

# ERP signatures of subject–verb agreement in L2 learning\*

LANG CHEN  
HUA SHU  
YOUYI LIU  
JINGJING ZHAO  
State Key Laboratory for Cognitive Neuroscience and Learning, Beijing Normal University  
PING LI  
*University of Richmond*

*In this study we examined ERP (event-related-potential) responses in the morphosyntactic processing of subject–verb agreements by L2 Chinese learners of English. Fifteen proficient L2 learners and fifteen native English speakers were presented with English sentences that varied in the grammaticality of the sentence with respect to subject–verb agreement. Our results indicate that late L2 learners show distinct ERP responses from native speakers in the processing of syntactic features that are absent in their L1, even when their behavioral patterns are similar to those of native speakers. The results are taken to support the proposal that language-specific experiences with L1 shape the neural structure of processing in L2.*

## Introduction

A casual observation of speech errors made by Chinese learners of English is that subject–verb (S–V) agreement is a perennial problem for these learners, reflected in the rampant errors in sentences which require that the subject and the verb agree in number (plural or singular). Subject–verb agreement is an important grammatical feature found in most Indo-European languages. In English, only number agreement is required in the present tense (*John speaks perfect Spanish* vs. *They speak perfect Spanish*). In many other Indo-European languages, number, gender, and case agreements may be part of the S–V agreement requirement. Given that the Chinese language does not have grammatical morphology for marking number, gender, and case, native speakers of Chinese often find it difficult to master the L2 grammar when the L2 is a morphologically rich language (for an earlier empirical study, see Liu, Bates and Li, 1992).

Subject–verb agreement has been an area of active research in language processing (see Bock and Miller, 1991; Eberhard, 1997; Pearlmuter, Garnsey and Bock, 1999), but it is only recently that it has attracted widespread attention when both behavioral and electrophysiological data have been collected and corroborated from normal monolingual speakers and brain-injured patients (Osterhout and Mobley, 1995; Wassenaar, Brown, and Hagoort, 2004; Rossi et al., 2005). Osterhout and Mobley (1995) showed, in an event-related-potential (ERP) study, that sentences that contain violations of

S–V agreement (e.g., *the elected officials hopes...*) would elicit different ERP components in the course of processing: a left anterior negativity (LAN) that peaks around 300–400 ms and a late positivity (P600) that peaks at around 500–700 ms after stimulus onset. Similar patterns have been reported in other sentence processing/reading studies (Roehm et al., 2005; Rossi et al., 2005). For example, Rossi et al. (2005) presented participants with sentences that contained S–V agreement violations and found a combination of LAN and P600 effects. Interestingly, adults with aphasic symptoms (e.g., Broca's patients) or children with developmental dyslexia do not consistently show the typical P600 responses in case of S–V agreement violations, as reported by Rispens, Been and Zwartz (2006) and Wassenaar et al. (2004). Thus, P600 (and LAN to a lesser extent) appears to be an important ERP signature for syntactic processing in the normal adult speakers, particularly for the processing of S–V agreement in sentences.

A number of studies have also used the ERP method to study syntactic and semantic processing in the bilingual or second language context. Weber-Fox and Neville (1996) asked a group of Chinese–English bilinguals to read sentences that contained three different types of syntactic violations (phrase structure, specificity constraint, and subadjacency constraint) as well as semantic violations. They found differences in the timing and distribution of the ERPs to both syntactic and semantic violations when comparing L2 learners to native speakers. ERP responses to syntactic violations by native speakers differed from those by L2 learners who were exposed to English after age 4, suggesting non-native-like neural patterns for syntactic processing in L2 learners. Moreover, early

\* This research was supported by grants from the Natural Science Foundation of China (#60534080, #30470574) to HS and a grant from the NSF (BCS-0131829) to PL.

Address correspondence

Ping Li, Department of Psychology, University of Richmond, Richmond, VA 23173, USA  
E-mail: pli@richmond.edu

bilinguals showed a typical P600 pattern (before age 10) when syntactic violations were processed, but late bilinguals (after age 11) showed a later latency of the positivity (700–900 ms). Hahne and Friederici (2001) further examined syntactic processing by late Japanese learners of German, and their findings were highly consistent with those from Weber-Fox and Neville (1996). In their study, native speakers and bilingual learners were presented with auditory (rather than visual) sentences that contained syntactic, semantic, or combined syntactic and semantic violations. Compared with correct sentences, the syntactic violations elicited a biphasic early left anterior negativity (LAN) and late positivity (P600) for native German speakers. However, such biphasic ERP responses were absent in the L2 learners. The LAN effects were observed only with the native German speakers. These patterns were taken to indicate that native and non-native syntactic processing may differ in fundamental ways, especially with respect to early automatic processes (presence or absence of anterior negativity) and late syntactic integration/reanalysis (latency and amplitude of P600) (see also Hahne and Friederici, 1998, 1999, for discussion of native patterns).

This “classical” view of the biphasic early negativities (LAN) and late positivity (P600) for native speakers as opposed to bilingual learners, however, should be taken with some caution. Recent research indicates that the picture may not be so clear-cut. For example, Hahne, Mueller and Clahsen (2006) pointed out that there are substantial topographic variations of the LAN effects in a number of ERP studies with native Italian, German, or Catalan speakers, where the anterior negativities were either bilateral or restricted to the right anterior sites (rather than left, as in LAN). While the detailed patterns may vary, still, it is clear that syntactic processing involves generally distinct patterns of neural activities for L1 and L2 speakers, either in latency or amplitude (as revealed by ERP studies) or in cortical site differences (as revealed by fMRI studies; e.g., Wartenburger et al., 2003; see Hernandez and Li, 2007 for review). Recent ERP evidence from training studies suggests that L2 learners can quickly pick up non-native syntactic and semantic features with minimal instructions (e.g., 1–4 months), especially when the features are also realized in some way in the L1 (Osterhout et al., 2006; see also discussion below). It is yet unclear, however, whether these early native-like patterns will eventually end up the same as native patterns, especially for syntactic features that are absent in the L1 (such as the S–V agreement studied here).

Given that syntactic processing, especially the morphosyntactic processing of grammatical agreement, has been a lively topic of research, it is surprising that few studies have been specifically designed to examine ERP responses to S–V agreement in the bilingual context. Only two previous studies have compared L2

learners’ processing of subject–verb agreement with the processing of other sentence structures. Osterhout et al. (2004; see also Osterhout et al., 2006) examined English learners of French in the processing of semantically and syntactically anomalous sentences, and one of the syntactically anomalous sentence types involved verb conjugation errors (S–V agreement). These authors found that while native French speakers showed clear N400 responses to semantic anomalies and P600 responses to syntactic anomalies, L2 learners initially showed N400-like effects to syntactic anomalies. Interestingly, among the two syntactically anomalous sentence types, P600 effects were observed (after an eight-month longitudinal training of the L2 learners) with the sentences involving verb conjugation errors but not those involving determiner–noun agreement errors. Osterhout et al. interpreted that these patterns as reflecting effects of the similarity in grammatical structures between the learner’s two languages: verb conjugation systems are similar in English and French while determiner–noun agreement exists only in French. Examining English learners of Spanish, Tokowicz and MacWhinney (2005) obtained comparable results. Subject–verb agreement, which is similar in English and Spanish, showed P600 patterns when the agreement was violated in L1 and L2; determiner–noun agreement, which is different between English and Spanish, elicited no P600 effects when the agreement was violated in the L2.

