Journal of Neurolinguistics 23 (2010) 615–630

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

Structure and meaning in Chinese: An ERP study of idioms

Youyi Liu [a,*], Ping Li [b], Hua Shu [a], Qirui Zhang [a], Lang Chen [c]

[a] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China
[b] Department of Psychology, Pennsylvania State University, PA, USA
[c] Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

ARTICLE INFO

Article history:
Received 28 December 2009
Received in revised form 20 May 2010
Accepted 1 June 2010

Keywords:
ERP
Language processing
Idioms
Semantics
Syntax
Chinese

ABSTRACT

Recent electrophysiological evidence suggests that the analysis of structure and the processing of meaning may differ across languages and across types of materials being processed. In this study we used event-related potentials (ERPs) to examine the interplay between structural analysis and meaning processing in Chinese idioms. Our results revealed that N400 effects reflect graded semantic distances in our experimental conditions involving synonyms, semantic violations, and combined semantic and syntactic violations. The P600 effects were uniform across these experimental conditions. There was no difference between the semantic only and the combined violation conditions with regard to either the N400 or the P600 component. These patterns suggest that in Chinese, unlike in other languages, meaning integration does not depend on the intactness of structural information. They also suggest, consistently with some previous studies, that P600 is not only an index of syntactic processes but may be a more general index of the processing of linguistic or perceptual well-formedness in structure in highly constraining context.

1. Introduction

The distinction between structure and meaning is fundamental to many linguistic theories. Chomsky (1957) argued most forcibly with his famous example "colorless green ideas sleep furiously", demonstrating that it is possible to construct a syntactically correct but semantically vacuous sentence (and consequently native speakers can judge the grammaticality of a sentence but may not interpret its

* Corresponding author.
E-mail address: psykylin@gmail.com (Y. Liu).

0911-6044/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jneuroling.2010.06.001

meaning). But whether structure and meaning are indeed separate, in representation and in processing, has been subject to much debate in linguistics, psycholinguistics, and neurolinguistics, especially in the last two decades (Frazier & Rayner, 1982; Friederici, 1995; Hagoort, 2003; Kaan, Harris, Gibson, & Holcomb, 2000; MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell & Tanenhaus, 1994; see Friederici & Weissenborn, 2007).

From a linguistic point of view, different languages appear to place emphasis differently on structure versus on meaning. Grammatical morphology is prominent in most Indo-European languages, but is not used in languages like Chinese (and used to a lesser extent in English). Grammatical morphology marks the agreements between the sentence subject, predicate, and object with respect to gender, number, case, and person. When grammatical devices are relatively weak, such as in English, word order can play an important role (Bates & MacWhinney, 1982, 1987). Such typological differences between languages have significant implications for the cognitive representation and neural processing of languages, as has been discussed in many empirical and theoretical studies (see recent syntheses in Li, Tan, Bates, & Tzeng, 2006 and Li & Shu, 2010). In particular, semantic and contextual cues are much more important than grammatical cues for Chinese speakers to determine who does what to whom during sentence comprehension (Li, 1996; Li, Bates, & MacWhinney, 1993).

In recent years, the event-related potentials (ERPs) technique has become a powerful tool in understanding the neural processing of language, largely due to the discovery of several components that reliably index syntactic and semantic on-line processes of language. The N400, a negatively-going brain waveform that peaks around 400 ms after stimulus onset and usually distributes centroparietally, is typically observed with larger amplitude when words cannot be easily integrated with the sentence context. Therefore, N400 is considered as a component sensitive to semantic integration (see Kutas & Federmeier, 2000, for a review; Kutas & Hillyard, 1980). Three other components, early left anterior negativity (ELAN), left anterior negativity (LAN), and late centroparietal positivity (P600) are found to be sensitive to different aspects of syntactic processing in some languages (e.g., German, Dutch and French) and are said to correspond to three functionally separate phases, according to Friederici (2002). Generally speaking, the ELAN, a negative waveform appearing at the early time window of 100–300 ms and normally distributing left-anteriorly on the scalp, is hypothesized to reflect an initial phrase of structure building process based on word category information. The LAN, also a negative component, has been found to be sensitive to morphosyntactic errors. It occurs approximately in the same time window as the semantic N400, but commonly has more anterior and left-lateralized scalp distribution. The P600 is a positively and centroparietally distributed shift with an onset at about 500 ms and is often considered relevant to more controlled syntactic processes, such as syntactic reanalysis or repair (see Friederici, 2002, for a review; Friederici, Pfeifer, & Hahne, 1993).

Just as Chomsky's example demonstrates, syntax in English and other Indo-European languages is relatively independent of semantics. Furthermore, both the neurocognitive model proposed by Friederici (2002) and the extended argument dependency model (eADM) by Bornkessel and Schlesewsky (2006) have assumed that that there is a primacy of syntax over semantics in sentence comprehension. It has been suggested that the primacy of syntax shows up in two respects. First and foremost is that the initial mental constructing of sentence comprehension is based on syntactic information only (e. g., word category), which is indexed by ELAN. Second is the suggestion that failed syntactic processing can block follow-up semantics integration. This suggestion has been based on ERP findings of syntax-related ELAN-P600 effects and the absence of semantics-related N400 effect when language users process sentences containing combined syntactic category and semantic violations, such as in the German version of *The door lock was in-the* ***eaten*** (Friederici, Gunter, Hahne, & Mauth, 2004; Friederici, Steinhauer, & Frisch, 1999; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001; Isel, Hahne, Maess, & Friederici, 2007). The evidence so far from European languages such as German and Dutch suggests that word category information plays a special important role during language comprehension.

Does the primacy of syntax hypothesis apply to other languages (e.g., Chinese) that are typologically distinct from the languages that have been tested in previous ERP studies? As discussed earlier, Chinese has few morphological markers to label word category (e.g., noun, verb, adjective, adverb etc.) or to label the agreement among sentence subject, predicate, and object with respect to tense, gender, number, case, and person. Considering the functional roles of ELAN and LAN as discussed above and the

important typological differences between European languages and Chinese, we can ask questions about whether the ELAN and LAN effects can be observed in Chinese and whether the relative importance of syntax and semantics during language comprehension may be different across languages.

So far only two published studies have investigated ERP effects of combined syntactic category and semantic violation in Chinese. Ye, Luo, Friederici, and Zhou (2006) observed an early-starting (50 ms after stimuli onset), long-lasting, widely-distributed negativity but no P600 for sentences with combined word category violation and semantic violation, such as in the object-deleted incongruous Chinese sentence *Mu4jiang4 xu1yao4 mu4cai2, ba3 **sao3**-le* (The carpenter needs timbers, *ba* [PREP] ***swept***, 'To get timber, the carpenter swept'). However, whether this broad negativity belongs to ELAN or N400 or both is difficult to judge, because it was not clear whether the overall semantic congruity of the semantic only violation and that of the combined syntactic and semantic violation sentences were comparable; the pre-stimulus baselines among the different conditions were also different, although the authors tried to compensate it with a long baseline interval during data analysis (1100 ms, e.g., "noun plus preposition '*ba*'" vs. "preposition '*ba*' plus noun"). In a recent study, Yu and Zhang (2008) used the same "*ba*" phrase structure, but with the substitution instead of the deletion method in constructing experimental materials so that pre-stimulus baselines were the same across conditions. Unlike the results from European language studies and those obtained by Ye et al. (2006), Yu and Zhang observed an N400 effect and a small P600 effect, but no ELAN effect for combined syntactic category and semantic violations, as in the Chinese sentence *Qing1jie2gong1 ba3 da4sha4 de chuang1hu4 quan2bu4 **tang2** le yi1bian4* (The cleaners *ba* [PREP] the building windows all ***sugar*** *le* [perfective marker] once, 'The cleaners sugar all the windows of the building once'), suggesting that semantic integration could proceed even when syntactic category processing fails during Chinese sentence reading. This suggestion raises an interesting possibility that the functional primacy of syntactic category over semantic processing might not be a universal principle cross-linguistically.

