading. But what about native speakers of the Chinese language that does not distinguish 'he' and 'she' in its oral form?

Dong and Li (2011) conducted a self-paced reading experiment with intermediate to advanced Chinese learners of English and recorded the reading time for each 'he' or 'she'

that matched or mismatched its corresponding antecedent. Antecedents were either celebrity names (e.g. Bill Clinton or Gong Li) or career names with prototype gender (e.g. nurse or plumber). Contrary to the literature, Dong and Li (2011) failed to find such a mismatch effect for either type of antecedents (celebrity names or career names with prototype gender). Further analysis of the data found a mismatch effect for only female celebrity names. Questionnaire data collected after the experiment indicated that participants found the gender of female celebrity names more salient than that of either male celebrity names or career names. The question is therefore whether the gender saliency of antecedents influences the presence of the mismatch effect for Chinese learners of English. The hypothesis is that the mismatch effect would appear if the gender information is made salient. The verification of the hypothesis suggests that the gender information of an antecedent is not as automatically activated for native speakers of Chinese as it is for native speakers of English.

The two experiments conducted in the current study, therefore, intended to verify the hypothesis that, the mismatch effect, consistently found with native speakers of English, would appear with Chinese learners of English only when the gender information of an antecedent is made salient. The first experiment, conducted in English using the same self-paced reading paradigm as in Dong and Li (2011), manipulated the saliency of the critical human antecedent by pairing the antecedent with either a gender-consistent human picture or the picture of a genderless object. The use of a picture is motivated by Lemm et al. (2005) who suggest that pictures of people can prime gender-related concepts.[2] The second experiment, without any manipulation of the saliency of the critical antecedent, consisted of two identical sub-experiments with one conducted in English and the other in Chinese. This second experiment without saliency manipulation served as a baseline for the first one with manipulation, but since materials for the second one came from the first one (only half of the materials in the first one were needed for the second one), the first one with saliency manipulation is reported here first.

Experiment One

Experiment One was intended to explore whether enhancing the gender saliency of an English antecedent would influence the presence or absence of the gender mismatch effect with Chinese-English learners. The prediction was that in a self-paced reading paradigm by Chinese-English learners, the mismatch effect would appear when the gender saliency of an antecedent was enhanced by a gender-consistent human picture, but there would be no mismatch effect if the antecedent was accompanied by a genderless non-human object picture.

Methods

Participants

Sixty-six undergraduate Chinese learners of English (33 females and 33 males) participated in return for a small amount of payment. They were third-year or fourth-year English majors in a

[2] According to a latest study (Palmiero et al. 2013), both visual and auditory mental image can be differentiated from semantic representation. A human picture added before the presentation of an antecedent, therefore, may be capable of augmenting features that are not clearly (or saliently) encoded in the semantic representation of the antecedent, and the biological gender of a human picture is one of the most basic human features that are automatically activated (Lemm et al. 2005). The augmented gender feature of the antecedent may thus aid the feature-check process in pronominal interpretation.

university in China and had learnt English for about 12 years (from the third grade in primary schools when they were 8 years old). Third-year and 4-year English majors are generally considered intermediate to advanced English learners. Since the rule of gender agreement, especially biological gender agreement, is a very simple piece of knowledge which every learner is told from the very beginning and, given the length of English instruction and the participants' major, we can therefore safely assume it was known by our participants. In consequence, there was no need to organize an objective proficiency test for these participants.

Materials and Design

The experiment was conducted using a self-paced paradigm with a design of 2 (Picture: human vs. non-human pictures) × 2 (Antecedent: male vs. female names) × 2 (Match: matched vs. unmatched). There are two sentences in each experimental sentence set, with the first sentence starting with a typical English name like Mark or Mary (i.e., antecedent) and the second starts with 'he' or 'she' (i.e., matched or not matched). Here is an example set for a male or female antecedent (with each chunk marked by a star and with the critical pronoun always in the 9th chunk): Mark/Mary *goes*to the zoo*to watch*animals*every day*after work*for a good rest.* He/She *considers it*the best way*to relax*and*maintain*a good mood. Before each sentence set, there was a picture. Half of the experimental sentence sets would have a human picture that matches its antecedent in gender, and the other half would have a non-human picture that matches some entity mentioned in the sentences. Like similar studies in the literature (e.g., Dong and Jia 2011; Kennison and Trofe 2003), the dependent variable is the RT to each 'he' or 'she' that matched or did not match its antecedent.

