Available online at www.sciencedirect.com
SciVerse ScienceDirect
www.elsevier.com/locate/brainres

Research Report

No semantic illusions in the "Semantic P600" phenomenon: ERP evidence from Mandarin Chinese

Wing-Yee Chow[a,*], *Colin Phillips*[a,b]

[a]*Department of Linguistics, University of Maryland, 1401 Marie Mount Hall, College Park, MD 20742, United States*
[b]*Neuroscience and Cognitive Science Program, University of Maryland, United States*

ARTICLE INFO

Article history:
Accepted 7 February 2013
Available online 17 February 2013

Keywords:
Event-related potential
Sentence processing
N400
P600
Semantic illusion
Mandarin Chinese

ABSTRACT

Recent observations of unexpected ERP responses to grammatically well-formed role-reversed sentences (the "Semantic P600" phenomenon) have been taken to bear directly on questions about the architecture of the language processing system. This paper evaluates two central pieces of evidence for accounts that propose a syntax-independent semantic composition mechanism, namely the presence of P600 effects and the absence of N400 effects in role-reversed sentences. Experiment 1 examined the relative contribution of the presence of an animacy violation and the semantic relations between words ('combinability') to the ERP responses to role-reversed sentences. Experiment 2 examined the ERP responses to role-reversed sentences that are fully animacy-congruous. Results from the two experiments showed that animacy-violated sentences with no plausible non-surface interpretation elicited the same P600 effect as both types of role-reversed sentences; additionally, semantically anomalous target words elicited no N400 effects when they were strongly semantically related to the preceding words, regardless of the presence of animacy violations. Taken together, these findings suggest that the presence of P600s to role-reversed sentences can be attributed to the implausibility of the sentence meaning, and the absence of N400 effects is due to a combination of weak contextual constraints and strong lexical association. The presence of a plausible non-surface interpretation and animacy violations made no unique contribution to the ERP response profiles. Hence, existing ERP findings are compatible with the long-held assumption that online semantic composition is dependent on surface syntax and do not constitute evidence for a syntax-independent semantic composition mechanism.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Surface syntax is critical in determining the meaning of a sentence. Two sentences with the same words ordered differently (e.g., (1) and (2)) can have drastically different meanings:

1. The rebels killed the king.
2. The king killed the rebels.

Given the ease with which we detect the difference in meanings in sentences like (1) and (2), it can perhaps be taken for granted that we use surface syntax to compute the meaning of a sentence. In fact, most models of human sentence processing (e.g., Ferreira and Clifton, 1986; MacDonald et al., 1994; Trueswell et al., 1994) have assumed that surface syntax is always used to guide online semantic composition.

*Corresponding author. Fax: +1 301 405 7104.
E-mail address: wychow@umd.edu (W.-Y. Chow).

0006-8993/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.brainres.2013.02.016

The assumption that semantic composition relies on surface syntax should not be confused with the "syntax-first" position in the debate over online syntactic analysis in the study of structural ambiguity resolution. Although there are disagreements over whether syntactic information has priority over other sources of information, such as lexical bias, in online syntactic analysis (Ford et al., 1982; Frazier, 1987; Pickering et al., 2000; Trueswell et al., 1993), it is commonly assumed that only analyses that are compatible with the surface syntax are ever considered. Similarly, the view that semantic interpretation combines word meanings in accordance with syntactic constraints is independent of claims that syntactic anomalies are more rapidly detected than semantic anomalies (Friederici, 1995; McElree and Griffith, 1995). The assumption that semantic interpretation is based on the syntactic structure of the sentence is related to the claim that syntactic anomalies block the detection of semantic anomalies (e.g., Friederici et al., 1999; Hahne and Friederici, 2002), but these are logically distinct claims.