In this study, we intend to test this possibility further with idioms in Chinese. Idioms are commonly used in daily Chinese speech and over ten thousands idioms are recorded in the Dictionary of Chinese Idioms (Yu & Sun, 2004), according to which the four-character idioms are the most popular form. Many of these idioms are deeply grounded in the context of a historical event, a legendary story, a play, or a description in a novel.[1] Although they took the form of condensed four characters in the history of the Chinese language, the four characters in idioms tend to preserve the basic syntactic structure of Chinese sentences, such as the predicate-object structure and the subject-predicate structure, both of which are typical language structures in modern Chinese. For example, the idiom 叶公好龙 (ye4-gong1-hao4-long2, literally 'Mr. Ye likes dragons') is composed of three components, namely, the subject 叶公 (*Ye4gong1*, a person named Mr. Ye), the predicate 好 (*hao4*, like) and the object 龙 (*long2*, dragon). In fact, most syntactic structures in modern Chinese can be found in idioms, according to syntactic analyses by Ni and Yao (1990). This property of idioms allows us to construct experimental conditions of semantic violation and word category violation, as has been successfully done in regular sentential context.

Two additional distinct properties make idioms ideally suited for our study. As compared with sentences, idioms have a more compact structure and integrated meaning. The vast majority (more than 97%) of idioms are in the fixed four-character form (corresponding to four morphemes mostly). And the order of the characters/morphemes cannot be changed, nor can any character be substituted without a total change in meaning. Idioms also carry great meaning integrity, going beyond the sum of the literal meanings of the individual components. Zhou, Zhou, and Chen (2004) compared the processing of correct Chinese idioms with that of incorrectly formulated idioms, in which the last character of the correct idiom was substituted with a semantically unrelated character. Zhou et al. (2004)

[1] For example, the idiom 叶公好龙 *ye4 gong1 hao4 long2*, literally 'Mr. Ye likes dragons', for example, comes from a story in the book by Liu Xiang in the Han Dynasty (25–220 AD), which describes a man who claimed to be fond of dragons, with paintings and carvings of dragons all over his house, but was scared to death when a real dragon came into his house. The four-character idiom has since been used to refer to a situation when someone claims to love something but does not really love it or is in fact afraid of it.

observed a significant P600 effect following an N400 effect. Unfortunately, the authors did not report whether the two characters at the end of the correct and incorrect idioms belonged to the same grammatical category or not, making it impossible to discern the nature of N400 and P600 effects in their study. Surprisingly, they concluded that the N400 was related to the integration of both semantic and syntactic analyses while the P600 to reanalysis and integration of word meaning.

Compared with the N400, the specific function of the P600 component is actually more controversial. While earlier studies suggested that the P600 is an independent component specialized for indexing syntactic processing (see Osterhout & Hagoort, 1999 for a review), more recent research argues for alternative views of the P600, such as a monitoring mechanism or a reassignment process during language perception (van Herten, Kolk, & Chwilla, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; see Kuperberg, 2007 for a review), or the general processing of rule-based temporal sequence information such as in music perception (Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). Our study may shed light on the nature of P600, given that there are some similarities between idiom processing and music perception. Nan, Friederici, Shu, and Luo (2009) asked Chinese native speakers to listen to familiar four-syllable Chinese phrases and four-note musical pieces, and found a left-lateralized P600 in tonal violated language phrases and a right-lateralized P600 in musical pitch. Other studies also show language-like P600 with violations on music pitch patterns (see Patel, 2003 for a review).

In the present study, we constructed experimental conditions that contained semantic only violation and combined semantic and grammatical violation by changing the final nouns of idioms. To use the idiom 笑里藏刀 (*xiao4-li3-cang2-dao1*, *hiding daggers behind smiles*) as an example: the last noun 刀 (dao1, *dagger*) after the verb 藏 (cang2, *hiding*) was changed to a semantically unrelated noun 房 (fang2, *room*), which creates a semantic violation condition, or to a semantically unrelated verb 投 (tou2, *throw*), which creates a word category violation as well as a semantic violation condition (thus, a combined violation condition). In order to investigate how meanings of the individual morphemes of the idiom modulate processing, we also added a synonym condition by changing the noun 刀 (dao1, *dagger*) to its synonymous noun 劍 (jian4, *sword*), which turns the idiom into a non-idiom but preserves the overall meaning of the phrase.

Following the logic of the above illustration, we could make several important predictions concerning the role of syntax and semantic during sentence processing. First, considering the dominance of semantics over syntax in Chinese and the findings from previous studies, we predict N400 effects for all the three violation conditions. We also predict that the amplitude of N400 could be modulated by semantics, that is, a graded N400 effect will be observed from congruous to synonym to semantic violation conditions. Previous ERP studies have already shown graded N400 effect in sentence contexts (see Kutas & Federmeier, 2000; Li, Shu, Liu, & Li, 2006). Second, no ELAN effect will be observed on combined word category plus semantic violation condition, given the limited roles of morphosyntactic cues in Chinese and the findings from previous studies (e.g., Yu & Zhang, 2008). Third, given the highly stable structure and the meaning integrity of Chinese idioms, we expect a P600 effect to follow an N400 effect for combined violation conditions. Although no systematic P600 effect has been previously found for word category violations in Chinese, Yu and Zhang (2008) did observe a small P600 following a large N400 when the verb was substituted by a noun in a fixed structure with high predictability on word category (i.e., *BA* + *noun* + *verb* structure). Finally, if we observe P600 effects for both combined violations and for synonym and semantic violations, we could speculate that the P600 component might be a general index of the high structural constraint in linguistic or perceptual systems.

2. Methods

2.1. Participants

Twenty-four undergraduate or graduate students (8 men and 16 women, mean age 22.4 years, range: 19–31) from Beijing Normal University participated in the experiment as paid volunteers. All were native Mandarin Chinese speakers. They reported normal or corrected-to-normal vision and no history of neurological disorder. Written consent was obtained from all participants before the experiment according to established guidelines of the review boards of the State Key Laboratory of

Cognitive Neuroscience and Learning at Beijing Normal University. Two of the participants were excluded from the following analyses due to their low accuracy (below 90%) on behavioral performance, and a third one was excluded due to high rejection rates of artifact EEG trials (33.8%).

2.2. Materials

The experimental stimuli consisted of 160 commonly used four-character Chinese idioms. Three criteria guided our selection of idioms: (a) the third character is a transitive verb; (b) the fourth character is a noun denoting a concrete object or thing; and (c) the noun is the direct object of the verb. The stimuli were divided into four groups (four "idiom types") according to the status of the fourth (last) character: (1) congruous condition (Con), i.e., the original correct idiom; (2) synonym condition (Syn), i.e., the fourth character of the idiom was substituted by its synonymous noun; (3) semantic violation condition (Sem), i.e., the fourth character was substituted by a semantically unrelated noun that cannot serve as the object of the verb; (4) combined semantic and syntactic violation condition (Com), i.e., the fourth character was substituted by a verb. The frequency and stroke number of the fourth character were matched across the four idiom types (Frequency: 433, 423, 430, and 432 times per million, respectively, according to Sun, 1998; Stroke number: 8.5, 8.7, 9.4, and 9.3 respectively). As the example below indicates, idiom types (2–4) are all incongruous conditions, in which the combination of the four characters does not make an idiom, and furthermore, conditions (3) and (4) do not make sense.

