Brain and Cognition 81 (2013) 223–236

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Processing biological gender and number information during Chinese pronoun resolution: ERP evidence for functional differentiation

Xiaodong Xu [a,b], Xiaoming Jiang [c], Xiaolin Zhou [c,d,*]

[a] Research Center for Learning Science and Key Laboratory of Child Development and Learning Science (Ministry of Education), Southeast University, Nanjing 210096, China
[b] School of Foreign Languages and Cultures, Nanjing Normal University, Nanjing 210097, China
[c] Center for Brain and Cognitive Sciences and Department of Psychology, Peking University, Beijing 100871, China
[d] Key Laboratory of Machine Perception and Key Laboratory of Computational Linguistics (Ministry of Education), Peking University, Beijing 100871, China

ARTICLE INFO

Article history:
Accepted 5 November 2012

Keywords:
Pronoun resolution
Gender agreement
Number agreement
Chinese
ERP
P600

ABSTRACT

There have been a number of behavioral and neural studies on the processing of syntactic gender and number agreement information, marked by different morpho-syntactic features during sentence comprehension. By using the event-related potential (ERP) technique, the present study investigated whether the processing of semantic gender information and the processing of notional number information can be differentiated and to what extent they might interact during Chinese pronoun resolution. The pronoun (with singular form in Experiment 1 and with plural form in Experiment 2) in a sentence matched its antecedent or mismatched it with respect to either biological gender or notional number or both. While the number mismatch elicited a P600 effect starting from 550 ms (for singular pronoun) or 400 ms (for plural pronoun) post-onset of the pronoun, the gender mismatch elicited an earlier (for singular) and larger (for both singular and plural) P600 effect. More importantly, the double mismatch produced a P600 effect identical to the effect elicited by the single gender mismatch. These results demonstrate that biological gender information and notional number information are processed differentially and have different processing priorities during Chinese pronoun resolution.

1. Introduction

1.1. Agreement processing

Agreement is a form of cross-reference between different parts of a sentence or phrase; it occurs when a word changes its form as a function of other words to which it relates. The agreement based on overt grammatical changes (e.g., subject–verb agreement) contrasts with the notional agreement that is based on meaning (e.g., stereotypical gender agreement, collective noun–verb agreement; Deutsch & Dank, 2009; Molinaro, Vespignani, Zamparelli, & Job, 2011; see Acuña-Fariña (2009) for a review). To comprehend a sentence or utterance properly, readers or listeners must compute dependencies between words at multiple levels of language processing, including morphosyntactic and semantic levels. The agreement relationship is assumed to be processed at any of these levels in terms of number, gender, person, or case features (Corbett, 2006). Previous psycholinguistic and neurocognitive studies of language concentrated mostly on agreement processing at the morphosyntactic level (Barber & Carreiras, 2005; Bock, Eberhard, & Cutting, 2004; Deutsch & Bentin, 2001; Faussart, Jacubowicz, & Costes, 1999; Friederici & Jacobsen, 1999; Gunter, Friederici, & Schriefers, 2000; Hagoort & Brown, 1999; Martín-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006; Molinaro, Barber, & Carreiras, 2011; Molinaro, Kim, Vespignani, & Job, 2008; Nevins, Dillon, Malhotra, & Phillips, 2007; Schmitt, Lamers, & Münte, 2002; Silva-Pereyra & Carreiras, 2007; Vigliocco, Butterworth, & Garrett, 1996). It remains unclear how agreement based on semantic properties is processed during sentence comprehension and, in particular, how different types of agreement would function together when they appear simultaneously to establish referential constraints. The main purpose of this study is to provide electrophysiological evidence for the differentiation and the interplay of semantically-based agreement features in pronoun resolution during Chinese sentence comprehension. Before we describe the design of the present study, we first summarize theoretical arguments and empirical evidence concerning agreement processing at the local morphosyntactic level and the long-distance semantic levels, respectively.

1.2. Agreement processing in local phrases

Evidence from a variety of languages seems to suggest that morphosyntactic features, such as number and grammatical gen-

* Corresponding author at: Department of Psychology, Peking University, Beijing 100871 China. Fax: + 86 10 6276 1081.
E-mail address: xz104@pku.edu.cn (X. Zhou).

0278-2626/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bandc.2012.11.002

der are processed differently. The functional difference between number and gender has been proposed by Greenberg (1963), according to which one important universal is the dependence of gender on number (Universal 36: "If a language has the category of gender, it always has the category of number", pp. 92–96). Number is a functional head and occupies the highest position in the syntactic tree (De Vincenzi & Di Domenico, 1999; Faussart et al., 1999). Further development of this hypothesis assumes that the morphological features encode conceptual features with different degrees of 'cognitive salience' (Person > Number > Gender), with cognitively more important or salient feature being more easily accessible by the parser during sentence comprehension (Carminati, 2005; Harley & Ritter, 2002; Noyer, 1992). Cognitive salience generally refers to the state or quality by which a particular entity stands out relative to its neighbors and thus captures attention or processing resources. Salience may be the result of emotional, motivational as well as cognitive factors. At the morphosyntactic level, for instance, number is less likely to be arbitrarily coded than grammatical gender as it signals the cardinality of the noun and its computation requires the instantiation of one (singular) vs. more than one (plural) entities in the reference context (Acuña-Fariña, 2009; Adani, van der Lely, Forgiarini, & Guasti, 2012).

The suggestion of hierarchical agreement processing based on structural rules leads to another proposal that agreement features with lower cognitive salience are processed less efficiently than features with higher cognitive salience (Antón-Méndez, Nicol, & Garrett, 2002; Carminati, 2005; Igoa, García-Albea, & Sánchez-Casas, 1999). This prediction seems to be supported by behavioral evidence from comprehension tasks. For example, Sagarra and Herschensohn (2010) found that beginners and intermediate learners of Spanish, in a grammaticality judgment task, performed less accurately to a sentence with an adjective disagreeing with its preceding noun in gender (e.g. *prototipio famoso* vs. **prototipo famosa*) than to a sentence with an adjective disagreeing in number (e.g. *prototipo famoso* vs. **prototipo famosos*).

At the neural level, however, the processing prediction of the Feature Hierarchy Theory received only partial support. If the processing of feature agreements with different cognitive salience is differentiated with respect to feature hierarchy, violation of agreements with features of higher cognitive salience should induce stronger neural responses than violation of those with features of lower salience (see Molinaro et al. (2011) for a review). Indeed two recent ERP studies showed that person violations elicit increased P600 responses compared to number violations (Mancini, Molinaro, Rizzi, & Carreiras, 2011; Nevins et al., 2007), indicating that person agreement is cognitively more salient than number agreement. A functional magnetic resonance imaging (fMRI) study comparing brain responses to the violations of local morpho-syntactic agreement in gender or number in determiner-noun and noun–adjective pairs also revealed more activations for number mismatch than for gender mismatch in the right intraparietal sulcus, although common activations were observed in the left premotor and left inferior frontal gyrus for the two types of mismatch (Carreiras, Carr, Barber, & Hernandez, 2010). However, as pointed out by Molinaro et al. (2011), the difference in P600 responses to mismatches in feature agreement could reflect the number of backward steps in reanalysis (Faussart et al., 1999), rather than the difference in cognitive salience of the features: the more the system has to regress to a previous stage of processing to carry out reanalysis, the larger the P600. Consistent with this view, Barber and Carreiras (2005) found a larger P600 amplitude for syntactic gender than for number mismatches, probably because the initial processing of gender information was performed at an earlier stage (i.e., lexical access) and the reanalysis for its mismatch needed more steps as compared to the reanalysis of number mismatch, which needs only access to inflectional features.

1.3. Agreement processing in pronoun–antecedent coindexation

The agreement processing is not restricted to local syntactic dependencies, however. In case of gender and number, the agreement also takes place in long-distance (cross-clause) dependencies or long-distance anaphoric relations (Gibson, 1998; Hammer, Jansma, Lammers, & Münte, 2008; Phillips, Kazanina, & Abada, 2005). Psycholinguistic models concerning anaphoric resolution (e.g. pronoun–antecedent agreement) have suggested that pronominal interpretation involves two processing stages (Callahan, 2008; Garrod & Sanford, 1990, 1994; Garrod & Terras, 2000). In the initial bonding stage, candidate antecedents are activated/retrieved and the activation is constrained by featural agreement constraints (e.g., morphosyntactic constraints like gender, number) and structural constraints. In the second resolution stage, the appropriate antecedent is selected and integrated with the anaphoric expression (pronoun); this integration is constrained by world knowledge and discourse factors. According to one proposal (Callahan, 2008), the failure in the initial stage of antecedent retrieval may be associated with an early left anterior negativity (LAN) or a sustained bilateral anterior negativity, reflecting the difficulty of establishing a link between anaphor and antecedent or retrieving the antecedent from working memory. The failure in the second stage of integration/resolution is usually manifested by a centro-parietal N400 or a P600, reflecting difficulty of semantic or syntactic integration difficulty, respectively.

Hammer et al. (2008) examined the biological and syntactic gender agreement between pronoun and its antecedent. The distance between pronoun-antecedent dependency was either short (adjacent dependency), or long with no gap (dependency separated by more words) or long with gap (dependency separated by a clause with a gap position where the antecedent could be activated due to gap filling). Gender agreement was either biological + syntactic (on *person* pronoun), or syntactic alone (on *thing* pronoun). The authors observed that the syntactic gender mismatch elicited a late-starting negativity on pronoun, regardless of the distance, whereas the double-gender mismatch elicited a N400-like effect (in the 200–400 ms window) for the short-distance condition and the long-with-no-gap condition but a P600 effect for the long with gap condition. Qiu, Swaab, Chen, and Wang (2012) investigated purely semantically-based biological gender agreement between a pronoun and its antecedent in Chinese, a language which does not encode syntactic gender information on pronoun. They found that the distance between a pronoun and its antecedent can influence the ERP response evoked by the pronouns. In comparison with the gender-matched pronoun, gender-mismatched pronoun evoked an N400 effect when the antecedent was adjacent but a P600 effect when a clause intervened. Other studies focused on the mismatch of pronoun-antecedent in biological gender agreement and consistently observed a centroparietally distributed late positivity (P600) for such violations (Hammer, Jansma, Lamers, & Münte, 2005; Lamers, Jansma, Hammer, & Münte, 2008; Osterhout, Bersick, & McLaughlin, 1997; Osterhout & Mobley, 1995), although a few studies also reported an additional N400 effect (Lamers, Jansma, Hammer, & Münte, 2006; Schmitt et al., 2002). It appears that processing both biological gender and syntactic gender disagreement is associated with the difficulty in the stage of antecedent-pronoun integration. The distance factor may play a role in determining whether the syntactic or semantic processes dominate when both the biological and syntactic gender agreement between pronoun and its antecedent is violated.