Capitalizing on the lack of S–V agreement in the Chinese language, and on the fact that L2 Chinese learners encounter significant difficulties with S–V agreement in English, our study attempts to reveal the neural patterns of ERP responses that may underlie L1 and L2 morphosyntactic processing. Previous research in morphosyntactic processing, including the ones reviewed above, has been conducted mainly with Indo-European languages, particularly English, Dutch, and German. Chinese represents a significantly different language from the languages that have been examined so far. In contrast to Indo-European languages, Chinese uses virtually no grammatical morphology to mark gender, number, or case, and this lack of grammatical morphology has clear consequences on sentence comprehension and language processing in general (for discussion on such consequences see Li, Bates and MacWhinney, 1993; Li, Jin and Tan, 2004; see also the chapters in Li, Tan, Bates and Tzeng 2006). Given the lack of grammatical morphology, the Chinese syntax does not require S–V agreement, such that any nominal subject can take any verb form. Thus, the acquisition of S–V agreement represents a significant obstacle to L2 Chinese learners, and it may well be the last obstacle that many learners have to overcome, as S–V agreement errors often linger on even when the learner has become proficient in the target L2 language.

Our general hypothesis, based on previous findings to date, is that L2 processing of S–V agreement will evoke distinct neural patterns from L1 morphosyntactic processing, even when behavioral measures (such as response times) show similar patterns. For example, key ERP signatures of morphosyntactic processes (e.g., P600) may be absent in L2 processing, while other ERP components may emerge and differ from those in L1 processing. Recent evidence suggests that the presence or absence of ERP signatures of morphosyntactic processing may depend on the similarity of the structures in the bilingual's two languages, as indicated by Osterhout et al. (2004) and Tokowicz and MacWhinney (2005). Given the significant differences between Chinese and English (unlike French and English or Spanish and English), our study also attempts to reveal the specific patterns of ERP responses that are associated with bilingual processing of the S–V agreement, and how the language-specific properties of Chinese influence learners' processing characteristics.

## Method

### Participants

Eighteen native Chinese speakers who had learned English as a second language participated in this study. They were undergraduate or graduate students (mean age 22 years, range 20–26; 8 males) from Beijing Normal University and Capital Normal University. They had learned English for more than 9 years (mean 10, range 9–14), and the average age when they started to learn English was 12 years (range 10–13). To ensure a more uniform level of proficiency, all participants had obtained the CET (College English Test) level 6 (the highest level), or had scored higher than 80 (out of 100) on the CET level 4. On a self-rated English reading proficiency scale (of the questionnaire by Li, Sepanski and Zhao, 2006), the participants' average value was 4.64 (1 = very poor; 7 = native-like). Data from three participants were excluded due to their low accuracy on the test sentences (<.75). Therefore, our analyses in this study involved data from fifteen instead of eighteen L2 learners.

An additional set of seventeen native English speakers served as the L1 controls, against which our L2 learners could be compared. The native speakers were exchange students from Australia, UK, and USA who were studying Chinese at Beijing Language and Culture University and the Second High School of Beijing Normal University (mean age 20 years, range 16–24; 9 males). All participants were right-handed, and had normal or corrected-to-normal vision. They were paid for their participation in the experiment. Data from two native English speakers were excluded from our analyses: one

participant had excessive body movements during the ERP recording, and another showed outlier performance in her ERP patterns. Thus, our analyses below involved data from fifteen instead of seventeen native speakers.

### Materials

One hundred and sixty-eight test sentences were constructed (see examples below, which represent the four versions of one sentence). Each sentence consisted of a subject noun modified by a preposition phrase (PP), followed by a verb phrase. The verb phrase consisted of a past-tense copula (*was* or *were*) plus an adjectival, so that lexical variations in semantic difficulty or visual complexity of the verb were minimized. The subject noun was always singular, and four versions of the sentence were constructed by varying the number of the noun in the PP (e.g., *car* or *cars*) and the number of the verb (*was* or *were*). All nouns in the sentences were count nouns, and their plural forms were regular. The length of the sentences ranged from 7 to 10 words. A sample is given below (see the Appendix for more examples).

- (a) *Grammatical, congruent (G-C)*  
The price of the car was too high.
- (b) *Grammatical, incongruent (G-I)*  
The price of the cars was too high.
- (c) *Ungrammatical, congruent (U-C)*  
\*The price of the cars were too high.
- (d) *Ungrammatical, incongruent (U-I)*  
\*The price of the car were too high.

The four versions of the sentence (henceforth “sentence type”) correspond to the four critical conditions in our experiment. Type (a) is a grammatical sentence, with congruency in number between the sentence subject (the head noun) and the verb and also between the noun in the PP (the local noun) and the verb (i.e., both singular). This type served as the baseline condition against which the following sentence types were compared. Type (b) differs from (a) in that the local noun is plural while the following verb is singular, hence the incongruence. Types (c) and (d) are ungrammatical versions of (a) and (b). In particular, for type (c) the local noun agrees with the verb but not with the head noun, and as such, the adjacency between the local noun and the verb can attract agreement errors in speech production or comprehension (see Franck, Vigliocco and Nicol, 2002 and Vigliocco and Nicol, 1998 for psycholinguistic studies of attraction errors). Our goal to include this type of sentences was to see if L1 and L2 speakers would also be more prone to such attraction errors as compared with both the baseline (a) and the ungrammatical (d) sentences.

Which type of the sentence was received by a given participant was determined in four separate lists of sentences, with each list containing one version of a

Table 1. Plausibility and grammaticality ratings of four sentence types.

| Sentence type  | Plausibility |      | Grammaticality |      |
|--|--------------|------|----------------|------|
|  | mean         | SD   | mean           | SD   |
| G-C <i>The price of the car was too high.</i>        | 4.50         | 0.53 | 4.31           | 0.51 |
| G-I <i>The price of the cars was too high.</i>       | 4.32         | 0.59 | 4.21           | 0.64 |
| U-C    * <i>The price of the cars were too high.</i> | 4.38         | 0.57 | 1.65           | 0.83 |
| U-I    * <i>The price of the car were too high.</i>  | 4.45         | 0.49 | 1.65           | 0.84 |

Note: \* = ungrammatical sentence

given sentence but an equal number of the four sentence types. Which list was given to which participant was determined randomly. This ensured that every participant received an equal number of (a) to (d) sentences (i.e., 42 sentences in each condition). A separate set of 100 sentences was selected as fillers, 60 of which were ungrammatical and contained different grammatical violations (including errors in past tense, word order, auxiliaries, etc, patterning Johnson and Newport, 1989). The length of filler sentences ranged from 4 to 9. Each participant was presented with the same filler sentences intermingled with the test sentences, thus receiving a total of 268 sentences in the entire test session. As a result, 54% of these sentences were grammatically incorrect, and the remaining 46% sentences were correct. Another group, of 24 students who did not participate in the ERP experiment, was asked to evaluate the plausibility and grammaticality of the test sentences. This was to ensure that the differences between the conditions were due to the S–V agreement and the number of the local noun, and not due to the plausibility of the sentence. The participants were asked to evaluate the sentences on a scale of 1 (definitely implausible or ungrammatical) to 5 (perfectly plausible or grammatical). They were explicitly instructed to judge the “plausibility” and “grammaticality” of sentences, and were given examples for what is plausible and what is implausible (no example for what is grammatical or ungrammatical, as that was easily understood).

For plausibility, the difference across the four sentence types was not significant when analyzed with repeated-measures ANOVA by subject,  $F_1(3, 69) = 2.43, p > .05$ , although it reached significance by item,  $F_2(3, 501) = 4.57, p < .01$ . Considering that the number of sentences was large and the mean values of plausibility were similar (see Table 1), we deemed the plausibility of different sentence types compatible. For grammaticality, the four types differed significantly when analyzed by subject and by item,  $F_1(3, 69) = 140.59, p < .001, F_2(3, 501) = 1185.99, p < .001$ . Post hoc analysis with LSD revealed that the grammatical conditions were rated much higher than the ungrammatical conditions. Therefore, the L2

participants were able to detect the S–V agreement violations, and to them the sentences in grammatical conditions differed from those in ungrammatical conditions.

