Semantics Does Not Need a Processing License From Syntax in Reading Chinese

Yaxu Zhang and Jing Yu
Peking University

Julie E. Boland
University of Michigan

Two event-related brain potential experiments were conducted to investigate whether there is a functional primacy of syntactic structure building over semantic processes during Chinese sentence reading. In both experiments, we found that semantic interpretation proceeded despite the impossibility of a well-formed syntactic analysis. In Experiment 1, we found an N400 difference between combined syntactic category and semantic violations and single syntactic violations. This finding is inconsistent with earlier German and French studies (e.g., Friederici, Gunter, Hahne, & Mauth, 2004; Friederici, Steinhauser, & Frisch, 1999; Hahne & Friederici, 2002) showing that semantic integration does not proceed for words of the wrong syntactic category. In Experiment 2, we used a design that was very similar to that used in earlier German and French studies, but semantic violations still evoked an N400, irrespective of a simultaneous syntactic category violation. We argue against processing models that do not allow for semantic integration of a word unless it can be grammatically attached to the developing phrase structure tree. Rather, language experience may modulate the mode of interplay between syntax and semantics.

Keywords: syntax, syntactic category, semantics, sentence processing, event-related potentials

One of the central questions in the study of sentence comprehension is whether syntactic processes are functionally prior to semantic processes, such that semantic interpretation can only occur after syntactic attachments have been made. According to serial, syntax-first theories of sentence comprehension (e.g., Ferreira & Clifton, 1986; Frazier, 1987; Frazier & Fodor, 1978; Frazier & Rayner, 1982; Friederici, 2002), initial syntactic analysis is independent of, and prior to, semantic analysis, and the semantic integration depends on the syntactic structure built by the modular parser. For example, in the three-phase neurocognitive theory proposed by Friederici (2002), initial syntactic processing of a constituent, that is, computing local phrase structure using syntactic category information (noun, verb, adjective, etc.), precedes semantic and morphosyntactic processing. Recently, Friederici and Weissenborn (2007, p. 54) even claimed that “during normal on-line processing, semantic integration of words does not take place for words which are not syntactically licensed (have wrong word category).”

In contrast, syntactic processes are not endowed with functional primacy in some other theories, including constraint-based lexicalist models (e.g., MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garsey, 1994); the concurrent model (Boland, 1997); Townsend and Bever’s (2001) analysis by synthesis account; and recent neuroscience models, such as the unification model (Hagoort, 2003, 2005) and the nonsyntactocentric, dynamic model (Kuperberg, 2007). For example, in the concurrent model, semantic analysis is permitted to get ahead of bottom-up syntactic analysis, assigning thematic roles and developing a provisional interpretation before a syntactic commitment has been made (Boland, 1997). Even so, syntactic constraints, when available, prevent adoption of a plausible, but syntactically ill-formed, interpretation. According to Kuperberg (2007), a semantic memory-based neural processing stream is responsible for computing the semantic features and semantic relationships between content words within a sentence (including associative relationships and semantic–thematic relationships driven by lexical semantic memory instead of syntax). One possibility proposed by Kuperberg is that “the meaning of a verb, argument(s) and other content words are first combined through pragmatic or inferential heuristic mechanisms into tentative propositions” (p. 37).

The measurement of event-related brain potentials (ERPs) has become a powerful tool for addressing the issue of the functional primacy of syntax, because of the fact that some distinct ERP responses to syntactic and semantic processes have been amply demonstrated in ERP studies of language. Semantic violations, as in He spread the warm bread with socks, evoke a larger N400, a 300–500-ms negativity peaking around 400 ms. The N400 effect is usually maximal over the centro-parietal areas (Kutas & Hillyard, 1980) but sometimes has been found to have a more anterior
distribution (Friederici & Frisch, 2000; Holcomb, 1993; Kutas & Van Petten, 1990). This negativity is believed to reflect the processes of semantic integration (see Kuperberg, 2007, for a brief overview of the functional significance of the N400 effects).

Syntactic category or local phrase structure violations, as in the German version of *The friend was in-the visited, usually elicit an (early) left anterior negativity (ELAN or LAN) either between 100 and 300 ms (e.g., Friederici, Pfeifer, & Hahne, 1993; Friederici, Steinhauer, & Frisch, 1999; Neville, Nicol, Barss, Forster, & Garrett, 1991) or between 300 and 500 ms (e.g., Friederici, Gunter, Hahne, & Mauth, 2004; Friederici, Hahne, & Mecklinger, 1996; van den Brink & Hagoort, 2004), followed by a late centro-posterior positivity labeled P600. It has been claimed that the latency of the anterior negativity depends on the modality, the presentation rate, and the visual contrast of the input, as well as the syntactic category uniqueness point (prefix or suffix) of the critical words (Friederici et al., 2004; Gunter, Friederici, & Hahne, 1999; Hagoort, Wassenaar, & Brown, 2003; van den Brink & Hagoort, 2004; see Friederici & Weissenborn, 2007, for a recent review). In addition, bilateral anterior negativity has been found for syntactic category violations in some studies (e.g., Hagoort et al., 2003; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001; Ye, Luo, Friederici, & Zhou, 2006).

When ERPs are used to test the functional primacy of syntax, the basic approach is to investigate whether there is a larger N400 for words constituting a combined violation of syntactic category and semantics either compared with correct words or to words embodying a syntactic category violation only. If syntactic category processing is functionally prior to the semantic processing and the failed syntactic category processing blocks semantic integration, there should be no N400 effects when the combined violation condition is compared with either the correct condition or the syntactic violation only condition. Contrarily, there should be an N400 effect if semantic integration proceeds even when syntactic category processing fails.¹

Using the approach just described, a series of ERP studies conducted in German and French have found that combined syntactic category and semantic violations, as in the German version of *The door lock was in-the eaten, elicit an ELAN or LAN, followed by a P600, but no N400 effects (Friederici et al., 2004, 1999; Hahne & Friederici, 2002, Experiment 1; Hahne & Jescheniak, 2001; Isel, Hahne, Maess, & Friederici, 2007), suggesting that failed syntactic category processing blocks the process of semantic integration. This finding has been shown to be independent of the modality of stimulus presentation (auditory or visual), the syntactic category uniqueness point (prefix or suffix), and some task demands (overall correctness judgment or probe verification). However, an N400 effect has been found when semantic integration was encouraged by instructions asking participants to judge the sentences only in terms of semantic coherence and to disregard structure violations (in German: Hahne & Friederici, 2002, Experiment 2; in English: Thierry et al., 2008). We take it to be an open question which task conditions correspond most closely to “normal” comprehension processes, and we return to this issue in the General Discussion section. For now, what is crucial is that the absence of N400 effects for the combined violation, under several task conditions and in several different languages, has been taken as evidence supporting the functional primacy of processes dependent upon syntactic category over semantic processes. That is, in the experiments described above, lexical-semantic integration of a given word seems to take place only if that word can be attached to the developing syntactic representation according to the rules of the grammar.

Note that although some studies have revealed an N400 effect for combined syntactic and semantic violations (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Gunter, Stowe, & Mulder, 1997; Hagoort, 2003; Martín-Loeches, Nigbur, Casado, Hohlfeld, & Sommer, 2006; Osterhout & Nicol, 1999; van den Brink & Hagoort, 2004), the N400 effects obtained in these studies cannot rule out the functional primacy of structure-building processes on the basis of syntactic category, as proposed by Friederici and colleagues (e.g., Friederici & Weissenborn, 2007). In each of these experiments, the syntactic anomaly was not a clear, immediate syntactic category violation. For example, in van den Brink and Hagoort’s (2004) auditory study, a verb was preceded by a semantically anomalous adjective, resulting in a combined violation of syntactic category and semantics. They observed an N400 effect at the verb. However, the semantically anomalous stem of the verb is a noun, which can be preceded by an adjective; it is only the word-final syllable de that makes the word a verb. Thus, there was actually no syntactic category violation before de was heard, and the N400 effects may reflect semantic processes occurring on the basis of a well-formed syntactic structure (see Isel et al., 2007, for comments).

For the other studies cited above, the so-called syntactic violation was either morphosyntactic (Gunter et al., 1997; Hagoort, 2003; Osterhout & Nicol, 1999; see Martín-Loeches et al., 2006, for a recent review) or the critical word has been claimed to be a possible grammatical continuation (Ainsworth-Darnell et al., 1998; for comments, see Friederici et al., 1999; Friederici & Weissenborn, 2007). In Friederici’s (2002) account, the distinction between syntactic category information and morphosyntactic information is important, with the former involved in parsing operations prior to the latter. Thus, evidence that the N400 effect is preserved in combined semantic and morphosyntactic violations is not necessarily a problem for Friederici and Weissenborn’s (2007) claim about syntactic primacy. However, as we discuss below, the affixes of a word often determine its syntactic category in Indo-European languages, and in such cases, it is not clear how syntactic category information could be available prior to morphosyntactic cues.