The processing of notional number agreement was based upon the interpretation of the number of the antecedent noun (with the help of the quantifier or the collective meaning of the noun) rather than the strict grammatical form of the noun. Some studies showed that when more than one potential antecedent was available in the discourse context, a frontally-distributed negative shift (*Nref*) was elicited for the referentially ambiguous nouns or pronouns (Nieuwland & Van Berkum, 2006, 2008; Van Berkum, Brown & Hagoort, 1999; Van Berkum, Brown, Hagoort, & Zwitserlood, 2003). Other studies demonstrated that when there were no suitable referents for an anaphor, a late positivity (P600) was induced on the anaphor (Kaan, Harris, Gibson, & Holcomb, 2000; Nieuwland & Van Berkum, 2006; Van Berkum, Koornneef, Otten, & Nieuwland, 2007). Additionally, a P600 effect was obtained for reflexive pronouns which disagree with their antecedents in number (Li & Zhou, 2010; Osterhout & Mobley, 1995) or for quantifiers which have to be interpreted as new referents (e.g., *Four ships appeared on the horizon, six had sunk*; Kaan, Dallas, & Barkley, 2007). The *Nref* effect observed in referentially ambiguous context might reflect the working memory cost associated with selecting or retrieving the right antecedent among multiple competitors while the P600 effect obtained in other studies can be interpreted as reflecting the difficulty in integrating the antecedent and the pronoun (noun, quantifier) in the long-distance dependency.

Only a few studies have manipulated biological gender and number agreement simultaneously. Osterhout and Mobley (1995) investigated both biological gender and number agreement dependencies between the reflexive pronoun (and pronoun) and its antecedent and observed a P600 for both gender and number violations, suggesting that mismatch in either gender or number agreement would cause difficulty in integration of antecedent and pronoun within long-distance dependency. However, neither did this study fully cross the two types of agreements nor did it compare the magnitude of gender and number mismatch effects (due to a between-item design). It is thus unclear which type of information (gender or number agreement) has higher cognitive salience and hence may play a more dominant role in the integration process in face of mismatch of both types.

1.4. The present study

In this study, we focused on the online processing of biological gender agreement and conceptual number agreement. By simultaneously manipulated these two agreement features, we were able to test whether there is a "salience" hierarchy for the semantically-based agreement features and whether this hierarchy is comparable to that for the morphosyntactic agreement features. We focused on Mandarin Chinese due to its following characteristics. Mandarin Chinese has essentially no morphological marking for either grammatical or biological gender information and relies heavily on semantic information to establish referential links. In written Chinese, biological gender information can be conveyed through orthographic forms, that is, the semantic radicals embedded in characters corresponding to the pronouns. For singular third-person pronoun, for example, the character 他 represents the masculine pronoun "he", and the character 她 the feminine pronoun "she", although both of them are pronounced as "/ta/". This differentiation in biological gender may affect pronoun resolution, at least during the comprehension of written sentences. In the pronoun-antecedent link, the referent of a word (e.g., an animate noun) generally has biological sex in the real world and gender information associated with a noun can be automatically activated (Osterhout et al., 1997).

Number agreement is typically taken as a morphosyntactic phenomenon in European languages. However, in Chinese, number agreement may take place at the semantic level. Mandarin Chinese adds a morpheme 们 (/men/) to a singular pronoun, like 我 (/wo/, "I"), 你 (/ni/, "you"), 他 (/ta/, "he") or 她 (/ta/, "she"), to form the collective form 我们 (/wo men/, "we"), 你们 (/ni men/, "you"), 他们 (/ta men/, "they$_{masculine}$", or 她们 (/ta men/, they$_{feminine}$). All of these pronouns are function words although they have concrete references in context. Linguists proposed that 们 (/men/) can only function as a collective marker (Iljic, 1994; Li, 1999), although others treat 们 (/men/) as an inflectional morpheme like –*s* in English (Li & Shi, 2000). Nevertheless, for the number agreement in the pronoun-antecedent link, the agreement is established by taking into account specific notional (or semantic) properties of their intended antecedents (Bock et al., 2004; Kreiner, Garrod, & Sturt, 2012). A plural pronoun and its antecedent represent the speaker's valuation of numerosity of the referent, a product of categorizing the intended referent as "more than one thing" vs. "one thing".

Given the characteristics of Chinese pronouns outlined above, it is possible for us to manipulate biological gender and notional number information simultaneously and to concentrate on the semantically-based agreement processing. Moreover, the electrophysiological investigation of agreement processing in Chinese pronoun resolution could advance the ongoing theoretical exploration of anaphoric processing. Although the two-stage model has been supported by evidence from studies on morphologically rich languages, with each stage being associated with a distinct ERP effect, it is unclear whether similar results would be observed in a language lacking morphosyntactic markers.

Table 1 presents exemplar sentences from Experiment 1, in which number agreement and gender agreement were crossed factorially, forming 4 experimental conditions. While the first clause presents the antecedent noun in either the singular or the plural form (e.g., through the determiner, i.e. numeral and classifier, preceding the noun, see also Zhou et al., 2010), the singular pronoun in the second clause, at the object position, agrees or disagrees with the antecedent in number and/or gender. Experiment 2 used a similar design but with a plural pronoun in the second clause.

The crucial ERP measurement was time-locked to the presentation of the pronoun in the second clause. The ERP effects for semantically-based agreement are predicted mainly based on the previous studies in morphologically-rich languages. Given the majority of findings on number agreement, we expected to observe a P600 effect for number mismatch as compared with the control condition with correct sentences. However, for the singular pronouns (Experiment 1), it is also plausible that an *Nref* effect

Table 1
Experimental conditions and exemplar sentences with approximate literal translations in Experiment 1.

Condition	Examples
Control	这位女患者情绪低落,医生/鼓励/她/振作/起来。 Zhewei nühuanzhe qingxudiluo, yisheng guli **ta$_{female}$** zhenzuo qilai. *This woman patient was in low spirits, doctors encouraged* ***her*** *to cheer up*
Number mismatch	这些女患者情绪低落,医生/鼓励/她/振作/起来。 Zhexie nühuanzhe qingxudiluo, yisheng guli **ta$_{female}$** zhenzuo qilai. *These women patients were in low spirits, doctors encouraged* ***her*** *to cheer up*
Gender mismatch	这位女患者情绪低落,医生/鼓励/他/振作/起来。 Zhewei nühuanzhe qingxudiluo, yisheng guli **ta$_{male}$** zhenzuo qilai. *This woman patient was in low spirits, doctors encouraged* ***him*** *to cheer up*
Double mismatch	这些女患者情绪低落,医生/鼓励/他/振作/起来。 Zhexie nühuanzhe qingxudiluo, yisheng guli **ta$_{male}$** zhenzuo qilai. *These women patients were in low spirits, doctors encouraged* ***him*** *to cheer up*

would be present when the number agreement was violated. According to Nieuwland and Van Berkum (2006), *Nref* is an anteriorly-distributed sustained negativity starting from about 300 ms post-onset of anaphoric phrases with ambiguous reference (e.g. *The chemist hit the historian while he*...; Nieuwland & Van Berkum, 2006, 2008; Van Berkum et al., 1999; Van Berkum et al., 2003). Here a singular pronoun could be taken as being ambiguous in referring to which of the multiple candidates in the plural antecedent. Moreover, as we reviewed earlier, the majority of studies on grammatical and biological gender mismatches generally found the P600 effect (Osterhout & Mobley, 1995; Barber, Salillas, & Carreiras, 2004; Molinaro et al., 2008; Osterhout et al., 1997) while small groups of studies on biological gender mismatch also reported a N400 effect (Lamers et al., 2006; Schmitt et al., 2002), the appearance of which may be modulated by the distance between the antecedent and the pronoun (Hammer et al., 2008; Qiu et al., 2012). Thus we might predict a P600 effect for the biological gender mismatch condition, given that the antecedent and pronoun in this study were rather distant (across clauses).

Importantly, this study aimed to compare the onset and magnitude of the two P600 effects for the notional number and biological gender mismatches in the long-distance dependency. If number agreement is cognitively more salient than biological gender agreement, we may expect the P600 effect elicited by number mismatch to be stronger than that for gender mismatch, due to the increased sensitivity to the failure of establishing an agreement relationship and to the increased integration difficulty. If, however, biological gender agreement is cognitively more salient than number agreement, the opposite pattern should be obtained. Moreover, according to the previous studies, for the double mismatch condition, if the processing of number agreement and of gender agreement interact, we would predict the combined P600 effect to be more similar to the P600 effect in one of the single mismatch conditions than to the effect in the other, assuming that the processing of one type of agreement dominates over the processing of the other in face of broken agreements. If, on the contrary, the processing of number agreement and the processing of gender agreement act in an additive manner, we would predict the combined P600 effect to be the sum of the effects for the two single mismatches.

2. Experiment 1

2.1. Method

2.1.1. Participants

Twenty-four native Chinese speakers (12 females, age ranging from 22 to 26 years with mean age of 24 years) were recruited from Southeast University, Nanjing, China and were paid for their participation. All of them were right-handed, neurologically healthy and had normal or corrected-to-normal vision. This study was approved by the Ethic Committee of the Department of Psychology, Peking University.

2.1.2. Design and materials

As indicated in Table 1, the experiment had a 2 $*$ 2 factorial design, with number and gender agreement between the antecedent in the main clause and the pronoun at the object position in the second clause being manipulated. There were 184 quartets of critical stimulus sentences. For each sentence, the first clause (the context clause) described the state that the protagonist(s) was in or the characteristics that the protagonist(s) had. The protagonist(s) was either male or female, with a preceding determiner of either singular or plural form (e.g., "这位, /zhewei/, this", "这些, /zhexie/, these") representing the singularity or plurality of the object noun. Information concerning the biological gender of the protagonist(s) was conveyed through the modifier 男(/nan/, *man*) or 女(/nü/, *woman*) in front of the otherwise neutral noun in one half of the stimuli and through a gender-defining noun (e.g. *mother, uncle*) in the other half of stimuli. Thus both the number and gender information concerning the potential antecedent of the subsequent pronoun was conveyed explicitly through lexical means. This protagonist(s) served as the antecedent of the subsequent pronoun. The second clause (the target clause) described another person's behavior or attitude towards the pronoun "他" (*he*) or "她" (*she*) which referred back to the antecedent. The pronoun acted as the object of the verb or the object of a preposition. There was always an intervening noun between the antecedent and the pronoun, acting as the subject in the second clause. However, this subject noun could not be an antecedent to the pronoun due to grammatical constraints.

Prior to the selection of the final set of sentences, the potential stimuli underwent two pretests: a sentence acceptability rating and an error correction test. The sentence acceptability rating was used to examine to what extent sentences in each condition were acceptable in natural language. The error correction test was used to examine how the reader would correct the sentences if he deemed them to be incorrect. In particular, results from this pretest could tell us whether a strategy of looking for a potential sentence-external referential entity would be used to re-interpret the inappropriate pronoun.