Sixty-four typical English male names (e.g. Mark) and sixty-four female ones (e.g. Mary) were selected as the antecedents according to the results in a rating study. In this preparatory study, twenty-seven participants from the same population were asked to rate 150 English first names (75 male names, 75 female names) on a 7-point Likert scale. On this rating questionnaire, '1' stands for 100 % sure that this name is a female one, while '7' stands for 100 % sure that it's a male name. The meaning rating for the 64 male names was 6.71 (range 6.1–7) and for 64 female names 1.17 (range 1–1.78). This is to exclude the possible confounding factor of participants' ignorance of English names used as antecedents in the experiment.

Sixty-four experimental sentence sets were created and counterbalanced across the eight conditions resulting from crossing the three variables, producing eight lists of 64 sets. Each participant would read only one list. To ensure that the eight lists were similar in all other respects, we conducted a baseline study with the original 64 sets, with all the three variables controlled, i.e., with only male names as antecedents and the matched pronoun 'he' as the critical pronoun. No picture was used in this baseline. Twenty-eight students from the same population described above participated for course credits (but they did not participate in the main experiments). The same self-paced chunk-by-chunk moving window was used, the procedure of which was exactly the same as in the formal experiment. Five participants failed to reach the 85 % accuracy rate in answering comprehension questions and their data was discarded. There was altogether 3.06 % of the original data deleted in data trimming (according to a standard of 2.5 SD). With the remaining data, the mean RT to the critical pronoun 'he' in each sentence was calculated, according to which we created eight groupings of sentence sets such that all groupings had the same average RT to the pronoun (within the range of 433.88ms to 435.41ms). That is, no two lists among the eight were different from each other in terms of the average RT to the critical pronoun.

In addition to the experimental sentence sets, 64 filler sentence sets were created. Consisting of two sentences, half of the fillers had *my* plus a human noun (e.g. *my teacher*) as the subject of the first sentence while the other half had the pronoun *I* as the subject. Neither English proper names nor single third-personal pronouns appeared in the fillers.

To ensure that the reading of the critical pronoun 'he' or 'she' was only affected by the variables manipulated, all the sentences constructed, including both experimental and filler sentences, were in a gender-neutral context. For each sentence set, a simple comprehension question was created to check whether participants understood the materials or not. The colored pictures used in the experiments were collected from the Internet and they were of 500×500 resolution.

Procedures

The experiment was conducted in the Bilingual Cognition and Education Lab, in which computers installed with E-prime were used to collect data. The experiment consisted of two blocks. The first block, a practice block of eight trials, helped participants to get familiar with the procedure. The second block, an experimental block of 128 sentence sets (64 experimental and 64 fillers), first presented participants with four fillers and then presented them randomly the rest of the 124 sets. Both blocks progressed in the same procedure. Each trial began with a picture located in the middle of the screen for 1,000 ms. Sentence sets were then displayed in a self-paced reading paradigm, in which participants read each set one chunk at a time by clicking the space bar. With the first click, several lines of dashes on the screen would appear. With the next click, the first dash would be replaced by the first chunk. Each subsequent clicking would turn the previous chunk into a dash and at the same time, present the next chunk. At the end of a trial, the participant was asked to answer a yes-or-no question by pressing a 'YES' or 'NO' key labeled on the keyboard. The next trial would start with another press of the spacebar. The entire experimental session lasted for about 35 minutes.

Results

Participants' RTs to the critical pronoun 'he' or 'she' were recorded as well as the first and second chunks after the pronoun (e.g. Clifton et al. 1997; Kennison and Trofe 2003). Two participants (one male and one female) whose accuracy rates for comprehension questions were below 85 % were excluded from further analyses (with all other participants scoring higher than 90 %), resulting in 4 male and 4 female participants taking each of the eight experimental lists. The mean accuracy rate for the remaining 64 participants was 94.69 % (SD 3.07). Regarding RTs, we took a standard of 2.5 SD for data trimming, with a data loss of 3.37 %.

Table 2 is a summary of the mean RTs for the eight combinations of variables.

For the critical pronoun region, a repeated-measures ANOVA was conducted on the three within-subject variables. No main effect was significant. The interaction between Antecedent and Match was significant ($F1(1, 63) = 26.15$, $p < .001$; $F2(1, 63) = 19.02$, $p < .001$). And the interaction between Picture and Match approached significance in the subject analysis ($F1(1, 63) = 3.46$, $p = .067$; $F2(1, 63) = 2.77$, $p = .101$). All the other interactions were not significant.