The goal of the present study is to gain insight into whether the functional primacy of syntactic category over semantic processing, which has been repeatedly demonstrated in German, is found in Chinese. If so, it would support the hypothesis that syntactic primacy is a universal property of the language comprehension

¹ These predictions assume that the N400 can be measured equally well, regardless of whether components associated with syntactic violations are also observed. For example, Ainsworth-Darnell, Shulman, and Boland (1998) found roughly the same sized N400 effects in a semantic only violation condition and a double violation condition, in which a P600 was also found. Because the latency of the LAN overlaps to a greater degree with the N400, there may be more risk that the LAN and the N400 cannot be measured independently. However, differences in the usual scalp distribution for the two components usually allow researchers to independently assess the presence of an N400 and an LAN (e.g., Kim & Osterhout, 2005, p. 212, Footnote 4).
SYNTAX AND SEMANTICS IN CHINESE

system. If not, then either the German and French results must be reexamined, or the presence (and/or degree) of syntactic primacy within a language community must depend upon language-specific properties of the input.

The possibility of cross-linguistic differences in syntactic primacy is plausible and could hinge upon morphology. Note that in German and some other Indo-European languages, grammatical morphology (e.g., inflectional or derivational affixes) and orthographic conventions (such as capitalizing all German nouns) mark the syntactic category of a word. In contrast, Chinese is a non-Indo-European language with very limited inflections (e.g., plural inflection of human nouns; perfective and progressive markers). There is no grammatical morphology that either marks syntactic category or marks syntactic features, such as person, number (except plural inflection of human nouns), gender, case, and tense. Furthermore, the predominance of homophonic morphemes means that identifying a word often depends upon sentence context. Grammatical relations in Chinese are not generally cued morphologically; rather, they are cued by word order, function words (such as “ba”), and lexical/semantic content. These linguistic properties of Chinese have lead to the conjecture that sentence comprehension in Chinese relies more on semantic analyses compared with Indo-European languages (e.g., Xu, 1997; also see Li, 1998; Ye et al., 2006).

Some evidence appears to support the conjecture that Chinese sentence comprehension relies on semantic cues to a greater degree than Indo-European languages. First, animacy plays a much more important role than word order in determining the agent of a sentence in Chinese (Li, Bates, & MacWhinney, 1993), whereas in English, the strongest cue to the identification of the agent or subject of a sentence is word order (Bates, McNew, MacWhinney, Devescovi, & Smith, 1982; MacWhinney, Bates, & Kliegl, 1984). Second, Ye et al. (2006) tested Friederici’s (2002) three stage processing model in a Chinese ERP experiment modeled on Hahne and Friederici’s (2002) German experiment. Whereas Hahne and Friederici found the same pattern of ERP components in their syntactic violation and combined violation conditions, Ye et al. found a broadly distributed negativity in their combined violation condition that they interpreted as including an N400. They concluded that semantic processing was dependent upon local structure building in German, whereas in Chinese, syntactic and semantic processing proceeds independently and in parallel during the early stages of analysis. Ye et al. describe Chinese as fitting a variant of Friederici’s account, with the same three processes ordered differently in Chinese compared with German and French. Of course, there are alternative theories of sentence comprehension that do not require ordering syntactic and semantic processing differently for different languages. As described above, the idea that semantic processing is partially independent of syntactic processing is a defining feature of many theories of sentence comprehension (Boland, 1997; Hagoort, 2003, 2005; Kuperberg, 2007; MacDonald et al., 1994; Townsend & Bever, 2001; Trueswell et al., 1994). It is therefore plausible across a broad range of theoretical accounts that there is no functional primacy of syntactic category over semantic processes, and semantic integration can proceed even when syntactic category processing fails during Chinese sentence comprehension.

Despite the obvious merit of investigating the functional primacy of syntactic category in Chinese, there are only two published studies examining the ERP effects of combined syntactic category and semantic violations, neither of which allows for a definitive conclusion (the aforementioned Ye et al., 2006; also see Yu & Zhang, 2008). In Ye et al.’s (2006) combined violation condition, the so-called syntactic category violation was relative to the immediately preceding preposition ba (i.e., the preposition ba was followed by a verb instead of a noun), and the semantic violation occurred relative to more distant preceding sentence context, as in the Chinese sentence Famugong kaicai senlin, ba cai-le (The timberjack exploiting the forest, ba [PREP] cut, “Exploiting the forest, the timberjack cut”). They found an ELAN and a broadly distributed negativity in the 300–500-ms time window for the combined violations. Surprisingly, they found no P600 for either the combined violations or the syntactic violations. Unfortunately, it is difficult to judge whether there was an N400 effect in the broad negativity that they obtained in the combined violation condition, both because it is sentence-final (see Hagoort, 2003; Osterhout & Nicol, 1999) and because of the large number of homophones in spoken Chinese. In fact, a recent review categorized this study together with some other studies in which the N400 disappeared in the combined violation condition (Martín-Loeches et al., 2006).

Very recently, Yu and Zhang (2008) found a larger broad negativity with a centro-parietal focus in the 300–500-ms time window for their combined syntactic category and semantic violation condition, compared with their correct condition. Given its topographic distribution and latency window, this broad negativity was identified as including an N400 effect. The authors therefore concluded that semantic integration proceeds even when syntactic category processing fails. It should be noted, however, that the critical word was a verb in the correct condition, as in the word-by-word Chinese version of The dustman ba (PREP) the edifice’s windows all wiped . . ., but it was a noun in the combined violation condition, as in the word-by-word Chinese version of The dustman ba (PREP) the edifice’s windows all sugar . . . Two earlier ERP studies have shown that compared with Chinese verbs, Chinese nouns elicit a larger negativity over anterior sites in the 300–500-ms time window (Ma, Guo, Peng, & Yan, 2007; Zhang, Ding, Guo, & Wang, 2003). Thus, it is not completely convincing to identify the centro-parietally focused negativity as including an N400 effect, the index of lexical-semantic integration.

In the present study, we aimed to investigate whether semantic integration proceeds even when syntactic category processing fails during Chinese sentence reading by examining the ERP effects of combined syntactic category and semantic violations. In Experiment 1, we compared the combined violation with single syntactic
violation. If semantic integration proceeds even when syntactic category processing fails, there should be a larger N400 for the combined violation compared with the single syntactic violation condition. On the contrary, if the unsuccessful syntactic processing blocks lexical-semantic integration, there should be no such N400 effects.

To preview, we did obtain an N400 effect for the comparison of the combined violation with the single syntactic violation, in seeming contrast to earlier studies that used German and French. To further investigate this discrepancy, in Experiment 2, we designed a combined violation condition that was more similar to that used in the German and French studies. Moreover, unlike in Experiment 1, in which a task of verification sentence judgment was used, in Experiment 2, we used an overall (grammatical and semantic) acceptability judgment task with which the previous German and French studies did not find an N400 effect for the combined violation of syntactic category and semantics.

Experiment 1

The goal of Experiment 1 was to overcome the methodological problems in previous Chinese studies and to test whether semantic integration truly proceeds even when syntactic category processing fails during Chinese sentence reading. We compared combined syntactic category and semantic violations with single syntactic violations. The default word order in Chinese is subject–verb–object (SVO), but in Experiment 1, we used the so-called ba construction that forces the order subject–ba–object–verb. The preposition ba marks the noun phrases (NPs) before and after ba as the subject and the object arguments of the main verb, respectively, as in sentence (1a) in Table 1. Furthermore, the object NP must immediately follow ba, with no intervening words. Thus, for this construction, the syntactic and broad semantic roles of the NPs can be assigned even before the main verb is read, by using the syntactic marker ba and the relevant word order constraint as well as semantic properties of the nouns. For example, the sequence “Wei Li ba pear . . .” allows for a sentence-level representation in which Wei Li acted upon the pear in some way.

As shown in Table 1, four critical conditions were designed for Experiment 1. The sentence in the correct condition consisted of the following sequence: subject NP, the preposition ba, an object NP, an adverb, a plausible transitive verb, the perfective marker le, and a sentence-final classifier or particle word. The words prior to the verb (including the subject, ba, the direct object, and the adverb) set up a sentence context in which a transitive verb was very likely because of the syntactic constraints just described.

To observe the spatiotemporal patterns of the N400 effects, we designed the semantic violation condition, in which the transitive verb in the correct condition was replaced with a transitive but semantically incongruent verb, as intimidated in (1b), resulting in a violation of verb’s selectional restriction for its object, for example, a pear cannot be intimidated.

In the syntactic violation (SYNTACTIC) condition, the main verb in the correct condition, peeled in (1a), was replaced by a “good” artifact noun, knife in (1c). Although a noun can follow an adverb, slowly in (1c), as in the Chinese sentence manmandi daozi zhaodao le (Slowly knife was found [perfective marker], “Slowly the knife was found”), the noun induced a syntactic violation because of the earlier syntactic context. Crucially, for the “good” artifact noun, at least one of its functional semantic features (e.g., knife is an artifact for peeling) can be described by a verb that can assign thematic roles (agent and patient) for the preceding subject and object NPs. In other words, a knife is a good instrument for an action event involving a person and a pear. As demonstrated by the normative data described below, it was relatively easy for Chinese readers to integrate the “good” artifact noun into the preceding sentence context semantically (i.e., taking the knife to be the instrument with which Wei Li acted upon the pear). Note that a weak semantic violation may nonetheless occur, given that an expected action semantic feature, which is crucial for building a semantic interpretation as to who does what to whom, was absent in the artifact noun.