In the acceptability rating, the critical sentences, together with filler sentences, were divided into four versions with a Latin-square procedure. Twenty-four students who did not participate in the ERP experiment were each randomly assigned to one of the four versions and were asked to judge the overall acceptability of each sentence (which concerned both semantic coherence and syntactic appropriateness) on a 7-point Likert Scale (1 indicating the least acceptable and 7 indicating the most acceptable). As can be seen from Table 3, sentences with gender mismatch, number mismatch or double mismatch had much lower acceptability than the control sentences. Interestingly, sentences with number mismatch were rated more acceptable than sentences with gender mismatch or double mismatch ($ps < 0.01$).

In the error correction test, the critical sentences were also divided into four lists with a Latin-square procedure. Another 28 students who did not participate in the ERP experiment were each randomly assigned to one of the four versions and were asked first to judge whether each sentence had an error. They were further asked to correct the error with a more appropriate expression if the sentence was deemed to be incorrect and to skip the sentence if the sentence was correct. As can be seen from Table 2, for the number mismatch condition, 91.1% changes were made by replacing the pronoun with a number congruent pronoun; for the gender mismatch condition, 94.8% changes were made by replacing the pronoun with a gender congruent pronoun; for the double mismatch condition, 86.9% changes were made by replacing the pronoun with a pronoun that was congruent in both number and gender, although the total correction rate was rather high across mismatching conditions (over 95%). Other corrections were made on the determiners of the antecedents or on other parts of the sentences. The large percentages of pronoun correction for sentences

Table 2
Mean scores and standard deviations in the two pretests of Experiment 1.

	Sentence acceptability		Error correction rate (%)	
	Mean	SD	Mean	SD
Control	5.86	0.1	–	–
Number mismatch	1.9	0.15	91.1	1.7
Gender mismatch	1.57	0.15	94.8	0.8
Double mismatch	1.28	0.09	86.9	2.2

with gender and/or number mismatch suggest that the reader tended to rationalize the sentences by revising the gender or number mismatch pronoun into one that matches the preceding antecedent, rather than finding an external entity as a possible antecedent of the pronoun. The former tendency even occured in an offline task in which enough time was given to allow the latter strategy.

In addition to the critical sentences, 172 filler sentences with structures similar to the critical ones were constructed, including 132 correct sentences and 40 incorrect sentences with various problems (e.g., plural pronoun referring to singular antecedent, lexical semantic anomalies, grammatical violation, etc.). Among the correct sentences, 40 included a plural protagonist in the first clause with half male and half female, followed by a co-referential plural pronoun ("他们", "them", masculine or "她们", "them", feminine) in the second clause. In such way, the participants' strategy of predicting a pronoun mismatch after reading a plural protagonist was minimized. The remaining correct fillers included a protagonist or an entity in the first clause, followed by a repetition or a new protagonist or entity in the second clause. The inclusion of this type of sentence reduced the possibility that a sentence always continued with a pronoun. For incorrect sentences, violations appeared always in the second clause, and if the subordinate clause contained a pronoun, the violation appeared always before or after the pronoun.

Each critical sentence in a quartet was assigned to a different test list with a Latin square procedure, such that in each list there were 46 sentences per experimental condition. The filler sentences were then added to each list and sentences in each list were pseudo-randomized, with the restriction that no more than three consecutive sentences were of the same condition and no more than three consecutive sentences were correct or incorrect. Participants were each randomly assigned to one of the four lists.

2.1.3. Procedures

The participants were seated comfortably in a dimly lit sound-attenuating and electrically shielded booth. They were instructed to read each sentence attentively. All the stimuli were displayed in white against a black background. Each trial began with a fixation point ("+") at the center of the screen for 500 ms, followed by an interval of blank screen for 500 ms. Then the first clause containing the antecedent was presented in whole on the screen. After finishing reading the first clause, the participant pressed the space bar to initiate the second clause, which was presented segment-by-segment at the center of the screen. Each segment was presented for 400 ms followed by a blank screen for another 400 ms. This presentation rate is natural and comfortable for Chinese readers (Jiang & Zhou, 2009; Ye, Luo, Friederici, & Zhou, 2006). The pronoun was at the third, fourth, or fifth position in the second clause, but was never at the clause-final position. After the display of the whole sentence, a line of question marks was presented and the participant was prompted to press one of the two keys if the sentence was acceptable and to press the other if the sentence was unacceptable. The assignment of hand to response type was counterbalanced across participants.

The participant performed a practice block of 15 sentences, which had similar structures as the test stimuli. The test stimuli were divided into five blocks and the participant had a break of about 3 min between each block. The test of each participant lasted about 2 h, including electrode preparation.

2.1.4. EEG recording and data analysis

EEG activity was recorded from 62 electrodes in a secured elastic cap (Electro-cap International). Vertical and horizontal electro-oculograms were recorded. The EEGs were referenced online to the left mastoid and were re-referenced offline to the linked mastoids. Electrode impedances were kept below 5 kΩ. EEG signals were filtered using a bandpass of 0.05–100 Hz, and digitized at a sampling rate of 500 Hz. The ERP epoch was extracted for the pronoun in the second clause for each critical sentence, with a pre-stimulus baseline of 100 ms and the ERP response to the pronoun for 800 ms. Trials with EEG maximal amplitude exceeding ±75 μV or with incorrect responses were eliminated from data analysis. The mean number of trials included for EEG analysis was 38.8 for the correct condition, 39.8 for the gender mismatch condition, 37.3 for the number mismatch condition, and 40.8 for the double mismatch condition, which did not differ between conditions, $F < 1$. Based upon visual inspection and research hypotheses, the analyses of variance (ANOVAs) were conducted on mean ERP amplitudes in selected time windows (250–400 ms for early negativity, 400–550 ms and 550–800 ms for P600; see Results), with number (correct vs. mismatch), gender (correct vs. mismatch) and topographical factors as within-participant variables. For the midline analysis, the topographic factor was electrode (six levels from the most anterior to the most posterior: Fz, FCz, Cz, CPz, Pz and POz). For the lateral analysis, the topographic factors were region (three levels: anterior vs. central vs. posterior) and hemisphere (two levels: left vs. right). The region and hemisphere were crossed, resulting in six regions of interest: left anterior (F1, F3, F5, FC1, FC3 and FC5), left central (C1, C3, C5, CP1, CP3 and CP5), left posterior (P1, P3, P5, PO3, PO5 and PO7), right anterior (F2, F4, F6, FC2, FC4 and FC6), right central (C2, C4, C6, CP2, CP4 and CP6) and right posterior (P2, P4, P6, PO4, PO6 and PO8). Mean amplitudes over electrodes in each region of interest entered into ANOVAs. Pairwise comparisons were planned for each mismatch condition and the control condition, The Greenhouse–Geisser correction was performed when appropriate. Bonferroni correction was used for multiple comparisons.

2.2. Results

2.2.1. Behavioral results

On average, the accuracy rate for the acceptability judgment was 92.0% for the control condition (SD = 1.4%), 85.5% for the number mismatch condition (SD = 2.9%), 94.4% for the gender mismatch condition (SD = 1.7%), and 97.0% for the double mismatch condition (SD = 1.0%). An ANOVA taking number and gender as two within-participant factors revealed a significant main effect of gender, $F(1,23) = 21.32$, $p < 0.001$, but no main effect of number, $F(1,23) = 1.28$, $p > 0.1$. The interaction between number and gender was significant, $F(1,23) = 5.55$, $p < 0.05$. Pairwise t-tests between the number mismatch and gender mismatch conditions showed a significantly higher accuracy for detecting gender violations than for detecting number violations, indicating that the reader was generally more sensitive to gender mismatch than to number mismatch.

2.2.2. Electrophysiological results

The grand average ERPs, time-locked to the onset of the critical pronoun, are shown in Fig. 1. The scalp topographies in Fig. 2 depict the differences, in two time windows, between the three mismatch conditions and the control condition. As can be clearly seen in Figs. 1 and 2, the pronoun elicited an anterior-central negativity peaking around 350 ms, which was claimed to be a pronoun N400 in Hammer et al. (2008) or a lexical processing negativity in King and Kutas (1998) and Steinhauer, Pancheva, Newman, Gennari, and Ullman (2001); a late, posterior positive deflection finally appeared. While all the mismatch conditions elicited more positive ERP responses than the control condition, the onset of the divergence between conditions differed substantially. While the gender mismatch and the double mismatch condition started to show more positive responses than the control condition at about 400 ms post-onset, the number mismatch condition started to show this trend only at about 550 ms post-onset. Moreover, the positive effects for the gender mismatch and double mismatch

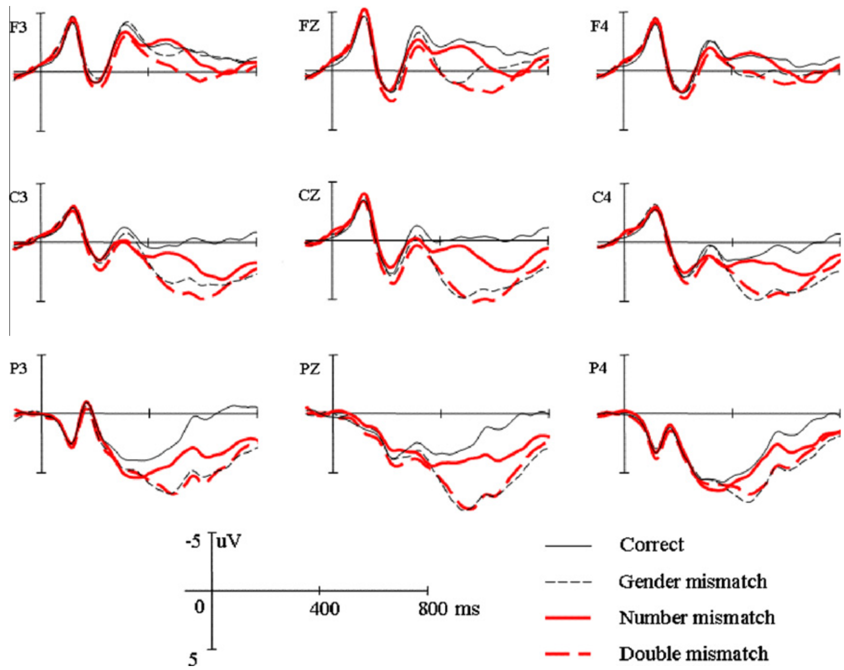

Fig. 1. Grand average ERPs time-locked to the critical singular pronoun for the control, the number mismatch, the gender mismatch and the double mismatch in Experiment 1.