(1) 笑里藏刀 (xiao4-li3-cang2-dao1, hiding a *dagger* behind one's smiles), Congruous condition (Con);
(2) 笑里藏剑 (xiao4-li3-cang2-jian4, hiding a *sword* behind one's smiles), Synonym condition (Syn);
(3) 笑里藏房 (xiao4-li3-cang2-fang2, hiding a *room* behind one's smiles), Semantic violation (Sem);
(4) 笑里藏投 (xiao4-li3-cang2-tou2, hiding *throw* behind one's smiles), Combined semantic and syntactic violation (Com).

Two separate groups of undergraduate students from Beijing Normal University were asked to complete a cloze pretest and a semantic relationship pretest, respectively. These students did not take part in the main experiment described below. In the cloze pretest, the participants were asked to write down the fourth character on paper given the first three characters for each idiom. In the semantic relationship pretest, the participants were asked to rate the meaning relationship between the last character in each of the three incongruous types and the last character in the correct idiom, on a 5-point scale (1 = not related at all, 5 = very highly related). The data from the cloze pretest showed that the averaged cloze probability for all the idioms reached 96%. Note that there is only one correct ending character for each idiom in the congruous condition. The data from the semantic rating pretest showed that the rated semantic relationship between the congruous and the three incongruous groups was 4.27 for the synonymous noun, 1.43 for the noun in the semantic violation condition, and 1.42 for the verb in the semantic and syntactic violation condition, confirming that the synonymous noun has the closest semantic relationship with the correct noun while the other two words in the violation conditions showed no difference (Syn > Sem, $p < 0.001$; Sem = Com, $p > 0.1$).

The 160 sets of items were Latin-square arranged into four lists according to the four idiom types so that each idiom could only appear once in the same list. Each list contains 40 items per type. Another 40 incongruous and 120 congruous idioms were added as fillers, resulting in a total of 320 items and the ratio of congruous to incongruous items equal to 1:1 in each list.

2.3. Procedure

We modeled our experimental procedure after the Zhou et al. (2004) study, which is illustrated in Fig. 1. Each trial consisted of a fixation, a four-character idiom, and a response mark. The inter-trial-interval (ITI) was 1900 ms. Each idiom was preceded by a crosshair fixation, which lasted 300 ms. After 200 ms blank screen, the 4-character idiom was presented visually, character-by-character at a rate of

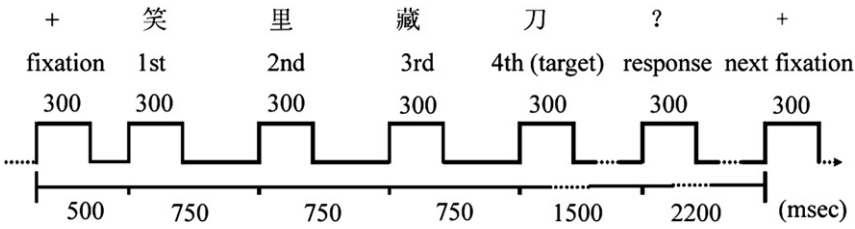

Fig. 1. Stimulus configuration of an example trial.

750 ms per character (duration = 300 ms, blank screen = 450 ms). Each idiom was followed by a 300-ms response mark "?" after a 1200 ms blank screen. All the stimuli were presented in 72-point *Song* font with a black (RGB 0, 0, 0) against a grey background (RGB 192, 192, 192) on a ViewSonic 22″ computer monitor (with 1024 × 768 resolution).

Participants were tested individually in a quiet experimental room. Each participant was seated about 1 m away from the display screen. Prior to the experiment, 18 practice items were presented to the participants; the practice items were similar to the test items but were not used in the real experiment. Each participant received only one of the stimulus lists, in a pseudo-randomized sequence such that no more than three congruous or incongruous trials appeared in succession. They were asked to read the characters carefully and to judge whether the fourth character was correct for the current idiom by pressing the "Yes" or "No" button on a keyboard when the response mark "?" appeared. They were instructed to respond to the stimuli as accurately as possible. To reduce the artifact influence of eye blinks, participants were asked to minimize their blinks during the presentation of stimuli, except during the ITI period. They were given a short break after 80 trials. The whole test session lasted about 35 min.

The presentation of the stimuli and the collection of behavioral data were completed by the use of the STIM software (NeuroScan Inc.).

2.4. EEG recording

EEG data were recorded using a 32-channel Quick-cap with Ag/AgCl electrodes (NeuroScan Inc.). EEG electrodes were placed according to the extended 10–20 system. All scalp electrodes were referenced to the left mastoid (P8 was placed at right mastoid to record the signal at that point) and off-line re-referenced by subtracting half the value of P8. In order to eliminate 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. One set of 32-channel Synamps amplifiers was used for EOG and EEG recordings. Band pass was 0.05–100 Hz and the sampling rate was 500 Hz.

2.5. Data analyses

Response accuracy and reaction times (RTs) were calculated for each idiom type and each participant. Trials on which the participant responded wrongly were excluded from the calculation of mean RTs.

For the preprocessing of the EEG data off-line, a blink-correction algorithm (SpatialSVD) was applied to reduce the artifact of blink. Epochs of 900 ms length were cut out from the continuously recorded data. The epochs started 100 ms before the onset of the fourth character. After whole-epoch-based linear detrend and baseline correction and filtering (band pass was 0.05–30 Hz), trials with potentials greater than 70 μV were rejected as artifacts. The mean trial rejection rates over all the remaining 21 participants for each idiom type were less than 1%, which means that the number of valid

trials is above 39 out of the total 40 (range from 37 to 40). The average of ERPs was computed for each participant on each idiom type.

Statistical analyses of the congruity effects consisted of several repeated-measures ANOVAs with mean amplitude values computed for each participant and each electrode in two time windows: a) 250–350 ms after the fourth character onset for the N400, b) 400–600 ms for the P600, based on the visual inspection of the difference waveforms between the congruous condition and the incongruous ones. For each time window, the results were first analyzed in an omnibus ANOVA across all four levels of the *Congruity* factor (Con, Syn, Sem and Com) with the 29-level *Electrode* factor. In addition to the omnibus ANOVA, a priori pairwise comparisons between the congruity conditions were tested using ANOVAs with a 2-level congruity factor. If the interaction between Congruity and Electrode reached significance, topographic analysis was subsequently explored in a $2 \times 2 \times 2$ ANOVA. In addition to the factor *Congruity Condition*, the other two factors are *Hemisphere* (Left: F7, F3, FT7, TP7, CP3, P3/Right: F8, F4, FT8, TP8, CP4, P4) and *Region* (Anterior: F7, F3, FT7, F8, F4, FT8/Posterior: TP7, CP3, P3, TP8, CP4, P4). Greenhouse-Geisser correction was applied when Mauchly's Test of Sphericity was violated, and the original degrees of freedom and adjusted p values were reported.

3. Results

3.1. Behavioral data

The judgment accuracy and reaction times (RTs) across the congruous condition, the synonym condition, and the semantic and the combined violation conditions were shown in Table 1.