In contrast, in the combined syntactic and semantic violation (COMBINED) condition, the verb in the correct condition was replaced by a “bad” artifact noun, such as piano in (1d). By “bad” we mean that none of the functional semantic features of this artifact noun can be described by a verb that can assign thematic roles for the preceding subject and object NPs. Thus, there is no way for Wei Li to use the piano to act upon the pear, making it

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct (CORRECT)</td>
<td>(1a) Wei Li ba (PREP)/ fresh pears slowly/ peeled/ le (PERF)/ two. (Wei Li peeled two fresh pears slowly.)</td>
</tr>
<tr>
<td>Semantic violation (SEMANTIC)</td>
<td>(1b) Wei Li ba (PREP)/ fresh pears slowly/ intimidated/ le (PERF)/ two. (Wei Li intimidated two fresh pears slowly.)</td>
</tr>
<tr>
<td>Syntactic violation (SYNTACTIC)</td>
<td>(1c) Wei Li ba (PREP)/ fresh pears slowly/ knife/ le (PERF)/ two. (Wei Li knife two fresh pears slowly.)</td>
</tr>
<tr>
<td>Combined syntactic and semantic violation (COMBINED)</td>
<td>(1d) Wei Li ba (PREP)/ fresh pears slowly/ piano/ le (PERF)/ two. (Wei Li piano two fresh pears slowly.)</td>
</tr>
</tbody>
</table>

Note. Examples are given in Chinese, with English literal glosses and translations. The critical words are in bold. PREP = preposition; PERF = perfective marker.
difficult to integrate this “bad” artifact noun into the preceding sentence context semantically or compute the semantic associations between this artifact noun and the preceding NPs.

At the artifact noun, if the detection of the syntactic category violation blocks the semantic integration between this noun and its preceding NPs, there should be no N400 differences between the COMBINED and the SYNTACTIC conditions. However, there should be a larger N400 for the COMBINED condition compared with the SYNTACTIC condition, if semantic integration proceeds even when the incorrect syntactic category results in an ill-formed syntactic structure.

Method

Participants. Sixteen students from Peking University gave their informed consent to participate in the experiment (mean age = 21 years, range = 19–25 years; 11 women). All were native speakers of Mandarin Chinese, were right-handed, had normal or corrected-to-normal vision, and had no known reading or neurological disorders. All were paid a nominal sum for their participation. The experiment was approved by the Academic Committee of the Department of Psychology, Peking University. One (female) participant was replaced because of a large number of slow voltage shift artifacts in the recordings.

Materials and normative measures. The critical materials consisted of 144 sets of Chinese sentences (see Table 1 for examples). These sentences were assigned to four experimental lists using a Latin square procedure. To counterbalance the number of correct and incorrect sentences in the experiment and to offset the ba structure used in the critical items, in each list, we mixed the 144 critical items with 144 filler sentences that consisted of 72 correct ba sentences, 36 correct SVO sentences, and 36 incorrect SVO sentences with semantic or combined syntactic category and semantic violations.

The correct main verbs and their anomalous replacements served as the critical words in the ERP experiment. At the critical word position, a transitive main verb was strongly expected because of the preceding syntactic context. This was confirmed by a cloze test, in which a separate group of 48 participants was presented with the sentence fragments up to but not including the main verb and was asked to complete the sentence fragments with the first words that come to mind. We randomly assigned 207 critical fragments (144 of them were used in the ERP experiment, on the basis of the results of all three pretests and the purpose of matching extraneous variables, see below) to three experimental lists, with 69 each. In each list, the 69 critical items were pseudorandomly mixed with 41 filler sentence fragments of other structures, such that at most three critical trials occurred consecutively. For each list, another version with a reverse order was formed to further counterbalance the order effects. Each participant received only one of the six experimental lists. For the 144 critical sentence fragments, participants completed them with an immediate transitive verb 98% of the time (the other immediate response words were adverbs or prepositions).

As mentioned earlier, the pure semantic violation was realized by replacing the transitive main verb in the correct condition with a transitive, but semantically incongruent, verb. The syntactic violation was realized by replacing the main verb in the correct condition with a “good” artifact noun, and the combined syntactic and semantic violation with a “bad” artifact noun.

To evaluate the degree of semantic violations, we conducted a pretest of sentence comprehensibility (i.e., meaningfulness), in which a separate group of 48 participants was asked to judge the comprehensibility of each sentence on a 5-point scale, with 1 indicating that the sentence was completely incomprehensible, and 5 indicating that the sentence was completely comprehensible. We assigned 199 sets of sentences to four experimental lists by using a Latin square procedure, with 199 sentences each (eight sets of sentences were not included in this pretest, on the basis of the results of the cloze test). In each list, the 199 items were presented in a pseudorandom order, such that at most three trials with the same condition occurred consecutively. For each list, another version with a reverse order was formed. Each participant received only one of the eight experimental lists. Table 2 shows the rating results for the 144 sets of items used in the ERP experiment. An analysis of variance (ANOVA) performed on the comprehensibility scores revealed a significant main effect of condition: by subjects, \( F(3, 141) = 659.23, p < .0005, \text{MSE} = 0.27 \); by items, \( F(3, 572) = 3,329.44, p < .0005, \text{MSE} = 0.11 \). Post hoc Newman–Keuls comparisons showed that the correct sentences were rated more comprehensible than all three types of sentences with violations (ps < .01). In addition, the sentences were more comprehensible in the SYNTACTIC-only condition compared with the SEMANTIC-only and the COMBINED conditions (ps < .01), whereas the sentences were rated equally incomprehensible for the latter two conditions (p > .10).

The difference in the comprehensibility between the SYNTACTIC and the other two violation conditions (SEMANTIC and COMBINED) was further confirmed by a sentence correction pretest, in which a separate group of 36 participants were presented with sentence fragments up to the anomalous words that were underlined and in bold, for example, *Wei Li ba (PREP) fresh pears slowly knife*. Participants were asked to read each sentence fragment and to

<table>
<thead>
<tr>
<th>Measure</th>
<th>Correct (CORRECT)</th>
<th>Semantic violation (SEMANTIC)</th>
<th>Syntactic violation (SYNTACTIC)</th>
<th>Combined syntactic and semantic violation (COMBINED)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>10.46 (16.82)</td>
<td>13.60 (59.52)</td>
<td>11.67 (40.68)</td>
<td>9.30 (30.11)</td>
</tr>
<tr>
<td>No. of strokes</td>
<td>16.94 (5.53)</td>
<td>17.64 (3.67)</td>
<td>16.82 (6.62)</td>
<td>17.42 (6.91)</td>
</tr>
<tr>
<td>Comprehensibility score</td>
<td>4.78 (0.23)</td>
<td>1.60 (0.31)</td>
<td>3.51 (0.43)</td>
<td>1.55 (0.32)</td>
</tr>
</tbody>
</table>

Note. Values in parentheses indicate standard deviations.
correct the anomalous word with the first grammatical and meaningful word or phrase that comes to mind. Forty-two sets of sentences were not included in this pretest, on the basis of the results of the sentence comprehensibility pretest. The remaining 157 triplets were assigned to three experimental lists by using a Latin square procedure, with 157 items each. In each list, the 157 items were presented in a pseudorandom order, such that at most three items in the same condition occurred consecutively. For each list, another version with a reverse order was formed. Each participant received only one of the six experimental lists. The correction was always replacement with a verb. Moreover, for the 144 items used in the ERP experiment, 78.2% (9.4/12) of the corrections were semantically related to the anomalous word in the SYNTACTIC condition, whereas the percentages were 0.2% and 0.5% for the SEMANTIC and the COMBINED conditions, respectively. These results may explain why the sentences were more comprehensible (meaningful) in the SYNTACTIC compared with the SEMANTIC and the COMBINED conditions.

The critical words were not ambiguous as to syntactic category, according to both the intuitions of the Chinese-speaking authors and the electronic dictionary that is based on the Corpus for Studies of Modern Chinese (Beijing Language and Culture University, 1995), which has 1.24 million words collected from a broad range of genres. In addition, all the critical words were matched for written lemma frequencies and number of strokes, as shown in Table 2. An ANOVA performed on the mean frequencies and number of strokes did not reveal any effect of condition ($F_s < 1$).