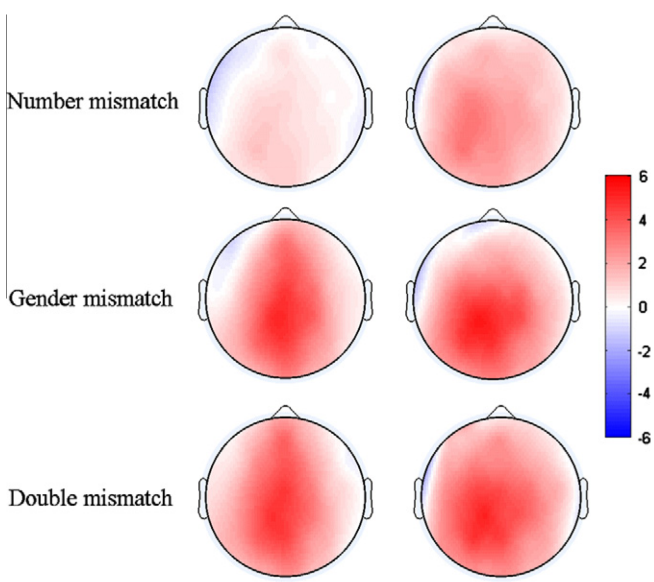

Fig. 2. Topographic maps for difference waves on the pronoun between each mismatch condition and the control condition in 400–550 ms window (the left column) and 550–800 ms window (the right column) in Experiment 1.

conditions, as assessed against the control condition, were apparently larger than the effect for the number mismatch condition.

2.2.2.1. ERP responses in the 250–400 ms time window. Repeated-measures ANOVA over the mean amplitudes in this window yielded a significant main effect of gender in the midline analysis (0.73 μV), $F(1,23) = 4.63$, $p < 0.05$, and a significant main effect of number in both the midline (0.83 μV) and the lateral (0.68 μV) analysis, $F(1,23) = 5.90$, $p < 0.05$, and $F(1,23) = 6.76$, $p < 0.05$, respectively. The interaction between gender and number was not significant, Fs < 1. Further comparisons showed that, compared with the control condition, the gender mismatch, number mismatch, and double mismatch all elicited less negative-going ERP responses: for the gender mismatch, $F(1,23) = 3.15$, $0.05 < p < 0.1$ in the midline and $F < 1$ in the lateral; for the number mismatch, $F(1,23) = 4.04$, $0.05 < p < 0.1$ in the midline and $F(1,23) = 3.34$, $0.05 < p < 0.1$ in the lateral; for the double mismatch, $F(1,23) = 6.99$, $p < 0.05$ in the midline and $F(1,23) = 3.37$, $0.05 < p < 0.1$ in the lateral.

2.2.2.2. ERP responses in the 400–550 ms time window. There was a significant main effect of gender in the midline analysis, $F(1,23) = 70.02$, $p < 0.001$, and in the lateral analysis, $F(1,23) = 48.13$, $p < 0.001$, with more positive ERP responses when gender was mismatched than when it was matched (3.61 μV in the midline and 2.23 μV in the lateral). There was also a significant interaction between gender and electrode in the midline, $F(5,115) = 4.49$, $p < 0.05$, and a marginal interaction between gender and region in the lateral, $F(2,46) = 3.89$, $0.05 < p < 0.1$. It is clear from Figs. 1 and 2 that the positivity effect for gender was larger in the posterior than in the anterior regions. Importantly, there was no main effect of number in the midline, $F(1,23) = 1.90$, $p > 0.1$, nor in the lateral, $F < 1$, indicating that the number mismatch had not elicited differential ERP responses at this time window. The interaction between number and gender was also not significant, $F(1,23) = 1.45$, $p > 0.1$ in the midline, and $F < 1$ in the lateral.

Pair-wise comparisons revealed a significant effect of sentence type between gender mismatch and control, $F(1,23) = 42.52$, $p < 0.001$ in the midline and $F(1,23) = 32.33$, $p < 0.001$ in the lateral, and between double mismatch and control, $F(1,23) = 50.99$, $p < 0.001$ in the midline and $F(1,23) = 46.08$, $p < 0.001$ in the lateral. The difference between number mismatch and control was, however, not significant, $F(1,23) = 2.49$, $p > 0.1$ in the midline and $F < 1$ in the lateral. Direct comparison between gender mismatch and number mismatch revealed significant differences in the midline, $F(1,23) = 49.64$, $p < 0.001$, and in the lateral analysis, $F(1,23) = 43.05$, $p < 0.001$, indicating that the gender mismatch elicited a stronger positivity (3.24 μV in the midline and 2.04 μV in the lateral) than the number mismatch.

2.2.2.3. ERP responses in the 550–800 ms time window. There was a significant main effect of gender, $F(1,23) = 44.44$, $p < 0.001$ in the midline and $F(1,23) = 41.07$, $p < 0.001$ in the lateral, and of number, $F(1,23) = 9.92$, $p < 0.005$ in the midline and $F(1,23) = 18.84$, $p < 0.001$ in the lateral, suggesting that the gender and the number mismatch conditions elicited increased P600 responses as compared with the match conditions. There was also a significant interaction between gender and number, $F(1,23) = 9.18$, $p < 0.01$ in the midline and $F(1,23) = 8.88$, $p < 0.01$ in the lateral, and three-way interactions between number, gender and electrode, $F(5,115) = 5.55$, $p < 0.01$ in the midline, and between number, gender and region, $F(2,46) = 4.39$, $p < 0.05$ in the lateral, indicating that the interaction between number and gender took place mainly at centroposterior regions.

Separate analysis showed that there was a significant difference between the gender mismatch and the control, $F(1,23) = 40.78$, $p < 0.001$ in the midline and $F(1,23) = 17.37$, $p < 0.005$ in the lateral analysis, between the number mismatch and the control, $F(1,23) = 42.69$, $p < 0.001$ in the midline and $F(1,23) = 12.58$, $p < 0.005$ in the lateral analysis, and between the double mismatch and the control, $F(1,23) = 38.02$, $p < 0.001$ in the midline and $F(1,23) = 47.36$, $p < 0.001$ in the lateral analysis. Importantly, as illustrated in Figs. 1 and 2, the number mismatch contributed little to the positivity effect in the double mismatch condition: there was no significant difference between the gender mismatch and the double mismatch in either the midline or the lateral analysis, Fs < 1. The dominance of gender processing over number processing can also be observed in the direct comparison between the two single mismatch conditions, with gender mismatch eliciting more positive responses than number mismatch in the midline analysis, $F(1,23) = 10.18$, $p < 0.005$, and in the lateral analysis, $F(1,23) = 7.13$, $p < 0.05$.

2.3. Discussion

In summary, in the 250–400 ms time window, all the three mismatch conditions elicited *less* negative-going responses as compared with the baseline, with the effect elicited by the double mismatch being the strongest. In the P600 time window, both gender and number mismatches elicited a late, centroposteriorly distributed positivity as compared with the control condition. However, the P600 effect emerged earlier (i.e., appearing in the 400–550 ms window) and was larger for gender mismatch than for number mismatch. Moreover, the effect for the double mismatch was dominated by the effect of the gender mismatch, as there were no differences between the effects elicited by double mismatch and by gender mismatch—in other words, the number mismatch made no significant contribution to the effect in the double mismatch condition.

The finding of less negative-going responses to the mismatching pronoun is consistent with an earlier study on the Chinese reflexive pronoun. Li and Zhou (2010) found that, compared with the reflexive pronoun *zij* (oneself) referring to a local antecedent, the pronoun referring to a long-distance antecedent would elicit a larger early positivity (300–400 ms) followed by a P600 effect. The authors interpreted the early positivity as reflecting a detection of incongruence between a local assignment (based on structural constraint) and a long-distant assignment (based on the property of the verb). In the current experiment, the context in the first clause strongly predicted a specific pronoun in the second clause. The mismatch between the actual input and the expected pronoun, in terms of gender and/or number, could lead to incongruence or conflict, the detection of which was reflected in the early positivity, or more accurately, the reduced negativity in the 250–400 ms window. We will return to this effect in Section 3.3.

The finding of a P600 effect for semantic/biological gender mismatch is consistent with most studies on pronoun resolution in morphologically marked languages (e.g., Hammer et al., 2005; Lamers et al., 2008) but inconsistent with some other studies which observed an N400 effect (Deutsch & Bentin, 2001; Hammer et al., 2008; Lamers et al., 2006; Schmitt et al., 2002). This effect may indicate the difficulty in semantically integrating the pronoun with its antecedent during Chinese sentence comprehension, although a slightly different pattern of effects was observed for semantic gender mismatch in another study (Qiu et al., 2012).

Importantly, the finding of the P600 effect for gender mismatch in the 400–550 ms and 550–800 ms windows and the finding of the P600 for number mismatch only in the 550–800 ms window suggest that biological gender information is available earlier than number information and that the system is more sensitive to gender mismatch than to number mismatch during pronoun resolution. Moreover, gender and number information is processed in a way such that constraints of gender agreement completely dominate over the constraints of number agreement when the two types of agreement are violated simultaneously. We will discuss this dominance in detail in Section 4.

There could be two different views concerning the functions of the P600 effects in the present study, both of which assume that detecting the mismatch between the input pronoun and its antecedent and correcting or rationalizing the pronoun cannot be a purely syntactic process. The first view is to attribute the P600 effect observed here to a purely semantic process (Jiang, Tan, & Zhou, 2009; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Zhou et al., 2010). Jiang et al. (2009) manipulated the congruency in number between the Chinese universal quantifier 都 (/*dou*/, *all*) and the antecedent entity. A sustained positivity effect was observed on the quantifier for the incongruent condition. This effect was taken as reflecting a second-pass integration process that links the universal quantifier with the preceding entity. Likewise, to recover from mismatch between pronoun and its antecedent in biological gender and/or number and to rationalize the current pronoun (i.e., to change the current singular form into the plural form, as suggested by the error correction test), the semantic relationship between the pronoun and its antecedent may have to be re-computed.

Another related view is to attribute the P600 effect to a more general conflict control process after the detection of mismatch between the expected pronoun and the actual input (Hammer et al., 2005; Van Berkum et al., 2007; see Ye and Zhou (2009) for a review). When the reader finds out that the input does not match the expectation based on context, whether this mismatch is due to number, semantics, or other factors, a process attempting to solve the mismatch is initiated (Kolk, Chwilla, van Herten, & Oor, 2003; van Berkum et al., 2007). The P600 effect observed here may reflect the effort to control the conflict (as reflected by early positivity) by rationalizing the current input.

We did not observe an *Nref* effect, on the singular pronoun when it mismatched its plural antecedent. An important difference between the present mismatch and the referential ambiguity in Nieuwland and Van Berkum (2006) is that the potential antecedents in sentences of the latter study (e.g. *The chemist hit the historian while* <u>*he*</u>...) were clearly individuated while the plural antecedent in the current experiment could still be seen as an integral group.