### Procedure

Participants were tested individually in a quiet experimental room. Each participant was seated about 1m away from the display screen. Prior to the experiment, 14 practice trials were given to the participants, and the practice sentences were similar to the test sentences but were not used in the real experiment. Each participant received only one of the stimuli lists, in a pseudo-randomized sequence such that no more than three correct or incorrect sentences appeared in succession. All the sentences were displayed visually word-by-word in the center of the screen. Before each sentence, a fixation was presented for 500 ms followed by a 500 ms blank. Then the first word of the sentence was on for 500 ms followed by a 200 ms blank, and the remaining words were displayed in the same way until the last word ended with a period. A blank screen followed the end of the sentence for 500 ms, and a string of yellow question marks was then presented on the screen for 2s, signaling the participant to make a grammatical judgment. The inter-trial-interval was set to 1000 ms between different sentence trials.

When reading the sentences, the participants were instructed to minimize their blinks and body movements, and to press one of the two buttons to make sentence grammaticality judgment on the presentation of yellow question marks. Participants were prompted to take a short break after 65 trials. The whole test session lasted about 50 minutes. After the experiment, all the subjects completed a language history questionnaire (Li, Sepanski and Zhao, 2006).

### ERP recordings

EEG data were recorded using a 64-channel Quick-cap with Ag/AgCl electrodes. EEG electrodes were placed

according to the extended 10–20 system. The following 21 sites entered our ERP analyses: Fz, Cz, Pz, F3, F5, F7, F4, F6, F8, C3, C5, T7, C4, C6, T8, P3, P5, P7, P4, P6, and P8. All scalp electrodes were referenced to the left mastoid and off-line re-referenced to linked left and right mastoids. In order to control for eye-movement artifacts, the electro-oculogram (EOG) was monitored: The vertical EOG (VEOG) was monitored with two electrodes located above and below the participant's left eye. The horizontal EOG (HEOG) was recorded from electrodes at the outer canthus of each eye. Electrode impedance was kept below 5 kΩ. Two sets of 32-channel amplifier (SYNAMPS, NeuroScan) were used for EOG and EEG recordings. Band pass was 0.05 Hz–100 Hz and the sampling rate was 500 Hz.

### **Data analysis**

#### **Behavioral data**

Accuracy was computed as the percentage of correct responses, for all test sentences. A correct response was a judgment of “acceptable” for the sentences in the two grammatical conditions and judgment of “unacceptable” for sentences in the two ungrammatical conditions. The RT (reaction time) data were computed only for correct responses.

#### **ERP data**

A blink-correction algorithm (SpatialSVD) was applied off-line. Epochs of 1000 ms length were cut out from the continuously recorded data. The epochs started 200 ms before the onset of the critical word (the copula *was* or *were*). After baseline correction and filtering (band pass was 0.05 Hz–30 Hz), trials with potentials greater than 70 μV were rejected. The average of ERPs was computed for each participant and each sentence type.

Although different time windows were used in previous studies to examine different ERP components (Weber-Fox and Neville, 1996; Friederici and Frisch, 2000; Hahne and Friederici, 2001; Hagoort, 2003; Ye et al., 2006), the following windows have been consistently used (see our previous review): 150–300 ms for early anterior negativity (EAN), 300–500 ms for the negativities (N400 or LAN), and 500–700 ms (or even later) for the late positivity (P600). In our study, we will examine all time windows but focus on 300–500 ms and 500–700 ms in our statistical analyses given the demonstrated relevance of these time windows to bilingual processing in previous studies.

We first selected three midline electrodes Fz, Cz, Pz for analysis, since these electrodes are most frequently reported. For the purpose of identifying topographic differences, we defined 6 Regions of Interest (ROIs), with three in each hemisphere. These ROIs were: left-anterior (LA: F3, F5, and F7), right-anterior (RA: F4, F6, and F8), left-central (LC: C3, C5, and T7), right-central (RC: C4,

C6, and T8), left-posterior (LP: P3, P5, and P7), right-posterior (RP: P4, P6, and P8). Therefore, we performed repeated-measures ANOVAs with three within-subject variables for the lateral electrodes: sentence type (G-I vs. G-C, U-C vs. G-C, or U-I vs. G-C), hemisphere (left and right) and region (anterior, central, and posterior). Whenever the interaction between the variable sentence type and a topographic variable reached a significance level of  $<.10$ , subsequent analysis was executed (as done in ERP L2 research, see Hahne and Friederici 2001). The Greenhouse-Geisser correction was applied whenever the assumption of sphericity was violated. Here we report only unadjusted degrees of freedom and adjusted probabilities.

### **Results**

#### **Behavioral data**

The grammaticality judgment accuracy by our L2 learners showed that they had excellent knowledge of the target language grammar and could detect agreement violations (accuracy on test sentences: 88%). Accuracy across the four sentence types showed no statistical difference: 87% (G-C), 86% (G-I), 90% (U-C), 89% (U-I). A repeated-measures ANOVA revealed no main effect of sentence type,  $F(3, 42) = .57, p > .05$ . The average RTs were 608 ms (G-C), 574 ms (G-I), 522 ms (U-C), and 518 ms (U-I). The ANOVA for RTs revealed a main effect of sentence type,  $F(3, 42) = 4.18, p < .05$ . Post hoc analysis LSD showed that the RTs in the two ungrammatical conditions (U-C and U-I) were significantly faster than the baseline grammatical sentences where there is local congruency (G-C). No significant difference was found for other comparisons.

These behavioral data helped to establish that our proficient L2 learners of English can detect S–V agreement violations, as shown in both their high level of accuracy during grammaticality judgment and their fast reaction times in response to ungrammatical sentences that involved S–V agreement violations. However, given that the participants were instructed to press the button to make a judgment after some time has elapsed since reading the sentence, the RT measure may not be a good index of their on-line processing of S–V agreement. It would therefore be important for us to examine their ERP responses to the various sentence types during on-line processing.

#### **ERP data from L2 learners**

On average, only a few segments were affected by artifacts and there were 95% ( $SD = 8\%$ ) artifact-free segments in the G-C sentence type, 96% ( $SD = 7\%$ ) in G-I, 95% ( $SD = 7\%$ ) in U-C, and 93% ( $SD = 11\%$ ) in U-I. Figures 1–3 present the grand average ERPs in the four conditions from 200 ms before the onset of the copula (*was*,

Table 2. ANOVA results for the time window 300–500 ms for midline and lateral electrodes.

| Sources                   | G-C vs. G-I       |                  | G-C vs. U-C       |                  | G-C vs. U-I       |                  |
|---------------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|
|                           | F value           | p value          | F value           | p value          | F value           | p value          |
| <i>Midline electrodes</i> |                   |                  |                   |                  |                   |                  |
| sentence type             | $F(1, 14) = 8.49$ | .01              | $F(1, 14) = .03$  | .86              | $F(1, 14) = .28$  | .60              |
| type × electrodes         | $F(2, 28) = .36$  | .59 <sup>#</sup> | $F(2, 28) = .70$  | .51              | $F(2, 28) = 5.95$ | .02 <sup>#</sup> |
| <i>Lateral electrodes</i> |                   |                  |                   |                  |                   |                  |
| sentence type             | $F(1, 14) = 6.19$ | .03              | $F(1, 14) = .56$  | .47              | $F(1, 14) = .80$  | .39              |
| type × hemi               | $F(1, 14) = 2.68$ | .12              | $F(1, 14) = .00$  | .96              | $F(1, 14) = 1.00$ | .34              |
| type × region             | $F(2, 28) = .44$  | .55 <sup>#</sup> | $F(2, 28) = .59$  | .49 <sup>#</sup> | $F(2, 28) = 5.72$ | .02 <sup>#</sup> |
| type × hemi × region      | $F(2, 28) = .93$  | .41              | $F(2, 28) = 1.29$ | .28 <sup>#</sup> | $F(2, 28) = .54$  | .54 <sup>#</sup> |