**Procedure.** Participants were seated in a comfortable chair approximately 1 m from the computer screen in a dimly lit and sound-attenuated room. They read the sentences sequentially, as each word (or sometimes a short phrase) appeared in the center of the screen. Each trial started with a central fixation cross presented for 800 ms, followed by a 500-ms blank screen. Each word or short phrase was presented for 400 ms, with an additional 100-ms interstimulus interval, during which the screen was blank. After the presentation of the last segment of the sentence, there was an 800-ms blank, followed one third of the time by a verification sentence to avoid forcing semantic integration in a syntactically or semantically anomalous sentence. Participants were asked to decide whether the sentence that was either a paraphrase of the sentence or a misparaphrase of it. Only grammatical and sensible critical and filler sentences were presented in a verification sentence to avoid forcing semantic integration in a syntactically or semantically anomalous sentence. Participants were asked to decide whether the verification sentence correctly expressed the content of the preceding sentence. For each condition, at most two trials in the same condition occurred consecutively (if responding correctly), and at most six $bu$ sentences occurred consecutively. For each of the four lists, another version with a reverse order was formed. Each participant received only one of the eight experimental lists. Prior to the experimental blocks, participants received a practice block of 32 trials (including 10 comprehension questions that all followed correct sentences). The experimental session lasted about 1 hr.

**ERP recording.** The electroencephalogram (EEG) was recorded from 62 Ag/AgCl electrodes mounted in an elastic cap (Quick-Cap, NeuroScan Inc., Herndon, VA). The electrodes were placed in the following sites: AF7, AF3, FP1, FPz, FP2, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, O2, and O3. Recordings were referenced to the left mastoid but were referenced to linked mastoids offline. The horizontal electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye, and the vertical EOG was recorded from electrodes placed above and below the participant’s left eye. Electro-impedances were kept below 5kΩ. The EEG and EOG were amplified with a band-pass from DC to 70 Hz and were recorded continuously with a digitization rate of 500 Hz. ERPs were additionally filtered offline with 30 Hz low pass for the plots only. All statistical analyses were performed on the original data.

**ERP data analysis.** ERPs time-locked to the critical words (the correct main verbs and their substitutes) were computed for each participant, condition, and electrode site. A detrending algorithm was applied to correct for a common linear component caused by the slow voltage shifts that are common for DC EEG recordings, as in Angrilli, Dobel, Rockstroh, Stegagno, and Elbert (2000); Fiebach, Schlesewsky, and Friederici (2002); and Phillips, Kazanina, and Abada (2005). We used a 15-s time interval, ranging from the onset of the critical words to 15 s after them, to estimate the linear component of any slow voltage shifts and to leave unaffected the relatively rapidly changing components of interest that were evoked by the words in our sentences (N400, LAN, and P600). The subsequent analyses were based on 1,200-ms epochs, ranging from 200 ms before the onset of the critical words to 1,000 ms after them. We used a 200-ms prestimulus baseline. All epochs were evaluated individually for EOG or other artifacts. Epochs with amplitudes exceeding ±100 μV were excluded from the averages through artifact rejection. The overall rejection rate was 12.6%, equal for all four conditions (CORRECT: 11.3%; SEMANTIC: 13.7%; SYNTACTIC: 12.0%; and COMBINED: 13.5%).

Three time windows were chosen on the basis of visual inspection and earlier studies: (a) 100–200 ms for possible ELAN effects, (b) 350–550 ms for LAN and N400 effects, and (c) 600–1,000 ms for P600 effects. All statistical analyses were performed on the mean amplitudes in the selected time windows. ERPs were analyzed separately for midline and lateral electrodes. Omnibus ANOVAs for midline electrodes included two within-subject factors: electrode (Fz/FCz/Cz/CPz/Pz) and condition (CORRECT/SEMANTIC/SYNTACTIC/COMBINED). Omnibus ANOVAs for lateral electrodes included three within-subject factors: hemisphere (left/right), region (anterior/central/posterior), and condition. Crossing the variables of region and hemisphere yielded six regions of interest, with six electrodes for each region of interest: left anterior (F3, F5, F7, FC3, FC5, and FT7), left central (C3, C5, T7, CP3, CP5, and TP7), left posterior (P3, P5, P7, PO3, PO7, and O1), right anterior (F4, F6, F8, FC4, FC6, and FT8), right central (C4, C6, T8, CP4, CP6, and TP8), and right posterior (P4, P6, P8, PO4, PO8, and O2).
Only effects involving the factor condition are reported. The Greenhouse–Geisser correction was applied when evaluating effects with more than one degree of freedom in the numerator. In these cases, the original degrees of freedom and the corrected mean square error and probability levels are reported.

**Results**

Participants had high accuracy on the YES/NO judgments of the verification sentences ($M = 96.46\%; \text{range} = 86.36\%–100\%$), suggesting that they read the sentences attentively. Figure 1 shows grand average ERPs elicited by the critical words for all four critical conditions (CORRECT, SEMANTIC, SYNTACTIC, and COMBINED) at 11 representative electrodes.

As shown in Figure 1, there was an LAN for both the SYNTACTIC and COMBINED conditions, beginning about 300 ms after the onset of the critical word. Beginning about the same time, there were some right anterior and central negativities, mainly for the COMBINED condition. The SEMANTIC condition was more negative compared with the CORRECT condition at the central-parietal midline and right posterior sites, and crucially, the COMBINED condition was more negative than the SYNTACTIC condition at right anterior and central sites. In addition, there was a P600 effect that was more pronounced for the COMBINED condition at Pz. These observations were statistically verified by ANOVAs performed on the mean amplitudes in the 100–200-ms, 350–550-ms, and 600–1,000-ms time windows, respectively. The results of the global ANOVAs are shown in Table 3.

The 100–200-ms time window. As shown in Table 3, at the midline electrodes, the global ANOVA revealed neither a main effect of condition nor an interaction of condition and electrode. At lateral electrodes, the global ANOVA revealed a $\text{Condition} \times \text{Region} \times \text{Hemisphere}$ interaction. Separate analyses limited to each region at each hemisphere revealed a main effect of condition only at the left central region, $F(3, 45) = 3.59, p = .038, MSE = 2.11$. Post hoc Newman–Keuls comparisons revealed a smaller negativity at the left central region for the SEMANTIC and COMBINED conditions compared with the SYNTACTIC condition ($ps < .05$). Note that the SEMANTIC and COMBINED conditions were the two conditions rated very low (incomprehensible) in the comprehensibility ratings (see Table 2).

It is tempting to view this early, left-lateralized negativity as a type of ELAN; however, an ELAN is typically elicited by conditions containing syntactic category violations (SYNTACTIC and COMBINED), not semantically anomalous conditions, as we observed. An alternative explanation is that the early negativities are an N1 effect, reflecting attentional effort: The more comprehensible the sentence was, the more attention the participants paid. Participants may have quickly reduced their attention to the SEMANTIC and COMBINED conditions because they learned that only correct sentences were followed by a comprehension question. Indeed, if participants learned to adjust their attentional effort, the learning is likely to have occurred during the 32 practice trials. A post hoc analysis with session (first vs. second half of the experiment) as an additional factor failed to find a difference in this very early effect between the first and second halves of the experiment (no interaction of condition with session, $ps > .26$). The possible reduced level of attention in the COMBINED condition relative to the SYNTACTIC condition would minimize the chance of observing the theoretically important N400 effect. However, the fact that a larger N400 was obtained for the COMBINED condition compared with the SYNTACTIC condition, as predicted (see below), suggests that participants still engaged in difficult semantic integration in the combined violation condition.

**Figure 1.** Grand average event-related brain potentials time locked to the onset of the critical words and from 11 scalp sites, for all four critical conditions in Experiment 1.
The 350–550-ms time window. At the midline electrodes, the global ANOVA revealed a Condition × Electrode interaction (see Table 3). Separate analyses limited to each electrode revealed a main effect of condition at Fz, F(3, 45) = 4.26, p = .014, MSE = 4.22; FCz, F(3, 45) = 3.65, p = .032, MSE = 5.02; and CPz, F(3, 45) = 3.86, p = .036, MSE = 7.25. Post hoc Newman–Keuls comparisons revealed a larger negativity at CPz for the SEMANTIC condition, a larger negativity at FCz for the SYNTACTIC condition, and a larger negativity at Fz and FCz for the COMBINED condition, all compared with the CORRECT condition (p < .05). In addition, the COMBINED condition was more negative compared with the SEMANTIC condition at Fz (p < .05).

At the lateral electrodes, the global ANOVA revealed a main effect of condition, a Condition × Region interaction, and a Condition × Region × Hemisphere interaction (see Table 3). Separate analyses limited to each region at each hemisphere revealed a main effect of condition at left anterior, left central, and all three right regions: left anterior, F(3, 45) = 10.49, p < .0005, MSE = 3.29; left central, F(3, 45) = 4.55, p = .012, MSE = 2.85; right anterior, F(3, 45) = 5.54, p = .005, MSE = 3.59; right central, F(3, 45) = 4.07, p = .021, MSE = 2.62; and right posterior, F(3, 45) = 3.48, p = .031, MSE = 2.67. Post hoc Newman–Keuls comparisons revealed a larger negativity at the right posterior region for the SEMANTIC condition (p < .05), a larger negativity at the left anterior and central regions for the SYNTACTIC condition (p < .01), and a larger negativity at the bilateral anterior and central regions for the COMBINED condition (p < .01), all compared with the CORRECT condition. In addition, compared with the SEMANTIC condition, the SYNTACTIC condition elicited a larger negativity at the left anterior region (p < .01), and the COMBINED condition elicited a larger negativity at the bilateral anterior regions (p < .01). Most importantly, compared with the SYNTACTIC condition, the COMBINED condition elicited a larger negativity at the right anterior region (p < .05).