However, this assumption has not gone unchallenged (e.g., Bever, 1970; Caramazza and Zurif, 1976; Ferreira et al., 2002; Slobin, 1966; Townsend and Bever, 2001; Jackendoff, 2002). In fact, many have argued that the recent discovery of the "Semantic P600" phenomenon in the electrophysiological literature directly challenges this assumption (e.g., Kim and Osterhout, 2005; Kolk et al., 2003). These studies used event-related potentials (ERPs) to examine brain responses to fully grammatical sentences that contradict stereotypical thematic relationships ("role-reversed sentences", e.g., a criminal arresting a policeman, as opposed to being arrested by a policeman). The amplitude of the N400, a centro-parietal negative-going waveform peaking at around 400 ms after stimulus onset, is generally modulated by the cloze probability and semantic/pragmatic congruity of the word in a given context (e.g., Kutas and Hillyard, 1980, 1984; Van Berkum et al., 2008; Van Berkum, 2009). The P600, on the other hand, is a late posterior positive-going ERP waveform that has been associated with the presence of grammatical anomalies and syntactic processing difficulty (e.g., Osterhout and Holcomb, 1992; Hagoort et al, 1993). Interestingly, although role-reversed sentences are clearly semantically anomalous, they typically fail to elicit a larger N400 than their canonical control (e.g., Hoeks et al, 2004; Kim and Osterhout, 2005; Kim and Sikos, 2011; Kolk et al., 2003; Kuperberg et al, 2003, 2006, 2007; Stroud and Phillips, 2012; Van Herten et al., 2005, 2006). Further, despite being fully grammatical and structurally unambiguous, role-reversed sentences consistently elicit a larger P600 compared to the canonical control condition.

In this paper, we will refer to the phenomenon that grammatically well-formed role-reversed sentences elicit (i) only a P600 effect and (ii) no N400 effects as the "Semantic P600" phenomenon. Various accounts of the phenomenon have proposed processing architectures that assume a semantic interpretation mechanism that is independent of surface syntax, i.e., an *independent semantic composition* mechanism, and thereby challenge the assumption that online semantic composition relies on surface syntax (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Hagoort et al, 2009; Hoeks et al., 2004; Kim and Osterhout, 2005; Kolk et al., 2003; Kuperberg, 2007; Van Herten et al., 2005, 2006; Van de Meerendonk et al., 2009). An influential study by Kim and Osterhout (2005) examined ERP responses to unambiguous, grammatically well-formed sentences that depict an anomalous thematic relation (e.g., (3) and (4)). They reported that semantically anomalous sentences with a "semantically attractive" predicate–argument combination (e.g., (3), in which *meal* is a likely Theme argument for *devour*) elicited only a P600 effect and no N400 effect. In contrast, semantic anomalies such as (4), where the predicate and its argument are not semantically attractive, elicited only an N400 effect and no P600 effects.

3. Semantic anomaly with a plausible non-surface interpretation: The hearty meal was *devouring*… (control: the hearty meal was *devoured*).
4. Semantic anomaly (no plausible non-surface interpretation): The dusty tabletops were *devouring*… (control: the hearty meal was *devoured*).

Kim and Osterhout (2005) present a two-part argument that online semantic composition can be independent of surface syntax. First, when the subject and the verb are semantically attractive, as in (3), the processor constructs a plausible semantic representation, i.e., *the hearty meal* as the Theme of *devour*, even if it contradicts what is unambiguously dictated by surface syntax, i.e., *the hearty meal* as the Agent of *devour*; henceforth a "non-surface interpretation". Therefore, the processor is blind to the semantic anomaly in the input (a 'semantic illusion', see also Hoeks et al., 2004) and hence no N400 effects are elicited. Meanwhile, since the surface syntax of the input conflicts with that of the semantic representation computed, the processor in turn perceives the sentence as ungrammatical, resulting in a P600 effect. Second, when the subject and the verb are not semantically attractive, as in (4), and therefore no plausible semantic interpretation can be constructed, even by altering the structure or word order of the sentence, the processor perceives the sentence as semantically anomalous and generates an N400 effect and no P600 effect. Taken together, Kim and Osterhout argued that these results show that the processing system uses the meaning of individual words to compute a plausible interpretation, even when surface syntax unambiguously conflicts with that interpretation.

In sum, both the presence of a P600 effect and the absence of N400 effects have been taken as evidence for an independent semantic composition mechanism. Below, we use evidence from Mandarin Chinese to evaluate these two key pieces of evidence in turn, and propose that (i) the presence of a P600 effect in role-reversed sentences may be attributed to factors that are independent from, but often confounded with, the presence of plausible non-surface interpretations; and (ii) the absence of N400 effects in role-reversed sentences is attributable to a combination of lexical priming and weak contextual constraints. Most of the comparisons presented in this study build upon previous studies, and our conclusions have precursors in the literature. The primary contribution of the current study is that it takes advantage of the

properties of Mandarin Chinese to better assess proposals for syntax-independent semantic composition and the impact of factors such as animacy and implausibility.