Simple effect analyses were then performed to see how the variable of Match was at work in different conditions of Picture. With human pictures, the RTs to the pronouns that matched their antecedents in gender were shorter than the RTs in the unmatched conditions

Table 2 Mean RTs (in ms with SD in brackets) for critical pronouns and the two post-pronoun regions in each combination of variables (picture * antecedent * match), and the presence of the mismatch effect in Experiment 1

Picture	Antecedent	Match	Pronoun	Post1	Post2
Human picture	Male	Match(he)	437 (55)	577 (118)	746 (189)
		Mismatch(*she)	472 (78)	626 (149)	752 (212)
	Female	Match(she)	461 (63)	574 (127)	740 (239)
		Mismatch(*he)	452 (71)	633 (181)	805 (267)
	Mismatch effect		Yes	Yes	No
Non-human picture	Male	Match(he)	439 (54)	608 (156)	730 (171)
		Mismatch(*she)	453 (64)	602 (161)	736 (224)
	Female	Match(she)	464 (77)	574 (132)	744 (199)
		Mismatch(*he)	445 (58)	611 (150)	764 (237)
	Mismatch effect		No	No	No

($F1(1, 63) = 4.35$, $p = .041$; $F2(1, 63) = 3.68$, $p = .060$). This contrast did not exist with non-human pictures ($F1(1, 63) = .16$, $p = .688$; $F2(1, 63) = .20$, $p = .656$).

For the first post-pronoun region, the same analyses were applied. The main effect of Match was significant ($F1(1, 63) = 10.17$, $p = .002$; $F2(1, 63) = 16.58$, $p < .001$). The interaction between Match and Picture was significant ($F1(1, 63) = 14.03$, $p < .001$; $F2(1, 63) = 17.47$, $p < .001$). All the others were not significant.

The same simple effect analyses were then performed for this first post-pronoun region, which produced the same result. With human pictures, the RTs to the pronouns that matched their antecedents in gender were shorter than the RTs in the unmatched conditions ($F1(1, 63) = 14.03$, $p < .001$; $F2(1, 63) = 17.47$, $p < .001$). This contrast did not exist with non-human pictures ($F1(1, 63) = 1.34$, $p = .251$; $F2(1, 63) = 2.12$, $p = .150$).

For the second post-pronoun region, the same analyses only found the main effect of Match was significant by items ($F1(1, 63) = 2.49$, $p = .119$; $F2(1, 63) = 5.03$, $p = .028$). Simple effect analysis found a weaker contrast here than found in the two regions preceding this one. With human pictures, the RTs to the pronouns that matched their antecedents in gender were shorter than the RTs in the unmatched conditions only by items analysis ($F1(1, 63) = 2.42$, $p = .125$; $F2(1, 63) = 5.84$, $p = .019$). The contrast did not exist with non-human pictures ($F1(1, 63) = .33$, $p = .570$; $F2(1, 63) = 1.01$, $p = .318$).

Table 3 presents the results of the contrast between the two conditions of the variable Picture. The mismatch effect (more time reading pronouns mismatching antecedents in gender) did not appear unless participants were first presented a human picture (that matched the antecedent in gender).[3]

[3] We also ran separate analysis with the two pronouns "he" and "she". For example, with each "he", the preceding antecedent may be matched or mismatched in gender. The results indicate that, with a gender-consistent human picture preceding the antecedent, a robust mismatch effect appeared for either "he" or "she" in the first segment after the pronoun [For "he": $F1(1, 63) = 7.820$, $p = .007$; $F2(1, 63) = 7.081$, $p = .01$; for "she": $F1(1, 63) = 7.327$, $p = .009$; $F2(1, 63) = 9.421$, $p = .003$]. No such mismatch effect appeared with a non-human object picture preceding the antecedent. This is further evidence supporting the conclusions. Readers may write to the corresponding author for the full results of this statistical analysis, which we omitted here for better readability.

Table 3 Summary of the mismatch effect in each position under either the human or non-human picture condition in Experiment 1

Picture	Pronoun	Post1	Post2
Human picture	F1(1, 63) = 4.35, $p = .041$; F2(1, 63) = 3.68, $p = .060$	F1(1, 63) = 14.03, $p < .001$; F2(1, 63) = 17.47, $p < .001$	F1(1, 63) = 2.42, $p = .125$; F2(1, 63) = 5.84, $p = .019$
Non-human picture	F1(1, 63) = .16, $p = .688$; F2(1, 63) = .20, $p = .656$	F1(1, 63) = 1.34, $p = .251$; F2(1, 63) = 2.12, $p = .150$	F1(1,63) = .33, $p = .570$ F2(1, 63) = 1.01, $p = .318$

Discussion

The prediction for Experiment One is fulfilled. That is, the mismatch effect appeared when gender saliency of an antecedent was enhanced by a human picture matched in gender, but there was no mismatch effect when the antecedent was accompanied by a non-human object picture. The hypothesis that the manipulation of gender saliency of an antecedent affects the presence of the mismatch effect in comprehension has been supported.