Note: hemi = hemisphere; <sup>#</sup> = adjusted value (see text)

or were) up to 800 ms on a subset of 9 electrodes. Negative amplitudes are plotted above the horizontal midline. As shown in these figures, the ERPs in the four conditions are very similar during the first 300 ms after the onset of the copula with a clear negative component (N100) and a positive component (P200). This pattern is consistent with previous studies that presented language stimuli visually (see Kutas and Van Petten, 1994, for a review). After 300 ms, ERPs of the grammatical incongruent (G-I) sentence type (e.g., *The price of the cars was too high*) showed a negative deflection followed by a positive wave compared to the baseline grammatical congruent (G-C) sentences (e.g., *The price of the car was too high*). ERPs of the two ungrammatical sentence types showed no difference from those for G-C, until a negative component occurred at about 600 ms.

#### Time window 300–500 ms

Statistical analyses for the first time window between 300 ms and 500 ms for midline and lateral electrodes are shown in Table 2. For midline electrodes, the difference between G-C and G-I sentences was significant,  $F(1, 14) = 8.49$ ,  $p < .05$ . Comparison of single electrodes revealed that G-I sentences showed a significant negative deviation from G-C sentences at Cz:  $F(1, 14) = 7.45$ ,  $p < .05$  and Pz:  $F(1, 14) = 8.83$ ,  $p < .01$ , but only a marginally significant difference at Fz:  $F(1, 14) = 3.71$ ,  $p = .08$ . There was no pronounced difference between G-C and U-C, or between G-C and U-I sentences. For lateral electrodes, the comparison between G-C and G-I yielded the main effect of sentence type,  $F(1, 14) = 6.19$ ,  $p < .05$ , and the difference was significant for all the six ROIs. The comparison between G-C and U-C led to no statistically significant difference. The comparison between G-C and U-I sentences showed a slight negative deflection in the left-posterior region,  $F(1, 14) = 4.24$ ,  $p = .06$ , and the right-posterior region,  $F(1, 14) = 6.08$ ,  $p < .05$ .

Both visual inspection of ERP patterns in Figure 1 and statistical analyses suggest that the negativity of the grammatical incongruent condition (G-I) is widely distributed and robust. There was no interaction among sentence type, hemisphere, and region, and therefore this negativity is unlikely to be the so-called left anterior negativity (LAN), which is usually more focused to the left anterior region or bilaterally to the frontal regions (Hagoort, 2003). We consider this negativity the N400, which indexes a semantic or conceptual incompatibility (Kutas and Hillyard, 1980; Hagoort et al., 2004).

#### Time window 500–700 ms

Statistical analyses for the time window between 500 ms and 700 ms for the midline and lateral electrodes are shown in Table 3 and Table 4. For midline electrodes, the comparisons between G-C and G-I and between G-C and U-C both yielded main effects of sentence type,  $F(1, 14) = 6.42$ ,  $p < .05$  and  $F(1, 14) = 7.46$ ,  $p < .05$ , respectively. There were also significant interactions between sentence type and electrode for the comparisons between the ungrammatical (U-C or U-I) and the grammatical (G-C) sentences. For the lateral electrodes, the comparison between the G-C and G-I, and between the G-C and U-C sentences all yielded significant main effects as well as interactions between sentence type and hemisphere (all  $p < .05$ ), but the comparison between G-C and U-I sentences failed to reach significance ( $p > .05$ ). ROI analyses (see Table 4) indicated that in the left hemisphere, the grammatical incongruent (G-I) sentences showed significant positive deviation from the congruent G-C sentences:  $F(1, 14) = 6.30$ ,  $p < .05$  in the left-anterior region,  $F(1, 14) = 12.80$ ,  $p < .01$  in the left-central region, and  $F(1, 14) = 8.55$ ,  $p < .05$  in the left-posterior region. No such differences were observed

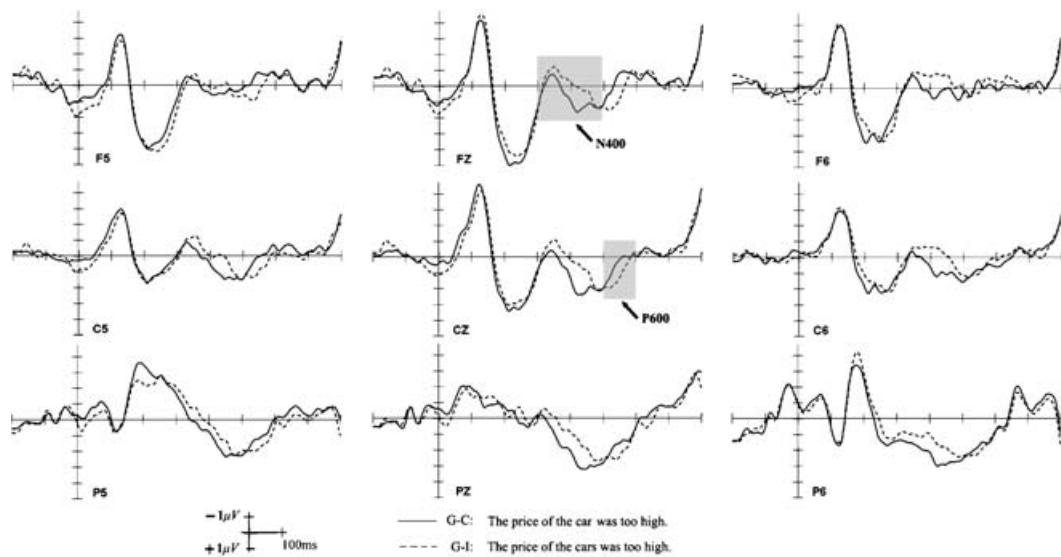


Figure 1. Averaged ERPs for the critical word (*was/were*) in G-C and G-I sentences.

Table 3. ANOVA results for the time window 500–700 ms for midline and lateral electrodes.

| Sources                   | G-C vs. G-I       |                  | G-C vs. U-C        |                  | G-C vs. U-I        |                  |
|---------------------------|-------------------|------------------|--------------------|------------------|--------------------|------------------|
|                           | F value           | p value          | F value            | p value          | F value            | p value          |
| <i>Midline electrodes</i> |                   |                  |                    |                  |                    |                  |
| sentence type             | $F(1, 14) = 6.42$ | .02              | $F(1, 14) = 7.46$  | .02              | $F(1, 14) = .29$   | .60              |
| type × electrodes         | $F(2, 28) = 0.10$ | .90              | $F(2, 28) = 5.56$  | .01              | $F(2, 28) = 11.02$ | .00              |
| <i>Lateral electrodes</i> |                   |                  |                    |                  |                    |                  |
| sentence type             | $F(1, 14) = 7.95$ | .01              | $F(1, 14) = 16.68$ | .00              | $F(1, 14) = 3.25$  | .09              |
| type × hemi               | $F(1, 14) = 9.37$ | .01              | $F(1, 14) = 21.06$ | .00              | $F(1, 14) = 3.89$  | .07              |
| type × region             | $F(2, 28) = .54$  | .59 <sup>#</sup> | $F(2, 28) = 7.98$  | .01 <sup>#</sup> | $F(2, 28) = 6.07$  | .02 <sup>#</sup> |
| type × hemi × region      | $F(2, 28) = 2.97$ | .07              | $F(2, 28) = .99$   | .39              | $F(2, 28) = 2.00$  | .28              |