3. Experiment 2

Results in Experiment 1 suggested a functional priority of gender over number processing in resolving agreement mismatch, with the P600 being equally large for double mismatch and gender mismatch and being larger for gender mismatch than for number mismatch. Partially different neural mechanisms might underlie the processing of gender and number information.

However, one might argue that the observed earlier and larger P600 for gender mismatch in Experiment 1 can be attributed to the lower grammatical acceptability of gender than of number mismatch, as indicated by the pretest, rather than to the generally higher cognitive saliency for gender than for number information. One potential reason for not being able to balance the acceptability of gender and number mismatch was the use of a singular pronoun as the target word. In Chinese, the singular pronoun could be treated as un-marked whereas gender information is explicitly marked on the pronoun (through the semantic radical embedded in the character; Li & Shi, 2000). It is plausible that the number disagreement is less salient than the gender disagreement in Experiment 1 because of the markedness of relevant information on the target pronoun.

A previous study (Eberhard, 1997) demonstrated that whether a noun was marked as singular or plural impacts the frequency of producing subject–verb agreement errors. In this study, participants were asked to complete sentence fragments like *the key to the cabinets*. Production errors with subject-number disagreement (e.g., *were*) occurred when the local and the head noun bore different plurality. However, such errors occurred less often when the head noun was marked as plural (e.g., *the keys to the cabinet*). The author argued that, compared with singular noun, the plural noun was more richly-marked and had stronger attraction in agreement processing (plural attraction), rendering stronger tendency to produce words consistent with the plural form. In a semantic plausibility judgment task, responses to the English local adjective-noun combination was faster when the noun was in plural than in singular form (Kennison, 2005), suggesting that the plural-marked noun is more easily integrated with the preceding adjective than the singular form.

Moreover, in Experiment 1 number mismatch was not perceived as anomalous as gender mismatch or double mismatch. It is plausible that the violation of number agreement was not perceived as a real "mismatch" but just as less acceptable, because participants could draw a within-context inference to bridge the gap between the singular pronoun and plural antecedent (e.g. treating the pronoun as referring to one member of the group antecedent). The residual acceptability could have delayed the appearance of the P600 effect.

In Chinese, the plurality for human nouns is marked by attaching a morpheme (a character, i.e., the collective marker "们, /*men*/"; Iljic, 1994; Li, 1999) to a pronominal and nominal word. To balance the markedness of number and gender information encoded by the pronoun, in Experiment 2 we used plural pronouns, rather than singular pronouns, as critical words. Here both the number and gender information were marked in the orthographic forms. The design, as illustrated in Table 3, was essentially the same as Experiment 1. The number mismatch conditions were created by making the plural pronoun refer to a singular antecedent.

Table 3
Experimental conditions and exemplar sentences with approximate literal translations in Experiment 2.

Condition	Examples
Control	这些女患者情绪低落,医生/鼓励/她们/振作/起来。 Zhexie nühuanzhe qingxudiluo, yisheng guli **ta-men$_{female}$** zhenzuo qilai *These women patients were in low spirit, doctors encouraged* ***them$_{female}$*** *to cheer up*
Number mismatch	这位女患者情绪低落, 医生/鼓励/她们/振作/起来。 Zhewei nühuanzhe qingxudiluo, yisheng guli **ta-men$_{female}$** zhenzuo qilai *This woman patient was in low spirit, doctors encouraged* ***them$_{female}$*** *to cheer up*
Gender mismatch	这些女患者情绪低落,医生/鼓励/他们/振作/起来。 Zhexie nühuanzhe qingxudiluo, yisheng guli **ta-men$_{male}$** zhenzuo qilai *These women patients were in low spirit, doctors encouraged* ***them$_{male}$*** *to cheer up*
Double mismatch	这位女患者情绪低落, 医生/鼓励/他们/振作/起来。 Zhewei nühuanzhe qingxudiluo, yisheng guli ta-men$_{male}$ zhenzuo qilai *This woman patient was in low spirit, doctors encouraged* ***them$_{male}$*** *to cheer up*

3.1. Method

3.1.1. Participants

Twenty-four right handed native speakers of Chinese (12 males, age ranging from 22 to 27 years with mean age of 24.1 years), who did not take part in Experiment 1, were recruited from Southeast University. All of them had normal or corrected-to-normal vision and were paid for their participation.

3.1.2. Design and procedures

One hundred and eighty-four quartets of sentences were constructed based on the critical sentences used in Experiment 1, with the singular pronouns being replaced by plural pronouns and with the singular determiners of the antecedent nouns being replaced by the plural determiners (or vice versa) in different conditions (see Table 3). For some of the quartets (22 quartets, about 12% of the critical sentences), conjunctive words (such as "因此/because", "所以/therefore") were added at the beginning of the second clause to increase the relational coherence between the two clauses. Filler sentences used in Experiment 1 were modified such that 40 incorrect sentences had singular pronouns.

3.1.3. EEG recording and data analysis

As in Experiment 1, two pretests, one acceptability rating and one error correction, were carried out, with 20 participants for each pretest. None of these participants attended the ERP study. As shown in Table 4, sentences with double mismatch were rated less acceptable than sentences with single mismatch, $ps < 0.02$; although sentences with gender or number mismatch were rated as equally unacceptable, $p > 0.4$. The error correction results showed that, for the mismatch conditions, more than 96% of sentences were corrected by replacing the incorrect pronouns with the corresponding correct forms.

Other procedures including assigning stimuli into test lists, presenting stimulus, collecting and analyzing the EEG data were conducted in the same way as in Experiment 1. After rejecting incorrectly judged trials and trials with artifacts, the number of trials accepted for ERP data analysis was 38.4 for the control condition, 39.8 for the number mismatch condition, 42.3 for the gender mismatch condition, and 43.1 for the double mismatch condition. In order to compare directly the ERP pattern shown in

Table 4
Mean scores and standard deviations in the two pretests of Experiment 2.

	Sentence acceptability		Error correction rate (%)	
	Mean	SD	Mean	SD
Control	5.86	0.13	–	–
Number mismatch	1.96	0.27	98.9	0.6
Gender mismatch	2.04	0.27	96.4	0.7
Double mismatch	1.55	0.2	97	0.7

this experiment and the pattern in Experiment 1, the same time windows were selected in performing statistical analysis.

3.2. Results

3.2.1. Behavioral results

The accuracy rate in the sentence acceptability judgment task was 87.6% for the correct condition (SD = 1.7%), 91.3% for the number mismatch condition (SD = 1.4%), 95.6% for the gender mismatch condition (SD = 1.1%), and 98.3% for the double mismatch condition (SD = 0.5%). ANOVA over gender and number revealed a significant main effect of gender, $F(1,23) = 42.04$, $p < 0.001$, and a significant main effect of number, $F(1,23) = 7.78$, $p < 0.01$, but no interaction between them, $F(1,23) < 1$. Compared to the match conditions, judgments of sentences were more accurate when pronouns mismatched their antecedent in either gender (96.9% vs. 89.5%) or number (94.8% vs. 91.6%).

3.2.2. Electrophysiological results

ERP responses to critical pronouns are shown in Fig. 3. Similar to Experiment 1, the pronouns elicited an anterior-central negative response which finally merged into a positive deflection. The scalp topographies of differences between the three mismatch conditions and the control condition are depicted in Fig. 4. For the purpose of consistency, we defined the time windows in the same way as Experiment 1.

3.2.2.1. ERP responses in the 250–400 ms time window. Repeated-measures ANOVA revealed a significant two-way interaction between number and gender in this time window, $F(1,23) = 7.82$, $p < 0.05$ in the midline analysis, $F(1,23) = 9{,}08$, $p < 0.01$ in the lateral analysis. Neither number nor gender showed a significant main effect, Fs < 1.

Pairwise comparisons revealed a significant effect of gender between the gender mismatch and the control in the lateral analysis, $F(1,23) = 5.04$, $p < 0.05$, suggesting that a larger negativity (-0.72 μV) was elicited by the gender mismatch as compared with the control. On the other hand, although the number mismatch and double mismatch elicited less negative (more positive responses) compared with the control, as in Experiment 1, the differences between conditions did not reach significance ($ps > 0.1$).

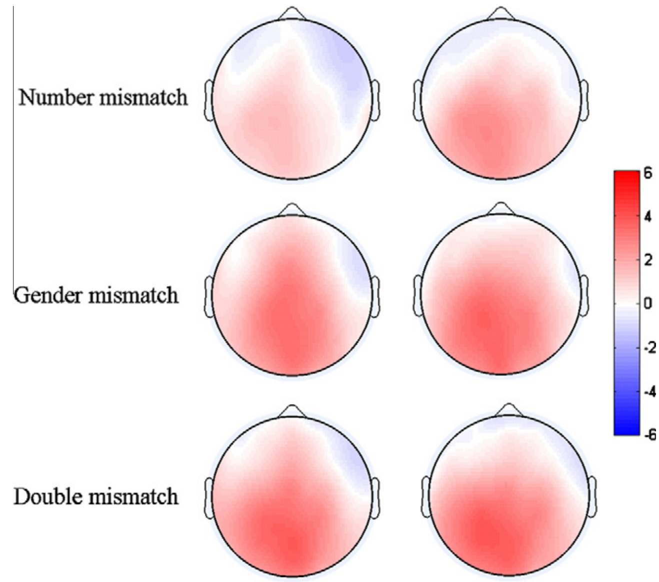

Fig. 4. Topographic maps for difference waves on the pronoun between each mismatch condition and the control condition in 400–550 ms window (the left column) and 550–800 ms window (the right column) in Experiment 2.

3.2.2.2. ERP responses in the 400–550 ms time window. As in Experiment 1, in the midline analysis there was a significant main effect of gender, $F(1,23) = 41.93$, $p < 0.001$, but no main effect of number, $F(1,23) = 1.72$, $p > 0.1$, nor interaction between gender and number, $F(1,23)=2.61$, $p > 0.1$. The lateral analysis obtained the same pattern of effects: a significant main effect of gender, $F(1,23) = 32.11$, $p < 0.001$, but no main effect of number, $F(1,23) < 1$, nor any interaction between gender and number, $F(1,23) < 1$.