On accuracy, repeated-measures ANOVA showed that the main effect of idiom type was significant [$F(3, 60) = 9.474$, $p < 0.001$]. Pairwise comparisons indicated that the accuracy of semantic violation and the combined violation conditions was higher than that of the congruous and the synonym conditions ($ps < 0.05$), but no difference was found between the former two or between the later two ($ps > 0.1$).

On RTs, a marginally significant main effect of the idiom type was revealed by ANOVA [$F(3, 60) = 2.296$, $p = 0.087$]. The only difference was found between the congruous and the synonym conditions ($p = 0.012$).

Taken together, this pattern of results suggests that it is more difficult for the participants to reject the synonymous noun than to reject other two violations for idioms, although the participants' responses were postponed by the question marker due to the design of our experiment.

3.2. ERP data

Fig. 2 displays the grand averaged waveforms by electrode sites time-locked to the onset of the fourth character. There are several important findings. First, there is little difference between the four idiom types before 200 ms, and no evidence for an ELAN effect on the combined violation condition as reported in previous studies in European languages. Second, compared to the congruous idioms, a negativity component at about 250–350-ms time window is apparent in all three incongruous idioms. It appears that the negativity is largest in the combined semantic and syntactic violation condition and is smallest in the synonym condition. Although the latency of the negativity is earlier than in previous studies on sentence materials, its pattern and distribution are similar to previously reported N400 effects. Third, at 400- to 600-ms time window, the waveforms of three incongruous idioms go more positively than that of congruous idioms, which is especially noticeable over the left hemisphere (see also Fig. 3).

Table 1
Reaction times (RT, ms) and accuracy across four types of material on the judgment task (with S.D. in parentheses).

	Congruous	Synonymous	Semantic violation	Combined violation
RT	747 (195)	779 (196)	758 (215)	762 (200)
Accuracy	0.96 (0.03)	0.96 (0.03)	0.99 (0.02)	0.98 (0.03)

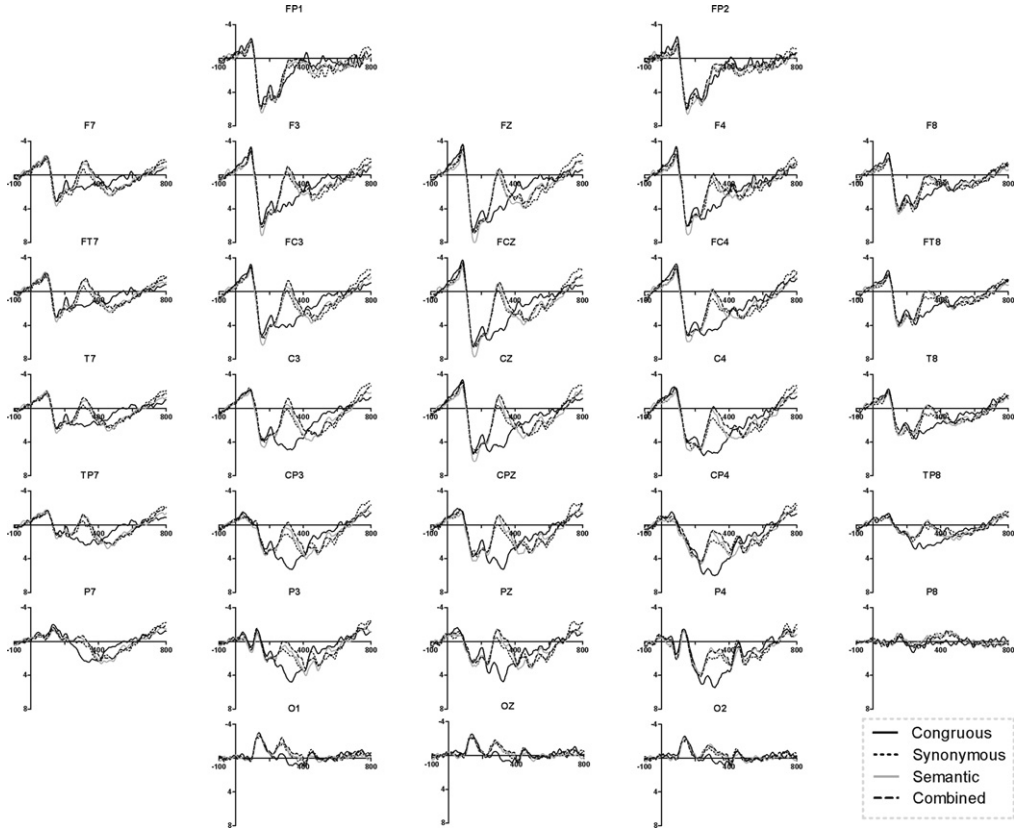

Fig. 2. Grand average ERPs from all 29 electrodes to the final character across congruous idiom (black solid line), synonymous (black dotted line), semantic violation (grey line) and combined semantic and syntactic violation (black dash/dot line), after baseline correction in the -100- to 0 ms pre-stimulus interval. Time 0 is the onset of the final character of idiom. The *X* axis (time) is in millisecond and the *Y* axis (amplitude) is in μV.

3.2.1. N400 effects: 250–350 ms

Table 2 displays the results of ANOVAs on the mean ERP amplitude in the 250- to 350-ms time window. The omnibus ANOVA indicated that the main effect of *Congruity* in the N400 time window was significant. Further analyses revealed that all three incongruous conditions elicited larger N400 than the congruous idioms (corresponding to effects: Syn – 2.20 μV, Sem – 2.65 μV, and Com – 2.90 μV, respectively). The main effect of *Electrode* was also significant [$F(28, 560) = 4.610$, MSE = 17.020, $p < 0.01$]. An interaction between *Congruity* and *Electrode* were also found. In addition, both the Sem and Com conditions evoked larger N400 than the Syn condition (corresponding to effects of 0.46 and 0.71 μV, respectively). No significant difference was found between Sem and Com conditions. Topographical analyses showed that all three congruity effects were significantly larger over posterior than anterior regions of the scalp (corresponding to effects of 2.53 vs. 1.66, 3.03 vs. 2.21, and 3.23 vs. 2.41 μV, respectively), and all three congruity effects were significant over both posterior and anterior regions (all $ps < 0.001$). No hemispheric difference was observed (see also Figs. 3 and 4).

3.2.2. P600 effects: 400–600 ms

Table 3 displays the results of ANOVAs on the mean ERP amplitude in the 400- to 600-ms time window. Similar to that of N400, the omnibus ANOVA indicated that the congruity effect in the P600 time window was significant. Further analyses revealed that all three incongruous conditions elicited