The absence of the mismatch effect in the non-human object picture condition is contrary to findings in relevant studies conducted with native speakers of languages that distinguish 'he' and 'she' (Cacciari et al. 1997; Carreiras et al. 1996; Foertsch and Gernsbacher 1997; Kennison and Trofe 2003). For example, Kennison and Trofe (2003), with no manipulation of gender saliency, found that the critical pronoun 'he' or 'she' resulted in longer reading time when its gender information mismatched the stereotyped gender information of its antecedent. However, the absence of the mismatch effect when gender saliency was not enhanced is consistent with the fact that there is no gender distinction in pronouns referring to people in oral Chinese, which suggests that L1 procedures may still dominate in L2 processing.

Experiment Two

It could be argued that the absence of the mismatch effect in Experiment One when the antecedent was accompanied by a non-human object picture was due to the existence or disturbance of the picture, not due to the particular nature of oral Chinese. To refute this possibility, two sub-experiments were conducted. The first one, conducted in English, was identical to Experiment One except that there was no picture accompanying the antecedent. The second one, conducted in Chinese, was identical to the first sub-experiment. The prediction was that the mismatch effect would not appear in either the English sub-experiment or the Chinese sub-experiment.

Methods

Participants

Thirty-two undergraduate Chinese learners of English (16 females and 16 males) participated in return for a small amount of payment. They were from the same population as the participants in Experiment One, but none of them participated in Experiment One.

Materials and Design

Experiment Two consisted of two sub-experiments with one in English and one in Chinese. The two sub-experiments were identical in their design but different in materials.

The experiment employed the self-paced paradigm within a 2 (Antecedent: male vs. female names) by 2 (Match: matched vs. unmatched) design. As in Experiment One, there were two sentences in each experimental sentence set, with the first sentence starting with a typical name like Mark or Mary (i.e., antecedent) and the second starting with 'he' or 'she' (i.e., matched or not matched). All the materials in the English sub-experiment, that is, 32 experimental sentence sets and 32 fillers, were randomly selected from Experiment One, and the Chinese materials were specially created for Experiment Two. Here is an example of a Chinese sentence set for a male or female antecedent (with each chunk marked by a star and with the critical pronoun always in the 9th chunk): 刘力超/韩可莹*得知*这次考试的*成绩*之后,*看起来*非常*不开心。*他/她*不想*和任何人*说话。 (Liu Lichao/ Han Keying seemed to be very upset after being informed of the result of the exam. He/She didn't want to talk to anybody). Again, the dependent variable is the RT to each 'he' or 'she' that matched or did not match its antecedent.

To create the Chinese sentences for the Chinese sub-experiment, we had to first collect 32 typical male names and 32 typical female names in a rating preparatory study. In this rating study, 29 participants who did not take part in the reading experiment were asked to rate 146 Chinese names (73 male names, 73 female names) on a 7-point Likert scale. Again as in Experiment One, '1' in the scale stands for 100 % sure that this name is a female one, while '7' stands for 100 % sure that it's a male name. According to the results of the ratings, 32 typical Chinese male names (e.g. Liu Lichao) and 32 female ones (e.g. Han Keying) were selected as the antecedents. The mean rating for the 32 male names was 6.66 (range 6.41–6.93) and for 32 female names 1.15 (range 1.03–1.24).

Thirty-two Chinese experimental sentence sets were created and counterbalanced across the four conditions resulting from crossing the two variables, producing four lists of 32 sets, the same arrangement as in the English sub-experiment. In addition to the experimental sentence sets, 32 Chinese filler sentences were created. Consisting of two sentences, half of the fillers had *my* plus a human noun (e.g. *my teacher*) as the subject of the first sentence while the other half had the pronoun *I* as the subject. Neither proper names nor single third-personal pronouns appeared in the fillers.

To ensure that the reading of the critical pronoun 'he' or 'she' was only affected by the variables manipulated, all the sentences constructed, including both experimental and filler sentences, were in a gender-neutral context. For each sentence set, a simple comprehension question was created to check whether participants understood the materials or not.