1.1 When do semantic anomalies elicit a P600 effect?

Among the accounts that assume a processing architecture with a syntax-independent interpretation mechanism, several of them maintain that certain semantic anomalies elicit a P600 effect because the processor computes plausible interpretations that are incompatible with the surface syntax (e.g., Kim and Osterhout, 2005; Kolk et al., 2003; Van Herten et al., 2005, 2006). The strongest evidence for this account involves arguments that the P600 response to semantic anomalies is selective. If semantic anomalies elicit a P600 response only if a plausible non-surface interpretation is available, then this suggests that the non-surface interpretation plays a role in the processing of the sentence. On the other hand, if the P600 effect is elicited by semantic anomalies regardless of the availability of a plausible non-surface interpretation, then the observation of P600 effects in role-reversed sentences is compatible with accounts that assign no role to computation of non-surface interpretations.

To date, however, evidence for such selectivity is rather limited. Many studies have shown that semantic anomalies can elicit a P600 effect regardless of the availability of a plausible non-surface interpretation (e.g., Hoeks et al., 2004; Kuperberg et al., 2006, 2007; Paczynski and Kuperberg, 2011; Stroud and Phillips, 2012; Van Herten et al., 2006). For example, Hoeks et al. (2004) found that, along with role-reversed sentences such as The javelin has the athletes *thrown* (Dutch: De speer heeft de atleten *geworpen*.), semantically anomalous sentences that lack a plausible non-surface interpretation, such as "*The javelin has the athletes summarized*." (Dutch: De speer heeft de atleten opgesomd.), also elicited a significant P600 effect. Similar findings have been reported in studies across different languages, consistently showing that the presence of P600 effects to semantic anomalies is not restricted to cases in which a plausible non-surface interpretation is available (e.g., English: Kuperberg et al., 2006, 2007; Paczynski and Kuperberg, 2011; Stroud, 2008; Dutch: Van Herten et al., 2006; Spanish: Stroud and Phillips, 2012; Japanese: Oishi and Sakamoto, 2010).

In light of the finding that the P600 is not selectively elicited by role-reversals, some authors have proposed that other factors can elicit semantic P600s. Some of these proposals still assume some form of syntax-independent semantic interpretation mechanism (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Kuperberg, 2007; Van de Meerendonk et al., 2010). For example, Kuperberg (2007) emphasized that P600 effects to semantic anomalies are not solely modulated by thematic role-reversals. She identified that implausibility, along with the presence of animacy violations, played a key role in evoking a P600 effect in semantically anomalous sentences (see also Paczynski and Kuperberg, 2011). Meanwhile, Van de Meerendonk et al. (2010) proposed that the P600 is modulated by the severity of the conflict between what is expected (i.e., likely to be true) and what is observed, and found evidence that deeply implausible sentences such as *The eye consisting of among other things a pupil, iris, sticker…* elicit a larger P600 response than mildly implausible sentences such as *The eye consisting of among other things a pupil, iris, eyebrow…*

On the other hand, others proposals do not assume a syntax-independent semantic composition mechanism and have argued that the P600's sensitivity to role-reversals can be fully attributed to surface properties of the materials (e.g., Brouwer et al., 2012; Stroud, 2008; Stroud and Phillips, 2012). For example, Stroud (2008) observed that much existing evidence of P600 effects to role-reversals comes from studies that have confounded role-reversals with animacy violations. For instance, the role-reversal anomaly in (3) also involves a violation of the verb's requirement for an animate Agent. Stroud (2008) suggested that such P600 effects are attributable to the detection of animacy violations and therefore should not be taken as evidence for independent semantic composition. Meanwhile, Van Petten and Luka (2012) suggested that the P600 reflects reanalysis processes that are triggered by the detection of implausibility, whereas Brouwer et al. (2012) proposed that the P600 reflects the process of integrating the lexical information activated by a word into the current mental representation.

In order to evaluate whether factors such as the availability of plausible non-surface interpretations, animacy violations and implausibility make a unique contribution to the P600, comparisons need to be made between ERP responses to independent manipulations of non-surface plausibility and animacy congruity. However, only two studies to date (one in Dutch: Van Herten et al., 2005; one in Mandarin Chinese: Ye and Zhou, 2008) have examined the effects of thematic role-reversals using fully grammatical and animacy-congruous sentences. Both these studies used clauses with a subject–object–verb (SOV) word order, e.g., (5a) vs. (5b), and reported that role-reversal anomalies elicit a P600 effect and no N400 effect.