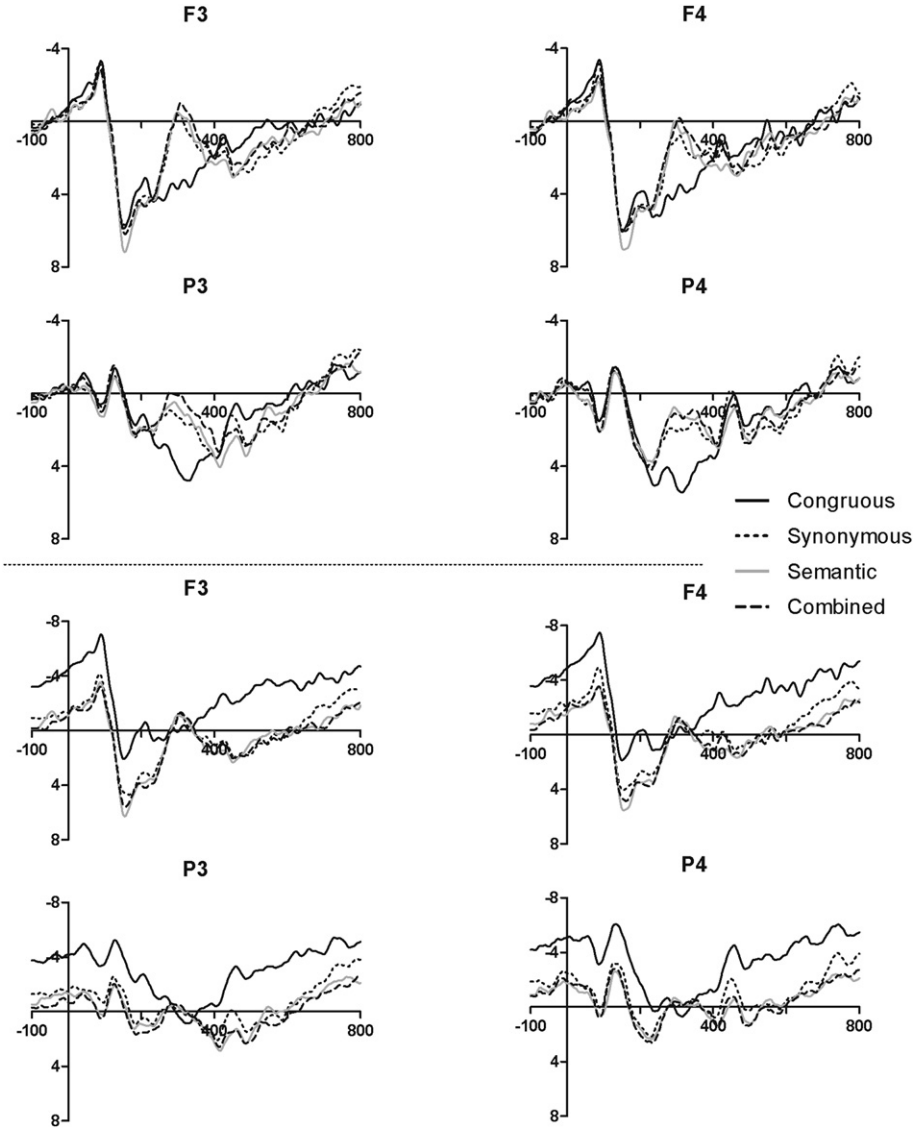

Fig. 3. Grand average ERPs from F3/F4, P3/P4 electrode to the final character across congruous idiom (black solid line), synonymous (black dotted line), semantic violation (grey line) and combined semantic and syntactic violation (black dash/dot line). Time 0 is the onset of the final character of idiom. The *X* axis (time) is in millisecond and the *Y* axis (amplitude) is in μV. Upper: baseline correction in the -100-0 ms pre-stimulus interval; Lower: baseline correction in the 250–350 ms post-stimulus interval.

larger P600 than the congruous idioms (corresponding to effects of Syn – 0.85 μV, Sem – 0.79 μV, and Com – 0.64 μV, respectively). The main effect of *Electrode* was also significant [$F(28, 560) = 10.376$, MSE = 3.544, $p < 0.001$]. An interaction between *Congruity* and *Electrode* was also found, but no significant difference was found between Syn, Sem and Com conditions. Topographical analyses showed that all three congruity effects were significantly larger over the left hemisphere than the right hemisphere (corresponding to effects of 1.13 vs. 0.40, 1.03 vs. 0.45, and 0.89 vs.0.30 μV), and the three congruity effects were significant only over the left hemisphere (all $ps < 0.001$) and not the right hemisphere (all $ps > 0.05$). No anterior–posterior difference was observed (see also Figs. 3 and 4).

Table 2
ANOVAs on mean amplitude in the 250- to 350-ms time window (N400).

	Source	*df*	*F*	MSE	*p*
Omnibus ANOVA (29 electrodes)					
Overall	Type	3, 60	52.24	33.24	.000***
	Type × El	84, 1680	9.03	7.62	.000***
Syn vs. Con	Type	1, 20	89.63	16.35	.000***
	Type × El	28, 560	11.70	5.84	.000***
Sem vs. Con	Type	1, 20	53.98	39.74	.000***
	Type × El	28, 560	11.27	7.03	.000***
Com vs. Con	Type	1, 20	97.29	26.39	.000***
	Type × El	28, 560	18.28	4.27	.000***
Sem vs. Syn	Type	1, 20	3.97	16.26	.06†
	Type × El	28, 560	.66	3.45	.627
Com vs. Syn	Type	1, 20	10.88	14.10	.004**
	Type × El	28, 560	1.01	3.16	.410
Com vs. Sem	Type	1, 20	1.91	9.95	.183
	Type × El	28, 560	.72	2.54	.602
Topographic ANOVA (12 electrodes)					
Syn vs. Con	Type × AP	1, 20	6.35	1.25	.020*
	Type × LR	1, 20	.08	.97	.781
Sem vs. Con	Type × AP	1, 20	5.54	1.28	.029*
	Type × LR	1, 20	.06	.76	.785
Com vs. Con	Type × AP	1, 20	12.02	.59	.002**
	Type × LR	1, 20	.45	.86	.480

Note: Type = congruity type; El = electrode; AP = anterior vs. posterior; LR = left vs. right. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$. In order to get the data presented more clearly and compactly, the effect of Electrode was not included in Table but could be found in text.

It should be noted that the amplitude of P600 could be affected by earlier differential effects in the N400 time window. We performed further analyses on the P600 component after baseline correction in the 250- to 350-msec interval as has been done in some previous studies (see also van den Brink, Brown, & Hagoort, 2001; van den Brink & Hagoort, 2004; Hagoort, 2003; Martin-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006). Table 4 displays the results of ANOVAs on the mean ERP

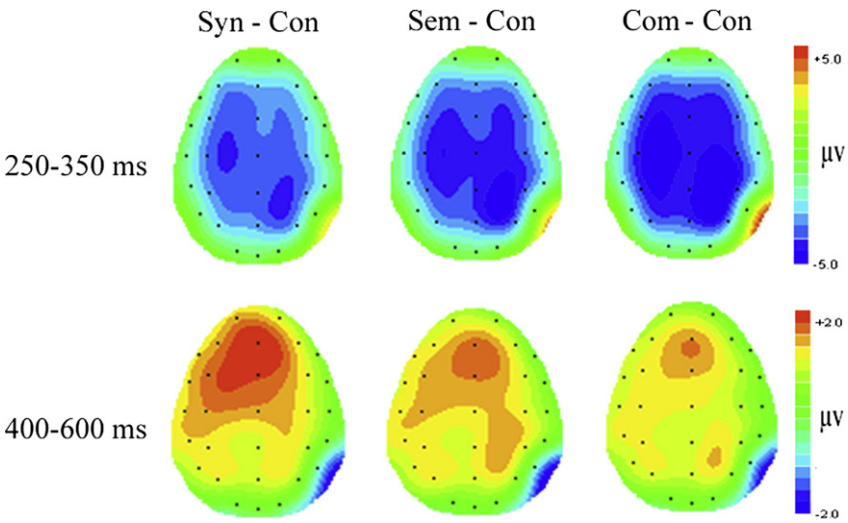

Fig. 4. Topographic maps of the N400 and P600 effects. Maps are computed from values resulting from the subtraction of the mean amplitude in the time windows of the grand ERPs between the incongruous conditions and the congruous condition. Con – congruous idiom, Syn – synonymous violation, Sem – semantic violation, Com – combined semantic and syntactic violation. The view is from top of head, and anterior is on top.