Note: hemi = hemisphere; <sup>#</sup> = adjusted value

Table 4. Effects of sentence types in the region of interests (ROI) analyses.

| ROIs            | 300–500 ms        |                   |             | 500–700 ms  |                   |             |
|-----------------|-------------------|-------------------|-------------|-------------|-------------------|-------------|
|                 | G-C vs. G-I       | G-C vs. U-C       | G-C vs. U-I | G-C vs. G-I | G-C vs. U-C       | G-C vs. U-I |
| Fz              | 3.71 <sup>Δ</sup> |                   |             | 4.44*       | 6.90*             | 6.20*       |
| Cz              | 7.45*             |                   |             |             | 12.80**           |             |
| Pz              | 8.83**            |                   |             | 5.39*       |                   | 8.05*       |
| left-anterior   |                   |                   |             | 6.30*       | 10.55**           | 7.04*       |
| right-anterior  | 3.13 <sup>Δ</sup> |                   |             |             | 19.61**           | 4.52*       |
| left-central    | 4.56 <sup>Δ</sup> |                   |             | 12.80**     | 7.55*             |             |
| right-central   | 8.44*             |                   |             |             | 27.76**           | 5.53*       |
| left-posterior  | 3.48 <sup>Δ</sup> | 4.24 <sup>Δ</sup> |             | 8.55*       |                   |             |
| right-posterior | 7.89*             |                   | 6.08*       |             | 3.70 <sup>Δ</sup> |             |

Note: Empty cells represent non-significant differences between sentence types.

<sup>Δ</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$

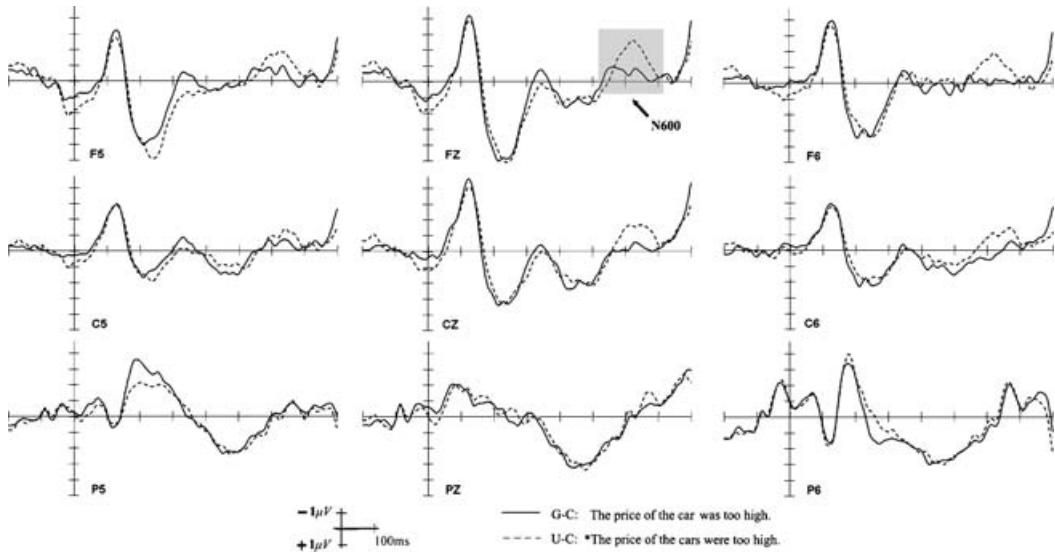


Figure 2. Averaged ERPs for the critical word (*was/were*) in G-C and U-C sentences.

in the right hemisphere. In addition, the G-C and U-I sentences differed in the two anterior regions but not the posterior regions, accounting for the significant interaction between sentence type and ERP region distribution,  $F(2, 28) = 6.07, p < .05$ .

In short, Figure 1 shows a clear positive shift for the grammatical incongruent (G-I) sentences in the left hemisphere as well as in the midline electrodes (but not the right hemisphere), which may be related to the P600. However, the two ungrammatical sentence types (U-C and U-I) had no such shift compared with the baseline grammatical G-C sentences, as indicated in Figures 2 and 3. Instead, the U-C and U-I sentences showed an unexpected significant negative shift that we label provisionally as the

“N600” in the anterior-central regions (though not in the posterior regions).

#### *ERP data from native controls*

Figure 4 presents the average ERP results from the fifteen native English speakers. These ERP patterns show a very different picture from those with the L2 learners. In contrast to the L2 learners, native English speakers showed a clear P600 response to the two ungrammatical sentence types in which S-V agreement was violated, especially toward the posterior sites. There was also an early negativity in the 300–500 ms time window for the ungrammatical sentences, which may be related to LAN,

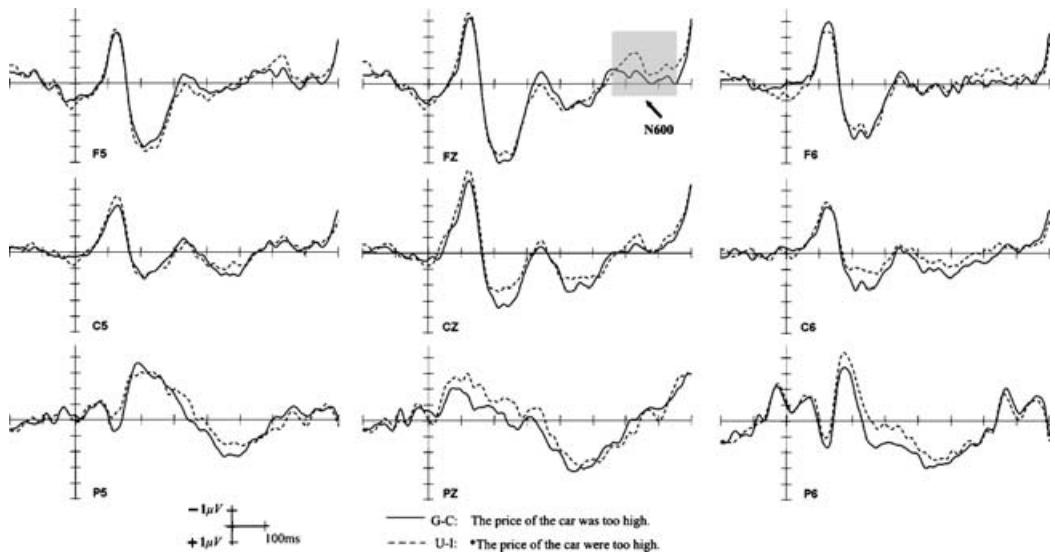


Figure 3. Averaged ERPs for the critical word (*was/were*) in G-C and U-I sentences.