Procedures

The experiment was again conducted in Bilingual Cognition and Education Lab, in which computers installed with E-prime were used to collect data. Half of the participants did the English sub-experiment first and then the Chinese one, with the other half completing the two in the reverse order. In between the two sub-experiments, participants were required to listen to a piece of tune that had no words in it and that lasted for a little more than one minute so that the next sub-experiment would not be influenced by what the participant had done before.[4] The procedure was similar to that of Experiment One. Each sub-experiment consisted of two blocks. The first block, a practice block of eight trials, helped participants to get familiar with the procedure. The second block, an experimental block of 64 sentence

[4] Possible influence of previous trials on later trials in the experiment would enhance participants' awareness of the purpose of the experiment and thus increase the possibility of noticing the presence of the mismatch effect.

Table 4 Mean RTs (in ms with SD in brackets) for critical pronouns and the two post-pronoun regions in each combination of variables (antecedent * match) in the *English* sub-experiment, with a summary of the main effect of the variable match in each position

Antecedent	Match	Pronoun	Post1	Post2
Male	Match(he)	413 (93)	503 (181)	594 (243)
	Mismatch(*she)	434 (99)	512 (183)	592 (246)
Female	Match(she)	429 (93)	519 (198)	578 (247)
	Mismatch(*he)	423 (88)	506 (171)	597 (255)
Mismatch effect		$F1(1, 29) = 2.62$, $p = .116$; $F2(1, 31) = 1.42$, $p = .242$	$F1(1, 29) = 0.01$, $p = .936$; $F2(1, 31) = .08$, $p = .775$	$F1(1, 29) = 0.12$, $p = .735$; $F2(1, 31) = .15$, $p = .228$

Table 5 Mean RTs (in ms with SD in brackets) for critical pronouns and the two post-pronoun regions in each combination of variables (antecedent * match) in the *Chinese* sub-experiment, with a summary of the main effect of the variable match in each position

Antecedent	Match	Pronoun	Post1	Post2
Male	Match(he)	360 (93)	328 (86)	359 (108)
	Mismatch(*she)	360 (96)	332 (88)	340 (96)
Female	Match(she)	359 (90)	325 (83)	345 (105)
	Mismatch(*he)	352 (82)	329 (86)	352 (115)
Mismatch effect		$F1(1, 30) = .85$, $p = .364$; $F2(1, 31) = .27$, $p = .609$	$F1(1, 30) = 1.54$, $p = .225$; $F2(1, 31) = .33$, $p = .572$	$F1(1, 30) = 1.24$, $p = .274$; $F2(1, 31) = .32$, $p = .577$

sets (32 experimental and 32 fillers), first presented participants with four fillers and then presented them randomly the rest of the 60 sets. Each block progressed in the same fashion as in Experiment one. The entire experimental session lasted for about 35 minutes.

Results

Participants' RTs to the critical pronoun 'he' or 'she' were recorded as well as the first and second chunks after the pronoun. For the English experiment, two participants (two male participants) whose accuracy rates for comprehension questions were below 85 % were excluded from further analyses. The mean accuracy rate for the remaining 30 participants was 92.71 % (SD 2.77). For data trimming, we adopted a standard of 2.5 SD, with a data loss of 3.37 %. For the Chinese experiment, one male participant was excluded because of an accuracy rate lower than 85 %. The accuracy rate for the remaining 31 participants was 95.77 % (SD 2.48) and 2.79 % of the data was excluded after data trimming.

Table 4 is a summary of the mean RTs for the four combinations of variables (antecedent by match) in the English sub-experiment, and Table 5 is a summary of corresponding data in the Chinese sub-experiment.

For each of the three regions in either the Chinese or English sub-experiment, a repeated-measures ANOVA was conducted on the two within-subject variables of Antecedent and

Match. The main effect of Match is of our concern, since it indicates whether the mismatch condition requires longer reading time than the match condition. The main effect for either Antecedent or Match in each region in each sub-experiment was not significant ($ps > .05$). The main effects of the variable Match for each of the sub-experiment are listed in Tables 4 and 5.[5]

Discussion

The prediction for the absence of the mismatch effect in either the English or Chinese sub-experiment is fulfilled. Generally speaking, this absence of the mismatch effect indicates that the Chinese learners of English were not sensitive to the gender information encoded in the antecedent, and that gender information in the antecedent was not automatically activated in the reading process. This finding is again contrary to what has been found in all relevant studies conducted with native speakers of languages that distinguish 'he'or 'she' (e.g., Cacciari et al. 1997; Carreiras et al. 1996; Foertsch and Gernsbacher 1997; Kennison and Trofe 2003; Duffy and Keir 2004; Sturt 2003).