5. Role-reversal anomaly in animacy-congruous sentences (dutch).
   a. De stroper die op de vos *joeg* slopen door het bos. The poacher$_{[singular]}$ that at the fox$_{[singular]}$ hunted$_{[singular]}$ stalked through the woods. "The poacher that hunted the fox stalked through the woods."
   b. De vos die op de stroper *joeg* sloop door het bos. The fox$_{[singular]}$ that at the poacher$_{[singular]}$ hunted$_{[singular]}$ stalked through the woods. "The fox that hunted the poacher stalked through the woods."

Since these studies differed from those that examined animacy-violated role-reversals in many respects (e.g., language, word order of the sentence, the grammatical category of the target word), it remains difficult to compare across studies to determine to whether the availability of plausible non-surface interpretations, animacy violations, and/or mere implausibility contribute uniquely to the P600 effects observed in role-reversed sentences. Therefore, in the present study we aim to provide a more rigorous test by comparing ERP responses to manipulations of animacy congruity and non-surface plausibility.

1.2 When do semantic anomalies fail to elicit an N400 effect?

Although it has attracted less attention than the P600 effects elicited by semantic anomalies, the N400's insensitivity to role-reversal anomalies is also surprising and central to arguments for independent semantic composition. Based on the functional interpretation of the N400 as reflecting the process of computing a coherent semantic representation by incorporating each new word into its context (e.g., Brown and Hagoort, 1993; Hagoort et al., 2004), several existing accounts have interpreted the lack of N400 effects in role-reversed sentences as evidence that the parser temporarily fails to detect the semantic anomaly in role-reversed sentences, i.e., a 'semantic illusion' (e.g., Kolk et al., 2003; Van Herten et al., 2005, 2006; Kim and Osterhout, 2005; Hagoort et al., 2009). For example, Van Herten et al., 2005 proposed that the lack of N400 effects reflects that comprehenders initially consider the interpretation that fits their world knowledge best. According to this hypothesis, a role-reversed phrase such as *the cat that fled from the mice* is initially interpreted as the assertion that the mice are fleeing from the cat, since "this describes a far more plausible real life event than the situation that the cat is fleeing from the mice" (p. 252). Meanwhile, Kuperberg (2007) proposed that the attenuation of the N400 in semantic P600 cases is driven by a "non-combinatorial semantic memory-based mechanism (that) computes the semantic features, associative relationships and other types of semantic relationships between content words (including verbs and arguments) within a sentence, and compares these relationships with those that are pre-stored within lexical semantic memory" (p. 37). Taken together, these accounts posit that the processor can ignore surface syntax to compute a plausible interpretation in role-reversed sentences and therefore is effectively (temporarily) blind to the semantic anomaly and thus experiences no difficulty in semantic interpretation.

It has also been proposed that animacy information makes a unique contribution to the N400's sensitivity to semantic anomalies (Bornkessel-Schlesewsky and Schlesewsky, 2008; Kuperberg et al., 2007; Paczynski and Kuperberg, 2011). For instance, Bornkessel-Schlesewsky and Schlesewsky (2008) noted that arguments with a dispreferred animacy feature (e.g., an inanimate subject, or an animate object) elicited larger N400 responses and proposed that the N400 reflects core argument interpretation based on prominence information such as animacy in addition to syntax-independent computation of plausible interpretations. Meanwhile, Kuperberg and colleagues observed that animacy-violated semantically incongruous sentences do not elicit an N400 effect and proposed that full semantic analysis, as indexed by the N400, can be 'switched off' when a reader's animacy-based expectations are violated (e.g., Kuperberg et al., 2007).

In this paper, we test the hypothesis that neither non-surface plausibility nor animacy violations make a unique contribution to the N400. We adopt a lexical access account of the N400, according to which N400 amplitude reflects the cost of access to a lexical entry in the lexicon (Deacon et al., 2000; Federmeier and Kutas, 1999; Kutas and Federmeier, 2000; Laszlo and Federmeier, 2009; Lau et al., 2008). We propose that the absence of N400 effects in role-reversed sentences reflects that the cost of accessing the target verb in the lexicon does not differ between the canonical and role-reversed conditions due to a combination of weak contextual constraint and strong lexical semantic association, and not due to the plausibility of a non-surface interpretation or to the presence of animacy violations. Based on previous findings regarding the effects of contextual constraint and lexical association on the N400, we aim to relate evidence of the N400's insensitivity in role reversals to other cases in which the N400 has been found to be insensitive to semantic anomalies.