Table 3
ANOVAs on mean amplitude in the 400- to 600-ms time window (P600).

	Source	df	F	MSE	p
Omnibus ANOVA (29 electrodes)					
Overall	Type	3, 60	7.85	15.26	.001***
	Type × El	84, 1680	3.72	3.94	.001***
Syn vs. Con	Type	1, 20	19.22	11.50	.000***
	Type × El	28, 560	9.07	2.37	.000***
Sem vs. Con	Type	1, 20	14.84	12.82	.001***
	Type × El	28, 560	3.80	3.51	.008**
Com vs. Con	Type	1, 20	12.87	9.74	.002**
	Type × El	28, 560	4.00	2.52	.003**
Sem vs. Syn	Type	1, 20	.07	15.46	.789
	Type × El	28, 560	1.32	3.08	.274
Com vs. Syn	Type	1, 20	.83	16.30	.374
	Type × El	28, 560	2.37	2.05	.068†
Com vs. Sem	Type	1, 20	1.28	5.28	.271
	Type × El	28, 560	.38	1.45	.861
Topographic ANOVA (12 electrodes)					
Syn vs. Con	Type × AP	1, 20	1.32	.40	.264
	Type × LR	1, 20	13.39	.41	.002**
Sem vs. Con	Type × AP	1, 20	.14	.58	.716
	Type × LR	1, 20	9.58	.37	.006**
Com vs. Con	Type × AP	1, 20	.26	.45	.615
	Type × LR	1, 20	8.16	.45	.010**
Com vs. Syn	Type × AP	1, 20	9.96	.12	.005**
	Type × LR	1, 20	.93	.21	.347

Note: Type = congruity type; El = electrode; AP = anterior vs. posterior; LR = left vs. right. †$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. In order to get the data presented more clearly and compactly, the effect of Electrode was not included in Table but could be found in text.

amplitude in the 400- to 600-ms time window after baseline correction in the 250- to 350-ms interval (see also Fig. 3).

As the analyses with the baseline correction in the -100- to 0 ms pre-stimuli interval, the omnibus ANOVA indicated that the congruity effect in the P600 time window was significant. Further analyses revealed that all three incongruous conditions elicited larger P600 than the congruous idioms (corresponding to effects of Syn – 3.03 μV, Sem – 3.41 μV, and Com – 3.51 μV, respectively). The main effect of *Electrode* was also significant [$F(28, 560) = 3.370$, MSE = 10.143, $p < 0.05$]. An interaction between *Congruity* and *Electrode* was also found. No significant difference was found between Sem and Syn conditions, or between Sem and Com conditions, but a marginally significant difference was observed between Syn and Com conditions. The topographical analyses confirmed that all three congruity effects were significantly larger over left hemisphere than right hemisphere (corresponding to effects of 3.29 vs. 2.42, 3.62 vs. 3.05, and 3.78 vs. 3.00 μV, respectively). In addition, we also observed anterior–posterior difference, which means the congruity effects were larger at posterior than anterior sites (corresponding to effects of 3.19 vs. 2.52, 3.79 vs. 2.88, and 3.87 vs. 2.91 μV, respectively).

To sum up, the two kinds of analyses showed very similar results on the P600 time window. All the three incongruous conditions elicited a significant left-hemisphere dominated P600 effect and there was no difference between the three, especially between the semantic violation and the combined violation conditions.

3.2.3. Scalp distribution analyses

Fig. 4 shows the voltage maps of mean amplitude differences between the three incongruous conditions and the congruous condition in the N400 and the P600 time windows. The figure and the statistical data above indicated a broadly distributed and centro-posterior dominated N400 effect and a left-hemisphere distributed P600 effect. In order to examine whether the three kinds of congruity effects at both N400 and P600 time windows still have statistically distinct distribution, we performed additional analyses separately for the two time windows following the normalization procedure

Table 4
ANOVAs on mean amplitude in the 400- to 600-ms time window (P600) after correction in the 250- to 350-ms interval.

	Source	df	*F*	MSE	*p*
Omnibus ANOVA (29 electrodes)					
Overall	Type	3, 60	61.94	45.63	.000***
	Type × El	84, 1680	12.65	11.07	.000***
Syn vs. Con	Type	1, 20	105.63	26.41	.000***
	Type × El	28, 560	21.36	6.08	.000***
Sem vs. Con	Type	1, 20	68.83	51.55	.000***
	Type × El	28, 560	13.16	12.49	.000***
Com vs. Con	Type	1, 20	95.52	39.30	.000***
	Type × El	28, 560	21.79	6.71	.000***
Sem vs. Syn	Type	1, 20	2.63	17.33	.120
	Type × El	28, 560	1.45	6.41	.239
Com vs. Syn	Type	1, 20	3.53	20.27	.075†
	Type × El	28, 560	1.91	2.90	.115
Com vs. Sem	Type	1, 20	.29	10.01	.597
	Type × El	28, 560	.75	3.81	.538
Topographic ANOVA (12 electrodes)					
Syn vs. Con	Type × AP	1, 20	4.79	.99	.041*
	Type × LR	1, 20	8.43	.93	.009**
Sem vs. Con	Type × AP	1, 20	4.35	1.98	.050*
	Type × LR	1, 20	3.69	.92	.069†
Com vs. Con	Type × AP	1, 20	11.94	.83	.003**
	Type × LR	1, 20	8.02	.78	.010**
Com vs. Syn	Type × AP	1, 20	3.50	.27	.076†
	Type × LR	1, 20	.18	.50	.678

Note: Type = congruity type; El = electrode; AP = anterior vs. posterior; LR = left vs. right. †$p < .1$, *$p < .05$, **$p < .01$, ***$p < .001$. In order to get the data presented more clearly and compactly, the effect of Electrode was not included in Table but could be found in text.

suggested by McCarthy and Wood (1985), which avoids the differential amplitude effects on distributional shape. In this procedure, difference ERPs were computed by subtracting the ERPs elicited in the congruous condition from those in the three incongruous conditions, and mean amplitudes within time windows of interest were measured on the difference ERP waveforms. Subsequently, the difference scores were *z*-transformed for each time window separately: the grand mean amplitude and standard deviation of all the electrode sites and all the participants were computed for each condition. The grand mean amplitude was subtracted from the amplitude at each electrode site for each participant, and the difference was divided by the standard deviation. The *z*-transformed values were entered into two ANOVAs that crossed the 3-level Incongruous Condition (IC) factor with the 29-level Electrode factor. The results showed, for N400, only the main effect of Electrode was significant [$F(28, 560) = 16.109$, $p < 0.001$], the main effects of IC and its interaction with Electrode were non-significant ($F < 1$ or $p > 0.1$). The same pattern was also observed for P600 [Electrode: $F(28, 560) = 6.628$, $p < 0.001$; IC: $F < 1$; Interaction: $p > 0.1$]. Taken together, the scalp distribution analyses confirmed that there was no topographic difference between the three kinds of incongruous stimuli.