Together with the findings from Experiment One, the absence of the mismatch effect in the English reading in Experiment Two suggests that the mismatch effect, common to native speakers of English, would not have been found if the gender saliency of the antecedent had not been enhanced. This particular phenomenon with Chinese learners of English is probably due to the particular feature of oral Chinese that does not distinguish 'he' and 'she', as confirmed by the absence of the mismatch effect in the Chinese reading in Experiment Two. Although written Chinese does distinguish 'he' and 'she', speech is primary and has an overriding influence on reading, especially for one's first language, since children start to learn to read after they have mastered at least basic oral communication skills. The absence of the mismatch effect in English in Experiment Two can be accounted for by the same reason since these participants started to learn English as a foreign language in school around 8 years old.

General Discussion

The self-paced reading experiments in the current study found that manipulating the gender saliency of an antecedent made a difference when Chinese learners of English responded to a pronoun that agreed or disagreed in gender with an antecedent. The gender-mismatch effect was only found when an antecedent's gender information was enhanced, which is different from what has been found in corresponding studies with native English speakers (Carreiras et al. 1996; Kennison and Trofe 2003; Duffy and Keir 2004; Sturt 2003).

Pronominal interpretation is commonly assumed to involve two stages: *bonding* in which candidate antecedents are activated under feature constraints (e.g. gender and number) and *resolution* in which contextual information and real-world knowledge are integrated to help identify the referent (Callanhan 2008). The absence of the mismatch effect in the current study may be interpreted as a failure to check the agreement of the gender feature at the first stage. The finding that this mismatch effect appeared when the gender of an antecedent was made salient by a gender-consistent picture suggests that native Chinese speakers probably do not automatically process gender information in an antecedent (or even in a pronoun) and that

[5] We also ran separate analysis with the two pronouns "he" and "she" in the 2nd experiment. No mismatch effect appeared in any position with either "he" or "she" in either the Chinese or English sub-experiment. Readers may write to the corresponding author for a report of the full results of this analysis.

they do not process biological gender information for linguistic purposes. The result is that the gender feature of an antecedent is not part of, or at least a salient part of the stored semantic representation derived from previous processing of the antecedent. When the corresponding pronoun has to be processed, the gender information is probably not checked, leading to an absence of the mismatch effect.

The hypothesis that Chinese-English learners do not automatically process biological gender information for linguistic purposes offers an adequate explanation for what has been found by Chen and Su (2011), that is, Chinese speakers were less accurate than English speakers when answering questions later about the gender of a protagonist in a story they had read or heard. But this hypothesis seems inconsistent with ERP findings from Qiu et al. (2012) and Xu et al. (2013). Xu et al. (2013) found a P600 effect when native Chinese speakers processed third-person singular pronouns (in written Chinese) that mismatched corresponding antecedents in gender, a typical finding in studies with native speakers of languages like English (e.g., Osterhout and Mobley 1995; Osterhout et al. 1997). With similar methods, Qiu et al. (2012) found the same P600 effect when the distance between the pronoun and the antecedent was relatively long, and a N400 effect when the distance was shortened. These two ERP studies seem to suggest that Chinese speakers process the feature of biological gender for linguistic purposes in the same way as English speakers, which is apparently contrary to the conclusions obtained in the current study and in Chen and Su (2011) that adopted behavioral methods. To explain this difference, two lines of research inquiry are needed. First, how might the use of different antecedents influence experimental results? The gender information in an antecedent may be salient or obscure depending on what particular antecedent is used (and depending on the context of that antecedent). The antecedents in the current study are typical male or female names in English or in Chinese. Xu et al. (2013) used noun phrases like "男教师" (male teacher) or "女患者" (female patient) as antecedents, for which we believe gender was made salient by the Chinese character "男" (male/man) or "女" (female/woman). However, Qiu et al. (2012) used typical male or female names in Chinese, similar to ours. Second, what different mechanisms are measured in ERP and behavioral studies? Data indexes in ERP and behavioral studies are quite different, and conclusions that are inconsistent across the two kinds of studies are not rare in the literature (e.g., Wen 2013; Foucart and Frenck-Mestre 2012). Wen (2013), for example, found that when Chinese learners of English were reminded of the gender mismatch between an English pronoun and its antecedent, their behavioral accuracy rate increased (to nearly 100 %), but their P600 effect remained insignificant (which is significant for native English speakers in the same gender violations). Foucart and Frenck-Mestre (2012) found evidence that native speakers of French and English-French learners process grammatical gender differently when employing ERP methodology but not eye-tracking methodology. The authors explained that these discrepant results are probably due to the fact that "the two methodologies in fact tap into different types of processing" (p246). More studies including replications are certainly needed to clarify these questions, including questions about possible differences between behavioral and neuropsychological methodologies.