To summarize, in both central-parietal midline (CPz) and right posterior regions, the SEMANTIC condition was more negative than the CORRECT condition, exhibiting the typical N400 pattern. An important theoretical question was whether a similar N400 would be observed for the COMBINED condition or whether semantic anomaly effects would be blocked by its ungrammaticality. In the right anterior region, we found that the COMBINED condition was more negative than the SYNTACTIC condition, suggesting that an N400 was indeed observed for the COMBINED condition, although with a more anterior distribution compared with the SEMANTIC condition.

In addition, we found anterior-central negativities for the conditions with syntactic anomalies, suggestive of an LAN effect. However, these negativities are difficult to interpret because they may reflect a combination of LAN effects caused by syntactic category violation, syntactic category effects in Chinese (compared with verbs, nouns often elicit a larger negativity over anterior sites in the 300–500-ms time window; Ma et al., 2007; Zhang et al., 2003), and/or an increased N400 reflecting difficulty in semantic integration.

The 600–1,000-ms time window. At the midline electrodes, the global ANOVA revealed a Condition × Electrode interaction (see Table 3). Separate analyses limited to each electrode revealed a main effect of condition at Pz, F(3, 45) = 3.12, p = .05, MSE = 8.92. Post hoc Newman–Keuls comparisons revealed a larger positivity at Pz for the COMBINED condition compared with the CORRECT condition (p < .05). A similar, but weaker, trend for the SYNTACTIC condition was not significant. Thus, a P600 effect was found only for the condition containing a combined syntactic and semantic violation. At the lateral electrodes, although the global ANOVAs revealed an interaction of condition and region (see Table 3), separate analyses limited to each region did not reveal any main effect of condition at each region (p > .10).

Discussion

To investigate whether semantic integration proceeds even when syntactic category processing fails, we compared the combined syntactic category and semantic violation with a syntactic violation alone, and we found a larger right anterior negativity in the 350–550-ms time window in our COMBINED condition compared with our SYNTACTIC condition. This larger negativity cannot be explained as reflecting a difference between the two conditions in syntactic category, word frequency, or number of strokes in the critical words; the critical words for both conditions were ungrammatical nouns, and their word frequencies and number of strokes were matched. Instead, the difference must be an N400 (semantic integration) effect. Note that although the classic semantic N400 has a centroparietal distribution, it has been also found with a more anterior distribution (Friederici & Frisch, 2000; Holcomb, 1993; Kutars & Van Petten, 1990).

The N400 observed for the combined violation condition compared with the syntactic-only violation condition is the strongest
evidence so far that words that are not syntactically licensed are nonetheless semantically integrated during Chinese ba sentence reading. If semantic integration had been blocked by the detection of a wrong syntactic category, as suggested in earlier German and French studies (e.g., Friederici et al., 2004, 1999; Isabel et al., 2007), there would be no N400 differences between the two syntactically anomalous conditions.

An alternative possibility is that the modulation of the N400 is caused by a difference in the contextual support for the critical words in the SYNTACTIC and COMBINED conditions, either because of expectations arising from the sentence context or because of semantic priming from words in the context. To address this possibility, we analyzed the data from the aforementioned cloze test and also conducted a word association test.

Recall that in the cloze test, 48 participants (16 for each item) were presented with the sentence fragments up to but not including the critical word and were asked to complete the sentence fragments with the first words that come to mind. We did not restrict participants’ responses to a single word. The material preceding the critical word was identical across conditions, so all four conditions could be evaluated simultaneously. We assessed the cloze value in two ways. First, we calculated the percentage of trials on which the participant supplied the critical word as the first word in the participant’s completion. This measure estimates participants’ expectation for the critical word at the point in the sentence where it actually occurred. Second, we included in the percentage any completion that contained the critical word, even if it was not the first word in the participant’s completion. In this way, we were able to evaluate whether the critical word was activated by the context, even if it was not expected at the point in the sentence when it occurred. As we anticipated, the CORRECT critical word was more predictable (37.0% of the time as the first word and 38.4% as any word in the completion) than any of the anomalous words, by either measure. None of the critical words in the three anomalous conditions ever occurred as the first word in the completion, but the critical word in the SYNTACTIC condition did occur later in the response 2.7% of the time (e.g., Wei Li ba fresh pears slowly WITH KNIFE PEELED LE [perfective marker], “Wei Li peeled the fresh pears with a knife slowly”).

In the association test, 50 participants (10 for each item) were asked to write down the first three words that come to mind when they saw a word list that consisted of the content words from the sentence context (except human names and very general adverbs, such as slowly). The association probability, that is, the proportion of participants who provided the critical word as one of the associates, was 1.0% for the critical words used in the SYNTACTIC condition and 0% for critical words used in the COMBINED and SEMANTIC conditions. The critical word was produced 2.2% of the time in the CORRECT condition.

In sum, although the CORRECT word occurred frequently in the cloze test, the SYNTACTIC critical word was no more expected to appear at the critical point in the sentence than the COMBINED critical word (in the most conservative cloze test measure). Nonetheless, the SYNTACTIC critical word (knife) may have been weakly activated by the words in the sentence context, as suggested by the 1%-3% response rates in the less conservative cloze measure and the association test. Whether such weak semantic associations could, themselves, be responsible for modulating the N400 effect is a matter that must be resolved by future research. Our concern in the current article is whether sentence-level semantic processing (also) modulated the N400 in the syntactically anomalous conditions—a possibility that we find quite plausible.

A semantic interpretation of who does what to whom could be built either by computing the semantic relationship between a verb and its arguments or by computing the semantic associations between NPs, especially if one of them is an artifact or instrument NP that may activate an associated verb. For the current sentences, the string “Wei Li ba fresh pears slowly” can be mapped onto the underspecified proposition: Wei Li acted upon the fresh pears slowly. If the next word is consistent with that proposition, such as “peeled” or “knife,” the word can be semantically integrated into the proposition fairly easily, even though “knife” is of the wrong syntactic category and will also result in a syntactic anomaly effect. On the other hand, neither “intimidated” nor “piano” could be easily integrated semantically. In short, a lexical semantic processing stream that is responsible for such computations, such as the one elaborated in Kuperberg’s (2007) study, could be the underlying mechanism of semantic integration when syntactic category or local phrase structure is violated.

Unlike Ye et al. (2006), who found no evidence of a P600 in their Chinese combined violation condition, we observed a P600 effect for the combined violation only. The absence of P600 effects for the syntactic-only violation has at least two possible explanations, both of which assume that Chinese readers were able to semantically integrate the syntactically anomalous word (knife). First, the P600 effect we obtained for the combined violation may reflect difficulty in the integration of syntactic and semantic information for final interpretation, a process corresponding to the third phase in a three-phase model (such as Friederici, 2002) rather than to pure syntactic processing. No difficulty would be expected during the third phase if semantic integration has been successful. Second, the P600 effect for the combined violation may reflect purely syntactic difficulty; however, because we used a sentence verification task that focuses on sentence content, the P600 response has been minimal when the semantics was well-formed (see Hahne & Friederici, 2002, for the influence of task requirements on the presence/absence of the P600). Either way, the P600 we observed is a syntax-related ERP response, reflecting either purely syntactic processing or difficulty in syntax-semantics integration.

Although the N400 effects for COMBINED versus SYNTACTIC provide evidence against the functional primacy of syntactic category over semantic processing during Chinese sentence reading, we sought additional evidence, using sentences and a task more similar to those used by Friederici and colleagues (e.g., Friederici et al., 2004, 1999; Hahne & Friederici, 2002) when they have observed syntactic primacy effects in German and French studies. In doing so, we hope to discover whether our semantic integration effects generalize beyond the Chinese ba construction and our sentence verification task.

First, in the German and French studies, whereas the syntactic category violations were very local, occurring relative to the immediately preceding word, the semantic violations occurred relative to more distant preceding words, as in the German version of *The door lock was in-the eaten* or in the French version of *The chair which is in the is sleeping*. This greater locality of the syntactic violation might cause the syntactic violations to be more salient or to capture more attention, relative to the semantic vio-
lations, resulting in no N400 effects for the combined violations (see Kutas, Van Petten, & Kluender, 2006, for an alternative interpretation). In contrast, in our Experiment 1, both syntactic category and semantic violations occurred relative to the same reference point in the preceding sentence context.

Second, in Experiment 1, we used a sentence verification task, which focuses on sentence content and which was not used in the previous German and French studies. Our choice of task might have both encouraged semantic integration in the syntactically anomalous conditions and modulated the P600 effect, as discussed above.

In Experiment 2, we used a design (including a task) that was very similar to that used by Friederici and colleagues in their German and French studies (e.g., Friederici et al., 2004, 1999; Hahne & Friederici, 2002). In addition, in contrast to Experiment 1, the critical word was always a noun.