4. Discussion

In this study we examined the interplay of structure and meaning with the most common four-character idioms in Chinese. The use of Chinese idioms provides us with the advantage of identifying the effects of semantic and syntactic processing in a fixed stable linguistic structure. We adopted the violation paradigm commonly used in ERP language research by comparing congruous with incongruous conditions that involve either synonyms of idioms, semantic violation, or semantic plus syntactic violation. Our results indicated that (1) no evidence for an ELAN or a LAN effect was found on the combined violation condition as reported in previous studies of European languages; (2) all three incongruous conditions elicited a widely-distributed N400 component when compared with the congruous, correct idioms, and within the incongruous conditions, the synonyms received a smaller

N400 than the semantic violation and the combined violation, while there was no significant N400 difference between the latter two conditions, and (3) all three incongruous conditions also elicited a left-hemisphere dominant P600 component when compared with the congruous condition, but there was no P600 difference among the three incongruous conditions, especially between the semantic violation and combined violation conditions.

The semantic manipulation across the four idiom types resulted in the graded N400 effects in the 250–350 ms time window. The negativity is broadly distributed and has a posterior maximum, which is different from the syntax-related negativity LAN which has a left-frontal distribution associated with morphosyntactic incongruity (see Friederici, 2002, for a review). It is much more likely that the negativity that we obtained is a classic N400 effect. As mentioned earlier, the fixed-form idioms in Chinese have a structural stability, in that the characters/morphemes for any given idiom are fixed and not subject to substitution or paraphrase. For our experiment, this means that the last character in each test material has an extremely high cloze probability (i.e., 96%, according to our pretest). In other words, items in the three incongruous conditions are equally unlikely to occur in the language. Consequently, the graded N400 effects observed in our experiment cannot be attributed to the predictability of a word in its context. Not coincidentally, the result from our semantic rating pretest mirrors this pattern of N400 effects. On these grounds, we suggest that the N400 effects for the three incongruous conditions may reflect the degree of meaning mismatch or the level of difficulty in integration, as has been suggested in some previous studies (Kutas & Federmeier, 2000; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kutas & Hillyard, 1980; Li, Shu, et al., 2006).

In contrast to the main findings from Indo-European languages, the combined word category and semantic violation in the present study did not elicit the biphasic ELAN-P600 as reported in several studies (See Friederici, 2002 for a review). In the literature, both the neurocognitive model (Friederici, 2002) and the eADM model (Bornkessel & Schlesewsky, 2006; Bornkessel-Schlesewsky & Schlesewsky, 2008) emphasize the role of identification of word category information in early local structure building, and ELAN has been thought to reflect this structure building process. Such computation of local phrase structure based on word category information was taken as the initial phase during on-line language comprehension, which is claimed to be task-irrelevant, entirely autonomous and even universal across languages (Friederici & Weissenborn, 2007). However, supporting evidence to this claim mostly comes from studies in Indo-European languages with rich morphosyntactic structures, such as German and Dutch. Ye et al. (2006) is the only study in Chinese that reported a very early ELAN (50 ms after onset) and a larger N400 on a combined word category and semantic violation condition, compared to a semantic only violation condition. The Ye et al. study suffered from stimulus selection problems, as discussed earlier. A better-controlled experiment by Yu and Zhang (2008) observed no ELAN but equal N400 effects on combined word category and semantic violation sentences when compared with semantic only violation sentences. Our data are consistent with the Yu and Zhang study with idioms in a fixed linguistic structure.

A possible explanation for why we did not observe EALN or LAN on the combined violation condition could be that the idioms in Chinese are stored and used as a unit like a word; there is no need to assign a grammatical role to each character, or to establish a thematic relationship among characters. As discussed earlier, most idioms in Chinese are based heavily on a historical and cultural context such as legends and tales. Linguists debate whether idioms in Chinese belong to phrases or compound words (Wang, 1983). On the one hand, idioms usually consist of several words. For example, in 笑里藏刀 (xiao4-li3-cang2-dao1, hiding daggers behind smiles), "笑 (smile)", "藏 (hide)", and "刀 (dagger)" are words and can be used independently in modern Chinese. On the other hand, as mentioned in Introduction, idioms are more compact in structure and more integral on meaning than regular phrases. In psycholinguistics, no serious endeavor has been made to investigate whether the processing of idioms in Chinese is more like regular phrases or compound words. In this study we argue that the processing of idioms is more like that of regular phrases or short sentences upon two considerations. First, the idioms used in present study have a special structure, i.e., the third and fourth characters are modern Chinese words which make up verb-object phrases. Second, our experiment displayed the stimuli character-by-character, which encouraged participants to process the individual characters in an idiom successively.

Another critical finding from our study is that the semantic processing was not influenced by the intactness of word category information. The evidence came from that a biphasic N400–P600 pattern

was observed on the combined word category and semantic violation in the present study, and there was no difference when compared with the semantic only violation condition on either the N400 or the P600 component. A widely accepted proposal is that syntactic processing may be independent of semantic context, but semantic integration is influenced by syntactic processing. For example, Friederici et al. observed P600 and ELAN effects, but no N400 effect on the combined violation condition (Friederici et al., 1999, 2004; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001; Isel et al., 2007). By contrast, Hagoort (2003) found a larger semantic N400 effect on combined semantic and syntactic violation compared to semantic only violation, and a comparable P600 effect compared to syntactic only violation. Although these two groups of researchers show different directions of the N400 amplitude change and the earliness of syntactic processing, they both accept that semantic integration is influenced by syntactic processing, but syntactic analysis is unaffected by semantic integration. Obviously, our findings are inconsistent with the "semantics is influenced by syntax" view. Similar findings regarding identical N400 effects on the semantic only violation and the combined violation were observed in a sentence context (e.g., Yu & Zhang, 2008). Based on these results, one can draw the conclusion that semantic integration is not affected by additional syntactic processing (e.g., processing of word category mismatch), at least not in the Chinese language. A recent study in Spanish (Martin-Loeches et al., 2006) showed a similar finding that N400 as an index to semantic violations was unaffected by additional syntactic violations. It should be noted, however, that Spanish is also a morphologically rich language, and the detection of semantic violation (root) is earlier than that of syntactic violation (suffix) in the experimental materials, although these were presented visually and the influence could be weak.

Taken together, our ERP patterns are consistent with the argument made in other behavioral and neural studies that semantic processing in Chinese can proceed even when syntactic or structural analysis fails (Li, 1996; Ye et al., 2006; Yu & Zhang, 2008). We cannot determine yet, in this study, whether syntactic analysis in Chinese will be influenced by semantic processing. As discussed earlier, there is no morphological marking of gender, number, case, or even word category in Chinese, and therefore it is difficult to construct test materials that contain only syntactic violations (without also changing the semantics) in Chinese as in morphologically rich languages. Our results suggest at least that there is no obvious influence of syntactic category processing on semantic processing in Chinese. This suggestion is consistent with the widely held conjecture that semantic analysis would naturally play a more dominant role in sentence comprehension in Chinese compared to Indo-European languages like German, given the structural properties of the Chinese language (see Li, 1996; Ye et al., 2006).

Finally, one intriguing question from our study is why the semantic violation elicits P600 effects. We observed identical P600 effects on synonym violation, semantic violation and combined violation conditions. Typically, P600 is reliably elicited by syntactic incongruency in sentences, including word category violation, morphosyntax violation, phrase structure violation, and ambiguous sentences (Neville, Nicol, Barss, Forster, & Garrett, 1991; Rossi, Gugler, Hahne, & Friederici, 2005; see Kuperberg, 2007 for a review). It seems that the P600 effects in the present study may not result entirely from the syntactic integration, repair, revision or complexity (see Friederici, Hahne, & Saddy, 2002; Hagoort, Brown, & Osterhout, 1999; Kaan & Swaab, 2003), since only one of the three incongruous conditions contained syntactic violations. It should be also noted that previous studies in Chinese reported weak or no P600 effects on word category violation in sentence context (e.g., Ye et al., 2006; Yu & Zhang, 2008).