The aim of the current study was to locate the underlying cause of biological errors of oral English pronouns produced by Chinese learners of English. Comprehension data from the two self-paced experiments indicate that the cause may at least partly lie in the conceptual processing of antecedents, which is consistent with conclusions from Antón-Méndez (2010a). Although comprehension and production differ in many ways, retrieving the biological gender of an animate antecedent seems essential for both proper comprehension and production in most languages. The processing failure found in comprehension in the current study may mirror to a large extent what occurs at the conceptual level when Chinese-English learners

produce pronouns. In short, the cause of English pronoun gender errors by Chinese-English learners is probably that these learners seldom process biological gender information for linguistic purposes, and therefore the locus of biological gender errors is probably at the conceptual level. This conclusion is consistent with previous research from L1 research by Slevc et al. (2007) indicating that biological gender agreement (different from grammatical or number agreement) is probably not a syntactically circumscribed process but relies more on information encoding at the conceptual level.

Findings from the current study may be considered as evidence for the idea of "L1 thinking for L2 speaking". Chen and Su (2011) attributed their findings that Chinese speakers were less accurate than English speakers about the gender of a protagonist in a story to linguistic relativity, that is, the linguistic device associated with a specific language orients its speakers to a particular aspect of the world and results in increased sensitivity to that aspect for the purpose of speaking. The current study, with direct manipulation of the saliency of biological gender, is further support for the idea of "L1 thinking for L2 speaking" because participants in the current study processed the L2 in the same way as the L1 unless the concept of biological gender was enhanced. In fact, much more research of linguistic relativity has been done on grammatical gender and the general finding is that grammatical gender conveys semantic information to speakers (e.g., Boroditsky et al. 2003; Konishi 1993; Sera et al. 2002; Vigliocco et al. 2005). For example, the noun 'key' is marked as feminine in Spanish and is often described as 'golden, intricate, little, lovely, shiny, and tiny' by Spanish speakers, while in German it is marked as masculine and is often described as 'hard, heavy, jagged, metal, serrated, and useful' by German speakers (Boroditsky et al. 2003).

The current study has pedagogical implications for the problem of gender errors by Chinese learners of English. First, since the problem may mainly originate from Chinese speakers' deficient processing of gender information for linguistic purposes, pedagogical measures to solve the problem should take this into consideration. In fact, Rodriguez-Fornells et al. (2005) points out that the influence of L1 on L2 can happen in any stage of language processing. Different measures should be taken for problems originating from different stages of language processing. Second, the mixed use of "he" and "she" by Chinese EFL learners is not just a casual mistake that could be neglected by both students and teachers. But the fact is, as far as we know, both students and teachers seldom take this mistake seriously. The reason is probably that Chinese EFL learners have few genuine communicative needs to use their English. They may practice talking to each other in make-believe situations in the classroom in which comprehension is not a problem and everyone believes they have made themselves clear. According to Johnson (1996), L2 learners stop making progress when their communicative needs are met. Kowal and Swain (1997) insist that as long as students are able to communicate their intended meaning to one another, there is little impetus for them to be more accurate in the form of the language they are using to convey their message.

Language proficiency may be a factor influencing how Chinese learners of English process the biological gender information of an antecedent in either comprehension or production. But since biological gender information is a very obvious piece of information and the participants in our previous and present studies (e.g., Dong and Jia 2011; Dong and Li 2011) were proficient in English (intermediate to advanced learners of English), we did not examine language proficiency in our study. We hypothesize that the most crucial factor is not general language proficiency but whether the learner is able to automatically process the biological gender information of a human antecedent in language comprehension and production. If learners make a deliberate effort to avoid such mistakes from the very beginning, they may learn to automatically process the gender information of a human antecedent in L2 processing

long before their general L2 proficiency reaches intermediate level. This hypothesis suggests there could be big individual differences among learners.

Acknowledgments We thank Prof. Ping LI from Pennsylvania State University, Dr. Guosheng DING from Beijing Normal University, Dr. Kalim Gonzales from Guangdong University of Foreign Studies, and the reviewers for their insightful suggestions and careful edits for previous versions of the manuscript. We also thank graduate students (esp. Mr. Zhibing YU) from the Bilingual Cognition and Education Lab for their help during the various stages of the study. This research was supported by the New Century Talents Program by the Chinese Ministry of Education.

References

Antón-Méndez, I. (2010a). Gender bender: Gender errors in L2 pronoun production. *Journal of Psycholinguistic Research*, *39*, 119–139.

Antón-Méndez, I. (2010). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, *14*(3), 318–331.

Boroditsky, L., Schmidt, L., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–80). Cambridge: MIT Press.