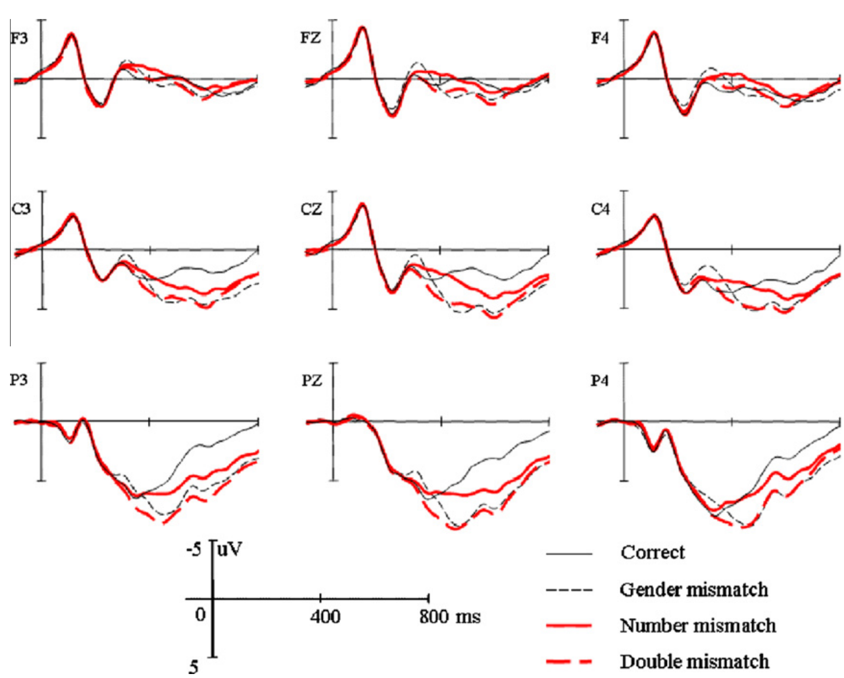

Fig. 3. Grand average ERPs time-locked to the critical plural pronoun for the control, the number mismatch, the gender mismatch and the double mismatch in Experiment 2.

Although the main effect of number did not reach significance, there was a significant interaction between number and electrode in the midline analysis, $F(5,115) = 3.90$, $p < 0.05$, and significant interactions between both number and hemisphere and number and region in the lateral analysis, $F(1,23) = 6.00$, $p < 0.05$ and $F(2,46) = 9.35$, $p < 0.005$, respectively. Further analyses showed that number mismatch did elicit more positive responses (0.56 μV) at posterior regions, $F(1,23) = 8.56$, $p < 0.01$.

3.2.2.3. ERP responses in the 550–800 ms time window. In the midline analysis, there was a significant main effect of gender, $F(1,23) = 26.22$, $p < 0.001$, a significant main effect of number, $F(1,23) = 9.81$, $p < 0.001$, and a significant interaction between gender and number, $F(1,23) = 7.09$, $p < 0.05$. The same pattern of effects was obtained in the lateral analysis: a main effect of gender, $F(1,23) = 25.46$, $p < 0.001$; a main effect of number, $F(1,23) = 6.90$, $p < 0.05$; and an interaction between number and gender, $F(1,23) = 5.43$, $p < 0.05$. These main effects suggest that both the gender and number mismatch between the pronoun and its antecedent elicited more positive responses on the pronoun in this time window (see Fig. 3).

The interaction between gender and number suggested that the system is differentially sensitive to the gender and number mismatch. Further analyses demonstrated that gender processing dominates over number processing. Firstly, there was no difference between the positivity effects elicited by gender mismatch and double mismatch, as there was no difference between the two conditions in ether the midline or lateral analysis, $F(1,23) < 1$. This finding suggested that the system has reduced sensitivity to number mismatch when pronoun concurrently mismatches the antecedent in gender. Secondly, a direct comparison between the gender mismatch and number mismatch conditions found more positive ERP responses to the former than to the latter in the midline (4.07 vs. 3.27 μV), $F(1,23) = 6.96$, $p < 0.05$, and in the lateral (3.52 vs. 2.73 μV), $F(1,23) = 9{,}64$, $p < 0.01$.

3.2.2.4. Combined analysis of ERP results in Experiments 1 and 2. Given that Experiments 1 and 2 used essentially the same design, we collapsed the ERP data in the two experiments and conducted statistical analyses for the mean amplitudes in different time windows.

ANOVA for the 250–400 ms time window revealed a significant main effect of gender in the midline analysis, $F(1,46) = 3.68$, $p < 0.05$, a significant main effect of number in both the midline and lateral analyses, $F(1,46) = 6.91$, $p < 0.05$, and $F(1,46) = 5.57$, $p < 0.05$, respectively. Neither gender nor number interacted with experiment, Fs < 1. Thus, across the two experiments, both the gender mismatch and the number mismatch elicited more positive (or less negative-going) responses than the baseline.

The interaction between gender and number was marginally significant in the lateral analysis, $F(1,46) = 3.09$, $0.05 < p < 0.1$. This interaction interacted further with experiment in the midline, $F(1,46) = 4.55$, $p < 0.05$, and in the lateral, $F(1,46) = 5.03$, $p < 0.05$. These interactions indicated that effects elicited by mismatch could have different patterns in Experiments 1 and 2. Further tests were conducted comparing each mismatch condition with the control condition, with experiment as a between-participant factor. For the double mismatch, the main effect was significant in the lateral analysis, $F(2,92) = 6.85$, $p < 0.05$. This effect did not interact with experiment in either the midline analysis, $F(5,230) = 1.13$, $p > 0.1$, or in the lateral analysis, $F(2,92) = 1.69$, $p > 0.1$, indicating that across the two experiments, the double mismatch elicited less negative-going responses in the 250–400 ms window. For the number mismatch, there was no significant main effect or interaction with experiment (Fs < 1), indicating that the number mismatch had essentially no effect in this time window. For the gender mismatch, although there was no significant main effect (Fs < 1), the interaction with experiment was significant in the midline, $F(5,230) = 3.83$, $0.05 < p < 0.1$, and in the lateral, $F(2,92) = 5.02$, $p < 0.05$. Clearly, while the number mismatch elicited less negative-going responses in Experiment 1, it elicited more negative-going responses in Experiment 2 (Figs. 1 and 3).

ANOVA for the 400–550 ms time window revealed a significant effect of gender mismatch in the midline, $F(1,46) = 111.84$, $p < 0.001$ and in the lateral, $F(1,46) = 80.01$, $p < 0.001$. There was a significant interaction between gender and electrode in the midline, $F(5,230) = 8.44$, $p < 0.005$, and between gender and region in the lateral, $F(2,92) = 11.45$, $p < 0.005$. These results suggested that the positive responses elicited by gender mismatch were shown mainly in the central and posterior regions. There was also a significant interaction between gender and experiment in the midline, $F(1,46) = 5.44$, $p < 0.05$, suggesting that the gender effect was larger in Experiment 1 than in Experiment 2.

The main effect of number was marginally significant in the midline, $F(1,46) = 3.55$, $0.05 < p < 0.1$, although the interaction between number and region was significant in the lateral analysis, $F(2,92) = 7.04$, $p < 0.01$. Thus, across the two experiments, the effect for number mismatch was present only in the posterior regions.

ANOVA for the 550–800 ms time window revealed a significant main effect of gender in the midline, $F(1,46) = 70.37$, $p < 0.001$, and in the lateral, $F(1,46) = 65.99$, $p < 0.001$, and a significant interaction between gender and electrode in the midline, $F(5,230) = 19.01$, $p < 0.001$, between gender and region in the lateral, $F(2,92) = 24.89$, $p < 0.01$. The interaction between gender and number was significant in the midline, $F(1,46) = 16.26$, $p < 0.001$, and in the lateral, $F(1,46) = 14.01$, $p < 0.005$, indicating that the P600 effect for the double mismatch was the same as the effect for the gender mismatch alone. Gender mismatch did not interact with experiment in the midline, $F(1,46) < 1$ and in the lateral, $F(1,46) = 1.43$, $p > 0.1$, suggesting that, across the two experiments, the P600 effect elicited by gender mismatch was larger in the posterior than in the anterior regions.

The main effect of number agreement was significant in the midline, $F(1,46) = 19.33$, $p < 0.001$, and in the lateral, $F(1,46) = 24.88$, $p < 0.001$. This effect interacted with region in the lateral analysis, $F(2,92) = 9.22$, $p < 0.01$, suggesting that the P600 effect elicited by number mismatch was larger in the posterior regions. There were also significant interactions between experiment, number and electrode in the midline, $F(5,230) = 5.86$, $p < 0.005$, and between experiment, number and region in the lateral, $F(2,92) = 9.39$, $p < 0.005$, suggesting that the posterior P600 effect elicited by number mismatch was larger in Experiment 2 than in Experiment 1. The interaction between gender and number indicated that the P600 effect elicited by number mismatch was smaller than the effect elicited by double mismatch.

Importantly, none of the effects elicited by gender mismatch, number mismatch or double mismatch interacted with experiment, indicating that the pattern of P600 effects in this time window was essentially the same for the two experiments.

3.3. Discussion

The pattern of effects in Experiment 2 was slightly different from that in Experiment 1. Compared with the control, the gender mismatch elicited *less* negative-going responses in the 250–400 ms window in Experiment 1 but *more* negative-going responses in Experiment 2, However, for the P600 effects, the two experiments had essentially the same patterns in the 400–550 ms and 550–800 ms windows, with the exception that the P600 effects appeared earlier and tended to be larger in Experiment 2 than in Experiment 1.

The most surprising finding in this experiment was that the gender mismatch elicited more negative-going responses, rather than less negative-going responses in Experiment 1, in the 250–400 ms window. This negativity was similar in temporal feature to the N400-like effect in Schmitt et al. (2002) and Lamers et al. (2006). Schmitt et al. (2002) consistently found a P600 effect on pronoun referring to a biological gender-mismatched antecedent, but an N400-like effect was observed only in sentences with non-diminutive antecedents (e.g., das Bübchen, 'the *little* boy'), but not with diminutive antecedents (e.g., der Bub, 'the boy'), suggesting that the appearance of the N400-like effect was modulated by the German diminutive suffix "-chen". Similarly, Lamers et al. (2006) found that the size of N400-like effect in Dutch was modulated by the difference in case marking assignment: the processing of a gender-mismatched pronoun evoked an early negativity (280–420 ms) when the pronoun was morphologically marked in terms of case marking than when it was not. It is possible that the negativity effect in Experiment 2 was due to the markedness of number information (i.e., the collective marker们, /men/) on the critical pronoun.

A second possible account could be that the shift from a positive effect in Experiment 1 to a negative effect in Experiment 2 in the 250–400 ms window could be due to the change of word length (Neville, Mills, & Lawson, 1992; Osterhout, Allen, & McLaughlin, 2002). Experiment 1 used a single-morpheme pronoun ((/*ta*/) while Experiment 2 used two-morpheme pronoun (/*tamen*/). Cross-experiment comparison showed that only the match conditions had differential ERP responses, $F(1,46) = 6.02$, $p < 0.05$ in the midline and $F(1,46) = 7.67$, $p < 0.01$ in the lateral, with the condition in Experiment 2 eliciting less negative responses than the condition in Experiment 1.

The third possible account could be that this shift is attributable to the difference in usage frequency of the singular pronoun (*ta*) and the plural pronoun (*tamen*) in Chinese. A previous study (King & Kutas, 1998) showed that the frequency of words affects the latency of an early negativity following P200, with words of low frequency evoking a delayed negativity than words of higher frequency. This account, however, does not fit the present findings as Experiments 1 and 2 obtained effects different in polarity and both pronouns are highly used in the language (more than 4000 per million).