**Experiment 2**

The goal of Experiment 2 was to discover whether we would again find an N400 effect for the combined syntactic category and semantic violation after modifying our experimental methodology and stimuli to more closely mirror the earlier German and French studies. Four conditions were formed by crossing the two factors local phrase (correct vs. incorrect) and verb’s selectional restriction (correct vs. incorrect).

As shown in Table 4, the correct sentence was a SVO sentence (CORRECT). The local phrase violation (SYNTACTIC), as in (2b), was realized by inserting the degree adverb hen (“very”) immediately before the first object noun of the correct sentence. In Chinese, hen can be followed by an adjective or a verb but not by a noun (Lü & Zhu, 1979/2002). The verb’s selectional restriction violation (SEMANTIC), as in (2c), was realized by replacing the verb in the CORRECT condition with a verb that was semantically incongruent with the object, inducing a violation of verb’s selectional restriction for its object. In (2d), the adverb hen was inserted immediately before the first object noun of the sentence with the selection restriction violation, inducing the local phrase and selection restriction violation (COMBINED). Note that the local phrase violation contained anomalies not only at the local phrase structure (syntactic category) level but also at the semantic level, given that the noun immediately following the degree adverb hen, as skirt in (2b), was not gradable. Thus, our SYNTACTIC violation could be construed as a weaker version of a combined violation. As demonstrated in the acceptability norms described below, our participants found the SYNTACTIC condition the most acceptable of the three violations conditions and the COMBINED violation the least acceptable.

The locality of the two violations in the COMBINED condition was designed to mirror the locality of the violations used by Friederici and colleagues in their German and French studies (e.g., Friederici et al., 2004, 1999; Hahne & Friederici, 2002), to see whether the N400 effect would disappear when the syntactic violation is more local than the semantic violation. Thus, our local phrase structure violation occurred relative to hen, the word immediately preceding the critical word. In contrast, the selection restriction violation occurred relative to the third word before the critical word, that is, the main verb, ate in (2d). Moreover, the semantic integration of the critical word occurred between a verb and noun (verb’s argument) instead of between NPs, as in Experiment 1.

If the semantic integration between the verb and its object argument is blocked by the local phrase structure violation, there should be an interaction between selectional restriction and local phrase violations, with respect to the N400 effect. That is, there should be an N400 effect to the selectional restriction violation when there is no local phrase structure violation (SEMANTIC vs. CORRECT), but there should be no N400 effects evoked by the selectional restriction violation when local phrase structure was violated (COMBINED vs. SYNTACTIC). On the contrary, a main effect of selectional restriction violation (no interaction with local phrase violation) would demonstrate that the semantic integration proceeds even when the phrase structure processing fails. Crucially, a larger N400 is predicted in the COMBINED condition compared with the SYNTACTIC condition. For example, upon reading “The girl bought/ate a very . . . ,” lexicosemantic constraints set up two kinds of expectations. The verb introduces an open thematic role corresponding to the thing that was bought or eaten, whereas “very” requires a gradable adjective. The noun “skirt” is thematically appropriate for “bought” but not for “ate,” predicting a difference in the magnitude of the N400, even though “skirt” is of the wrong syntactic (and semantic) category following “very.” Together with the results of Experiment 1, this pattern of results would provide evidence against the functional primacy of phrase structure building over semantic processes, at least in Chinese. Unlike Experiment 1, in Experiment 2, we used an overall acceptability judgment task, which has been used in some of the earlier studies that did not find N400 effects for the combined violation (Hahne & Friederici, 2002, Experiment 1; Isel et al., 2007).

**Method**

**Participants.** Twenty native Chinese-speaking students from Peking University gave informed consent to participate in the experiment (mean age = 22 years, range = 18–24 years; 10 women). All were right-handed, with normal or corrected-to-normal vision and with no known reading or neurological disorders. They were paid a nominal sum for their participation. The experiment was approved by the Academic Committee of the Department of Psychology, Peking University.

**Materials.** The critical materials consisted of 160 sets of sentences (see Table 4 for examples). Four experimental lists were created by rotating the conditions of the critical items. For each list, 160 critical items (40 for each condition) were pseudorandomly mixed with 360 filler items, which consisted of 240 correct sentences and 120 incorrect sentences containing syntactic (aspects) or semantic anomaly. The participants therefore received 280 correct and 240 incorrect sentences in total. Within the correct filler items, 80 contained the adverb hen, which was immediately

---

4 One might argue that hen can be followed by some nouns, for example, hen yishu (“very art”), hen nanren (“very man”), or hen shunv (“very gentlewoman”). At least for some Chinese linguists (e.g., Hu, 1992; Lü, 1989; Xing, 1962), however, words such as yishu (“art”), nanren (“man”), or shunv (“gentlewoman”) include syntactic category ambiguity (noun/adjective), and they are adjectives instead of nouns when they occur after hen.
followed by an adjective before the appearance of the object NP, to prevent participants from predicting the presence of a subsequent violation on the basis of the reading of *hen*.

The first object noun, which was constant across all four critical conditions, as *skirt* in Table 4, served as the critical word. To avoid any potential spillover effects, we matched the verbs between the CORRECT/SYNTACTIC and the SEMANTIC/COMBINED conditions for both written lemma frequencies (16 and 19 per million, respectively),  0.11. and number of strokes (17.4 and 17.7, respectively),  0.554.

To ensure that the sentences in the CORRECT condition were truly acceptable and that the sentences in the SYNTACTIC, SEMANTIC, and COMBINED conditions were unacceptable, we conducted a sentence acceptability pretest in which a separate group of 56 participants was asked to judge each sentence for acceptability on a 5-point scale (1 indicating that the sentence was completely unacceptable, and 5 indicating that the sentence was completely acceptable). We randomly assigned 180 sets of critical sentences (20 sets of them were not used in the ERP experiment, on the basis of the results of this pretest) to four experimental lists, with 180 items each. In each list, the critical items were pseudorandomly mixed with 40 correct filler sentences, such that at most three items in the same condition occurred consecutively. For each list, another version with a reverse order was formed to counterbalance any order effects. Each participant received only one of the four experimental lists. For the 160 sets of critical items, the correct sentences were highly acceptable (  4.48), and those with violations were highly unacceptable (  1.91, 1.56, and 1.27, for SYNTACTIC, SEMANTIC, and COMBINED, respectively). ANOVAs performed on these scores revealed a main effect of selectional restriction violation—  0.37;  4.642.28,  0.0005,  0.11—and local phrase violation—  385.93,  0.0005,  0.29;  3.004.98,  0.0005,  0.11—and an interaction—  260.32,  0.0005,  0.27;  1.906.12,  0.0005,  0.11.

Procedure. The procedure was the same as that used in Experiment 1, except for two changes. First, after the 800-ms blank that followed the last segment of each sentence, there was a row of question marks (“?????”) reminding participants to judge the overall correctness of each sentence by pressing one of two buttons. This cue remained on the screen until the participant had responded or for a maximum of 3,000 ms. Second, the intertrial interval was 1,000 ms.

Each participant received only one experimental list, in which the total of 520 sentences was divided into five blocks of 104 sentences each. Prior to the experimental blocks, participants completed a practice block of 26 trials. The experimental session lasted about 100 min.

### ERP recording and data analysis

The method of ERP recording and data analyses was the same as that used in Experiment 1, except for the following four changes. First, a 100-ms poststimulus baseline was used because the word immediately preceding the critical word differed between the conditions containing local phrase violation and the conditions containing no such violation (see Table 4; for the method of baseline selection, see Friederici et al., 2004; Isel et al., 2007). Second, a 300–500-ms and 500–1,000-ms time window was chosen for LAN/N400 and P600 effects, respectively (both effects started earlier in this experiment compared with Experiment 1). Third, the overall rejection rate due to artifacts and incorrect behavioral responses was 14.0%. Although the rejection rate was higher for CORRECT (16.4%) and SYNTACTIC (14.0%) compared with SEMANTIC (12.9%) and COMBINED (12.9%), the unequal rejection rate was not considered a problem because the rate was relatively low for an ERP experiment, even for CORRECT. Fourth, omnibus ANOVAs for midline electrodes included three within-subject factors: electrode (Fz/FCz/Cz/CPz/Pz), SYNTACTIC (local phrase: correct/incorrect), and SEMANTIC (verb’s selectional restriction: correct/incorrect). Omnibus ANOVAs for lateral electrodes included four within-subject variables: hemisphere (left/right), region (anterior/central/posterior), SYNTACTIC, and SEMANTIC.

### Results

Behavioral data. Average accuracy was computed as the percentage of correct responses. A correct response was the judgment of “acceptable” for the correct sentences and “unacceptable”
for the sentences with violations. Average accuracy was 96.64% ($SD = 5.24\%$) across all four conditions (CORRECT: 93.25\%, $SD = 6.02\%$; SYNTACTIC: 98.03\%, $SD = 2.76\%$; SEMANTIC: 95.90\%, $SD = 6.76\%$; and COMBINED: 99.38\%, $SD = 1.38\%$), suggesting that participants read the sentences attentively and that they distinguished the acceptable from the unacceptable sentences. An ANOVA performed on the accuracy data revealed both a main effect of local phrase violation, $F(1, 19) = 28.31, p < .0005, MSE = 12.02$, and a main effect of selectional restriction violation, $F(1, 19) = 4.18, p = .055, MSE = 19.15$. That is, participants were more accurate at rejecting sentences containing local phrase or selectional restriction violations than they were at accepting sentences containing no violations. This suggests a modest bias toward judging the experimental sentences to be unacceptable.