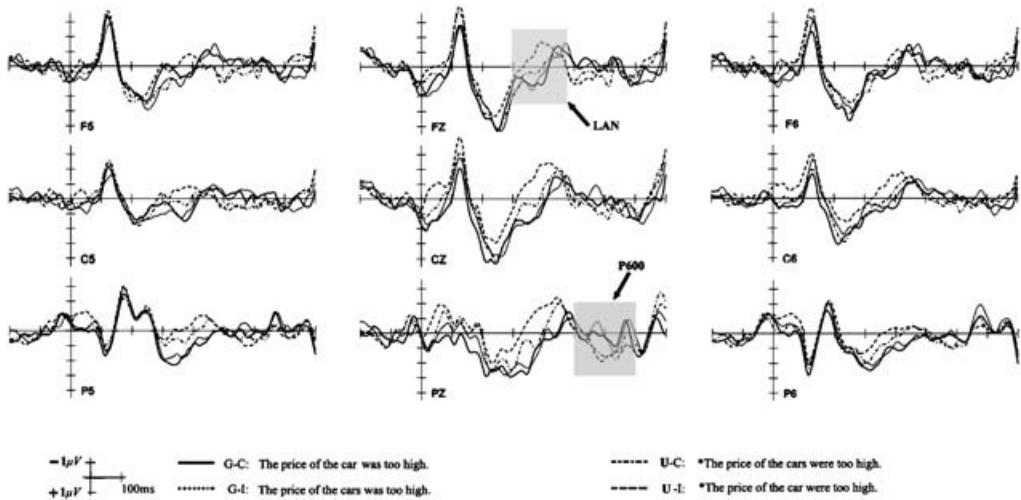


Figure 4. Averaged ERPs for the critical word (*was/were*) in all sentence types from native English speakers.

a component that is believed to be associated with a first-pass, perhaps automatic, morphosyntactic analysis (see Hahne and Friederici, 2001; Osterhout and Mobley, 1995).

#### Time window 300–500 ms

Comparing the two grammatical (G-C and G-I) sentence types in the 300–500 ms time window, we found no significant difference in the native speakers' results. This differed from the patterns with the L2 learners, where significant differences were found between G-C and G-I sentences for both the midline and lateral electrodes. By contrast, comparing the grammatical G-C and ungrammatical U-I sentences, we found a significant main effect of sentence type on midline electrodes,  $F(1, 14) = 13.77, p < .01$ , and a significant main effect on lateral electrodes,  $F(1, 14) = 15.23, p < .01$ . However, comparison of the grammatical G-C and ungrammatical U-C sentences yielded no significant main effects (midline electrodes:  $F(1, 14) = 1.19, p > .05$ ; lateral electrodes:  $F(1, 14) = 2.19, p > .05$ ). This non-significant difference might be due to the "attraction errors" inherent in the U-C sentences (e.g., *The price of the cars were too high*), where the intervening local noun differs from the head noun in number but agrees with the verb, and the adjacency between the preverbal local noun and the verb easily attracts agreement errors in speech production or comprehension (Vigliocco and Nicol, 1998; Franck, Vigliocco and Nicol, 2002). Thus, on a first-pass early analysis, our native participants may not be sensitive to the disagreement errors in the U-C sentences (but they showed significant differences from L2 learners in later analysis; see discussion below).

#### Time window 500–700 ms

Comparisons between the G-C and G-I sentences in this time window showed no significant difference on both

the midline and lateral electrodes. This pattern contrasted with that of the L2 learners, where a P600-like effect was observed for the grammatical incongruent G-I sentences (see Figure 1 above). For the comparison between G-C and U-C sentences, however, there was a significant main effect of sentence type on midline electrodes,  $F(1, 14) = 9.03, p < .01$ , and on the lateral electrodes,  $F(1, 14) = 7.21, p < .05$ . For the comparison between G-C and U-I sentences, there was a significant main effect of sentence type on midline electrodes,  $F(1, 14) = 6.46, p < .05$ , and a significant main effect on lateral electrodes,  $F(1, 14) = 4.77, p < .05$ . Thus, each of the ungrammatical sentence types differed in positivity from the grammatical baseline condition for native speakers within this time window, and the differences appeared robust on both the midline and lateral electrodes.

To further identify the differences between the ERP patterns of native speakers and those of L2 learners, we computed the difference waveforms for the two groups of participants between each of the three sentence types and the grammatical congruent sentences (the baseline condition). Figure 5 presents these difference waveforms based on the Fz and Pz electrodes, since robust differences between the two groups of participants were found on these sites. Table 5 also presents the group statistical analyses. Three key differences can be observed here: (1) For the L2 learners, the two grammatical sentence types elicited major ERP contrasts: G-I sentences showed a pronounced N400 effect, followed by a P600 component; for the native speakers, the two grammatical conditions showed no significant difference. Across groups, no significant difference was observed for the negativities during 300–500 ms on both midline and lateral electrodes, but the positivity during 500–700 ms was significant on lateral electrodes,  $F(1, 28) = 4.98, p < .05$ . (2) For the L2 learners, ungrammatical sentences

Table 5. ANOVA results for ERP effects across the two groups (native speakers vs. L2 learners).

| Sources                   | df   | G-C vs. G-I |                   | G-C vs. U-C |            | G-C vs. U-I       |            |
|---------------------------|------|-------------|-------------------|-------------|------------|-------------------|------------|
|                           |      | 300–500 ms  | 500–700 ms        | 300–500 ms  | 500–700 ms | 300–500 ms        | 500–700 ms |
| <i>Midline electrodes</i> |      |             |                   |             |            |                   |            |
| Fz (group)                | 1,28 |             |                   |             | 16.47**    | 3.40 <sup>Δ</sup> | 11.77**    |
| Pz (group)                | 1,28 |             | 3.97 <sup>Δ</sup> |             | 9.89**     | 3.57 <sup>Δ</sup> |            |
| <i>Lateral electrodes</i> |      |             |                   |             |            |                   |            |
| group                     | 1,28 |             | 4.98*             |             | 21.19**    |                   | 7.90**     |
| group × hemi              | 1,28 |             | 3.64 <sup>Δ</sup> |             |            |                   |            |
| group × region            | 2,56 |             |                   |             | 6.50**     |                   | 4.66*      |
| group × hemi × region     | 2,56 |             |                   |             |            |                   |            |

Note: Empty cells represent non-significant differences between sentence types.

hemi = hemisphere; <sup>Δ</sup> $p < .10$ , \* $p < .05$ , \*\* $p < .01$

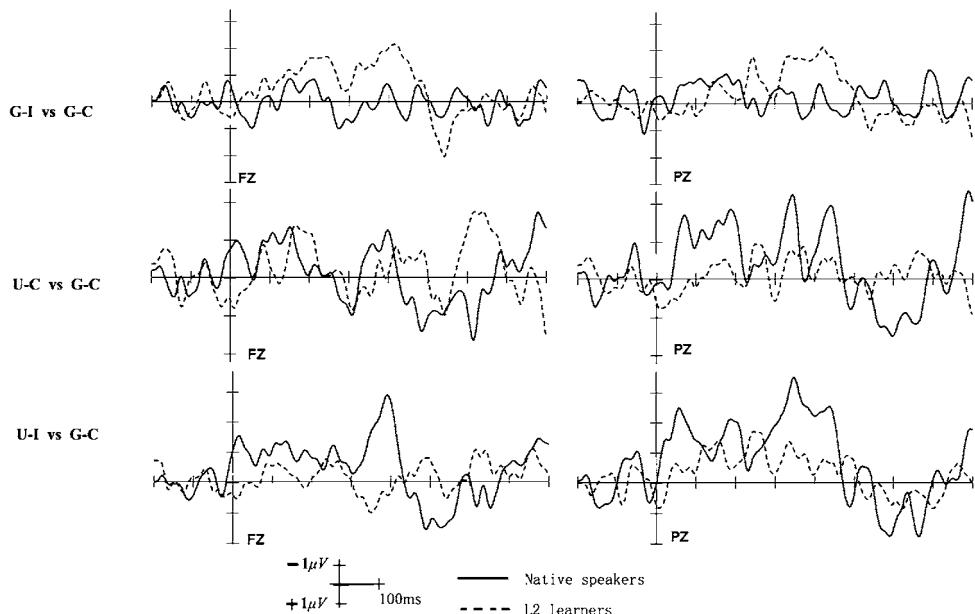


Figure 5. Difference waveforms: Native speakers vs. L2 learners.