A number of previous studies have found evidence of the N400's insensitivity to the compositional semantic meaning of a sentence. But these findings have previously been analyzed as independent phenomena. For example, Fischler and colleagues examined ERP responses to semantic anomaly in affirmative and negated sentences (Fischler et al., 1983). They observed that, for affirmative sentences like (6), false sentences elicited a larger N400 compared to true sentences. However, in negated sentences like (7) it was the true sentences that elicited a larger N400. Based on the assumption that the N400 reflects sentence meaning computation, the authors suggested that their results support a two-step theory of negation (e.g., Carpenter and Just, 1975), according to which the meaning of a proposition such as *A robin is not a bird* is hypothesized to be computed initially without the negation as *A robin is a bird*, and the semantic effect of negation is only computed in a second step. Under this account the N400 reflects only the first of these two steps.

6. Affirmative sentences: A robin is a *bird/tree*.
7. Negated sentences: A robin is not a *tree/bird*.

More recently Urbach and Kutas (2010) reported that the N400 is insensitive to semantic incongruity in sentences with certain types of quantifiers. They examined ERP responses to sentences such as (8) and (9) and observed that the atypical object (e.g., *worms*) elicited a larger N400 than the typical object (e.g., *crops*) in all cases, despite the fact that the relative semantic congruity in the *most/often* sentences is reversed in the *few/rarely* sentences. That is, in the *most/often* sentences the N400 amplitude was larger in the semantically incongruous conditions than in the congruous conditions, but in the *few/rarely* sentences the N400 amplitude was in fact smaller in the semantically incongruous conditions than in the congruous conditions. Based on this pattern of results, the authors suggested that semantic processing of quantifiers such as *most* and *often* occurs rapidly and incrementally, whereas quantifiers such as *few* and *rarely* are processed more slowly.

8. Sentences with noun phrase quantifiers:
(a) most farmers grow *crops/worms*.
(b) Few farmers grow *crops/worms*.
9. Sentences with adverbial quantifiers:
(a) farmers often grow *crops/worms*.
(b) Farmers rarely grow *crops/worms*.

One important similarity between these studies and previous studies on role-reversals may be the relatively low predictability of the target word in congruous and incongruous conditions alike, given that their sentence contexts are often minimally predictive. For example, in the case of negated sentences, given a context like "*A robin is not a …*", the range of possible continuations is very broad, and hence an incremental processor might not expect the congruous target word *tree* any more than the incongruous target word *bird*. The sentence contexts in these studies do not provide sufficient information to facilitate access to the congruous target word relative to the incongruous target word. Under these circumstances it should not be surprising that the amplitude of the N400 is not reduced in the congruous condition relative to the incongruous condition.

In fact, a recent study by Nieuwland and Kuperberg (2008) contrasted ERP response profiles for sentences in which negation was pragmatically licensed (e.g., "With proper equipment, scuba-diving isn't very *dangerous/safe…*") vs. those in which negation was pragmatically unlicensed (e.g., "Bulletproof vests aren't very *dangerous/safe…*"). They found that in the conditions with pragmatically unlicensed negation, which were compatible with many possible continuations, N400 amplitudes were not reduced in the congruous condition. But in the conditions with pragmatically licensed negation, which more tightly constrains the likely continuations, the N400 was reduced in the congruous condition relative to the incongruous condition.

Further, a recent study by Bornkessel-Schlesewsky et al. (2011) examined the effects of role-reversals by swapping the case marker or word order of an animate and an inanimate argument in verb-final sentences in Turkish and Mandarin Chinese. They found that the verb sometimes elicited a larger N400 in the role-reversed condition than in the canonical control condition. Although the authors attributed the contrast between the presence of an N400 effect in their studies and the absence of N400 effects in previous studies to whether the language studied has rigid or flexible word order, it is plausible that the N400 effect reflected that the processor uses the animacy feature of the arguments to predict different verbs in the canonical vs. role-reversed sentences, since the canonical sentences in these studies always had an animate Agent and an inanimate Theme and the opposite is true for the role-reversed sentences.