**ERP data.** Figure 2 shows grand average ERPs elicited by the critical words for all four critical conditions (CORRECT, SYNTACTIC, SEMANTIC, and COMBINED) at 11 representative electrodes. As shown in Figure 2, the selectional restriction violation evoked a broadly distributed negativity beginning about 300 ms after the onset of the critical word, regardless of whether the local phrase was violated simultaneously. During the same time period, compared with the conditions without local phrase violations (CORRECT and SEMANTIC), the conditions with such violations (SYNTACTIC and COMBINED) elicited a left-lateralized negativity. Beginning about 500 ms postset, the conditions with local phrase violations elicited a larger posterior positivity and a larger anterior negativity. In addition, the conditions with local phrase violations elicited an early negativity around 150 ms postset, similar to early negativities observed in Experiment 1. These observations were statistically verified by ANOVAs performed on the mean amplitudes in the 100–200-ms, 300–500-ms, and 500–1,000-ms time windows, respectively. The results of the global ANOVAs are shown in Table 5.

The 100–200-ms time window. As shown in Table 5, the ANOVA revealed a main effect of local phrase violation at both the midline and the lateral electrodes, with a larger negativity for the conditions with local phrase violations compared with the conditions without such violations. An early anterior negativity elicited by local phrase violations is often described as an ELAN effect, reflecting local phrase structure building (Friederici & Weissenborn, 2007). However, our early negativity is not just an ELAN effect, given its equal distribution over the anterior, central, and posterior electrodes. Instead, it may include an N1 effect reflecting differences in attentional effort, as we hypothesized in Experiment 1. In this experiment, the negativity may indicate that more attention was captured by the critical nouns exhibiting a local phrase violation compared with those without such violation, possibly driven by the sentence acceptability task used in the experiment: The local phrase violation should be a rapid cue for performing this task. In contrast, the verb’s selectional restriction violation, which occurred relative to a more distant word (the main verb), did not elicit an N1 effect. As in Experiment 1, further analyses with session (first vs. second half of the experiment) as an additional factor failed to find the difference in this very early effect between the first and second halves of the experiment, although this early effect was more frontal over the midline electrodes for the second half compared with the first half, resulting in a Local Phrase Violation × Electrode × Session interaction, $F(4, 76) = 4.41, p = .03, MSE = 0.99$. Thus, if the hypothesized attentional effect was learned during the experiment, it was likely learned during the practice trials. If participants were relying heavily on the local phrase violation as a rapid cue for rejecting a sentence, that strategy would minimize the possibility of observing a semantic incongruity effect in the combined violation condition. Nonetheless, an additional semantic violation (of the verb’s selectional restriction) actually still elicited an
N400 effect (see below), suggesting that participants still engaged in difficult semantic integration in the combined violation condition despite the possible increased level of attention to the local phrase violation.

**The 300–500-ms time window.** As shown in Table 5, the ANOVA revealed a main effect of selectional restriction violation at both the midline and the lateral electrodes, with a larger negativity for the conditions with a selectional restriction violation compared with the conditions without such a violation. At the lateral electrodes, there was also an interaction of a local phrase violation with hemisphere. Separate analyses limited to each hemisphere showed that the local phrase violation effect was significant at left hemisphere only—left, \( F(1, 19) = 6.82, p = .017, \text{MSE} = 1.45 \); right, \( F < 1 \)—with a larger negativity for the conditions with a local phrase violation compared with the conditions without such a violation. Most importantly, there was neither a Selectional Restriction \( \times \) Local Phrase Violation interaction nor an interaction of Selectional Restriction, Local Phrase Violation, and any location variable (Electrode, Region, or Hemisphere), indicating that the effect of a semantic violation was independent of whether local phrase structure was violated.

In sum, the selectional restriction violation evoked a broadly distributed negativity. This 

**The 500–1,000-ms time window.** As expected, there was a Local Phrase Violation \( \times \) Electrode interaction indicative of a P600 effect at the midline electrodes (see Table 5). Separate analyses limited to each electrode showed that the conditions with a local phrase violation elicited a larger positivity at CPz and Pz, that is, a P600, and a larger negativity at Fz; Fz, \( F(1, 19) = 8.51, p = .009, \text{MSE} = 2.44 \); FCz, \( F(1, 19) = 2.02, p = .171, \text{MSE} = 1.97 \); Cz, \( F(1, 19) = 2.33, p = .143, \text{MSE} = 1.68 \); CPz, \( F(1, 19) = 18.79, p < .0005, \text{MSE} = 1.89 \); and Pz, \( F(1, 19) = 66.63, p < .0005, \text{MSE} = 0.77 \).

At the midline electrodes, there was also a main effect of Selectional Restriction Violation, with a larger negativity for the conditions with selectional restriction violations compared with the conditions without such violations. At the lateral electrodes, the ANOVA revealed a Local Phrase Violation \( \times \) Region interaction. This interaction was due to a larger anterior negativity and a larger posterior positivity for both conditions with a local phrase violation (SYNTACTIC and COMBINED) and the absence of a reliable local phrase violation effect at the central region: anterior, \( F(1, 19) = 6.39, p = .021, \text{MSE} = 1.64 \); central, \( F(1, 19) = 3.11, p = .094, \text{MSE} = 0.93 \); and posterior, \( F(1, 19) = 6.29, p = .021, \text{MSE} = 1.85 \).

In sum, the conditions with local phrase violations elicited a larger posterior (including CPz and Pz) positivity and a larger anterior (including Fz) negativity. In addition, the conditions with selectional restriction violations elicited a larger negativity at the midline electrodes.

**Discussion**

Compared with Experiment 1, this experiment resulted in clearer examples of the typical components (N400 and P600) assumed to reflect syntactic and semantic effects. This puts us in an even stronger position to evaluate the hypothesis that semantic integration will occur even if the local syntactic phrase structure has been violated. As in Experiment 1, we conclude that semantic processing was not prevented by a phrase structure violation.

In Experiment 2, the syntactic category violations occurred relative to the immediately preceding word (the degree adverb *hen*), inducing a local phrase violation. For this violation, we
observed a left-lateralized negativity in the 300–500-ms time window, followed by a late (500–1,000-ms) posterior positivity (P600). In contrast, the verb selectional restriction violations did not elicit a P600 effect. The P600 effect for the syntactic violation conditions makes it clear that participants detected the syntactic category violation while reading the noun; P600 effects have often been observed for syntactic category violations in earlier studies in Indo-European languages (e.g., Friederici et al., 2004, 1996; Hagoort et al., 2003; Thierry et al., 2008; van den Brink & Hagoort, 2004).

The left-lateralized negativity elicited by the local phrase violation may be a combination of LAN and N400, given its scalp distribution. It is also possible that this negativity is purely an N400, with a broad distribution that includes the anterior, central, and posterior regions of the left hemisphere. If so, it would reflect difficulty in semantic integration caused by the degree adverb *hen* (“very”) being followed by a noun that cannot be gradable.

In addition, we found a larger broad negativity (N400) in the 300–500-ms time window for the verb selectional restriction violations. Crucially, this effect occurred independently of whether there was a simultaneous local phrase structure violation. This finding suggests that semantic integration between verb and noun (verb’s argument) proceeds even when there is a syntactic category violation at the noun.

Unexpectedly, both verb selection restriction violations and local phrase violations elicited a larger late (500–1,000-ms) negativity, over five midline and anterior (including Fz) electrodes, respectively. This negativity may reflect secondary semantic integration for the conditions containing syntactic violation (SYNTACTIC and COMBINED). Interestingly, Friederici et al. (1999) also observed a late (600–850-ms) negativity, although with a wide distribution over the right hemisphere, for both their semantic violation condition and their combined (syntactic category and semantic) violation condition.

Also unexpected, there was a broadly distributed, early negativity for local phrase violations. As discussed earlier, this effect may be an N1, indicating that the noun occurring after *hen* (the SYNTACTIC and COMBINED conditions) captured more attention compared with the same noun without a preceding degree adverb *hen* (the CORRECT and SEMANTIC conditions; see Table 4). Even if attention was modulated in this way, the theoretically important N400 effects elicited by the verb’s selectional restriction violation occurred even when the local phrase was violated.

### General Discussion

In conclusion, the present study provides evidence for the independence of semantic integration during Chinese sentence reading: semantic processing does not need a license from syntactic category processing for Chinese readers, even under conditions in which the reader judges the sentence to be incorrect, as in the violation conditions of our Experiment 2.