(i.e., sentences that contain S–V violations) showed no anterior negativity as compared with the grammatical baseline, but the negativity occurred later on (500–700 ms); for the native speakers, LAN was observed with both ungrammatical sentence types (U-C and U-I) but only the difference between U-I and the baseline G-C reached statistical significance. (3) For the L2 learners, there was no P600 effect associated with the processing of the S–V (dis)agreement in the ungrammatical sentences, whereas for the native speakers, this ERP component was transparent for both the U-C and U-I sentences when compared with the grammatical G-C sentences. Across groups, significant differences were found on both the

midline and lateral electrodes for the time window 500–700 ms. In short, for L2 learners the most transparent difference was observed between the two grammatical sentence types whereas for native speakers, the most transparent difference was between the ungrammatical and the grammatical sentences.

### General discussion

Our study is the first systematic attempt to examine ERP signatures of the processing of subject–verb agreement in the bilingual context. Our results indicate that proficient L2 learners, who might be able to detect morphosyntactic

violations in behavioral measures, nevertheless show distinct neural responses from native speakers in their processing of morphosyntactic features that are absent in their L1. The large typological differences between Chinese and English in grammar and syntax have allowed us to examine an interesting bilingual population on a theoretically significant and practically important topic. Given an L1 that does not encode grammatical morphology, the learning of a syntactic agreement system in an L2 presents a major obstacle to Chinese learners of English as a second language. The general findings here are consistent with those from previous studies of morphosyntactic processing by French–English and Spanish–English bilinguals (Osterhout et al., 2004; Tokowicz and MacWhinney, 2005; see also Chan et al., 2007 for distinct fMRI responses in the bilingual's processing of nouns and verbs in Chinese and English).

Although our study does not examine the S–V agreement errors actually made by L2 learners, our ERP results reveal important differences between L2 learners and native speakers in the processing of S–V agreement. It is particularly interesting to note that our L2 learners can make accurate grammatical judgments to the ungrammatical sentences that contain S–V agreement violations. In other words, they can detect the violations in comprehension, at least immediately after reading the sentence. However, their brain responses show distinct patterns from those of native speakers.

For native speakers, an early left anterior negativity is evoked in response to an S–V agreement violation, and according to some hypotheses (e.g., Hahne and Friederici, 1998, 1999, 2001), this anterior negativity represents a first-pass, automatic morphosyntactic analysis. In addition, native speakers also show a late positivity in response to the S–V agreement violation, which may represent a more fine-grained syntactic analysis/reanalysis or integration. Such a biphasic LAN–P600 syntactic processing profile found with our native speakers is highly consistent with results from previous studies in syntactic processing (e.g., Osterhout and Mobley, 1995; Weber-Fox and Neville, 1996; Hahne and Friederici, 1999), and it attests to the general validity of our methodology and the reliability of our native speakers as the control population.<sup>1</sup>

For the L2 learners, however, such a biphasic syntactic processing profile is absent in the ERP responses to ungrammatical sentences that contain S–V agreement violations. Instead, they show a pattern of negativity in the 500–700 ms time window. The nature of this negativity

is currently unknown, as no previous ERP studies have reported on this negativity as indexing aspects of sentence processing. The fact that this negativity occurs relatively late in the time course, and distinct from a positivity that is associated with syntactic processing, we suspect, is evidence that processing of S–V agreement by non-native speakers is sufficiently different from that by native speakers.

A different kind of biphasic processing is observed with the L2 learners in our study. The learners show an N400 followed by a P600 when processing grammatical sentences that contain an incongruity in number between the two nouns and the verb (e.g., *The price of the cars was too high*), as compared with the grammatical congruent sentence (e.g., *The price of the car was too high*). The N400 pattern might reflect our L2 learners' early analysis of the semantic or conceptual incompatibility of the two nouns (e.g., perhaps thinking that one price cannot apply to many cars; hence N400). The P600 probably reflects the L2 learners' focus on the local grammatical incongruence between the local noun and the verb. In any case, this biphasic pattern is not the same as the early-automatic, late-integrative syntactic analysis by native speakers. This limited scope semantic-syntactic analysis also explains why our L2 learners do not focus on the global grammatical or ungrammatical correspondence between the subject noun and the verb, and therefore do not show sensitivity to ungrammatical sentences (or show delayed sensitivity if they do – assuming that the N600 is an index of this sensitivity in some way).

How do we reconcile these findings with recent evidence from longitudinal training studies that show that L2 learners can quickly pick up non-native syntactic and semantic features with minimal instructions (Osterhout et al., 2004, 2006)? First, it is clear that cross-linguistic similarities or overlap in the bilingual's two languages play an important role here (see Hernandez and Li 2007 for a discussion of the role of language overlap in contrasting neural patterns). As Osterhout and colleagues pointed out, bilingual learners show native-like ERP patterns only for the L2 structures that are also somehow realized in the L1 (e.g., S–V person agreement in English and French). The S–V agreement feature exists in English but is totally absent in Chinese, and therefore it is not surprising that our L2 learners show very different processing patterns here. Second, there might also be a syntactic versus semantic difference with respect to how fast L2 learners can pick up the non-native patterns (see also Hernandez and Li 2007 for a discussion regarding this difference). McLaughlin, Osterhout and Kim (2004) showed that an N400 semantic effect was present after only 14 hours of L2 instruction, whereas a P600 syntactic effect requires about eight months to show up (Osterhout et al., 2004). S–V agreement might be the most difficult

<sup>1</sup> One could question whether the native speakers' knowledge of Chinese might have affected their responses in English. It appears that this was not a problem in our study, given that their responses were generally consistent with those from previous studies of native speakers.

and the last morphosyntactic obstacle to the Chinese L2 learners of English. Finally, as mentioned earlier, it is yet unclear whether the early native-like patterns reported by Osterhout and colleagues will eventually end up the same as native patterns, across the board for a variety of syntactic and semantic structures, especially for structures that are absent in the L1 (such as the S-V agreement studied here).

Before we conclude, it would be useful to point out several important directions for future research. First, as in all domains of language acquisition, there is huge individual variation with respect to the outcome of learning. Preliminary analyses of our ERP data show that such individual differences can be observed even with our limited number of participants. In future studies, we plan to examine a larger number of L2 learners to identify such individual differences and, more importantly, mechanisms underlying the differences (e.g., working memory, selective attention, etc.). Second, although we have studied only late proficient L2 learners, all of these learners reside in a non-L2 country (i.e., China). It would be interesting to see if proficient L2 speakers in the target language country show different processing patterns (assuming that immersion in the language environment has a more direct impact on learning outcomes). Third, although we have included the “attraction error” sentences in our study to manipulate the congruency between the local noun and the verb, we have not systematically compared the local congruency with global congruency (head noun and verb) for L1 and L2 speakers, due the fact that all of our head nouns are singular. In future studies we will examine the processing of sentences that vary in consistency both locally and globally by including sentences that have plural head nouns (and other verb types, in addition to copulas). One could hypothesize that native speakers are more sensitive to global congruency than are L2 learners. Finally, it would be important to correlate, in a more detailed study, the relationship between age of acquisition, levels of proficiency, and patterns of processing. Because all of our participants in this study are late learners, we do not know what neural patterns early learners will show in L2 morphosyntactic processing.

In sum, ERPs provide us with useful and powerful indices of grammatical, semantic, and syntactic processing in the bilingual context, as many previous studies have demonstrated. Our study is specifically designed to investigate the neural response patterns between native speakers and L2 learners in morphosyntactic processing, confirming our general hypothesis that there are significant differences between the two populations. Findings from our study are consistent with the general proposal that language-specific experiences help to shape the functional and neural structures of the brain (Bates, 1999; Nelson, 1999; Jeong et al., 2007 – this volume; Zhang

and Wang, 2007 – this volume). In particular, previous psycholinguistic studies have shown that language-specific properties of Chinese can affect patterns of representation, on-line processing, and acquisition in this language (see Li et al., 1993; Li, 1996; and chapters in Li et al., 2006). There is also mounting evidence that the neural systems of reading and speaking are modulated by the specific linguistic experiences with the written and spoken properties of Chinese (e.g., Gaudour 2006; Li et al., 2004; Siok et al., 2004). Clearly, our late L2 learners have already formed a consolidated linguistic representation in Chinese by the time they start to learn English, and their early learning of Chinese must have had a cascading effect on their later acquisition of English as a second language (Hernandez and Li, 2007; Hernandez, Li and MacWhinney, 2005).