The earlier appearance of the P600 effect at the posterior regions for number mismatch, as compared with Experiment 1, could be due to the explicit orthographic marking of the plural form (i.e., with "们, /men/" as a suffix or a compound constituent). The presence of formal cues, such as markings for number, may facilitate antecedent-pronoun co-indexation (Arnold, Eisenband, Brown-Schmidt, & Trueswell, 2000; Ehrlich, 1980; Garnham & Oakhill, 1985; Garnham, Oakhill, Ehrlich, & Carreiras, 1995), making the agreement conveyed through the cues cognitively more salient (see, for example, Eberhard, 1997; Kennison, 2005). Alternatively, the appearance of an earlier P600 effect for number mismatch, in Experiment 2, as compared with Experiment 1, was due to the fact that it is implausible for the system to link the plural pronoun to the singular antecedent through the type of inference process suggested for the singular pronoun with the plural antecedent in Experiment 1. Either way, the processing of the semantically-based number agreement is susceptible to the markedness of the pronoun.

Importantly, Experiment 2 showed that even with the explicit marking of the plural form on the pronoun and the absence of the "bridging inference" strategy, the P600 effect for processing number agreement was still smaller than that for processing gender agreement, as reflected in the magnitude of effect and in the statistical interaction between the two effects. The differentiation of the P600 effects between gender and number mismatch cannot be explained by a task-related view of the P600 (Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg, 2007), since the two mismatches were equally unacceptable (as demonstrated by pretest) and the involvement of the process of categorizing a sentence into correct or incorrect should be the same between the two conditions. Consistent findings of the dominance of gender agreement processing over number agreement processing may lead us to argue for differential functional significances of biological gender and number features in pronoun resolution (see Section 4).

4. General discussion

In this study, we investigated the processing of two types of semantically-based agreement features, biological gender and notional number, by comparing ERP responses to pronouns that match or mismatch with their antecedents in gender and/or number. Two experiments using singular (Experiment 1) and plural (Experiment 2) pronouns as critical words obtained essentially the same pattern of effects on the P600 for either gender or number or double mismatch, although the two experiments were somehow divergent in the ERP effects in earlier time windows. An important finding was that the P600 effect for gender mismatch appeared earlier (Experiment 1) and had a larger magnitude (Experiments 1 and 2) than the effect for number mismatch. Moreover, when gender and number were crossed to create a double mismatch, we found that the P600 effect for double mismatch was the same as for the single gender mismatch, with no apparent contribution from number mismatch. In the following discussion we concentrate on the functional hierarchy of semantic-based agreement features and its implications for the theory of pronoun resolution.

4.1. The cognitive salience of semantic gender and number agreement processing

The indistinguishable P600 effects for gender and double mismatch demonstrate the dominant role of gender information in pronoun resolution. The higher accuracy in grammatical judgment (Experiment 1) and the larger P600 effect for the gender than for the number mismatch (both Experiments 1 and 2) are inconsistent with previous studies on processing morphosyntatic agreement features. In one previous study, Faussart et al. (1999) showed that information concerning number agreement was processed more quickly than information concerning syntactic gender. In another study, Sagarra and Herschensohn (2010) showed that number agreement mismatch was identified more accurately than gender agreement in grammaticality judgment. One possible explanation for these results, as claimed in Faussart et al. (1999), is that number agreement is cognitively more salient than syntactic gender agreement. However, the present findings seem to suggest a reversed cognitive priority for semantic gender and number agreement processing in Chinese.

There could be a number of reasons for the difference in hierarchical relations for semantic-based vs. syntactic-based agreement features. For example, studies on syntactic agreement processing focused mostly on the local combinations (e.g. determiner–noun or noun–adjective relations), with the left anterior negativity and/or P600 effect as the common ERP manifestations of local morphosyntactic mismatch (see Molinaro et al. (2011) for a review); in contrast, studies on semantic agreement processing focused on constituents with long-distance dependency (e.g., anaphoric relations between antecedents and anaphors), with the N400 or P600 effect as the ERP correlate of semantic disagreement. The locality between constituents has been found to modulate the neural processing of semantic and/or syntactic features (Hammer et al., 2008;

Phillips et al., 2005; Qiu et al., 2012). It is also possible that the reversed hierarchy for semantic gender and number agreement features observed in this study is restricted only to the antecedent-pronoun relationship, whose processing may engage agreement processes structurally and/or temporally different from the processing of other types of agreement relations (Kreiner et al., 2012). Systematic studies are need to disentangle factors (e.g., syntactic-based vs. semantic-based agreement; local agreement vs. long-distance dependency; syntagmatic combination vs. anaphoric relation) that could affect the hierarchical relations between different agreement features. Nevertheless, it is important to note that, at least for pronoun resolution during online reading comprehension of Chinese sentences, semantic gender information has higher cognitive salience than notional number information.

The pattern of P600 responses for the present semantic gender and number mismatches is reminiscent of an ERP study (Nieuwland & Van Berkum, 2008) which examined how semantic and referential aspects of anaphoric noun phrase resolution interact during discourse comprehension. The authors found that anaphors that were both semantically incoherent and referentially ambiguous elicited an ERP pattern resembling the pattern elicited by anaphors that were only semantically incoherent. It appears that, in establishing co-referential relationships between a pronoun and its possible antecedents, the semantic, identity information (gender) plays a more dominant role than the quantity information (number) in anaphoric resolution, whether the number mismatch is conveyed through a singular noun that is ambiguously referring to more than one possible antecedents (Experiment 1) or through a plural pronoun referring to a single person (Experiment 2).

Two types of accounts could be proposed for the equivalent neural (P600) responses to double-mismatch and to gender mismatch. The first account assumes that the processing of gender information and the processing of number information are conducted in parallel and in horse-racing. Due to the functional precedence of gender information, the processing system is more sensitive to gender mismatch than to number mismatch; the system may discontinue the processing of number information as soon as it comes across gender mismatch, rendering the neural responses to double-mismatch similar to the responses to gender mismatch alone. The second account assumes that gender information processing and number information processing are conducted interactively, with the former contributed more to the combined neural responses than the latter (given the larger P600 effect for gender mismatch alone than for number mismatch alone). In either the account, it is the processing of biological gender agreement that dominates the pronoun resolution process.

There are at least two potential reasons for why semantic gender agreement has functional precedence over number agreement processing. Firstly, it has been argued that animate/inanimate categorization is fundamental to human cognition (Chan, Sze, & Cheung, 2004), serving as an anchoring point in social perception (Langacker, 1993), and the differentiation of biological gender is a core part of this categorization. Developmental studies showed that classifying humans into different sexes is more common and more useful than classifying people into a single person or a group of persons (Audring, 2008; Fagot & Leinbach, 1993; Martin, 1993). Both online and off-line studies have shown that 4- and 5-year-old children performed faster and more accurately in understanding pronouns that are disambiguated by gender, compared with those disambiguated by number (Arnold, Brown-Schmidt, & Trueswell, 2007; Tyler, 1983; Wykes, 1981).

Secondly, poverty of morphology in a given language would make it more susceptible to semantic influence in language processing (Acuña-Fariña, 2009; Berg, 1998). Mandarin Chinese has almost no inflectional morphology and has less restrictive syntactic structures than many Western languages. It has been argued that Chinese readers or listeners rely more on semantic than syntactic information in sentence comprehension (Ye et al., 2006; Yu & Zhang, 2008; Zhang, Yu, & Boland, 2010). This prevalence of semantic processing may extend to pronoun resolution such that the system relies more on agreement features closely related to semantics (e.g., biological gender information) than on features less related to semantics (e.g., the more suffix-like morpheme for plural).

4.2. Implications to the two-stage theory of pronoun resolution

As was introduced previously, different stages of pronoun resolution (Garrod & Terras, 2000) may be subserved by different ERP components (Callahan, 2008): the bonding stage is related to the manifestation of LAN or sustained negativity while the resolution stage is related to the appearance of the N400 or P600. Although this hypothesis was proposed mainly on the basis of morphosyntatic agreement processing, findings in the present study are largely consistent with this proposal.

For the pronoun resolution based on semantic gender and/or number features, we did not observe a LAN or sustained negativity effect on the mismatch pronouns. However, we did obtain ERP effects in the 250–400 ms window for the mismatch conditions. It is plausible that these effects are related to the processes in the bonding stage of pronoun resolution, although the effects in the present study were not consistent across conditions or across experiments. Further evidence is needed to solve the inconsistency.

The more solid P600 effects in this study are likely the manifestation of the resolution stage of pronoun processing. Previous studies on syntactic and semantic gender agreement revealed larger P600 effects on pronouns that mismatch both syntactic and semantic constraints than those that only mismatch one type of constraints (Schmitt et al., 2002), suggesting that the difficulty of resolving the mismatch increased when there were more mismatches. The P600 also increased when an anaphor and its antecedent had a long-distance dependency or were separated by a more complex structure, as compared with the situation in which the anaphora was local to its antecedent (Gouvea, Phillips, Kazanina, & Poeppel, 2010; Kaan et al., 2000; Phillips et al., 2005). Consistent with these findings, the present study observed stronger P600 responses to semantic gender mismatch on the pronoun than to number mismatch. The presentation of a pronoun triggers a linking (or retrieving) process for an agreeing antecedent in the previous contexts to form a co-reference. However, in cases where there is no appropriate antecedent available due to featural mismatch, the linking (or retrieving) process would fail and a reprocessing operation would be initiated. The difference in the size of the P600 effect for gender mismatch and number mismatch may reflect the relative difficulty in revising the failed link and finding an alternative. Thus the larger P600 effect for gender mismatch than for number mismatch could be taken as evidence that reanalysis or discourse integration process for gender information is more demanding than for number information in pronoun resolution (see also Barber & Carreiras, 2005).

5. Conclusion

By simultaneously manipulating the agreement between a pronoun and its antecedent in terms of biological gender and notional number, this ERP study demonstrated that both gender and number mismatches elicit larger late positive responses on the pronoun. The P600 effect, however, appears earlier and has a larger magnitude for gender mismatch than for number mismatch. Moreover, when a pronoun simultaneously mismatches its antecedent in both gender and number, it is the gender mismatch that determines the magnitude and latency of the P600 effect. These findings demonstrate differential mechanisms underlying the processing of different semantic features during (Chinese) pronoun resolution in sentence comprehension.