Although there is a likelihood that the explicit task used by some previous studies (judgment on the correctness of sentences) may have led to P600 (for a review see Kuperberg, 2007), we would like to entertain the possibility that the P600 effects in the present study could be attributed to the use of fixed-form idioms which have stable linguistic structure (as opposed to the use of more flexible sentence contexts). First, these idioms are quite common in Chinese daily life, so it is very easy for a native undergraduate or graduate students to accomplish this task. Compared to situations in sentence comprehension, especially some in clause contexts, this task on idioms is much easier. Second, P600s on semantic violation were only observed in some studies but not others. Ye et al. (2006) also used an explicit task – overall judgment on correctness of sentences, but failed observing significant P600 effects on both semantic and syntactic violation in Chinese sentences. Third, two other existing studies using Chinese idioms as experimental materials also observed P600 effects, as discussed earlier,

although they used a similar explicit judgment task like Ye et al. (Zhou et al., 2004, and Nan et al., 2009). Finally, a recent study (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007) observed frontal positivity from 500 to 900 ms post-stimulus-onset for unexpected words in highly constraining context, but not for unexpected words in weakly constraining sentences. These results from elsewhere, along with our findings, suggest that the P600 component might be sensitive to the structure in highly constraining context. The highly fixed structure of Chinese idioms cannot be altered in daily language use. And the disruption of the idiom structure with any change in semantics, syntax, or both could cause discontinuity in the processing of the materials as idioms. In the language comprehension domain, this late positivity can also be differentiated into different kinds, as suggested in the eADM model – generalized mapping P600 (or semantic P600) and well-formedness P600 (Bornkessel-Schlesewsky & Schlesewsky, 2008). In addition, a few musical perception studies also observed language-like P600 effects, as discussed earlier (e.g. Besson & Schon, 2001; Nan et al., 2009; Patel et al., 1998; see Patel, 2003 for a review). These data point to the possibility that the P600 does not necessarily reflect syntactic or grammatical processing, but rather serves as an index of the structural incongruence that can occur at various levels of linguistic and perceptual processes, including prosodic, phonological, and musical pattern incongruence.

Acknowledgements

This research was supported by China Postdoctoral Science Foundation (#20070410042) to YL, grants from Natural Science Foundation of China (30900392, 30870758) to YL and HS, from National Science Foundation of USA (#BCS-0642586) to PL, and from the fund for Foreign Scholars in University Research and Teaching Programs (B07008) to both PL and HS.

References

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner, & L. Gleitman (Eds.), *Language acquisition: The state of the art*. New York: Cambridge University Press.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.

Besson, M., & Schon, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences, 930*, 232–258.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*(4), 787–821.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews, 59*(1), 55–73.

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*, 967–985.

van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*(6), 1068–1084.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally-ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. *Brain & Language, 50*, 259–281.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*, 78–84.

Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *Neuroreport, 15*(1), 165–169.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research, 31*, 45–63.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183–192.

Friederici, A. D., Steinhauer, K., & Frisch, S. (1999). Lexical integration: sequential effects of syntactic and semantic information. *Memory & Cognition, 27*, 438–453.

Friederici, A. D., & Weissenborn, Y. (2007). Mapping sentence form onto meaning: the syntax–semantic interface. *Brain Research, 1146*, 50–58.

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., Brown, C. M., & Osterhout, L. (1999). The neurocognition of syntactic processing. In C. M. Brown, & P. Hagoort (Eds.), *Neurocognition of language* (pp. 273–317). Oxford: Oxford University Press.

Hagoort, P., Hald, L., Bastiaansen, M. C. M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science, 304*(5669), 438–440.
Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*, 339–356.
Hahne, A., & Jescheniak, J. D. (2001). What's left if the Jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research, 11*(2), 199–212.
van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241–255.
Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology, 74*(3), 337–346.
Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159–201.
Kaan, E., & Swaab, T. Y. (2003). Repair, revision, and complexity in syntactic analysis: an electrophysiological differentiation. *Journal of Cognitive Neuroscience, 15*(1), 98–110.
Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565–1577.
Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. W. (2003). Structure and limited capacity in verbal working memory: a study with event-related potentials. *Brain & Language, 85*, 1–36.
Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research, 1146*, 23–49.
Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive. Science, 4*(12), 463–470.
Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potential reflect semantic incongruity. *Science, 207*, 203–205.
Li, P. (1996). The temporal structure of spoken sentence comprehension in Chinese. *Perception & Psychophysiology, 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 & Language, 32*, 169–192.
Li, P., & Shu, H. (2010). Language and the brain: computational and neuroimaging evidence from Chinese. In M. Bond (Ed.), *The Oxford Handbook of Chinese psychology* (pp. 69–92). Oxford, UK: Oxford University Press.
Li, X., Shu, H., Liu, Y. Y., & Li, P. (2006). Mental representation of verb meanings: behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 18*, 1774–1787.
Li, P., Tan, L.-H., Bates, E., & Tzeng, O. (2006). *The Handbook of East Asian Psycholinguistics*. In *Chinese, Vol. 1*. Cambridge, UK: Cambridge University Press.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*, 676–703.
Martin-Loeches, M., Nigbur, R., Casado, P., Hohlfeld, A., & Sommer, W. (2006). Semantics prevalence over syntax during sentence processing: a brain potential study of noun-adjective agreement in Spanish. *Brain Research, 1093*, 178–189.
McCarthy, G., & Wood, C. C. (1985). Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology, 62*, 203–208.
Nan, Y., Friederici, A. D., Shu, H., & Luo, Y. J. (2009). Dissociable pitch processing mechanisms in lexical and melodic contexts revealed by ERPs. *Brain Research, 1263*, 104–113.
Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.
Ni, B. Y., & Yao, P. C. (1990). *Nine chapters of Chengyu*. 成语九章. Zhejiang, China: Zhejiang Education Press. (in Chinese).
Osterhout, L., & Hagoort, P. (1999). A superficial resemblance does not necessarily mean you are part of the family: counterarguments to Coulson, King and Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes, 14*(1), 1–14.
Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience., 6*, 674–681.
Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717–733.
Rossi, S., Gugler, M. F., Hahne, A., & Friederici, A. D. (2005). When word category information encounters morphosyntax: an ERP study. *Neuroscience Letters, 384*(3), 228–233.
Sun, M. S. (1998). *The Chinese written language corpus*. Beijing, China: Tsinghua University. Available on-line at. http://www.lits.tsinghua.edu.cn/ainlp/source1.htm.
Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist resolution framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.
Ye, Z., Luo, Y. J., Friederici, A. D., & Zhou, X. L. (2006). Semantic and syntactic processing in Chinese sentence comprehension: evidence from event-related potentials. *Brain Research, 1071*, 186–196.
Yu, J. C., & Sun, M. M. (2004). *Dictionary of Chinese idiom*. 汉语成语词典. Beijing: The Commercial Press. (in Chinese).
Yu, J., & Zhang, Y. X. (2008). When Chinese semantics meets failed syntax. *Neuroreport, 19*(7), 745–749.
Zhou, S., Zhou, W., & Chen, X. (2004). Spatiotemporal analysis of ERP during Chinese idiom comprehension. *Brain Topography, 17*(1), 27–37.
Wang, D. (1983). 词汇学研究 *(Lexicology Study in Chinese)*. Shandong: Shandong Education Press.