#### **Appendix: Sample sentences used in the experiment**

##### **(a) Grammatical, congruent (G-C)**

- The price of the car was too high.
- The report from the agency was very encouraging.
- The violinist with the singer was an excellent performer.
- The vase with the sunflower was bought in France.
- The actor in the advertisement was good at singing too.
- The book by the author was sold well.
- The mother of the boy was cooking the dinner.
- The table for the guest was placed in another room.
- The idea of the psychologist was widely supported.
- The candy from the stranger was really sweet.

##### **(b) Grammatical, incongruent (G-I)**

- The price of the cars was too high.
- The report from the agencies was very encouraging.
- The violinist with the singers was an excellent performer.
- The vase with the sunflowers was bought in France.
- The actor in the advertisements was good at singing too.
- The book by the authors was sold well.
- The mother of the boys was cooking the dinner.
- The table for the guests was placed in another room.
- The idea of the psychologists was widely supported.
- The candy from the strangers was really sweet.

##### **(c) Ungrammatical, congruent (U-C)**

- \*The price of the cars were too high.
- \*The report from the agencies were very encouraging.
- \*The violinist with the singers were an excellent performer.
- \*The vase with the sunflowers were bought in France.
- \*The actor in the advertisements were good at singing too.

- \*The book by the authors were sold well.
- \*The mother of the boys were cooking the dinner.
- \*The table for the guests were placed in another room.
- \*The idea of the psychologists were widely supported.
- \*The candy from the strangers were really sweet.

(d) *Ungrammatical, incongruent (U-I)*

- \*The price of the car were too high.
- \*The report from the agency were very encouraging.
- \*The violinist with the singer were an excellent performer.
- \*The vase with the sunflower were bought in France.
- \*The actor in the advertisement were good at singing too.
- \*The book by the author were sold well.
- \*The mother of the boy were cooking the dinner.
- \*The table for the guest were placed in another room.
- \*The idea of the psychologist were widely supported.
- \*The candy from the stranger were really sweet.

## References

- Bates, E. (1999). Plasticity, localization and language development. In S. H. Broman & J. M. Fletcher (eds.), *The changing nervous system: Neurobehavioral consequences of early brain disorders*, pp. 214–253. New York: Oxford University Press.
- Bock, K. & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, 23, 45–97.
- Chan, A., Luke, K., Li, P., Li, G., Yip, V., Weekes, B. S. & Tan, L. (2007). Neural correlates of nouns and verbs in early bilinguals. *Annals of the New York Academy of Sciences* (in press).
- Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language*, 36, 147–164.
- Franck, J., Vigliocco, G. & Nicol, J. L. (2002). Subject–verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, 17, 371–404.
- Friederici, A. D. & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language*, 43, 476–507.
- Gandour, J. T. (2006). Brain mapping of Chinese speech prosody. In Li et al. (eds.), pp. 308–319.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15, 883–899.
- Hagoort, P., Hald, L., Bastiaansen, M. & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304, 438–441.
- Hahne, A. & Friederici, A. D. (1998). ERP-evidence for autonomous first-pass parsing processes in auditory language comprehension. *Journal of Cognitive Neuroscience, Supplement*, 125.
- Hahne, A. & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis: early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, 11, 193–204.
- Hahne, A. & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–141.
- Hahne, A., Mueller, J. & Clahsen, H. (2006). Morphological processing in a second language: Behavioural and event-related potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, 18, 121–134.
- Hernandez, A. & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, 133 (4), 1–13.
- Hernandez, A., Li, P. & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, 9, 220–225.
- Jeong, H., Sugiura, M., Sassa, Y., Yokoyama, S., Horie, K., Sato, S., Taira, M. & Kawashima, R. (2007). Cross-linguistic influence on brain activation during second language processing: An fMRI study. *Bilingualism: Language and Cognition*, 10 (2), 175–187.
- Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.
- Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–205.
- Kutas, M. & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (ed.), *Handbook of psycholinguistics*, pp. 83–143. San Diego, CA: Academic Press.
- Li, P. (1996). The temporal structure of spoken sentence comprehension in Chinese. *Perception and Psychophysics*, 58, 571–586.
- Li, P., Bates, E. & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language*, 32, 169–192.
- Li, P., Jin, Z. & Tan, L. (2004). Neural representations of nouns and verbs in Chinese: An fMRI study. *NeuroImage*, 21, 1533–1541.
- Li, P., Sepanski, S. & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods*, 38, 202–210.
- Li, P., Tan, L.-H., Bates, E. & Tzeng, O. J. L. (eds.) (2006). *The handbook of East Asian psycholinguistics*, vol. 1: Chinese. Cambridge: Cambridge University Press.
- Liu, H., Bates, E. & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied Psycholinguistics*, 13, 451–484.
- McLaughlin, J., Osterhout, L. & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, 7, 703–704.
- Nelson, C. A. (1999). Change and continuity in neurobehavioral development: Lessons from the study of neurobiology and

- neural plasticity. *Infant Behavior and Development*, 22, 415–429.
- Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R. & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*, pp. 271–308. New York: Psychology Press.
- Osterhout, L., McLaughlin, J., Pitkanen, I., Frenck-Mestre, C. & Molinaro, N. (2006). Novice learners, longitudinal designs, and event-related potentials: A means for exploring the neurocognition of second language processing. *Language Learning*, 56, 199–203.
- Osterhout, L. & Mobley, L. A. (1995). Event-related potentials elicited by failure to agree. *Journal of Memory and language*, 34, 739–773.
- Pearlmutter, N. J., Garnsey, S. M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and language*, 41, 427–456.
- Rispens, J. E., Been, P. H. & Zwarts, F. (2006). Brain responses to subject–verb agreement violations in spoken language in developmental dyslexia: An ERP study. *Dyslexia*, 12, 134–149.
- Roehm, D., Bornkessel, I., Haider, H. & Schlesewsky, M. (2005). When case meets agreement: Event-related potential effects for morphology-based conflict resolution in human language comprehension. *NeuroReport*, 16, 875–878.
- Rossi, S., Gugler, M. F., Hanhe, A. & Friederici, A. D. (2005). When word category information encounters the morphosyntax: An ERP study. *Neuroscience Letters*, 384, 228–233.
- Siok, W. T., Perfetti, C. A., Jin, Z. & Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature*, 431, 71–76.
- Tokowicz, N. & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, 27, 173–204.
- Vigliocco, G. & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity concord syntactic or linear? *Cognition*, 68, B13–B29.
- Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A. & Perani, D. (2003). Early setting of grammatical processing in bilingual brain. *Neuron*, 37, 159–170.
- Wassenaar, M., Brown, C. M. & Hagoort, P. (2004). ERP effects of subject–verb agreement violations in patients with Broca's aphasia. *Journal of Cognitive Neuroscience*, 14, 553–576.
- Web-Fox, C. M. & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256.
- 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.
- Zhang, Y. & Wang, Y. (2007). Neural plasticity in speech acquisition and learning. *Bilingualism: Language and Cognition*, 10 (2), 147–160.

**Received 3 November 2006****Revision received 17 February 2007****Accepted 23 February 2007**