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] R Cn L nn n n K L C l D l n n L nn n (Mn Eu n), u n , Nn n 210096, C n
[b] l F n Ln u n C lu , Nn n N l n , Nn n 210097, C n
[c] Cn B n n C n n n D n P l , P n n , B n 100871, C n
[d]

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. vs. *) than to a sentence with an adjective disagreeing in number (e.g. vs. *).

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 with2005; Igoa, a2tipo 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 (*N*) 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., *Fʋ* *n* - *n* , *ʋn* ; Kaan, Dallas, & Barkley, 2007). The *N* 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 would be present when the number agreement was violated. According to Nieuwland and Van Berkum (2006), *N* is an anteriorly-distributed sustained negativity starting from about 300 ms post-onset of anaphoric phrases with ambiguous reference (e.g. *n w l* __...; 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 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 ("□ □", "them", masculine or "她□", "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. P u

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 n n n l

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

2.2.2.1. ERP n n 250 400 wn w. Repeated-measures ANOVA over the mean amplitudes in this window yielded a significant main effect of gender in the midline analysis (0.73 μV), $F(1,23) = 4.63$, < 0.05, and a significant main effect of number in both the midline (0.83 μV) and the lateral (0.68 μV) analysis, $F(1,23) = 5.90$, < 0.05, and $F(1,23) = 6.76$, < 0.05, *2.2.2.3. ERP n n 550 800 wn w.* There was a significant main effect of gender, $F(1,23) = 44.44$, < 0.001 in the midline and $F(1,23) = 41.07$, < 0.001 in the lateral, and of number, $F(1,23) = 9.92$, < 0.005 in the midline and $F(1,23) = 18.84$, < 0.001 in the lateral, suggesting that the gender and the number mismatch conditions elicited increased P600 responses as com-

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

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

3.2.2.3. ERP n n 550 800 wn w. In the midline analysis, there was a significant main effect of gender, $F(1,23) = 26.22$, < 0.001, a significant main effect of number, $F(1,23) = 9.81$, < 0.001, and a significant interaction between gender and number, $F(1,23) = 7.09$, < 0.05. The same pattern of effects was obtained in the lateral analysis: a main effect of gender, $F(1,23) = 25.46$, < 0.001; a main effect of number, $F(1,23) = 6.90$, < 0.05; and an interaction between number and gender, $F(1,23) = 5.43$, < 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), $F(1,23) = 6.96$, < 0.05, and in the lateral (3.52 vs. 2.73 μV), $F(1,23) = 9,64$, < 0.01.

3.2.2.4. C n n l ERP ɥ l n E n 1 n 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, $F(1,46) = 3.68$, < 0.05, a significant main effect of number in both the midline and lateral analyses, $F(1,46) = 6.91$, < 0.05, and $F(1,46) = 5.57$, < 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, $F(1,46) = 3.09$, $0.05 < \; < 0.1$. This interaction interacted further with experiment in the midline, $F(1,46) = 4.55$, < 0.05, and in the lateral, $F(1,46) = 5.03$, < 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, $F(2,92) = 6.85$, < 0.05. This effect did not interact with experiment in either the midline analysis, $F(5,230) = 1.13$, > 0.1, or in the lateral analysis, $F(2,92) = 1.69$, > 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, $F(5,230) = 3.83$, $0.05 < \; < 0.1$, and in the lateral, $F(2,92) = 5.02$, $<$ [illegible]

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 *l l* 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 □, /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 ((/ /) while Experiment 2 used two-morpheme pronoun (/ *n* /). Cross-experiment comparison showed that only the match conditions had differential ERP responses, $F(1, 46) = 6.02$, < 0.05 in the mid-F

[illegible]

genders-259.536.possle-mu-416.54Si4 8possplu-421.54tecedentsExperiment1J.4(7.)-ri

by

pronoun.

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 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.

A[illegible]

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.

[illegible]

Acuña-Fariña, J. C. (2009). The linguistics and psycholinguistics of agreement: A tutorial overview. *ln u*, *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. *ln u*, *120*, 2148–2166.

Antón-Méndez, I., Nicol, J. L., & Garrett, M. F. (2002). The relation between gender and number agreement processing. *n*, *5*, 1–25.

Arnold, J. E., Brown-Schmidt, S., & Trueswell, J. (2007). Children's use of gender and order-of-mention during pronoun comprehension. *Ln u n C n P*, *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. *C n n*, *76*, B13–B26.

Audring, J. (2008). Gender assignment and gender agreement: Evidence from pronominal gender languages. *M l*, *18*, 93–116.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: an ERP comparison. *Ju n l C n Nu n*, *17*, 137–153.

Barber, H., Salillas, E., & Carreiras, M. (2004). Gender or genders agreement? In M. Carreiras & C. Clifton (Eds.), *@ -h u n n n n ; - n , ERP n n* (pp. 309–327). Brighton, UK: Psychology Press.

Berg, T. (1998). The resolution of number agreement conflicts in English and German agreement patterns. *ln u*, *36*, 41–70.

Bock, K., Eberhard, K. M., & Cutting, J. C. (2004). Producing number agreement: How pronouns equal verbs. *Ju n l M n Ln u*, *51*, 251–278.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *B n R R w*, *59*, 55–73.

Callahan, S. M. (2008). Processing anaphoric constructions: Insights from electrophysiological studies. *Ju n l Nu h u*, *21*, 231–266.

Carminati, M. N. (2005). Processing reflexes of the Feature Hierarchy and implications for linguistic theory. *ln u*, *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. *Nu*, *49*, 1741–1749.

Chan, A. S., Sze, S. L. M., & Cheung, M. (2004). Neuroanatomical basis in the temporal lobes for processing living things. *Nu 749.*

Martin, C. L. (1993). New directions for investigating children's gender knowledge. *D l n l R w*, *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. *B n R* , *1093*, 178–189.

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

Molinaro, N., Kim, A., Vespignani, F., & Job, R. (2008). Anaphoric agreement