However, the low predictability of the target words alone does not explain why N400 amplitude was in fact larger in the congruous condition than in the incongruous condition in the studies by Fischler et al. (1983) and Urbach and Kutas (2010). Both these studies compared ERP responses to lexical items that differed in terms of their semantic relatedness to the words in the preceding sentence context. For example, in sentences such as (8) and (9), the typical object "*crops*" is more closely associated to the context words "*farmers*" and "*grow*" than the atypical object "*worms*" is. The N400 amplitude is known to be reduced by semantic priming in word lists (e.g., Rugg, 1985) as well as in sentences (e.g., Camblin et al., 2007; Ditman et al., 2007). Therefore, in a situation where the compositional meaning of the sentence context does not make one target word more expected than the other, it is unsurprising that the N400 amplitudes are modulated by effects of lexical relatedness (Nieuwland and Kuperberg, 2008). In previous studies of role-reversals, on the other hand, the canonical and role-reversed sentences differed only in either voice (active vs. passive) or word order, and so the lexical items were perfectly matched between conditions. The fact that the target words were therefore lexically associated to the same degree across conditions is consistent with the absence of N400 effects in these studies.

This brief survey of different cases in which the N400 is insensitive to semantic incongruity highlights the commonalities among them and suggests the following generalization: the amplitude of an N400 response to a word is modulated by the processor's expectation for that word, which in turn is mediated by the compositional meaning of the sentence context as well as by semantic association among words in the sentence. Therefore, in the present study we aimed to examine how the N400's sensitivity to semantic anomalies is modulated by lexical semantic association and whether non-surface plausibility and animacy congruity make any unique contribution to the N400.

1.3. *The present study*

The present study aimed to clarify the implications of the Semantic P600 phenomenon for architectural questions about the relations between syntax and online semantic interpretation. To this end, we devised two ERP experiments in Mandarin Chinese in tandem to examine the contributions of plausible non-surface interpretations, animacy violations, lexical association and mere implausibility to the ERP responses to role-reversals. We first explain the design of both experiments and then discuss the predictions of different hypotheses for the two experiments.

Both experiments examined the ERP responses to role-reversals. The role-reversals in Experiment 1 co-occurred with an animacy violation (e.g., *the student baffled the math problem*). The role-reversals in Experiment 2 were fully animacy-congruous (e.g., *the suspect arrested the inspector*). Due to practical constraints on generating fully animacy-congruous role-reversed sentences in sentences with a SVO word order, and in order to allow comparisons between the current study and previous studies on both kinds of role-reversals, all our experimental sentences had a SOV word order. Despite having a SVO basic word order, Mandarin Chinese has a highly frequent SOV *Ba*-construction. This construction requires a transitive verb, and the coverb *Ba* always follows the Actor argument and immediately precedes the Patient argument. Therefore, in this construction unambiguous and reliable cues about the arguments' syntactic roles are present in advance of the verb. Further, the fact that a clear role-reversal manipulation can be achieved by simply reversing the order of the arguments allowed us to avoid the ambiguity that occurs when role-reversed sentences are also morpho-syntactically anomalous in sentences such as "*The meal was devoured/devouring…*". Lastly, in order to maximize comparability among conditions across the two experiments, sentences in both experiments were intermixed and presented within a single experimental session.

In Experiment 1 we orthogonally manipulated animacy-congruity and the 'combinability' of the verb and its arguments

Table 1 – Experimental conditions and example sentences in Experiment 1. The target word is underlined.

Experimental condition	Sample materials
1. Animacy-congruous, Combinable (Control)	gaocaisheng ba shuxueti jieda-le student BA math problem *solve-ASP* "The student *solved* the math problem"
2. Animacy-violated, Combinable (Role-reversed)	gaocaisheng ba shuxueti nandao-le student BA math problem *baffle-ASP* "The student *baffled* the math problem"
3. Animacy-congruous, Non-combinable	gaocaisheng ba shuxueti guaqi-le Student BA math problem *hang-ASP* "The student *hung* the math problem"
4. Animacy-violated, Non-combinable	gaocaisheng ba shuxueti kunzhu-le student BA math problem *restrain-ASP* "The student *restrained* the math problem"
这个作家 把 这本小说 {读完了/感动了/打破了/剪伤了}	