The purpose of the present experiments was to investigate whether there is a functional primacy of syntactic category over semantic processes, that is, whether semantic integration is blocked when syntactic phrase structure construction fails, during Chinese sentence reading. In Experiment 1, we compared combined syntactic category and semantic violations with syntactic violation alone. An N400 effect was obtained for this comparison, suggesting that semantic integration proceeds even when it is not licensed by syntactic processing. This finding differs from observations in earlier German and French studies, which have suggested that failed syntactic category processing blocks the process of lexical semantic integration, resulting in no N400 effects for combined violations. As discussed earlier, although Experiment 1 provided evidence against the functional primacy of syntactic category processing, we considered the possibility that the discrepancy between the results of our Experiment 1 and the earlier German and French studies was due to differences in violation reference points and/or the task.

Therefore, in Experiment 2, we used a design similar to that used in the earlier German and French studies. That is, we manipulated the correctness of the local phrase and selectional restrictions in the combined violation condition, such that the syntactic category violation and the semantic violation occurred relative to the immediately preceding word and the third word before it, respectively. The stimuli we used were also very closely matched to those used in previous German studies (e.g., Friederici et al., 2004, 1999; Hahne & Friederici, 2002). Moreover, we used the sentence acceptability judgment task with which the earlier German and French studies found no N400 effects for the combined violation. As it turned out, this experiment yielded even clearer evidence of the standard syntactic and semantic violation effects and greatly strengthened the case against syntactic primacy in Chinese sentence processing.

As in the German and French studies, we obtained a P600 effect for the local phrase violation in Experiment 2. This finding is very important, because it provides direct evidence for the failed syntactic processing. Like the P600 effect observed in Experiment 1 for the combined violation, the P600 effect in Experiment 2 reflects difficulties associated with either pure syntactic processing or syntax-semantics integration (recall that our local phrase violation also contained semantic anomaly, which was caused by the degree adverb *hen* being followed by a noun that cannot be gradable).

Crucially, in Experiment 2, we found a main effect of the semantic violation, instead of its interaction with local phrase violation. The lack of an interaction suggests that semantic integration between the verb and noun occurs independently of the success of syntactic category processing or the computation of local phrase structure.

In a very early ERP study in English, an N400 effect was found for a condition containing a syntactically incorrect (random word order) context (van Petten & Kutas, 1991). At first glance, this result seems similar to that of the present study. However, as Friederici et al. (1999) have commented, the syntactically incorrect context used by Van Petten and Kutas (1991) contained no syntactic structure at all because of the fact that all words were scrambled. The parser therefore could be “switched off” and no longer guide lexical integration, such that the obtained N400 effect could reflect just semantic integration in word lists. Interestingly, even such semantic integration seems to be blocked by the phrase structure violation in the German and French studies, suggesting a very strong functional primacy of syntactic category processing under several different task conditions.

If we are correct, and semantic integration can occur in the absence of local phrase structure computation, at least in Chinese, how does this occur? One possibility is that the locally ill-formed
syntactic structure was repaired (e.g., by deleting hen ["very"] in Experiment 2), allowing semantic integration to take place on the repaired syntactic structure and manifest itself as the observed N400 effect. Indeed, we obtained a P600 effect in both experiments, which could reflect syntactic repair. However, this explanation is difficult to reconcile with the apparently similar time window of the N400 effects for semantic violations, regardless of whether syntax was violated simultaneously. Furthermore, this explanation does not account for the fact that no such N400 effect (reflecting the semantic processes occurring on an already repaired syntactic structure) was obtained in previous studies in which the syntactic repair could also have occurred, as in the German version of *The door lock was in-the eaten, in which the preposition-determiner contraction could be deleted (e.g., Friederici et al., 1999; Hahne & Friederici, 2002, Experiment 1). Thus, although a syntactic repair may well have occurred, it is unlikely that the N400 effect that we observed reflects semantic processes occurring on an already repaired syntactic structure.

In the discussion of Experiment 1, we considered whether the observed N400 effects might reflect the predictability of the critical words either on the basis of a highly constraining preceding sentence context or on the basis of semantic priming (see Kutas & Federmeier, 2000, for a review of such N400 effects) rather than sentence-level semantic integration of an unexpected word. In Experiment 1, it was crucial that the N400 was larger in the COMBINED condition, compared with the SYNTACTIC condition, even though both conditions used a syntactically anomalous noun as the critical word. We concluded that sentence-level semantic comprehension processes continued, even as syntactic processing was disrupted. However, as discussed above, it is also possible that the SYNTACTIC critical word was subtly activated via semantic associations, whereas the COMBINED critical word was not. The crucial N400 effects in Experiment 2 are not complicated by differences in contextual support because the most important finding was that an N400 that was induced by a selectional restriction (you cannot eat a skirt) was not modulated by the presence/absence of a local phrase violation (the presence/absence of a degree adverb).

In sum, an explanation for the modulation of N400 effects across the two experiments cannot be based solely on semantic associations or predictability. Instead, participants in our experiments most likely computed semantic relationships between content words within a sentence (the NPs in Experiment 1, such as Wei Li, pears, and knife, and the verb and its arguments in Experiment 2, such as bought and skirt), using a so-called “integrative” mechanism (Federmeier & Kutas, 1999; for review, see Kutas & Federmeier, 2000), in which the incoming critical words (nouns) were checked with the context information. A lexical semantic processing stream that is responsible for such computations, such as the one elaborated in Kuperberg’s (2007) study, provides a possible underlying mechanism of semantic integration when syntactic category or local phrase structure is violated.

Of course, under most circumstances, the syntactic structure of a sentence constrains the semantic interpretation of that sentence. For example, syntactic analysis prevents the adoption of a plausible, but syntactically ill-formed, interpretation in most cases (see Boland, 1997). However, as Kim and Osterhout (2005) have argued, combinatory semantic processing or “semantic attraction,” as in The meal was devouring . . . , can override morphosyntactic constraints under some circumstances. Indeed, they found that the main verbs (e.g., devouring) in such sentences elicited a P600 rather than an N400 effect (see Kuperberg, 2007, for a review of studies observing such pattern), suggesting that the theme interpretation of the subject noun, which was signaled by the semantic rather than syntactic cues, rendered the syntactically well-formed main verb syntactically anomalous. Our N400 finding suggests that the syntactic constraints on semantic processing do not necessarily mean that failed syntactic category processing blocks semantic interpretation.

Although we find Kuperberg’s (2007) theory appealing, it remains to be seen whether it can account for all the data. For example, why does semantic integration come and go, depending on the task, in German, whereas semantic integration seems to occur consistently in Chinese? Kuperberg’s lexical integration system could be responsible for the N400s observed following local phrase violations in German and English, when the task emphasized semantic coherence (Hahne & Friederici, 2002, Experiment 2) or meaningfulness (Thierry et al., 2008) at the expense of grammaticality. However, such system/processing stream appears to be overridden or blocked by failed syntactic category processing in German and French under other task conditions (overall correctness judgment or probe verification; Friederici et al., 2004, 1999; Hahne & Friederici, 2002, Experiment 1; Isel et al., 2007).

There seems to be an underlying assumption among some scholars (e.g., Hahne & Friederici, 2002) that judging overall (syntactic and semantic) correctness is a more neutral task than focusing on meaning, in terms of balancing the proportion of syntactic and semantic processing. However, one could make an alternative argument that acceptability judgments are less natural than focusing on meaning. In the comprehension of spoken language and informal electronic text, in which disfluencies often render the input string ungrammatical, it is quite natural to focus on semantic coherence while ignoring at least some types of syntactic anomalies. Furthermore, in all forms of written and spoken comprehension, the primary goal of the reader/listener is to recover the intended message.

What then, are we to make of the contrast between the syntactic primacy results found in German and French and the lack of syntactic primacy we find in Chinese, when all participants were asked to judge the overall correctness of the sentence? One possibility is that speakers of any language can operate in “syntactic primacy mode” or “independent semantics mode,” but there are consistent cultural differences in how language communities interpret the experimental task, for example, an overall acceptability judgment. If our Chinese participants were more focused on semantic coherence and meaningfulness than the French and German participants who produced the syntactic primacy findings, then that would provide a coherent account of all the results. Further research is necessary to see whether the Chinese participants can be forced into syntactic primacy mode.

A second possibility is that properties of the respective languages cause comprehenders to rely more or less on the computation of syntactic structure as laying the essential framework for interpretation. As we noted at the outset, some scholars have proposed that sentence comprehension in Chinese relies more on semantic analyses compared with Indo-European languages (e.g., Xu, 1997; also see Li, 1998; Ye et al., 2006). The inconsistent
findings between the present Chinese study and the earlier German and French studies imply that the degree to which syntactic category is made explicit in the morphology of the language might be one property that is relevant. The more you have to rely upon semantics to identify the syntactic categories of the words in the input, the more likely you are to find that semantic integration can proceed without a well-formed syntactic analysis. If so, language experience may modulate the mode of interoperable syntax and semantics. This hypothesis could be further tested by using Hahne and Friederici’s (2002) paradigm across a wider range of languages that differ in the degree to which syntactic category is marked morphologically. Again, we look to future research to assess this admittedly speculative hypothesis.

Either way, there is no need to favor processing models that assert syntactic primacy relative to semantic interpretation. Quite the opposite. To account for all the cross-linguistic data, a processing model must have a means for semantic integration in the face of severe grammatical violations.

References