Acknowledgments

This study was supported by grants from the Natural Science Foundation of China (30970889, 3011097) and the Ministry of Science and Technology of China (2010CB833904) to Xiaolin Zhou, by grants from China Post-doctoral Science Foundation (20100480150, 2012T50005) to Xiaoming Jiang, by grants from Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (12KJB180007) to Xiaodong Xu, by the Natural Science Foundation of China (31100814) to Qingrong Chen, and by the Academic Development Priority Program of Jiangsu Higher Education Institutions awarded to School of Foreign Languages and Cultures, Nanjing Normal University. Xiaodong Xu and Xiaoming Jiang equally contributed to this work. We thank Dr. Horacio A. Barber, Mr. Stephen Politzer-Ahles, Dr. Nicola Molinaro and other two anonymous reviewers for their comments on earlier versions of the manuscript. Electronic mail concerning this study should be addressed to Dr. Xiaolin Zhou, xz104@pku.edu.cn.

References

Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *Lingua, 119*, 389–424.
Adani, F., van der Lely, H. K. J., Forgiarini, M., & Guasti, M. T. (2012). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua, 120*, 2148–2166.
Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *Syntax, 5*, 1–25.
Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Language and Cognitive Process, 22*, 527–565.
Arnold, J., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. (2000). The rapid use of gender information: evidence of the time course of pronoun resolution from eye-tracking. *Cognition, 76*, B13–B26.
Audring, J. (2008). Gender assignment and gender agreement: Evidence from pronominal gender languages. *Morphology, 18*, 93–116.
Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Journal of Cognitive Neuroscience, 17*, 137–153.
Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *On-line study of sentence comprehension; eye-tracking, ERP and beyond* (pp. 309–327). Brighton, UK: Psychology Press.
Berg, T. (1998). The resolution of number agreement conflicts in English and German agreement patterns. *Linguistics, 36*, 41–70.
Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Journal of Memory and Language, 51*, 251–278.
Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews, 59*, 55–73.
Callahan, S. M. (2008). Processing anaphoric constructions: Insights from electrophysiological studies. *Journal of Neurolinguistics, 21*, 231–266.
Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy and implications for linguistic theory. *Lingua, 115*, 259–285.
Carreiras, M., Carr, L., Barber, H. A., & Hernandez, A. (2010). Where syntax meets math: Right intraparietal sulcus activation in response to grammatical number agreement violations. *Neuroimage, 49*, 1741–1749.
Chan, A. S., Sze, S. L. M., & Cheung, M. (2004). Neuroanatomical basis in the temporal lobes for processing living things. *Neuropsychology, 18*, 700–709.
Corbett, G. G. (2006). *Agreement*. Cambridge: Cambridge University Press.
De Vincenzi, M., & Di Domenico, E. (1999). A distinction among phi-features: The role of gender and number in the retrieval of pronoun antecedents. *Italian Journal of Linguistics, 11*, 41–74.
Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*, 200–224.
Deutsch, A., & Dank, M. (2009). Conflicting cues and competition between notional and grammatical factors in producing number and gender agreement: Evidence from Hebrew. *Journal of Memory and Language, 60*, 112–143.
Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language, 36*, 147–164.
Ehrlich, K. (1980). Comprehension of pronouns. *The Quarterly Journal of Experimental Psychology, 32*, 247–255.
Fagot, B. I., & Leinbach, M. D. (1993). Gender-role development in young children: From discrimination to labeling. *Developmental Review, 13*, 205–224.
Faussart, C., Jacubowicz, C., & Costes, M. (1999). Gender and number processing in spoken French and Spanish. *Rivista de Linguistica, 11*, 75–101.
Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of Psycholinguistic Research, 28*, 467–484.
Garnham, A., & Oakhill, J. (1985). On-line resolution of anaphoric pronouns: Effects of inference making and verb semantics. *British Journal of Psychology, 76*, 385–393.
Garnham, A., Oakhill, J., Ehrlich, M.-F., & Carreiras, M. (1995). Representations and processes in the interpretation of pronouns: new evidence from Spanish and French. *Journal of Memory and Language, 34*, 41–62.
Garrod, S., & Sanford, A. J. (1990). Referential processing in reading: Focusing on roles and individuals. In D. A. Balota, G. B. Flores d'Arcias, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 465–486). Hillsdale, NJ: Erlbaum.
Garrod, S., & Sanford, A. J. (1994). Resolving sentences in a discourse context: How discourse representation affects language understanding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 675–698). New York: Academic Press.
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.
Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*, 1–76.
Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes, 25*, 149–188.
Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of grammar* (pp. 73–113). Cambridge, MA: M.I.T.
Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556–568.
Hagoort, P., & Brown, C. M. (1999). Gender electrified: ERP evidence on the syntactic nature of gender processing. *Journal of Psycholinguistic Research, 28*, 715–728.
Hammer, A., Jansma, B. M., Lamers, M., & Münte, T. F. (2005). Pronominal reference in sentences about persons or things: An electrophysiological approach. *Journal of Cognitive Neuroscience, 17*, 227–239.
Hammer, A. H., Jansma, B. M., Lammers, M., & Münte, T. F. (2008). Interplay of meaning, syntax and working memory during pronoun resolution investigated by ERPs. *Brain Research, 1230*, 177–191.
Harley, H., & Ritter, E. (2002). Person and number in pronouns: motivating a feature-geometric analysis. *Language, 78*, 482–526.
Igoa, J. M., García-Albea, J. M., & Sánchez-Casas, R. (1999). Gender–number dissociation in sentence production in Spanish. *Rivista di Linguistica, 11*, 163–193.
Iljic, R. (1994). Quantification in Mandarin Chinese: two markers of plurality. *Linguistics, 32*, 91–116.
Jiang, X. M., Tan, Y. Y., & Zhou, X. L. (2009). Processing the universal quantifier during sentence comprehension: ERP evidence. *Neuropsychologia, 47*, 1799–1851.
Jiang, X. M., & Zhou, X. L. (2009). Processing different levels of syntactic hierarchy: An ERP study on Chinese. *Neuropsychologia, 47*, 1282–1293.
Kaan, E., Dallas, A. C., & Barkley, C. M. (2007). Processing bare quantifiers in discourse. *Brain Research, 1146*, 199–209.
Kaan, E., Harris, A., Gibson, G., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201.
Kennison, S. M. (2005). Different time courses of integrative semantic processing for plural and singular nouns: implications for theories of sentence processing. *Cognition, 97*, 269–294.
King, J. W., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters, 244*, 61–64.
Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain and Language, 85*, 1–36.
Kreiner, H., Garrod, S., & Sturt, P. (2012). Number agreement in sentence comprehension: The relationship between grammatical and conceptual factors. *Language and Cognitive Processes*. http://dx.doi.org/10.1080/01690965.2012.667567.
Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.
Lamers, M. J. A., Jansma, B. M., Hammer, A., & Münte, T. F. (2006). Neural correlates of semantic and syntactic processes in the comprehension of case marked pronouns: Evidence from German and Dutch. *BMC Neuroscience, 7*, 1–12.
Lamers, M. J., Jansma, B. M., Hammer, A., & Münte, T. F. (2008). Differences in the processing of anaphoric reference between closely related language: Neurophysiological evidence. *BMC Neuroscience, 9*, 1–13.
Langacker, R. W. (1993). Reference-point constructions. *Cognitive Linguistics, 4*, 1–38.
Li, Y. H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics, 8*, 75–99.
Li, Y. H. A., & Shi, Y. Z. (2000). The story of men. *Modern Linguistics, 1*, 61–62 (in Chinese).
Li, X., & Zhou, X. (2010). Who is ziji? ERP responses to the Chinese reflexive pronoun during sentence comprehension. *Brain Research, 1331*, 96–104.
Mancini, S., Molinaro, N., Rizzi, L., & Carreiras, M. (2011). A person is not a number: discourse involvement in subject–verb agreement computation. *Brain Research, 1410*, 64–76.

Martin, C. L. (1993). New directions for investigating children's gender knowledge. *Developmental Review, 13*, 184–204.

Martín-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantic prevalence over syntax during sentence processing: A brain potential study of noun–adjective agreement in Spanish. *Brain Research, 1093*, 178–189.

Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*, 908–930.

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement violation: An ERP analysis of its interpretation. *Cognition, 106*, 963–974.

Molinaro, N., Vespignani, F., Zamparelli, R., & Job, R. (2011). Why brother and sister are not just siblings: Repair processes in agreement computation. *Journal of Memory and Language, 64*, 211–232.

Neville, H. J., Mills, D. L., & Lawson, D. S. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2*, 244–258.

Nevins, A., Dillon, B., Malhotra, S., & Phillips, C. (2007). The role of feature-number and feature-type in processing Hindi verb agreement violations. *Brain Research, 1164*, 81–94.

Nieuwland, M. S., & Van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research, 1118*, 155–167.

Nieuwland, M. S., & Van Berkum, J. J. A. (2008). The interplay between semantic and referential aspects of anaphoric noun phrase resolution: Evidence from ERPs. *Brain and Language, 106*, 119–131.

Noyer, R. R. (1992). *Features, positions, and affixes in autonomous morphological structure*. Cambridge, MA: MIT dissertation.

Osterhout, L., Allen, M., & McLaughlin, J. (2002). Words in the brain: Lexical determinants of word-induced brain activity. *Journal of Neurolinguistics, 15*, 171–187.

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

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

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*, 407–428.

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.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*, 2022–2039.

Schmitt, B., Lamers, M., & Münte, T. F. (2002). Electrophysiological estimates of biological and syntactic gender violation during pronoun processing. *Cognitive Brain Research, 14*, 333–346.

Silva-pereyra, J. F., & Carreiras, M. (2007). An ERP study of agreement features in Spanish. *Brain Research, 1185*, 201–211.

Sitnikova, T., Holcomb, P., Kiyonaga, K., & Kuperberg, G. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real world events. *Journal of Cognitive Neuroscience, 20*, 2037–2057.

Steinhauer, K., Pancheva, R., Newman, A. J., Gennari, S., & Ullman, M. T. (2001). How the mass counts: An electrophysiological approach to the processing of lexical features. *Neuroreport, 12*, 999–1005.

Tyler, L. K. (1983). The development of discourse mapping processes: The on-line interpretation of anaphoric expressions. *Cognition, 13*, 309–341.

Van Berkum, J. J. A., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language, 41*, 147–182.

Van Berkum, J. J. A., Brown, C. M., Hagoort, P., & Zwitserlood, P. (2003). Event-related brain potentials reflect discourse-referential ambiguity in spoken language comprehension. *Psychophysiology, 40*, 235–248.

Van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*, 158–171.

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: The role of conceptual factors. *Cognition, 51*, 261–298.

Wykes, T. (1981). Inferences and children's comprehension of pronouns. *Journal of Experimental Child Psychology, 32*, 264–278.

Ye, Z., Luo, Y., Friederici, A. D., & Zhou, X. (2006). Semantic and syntactic processing in Chinese sentence comprehension: Evidence from event-related potentials. *Brain Research, 1071*, 186–196.

Ye, Z., & Zhou, X. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews, 33*, 1168–1177.

Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *NeuroReport, 19*, 745–749.

Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 765–781.

Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia, 48*, 1551–1562.