Cacciari, C., Carreiras, M., & Cionini, C. B. (1997). When words have two genders: Anaphor resolution for Italian functionally ambiguous words. *Journal of Memory and Language*, *37*(4), 517–532.

Callanhan, S. M. (2008). Processing anaphoric constructions: Insights from electrophysiological studies. *Journal of Neurolinguistics*, *21*(3), 231–266.

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotype gender information in constructing a mental model: Evidence from English and Spanish. *Quarterly Journal of Experimental Psychology*, *49A*, 639–663.

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*(3), 195–203.

Chen, X. (2004). *Personal referential strategies in Chinese EFL learners' oral narratives—A constrictive study based on multinatioanl learner corpus*. Unpublished MA Thesis, South China Normal University.

Clifton, C. J., Kennison, S. M., & Albrecht, J. (1997). Reading the words her, his, him: Implications for parsing principles based on frequency and on structure. *Journal of Memory and Language*, *36*(2), 276–292.

Dong, Y., & 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., & Li, Q. (2011). Zhongguo yingyu xuexizhe he/she hunyong cuowu tanyuan: Qianyanyu yuxin bianmazhong de daici xingbie xinxi queshi (The Mixed Use of He/She for Chinese Learners of English). *Zhongguo Waiyu (Foreign Languages in China)*, *41*(3), 22–29.

Duffy, S., & Keir, J. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory and Cognition*, *32*(4), 551–559.

Felix, S., & Hahn, A. (1985). Natural process in classroom second-language learning. *Applied Linguistics*, *6*, 223–238.

Foertsch, J., & Gernsbacher, M. (1997). In search of gender neutrality: Is singular they a cognitively efficient substitute for generic he. *Psychological Science*, *8*((2 Foertsch, Julie; Gernsbacher, Morton)), 106–111.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features: Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*, 226–248.

Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory and Cognition*, *30*(3), 439–446.

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. White Plains, NY: Addison-Wesley Longman.

Johnson, K. (1996). *Language teaching and skill learning*. Oxford: Blackwell.

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *32*(3), 355–378. doi:10.1023/a:1023599719948.

Khalil, A. (1985). Communicative error evaluations: Native speakers' evaluation and interpretation of written errors of Arab EFL learners. *TESOL Quarterly*, *19*, 225–351.

Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, *22*(5), 519–534. doi:10.1007/bf01068252.

Kowal, M., & Swain, M. (1997). From semantic to syntactic processing: How can we promote it in the immersion classroom? In K. Johnson & M. Swain (Eds.), *Immersion Education: International Perspectives* (pp. 284–309). Cambridge: Cambridge University Pres.

Lemm, K. M., Dabady, M., & Banaji, M. R. (2005). Gender picture priming: It works with denotative and connotative primes. *Social Cognition*, *23*(3), 218–241.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory and Cognition, 25*(3), 273–285.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language, 34*(6), 739–773.

Palmiero, M., Di Matteo, R., & Belardinelli, M. O. (2013). The representation of conceptual knowledge: Visual, auditory, and olfactory imagery compared with semantic processing. *Cognitive Processing*. doi:10.1007/s10339-013-0586-9.

Qiu, L., Swaab, T. Y., Chen, H.-C., & Wang, S. (2012). The role of gender information in pronoun resolution evidence from Chinese. *PLoS One*, *7*(5), e36156.

Rodriguez-Fornells, A., van der Lugt, A., Rotte, M., Britti, B., Heinze, H.-J., & Münte, T. F. (2005). Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence. *Journal of Cognitive Neuroscience*, *17*(3), 422–433.

Sera, Maria D., Elieff, C., Forbes, J., & Burch, M. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, *131*(3), 377–397.

Slevc, L. R., Lane, L. W., & Ferreir, V. S. (2007). Pronoun production: Word or world knowledge. *MIT Working Papers in Linguistics*, *53*, 191–203.

Sturt, P. (2003). The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, *48*, 542–562.

Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: Implications for language learning and language use. *Journal of Experimental Psychology: General*, *134*(4), 501–520.

Wen, Y. (2013). *English gender processing among Chinese EFL learners: An ERP study*. Unpublished MA thesis, Guangdong University of Foreign Studies.

Xu, Y. (2000). Yinghan zhicheng ciyu biaoda de kejixing (Refering expressions and their accessibility in Chinese and English). *Waiyu jiaoxue yu yanjv (Foreign Language Teaching and Research),* (5), 321–328.

Xu, X., Jiang, X., & Zhou, X. (2013). Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation. *Brain and Cognition*, *82*(2), 223–236.