(see Table 1 for a sample set of experimental materials). Using test sentences that had an animate subject and an inanimate direct object, animacy-congruity was manipulated by using verbs that can or cannot take an inanimate object. For purposes of the current study a verb was considered 'combinable' with its arguments if they can be combined to form a plausible sentence. For example, in the example in Table 1, the verb "hang" cannot be combined with the NPs "student" and "math problem" in a simple sentence to describe a plausible scenario, and therefore this verb–argument triplet was classified as non-combinable. We considered 'combinability' as a more restrictive criterion than mere lexical association, since verbs that are combinable with their arguments are likely also lexically related to the arguments (e.g. doctor–patient–cure), but lexically related verb–argument triplets might not be combinable (e.g., doctor–nurse–cure). Since accounts that assume independent semantic composition mechanisms predict that the N400 is modulated by the presence of a plausible non-surface interpretation (combinability) and not just lexical association, we manipulated combinability in the current study (see Kuperberg et al., 2006 for a discussion about the relative contribution of these factors). This approach allowed us to evaluate these accounts and our proposal at the same time, because combinability and lexical association are correlated (see Experimental procedures). The manipulations of animacy-congruity and combinability resulted in a fully crossed 2×2 within-subjects design. Importantly, all sentences in the animacy-violated and combinable condition were role-reversed (i.e., they had a plausible non-surface interpretation), but the design of the experiment was such that the role reversal was simply a consequence of the two independent factors. This design made it possible to assess whether the presence of a plausible non-surface interpretation made any unique contribution to the observed ERP effects, as predicted by accounts that assume independent semantic composition mechanisms.

As shown in Table 1, the four conditions in each item set had the same subject and object arguments and only differed in the target verb. Further, verbs were shuffled among item sets to appear in different experimental conditions, thereby minimizing lexical confounds (see Experimental procedures). Therefore, all comparisons were made between sentences with the same pre-target context and different target verbs. A related experimental design was used by Kuperberg et al. (2007), who also manipulated animacy congruity and lexical association. However, due to the constraints of SVO word order in English, comparisons in that study had to be made between sentences that differed in multiple ways. Animacy-congruity was manipulated by varying the subject noun while holding the verb constant (e.g., *For breakfast the boys/the eggs would eat*…). But animacy-congruous violations ('pragmatic violations' in Kuperberg's terminology) were created by combining a plausible subject–verb pair with an incongruous adverbial (e.g., *For breakfast the boys would plant* …). The lexically unrelated animacy violations were created by combining a lexically associated adverb–subject sequence with an unassociated verb (e.g., *For breakfast the eggs would plant*…). The SOV word order of the Chinese BA construction made it possible to tighten the manipulations, and also to provide closer comparisons with previous findings from languages with SOV order.

In Experiment 2 we manipulated the structural role of the arguments in simple BA-construction sentences (see Table 2 for a sample set of experimental materials). Unlike Experiment 1, both pre-verbal arguments in these sentences were animate NPs and therefore this role-reversal manipulation never co-occurred with an animacy violation. This manipulation was related to experimental designs used by Kolk and colleagues (e.g., Kolk et al., 2003; Van Herten et al., 2005) in Dutch, and Ye and Zhou (2008) in Mandarin Chinese. However, unlike the materials used by Kolk and colleagues, where the second arguments are prepositional phrases (e.g., *at the fox*; Dutch: *op de vos*), both arguments are noun phrases in the BA-construction in Mandarin Chinese. Further, unlike the materials used by Ye and Zhou (2008), where half of the sentences were in the active BA-construction while the other half were in the passive BEI-construction, the current study only used the active BA-construction in the experimental materials to ensure that the structural roles of the arguments were unequivocal to comprehenders.

Based on previous results, role-reversed sentences in both experiments were expected to elicit a P600 effect and no N400 effects relative to their canonical counterparts. However, competing theoretical accounts make different predictions in the other conditions. Specifically, if non-surface plausibility makes a unique contribution to the P600 (e.g., Kim and Osterhout, 2005), then the P600 effect should be largest in the role-reversed conditions in both experiments. If animacy violations make a unique contribution to the P600 (Stroud and Phillips, 2012), then the P600 effect should be larger in the

Table 2 – Experimental conditions and example sentences in Experiment 2. The target word is underlined.

Experimental condition	Sample materials
5. Canonical control	chen-tanzhang ba zhege-yifan jubu-le Inspector Chen BA the suspect arrest-ASP "Inspector Chen arrested the suspect"
6. Role-reversed (animacy-congruous)	zhege-yifan ba chen-tanzhang jubu-le Inspector Chen BA the suspect arrest-ASP "The suspect arrested Inspector Chen"

animacy-violated conditions than in the animacy-congruous conditions, including the role-reversed condition in Experiment 2. If both these factors uniquely contribute to the P600 and their effects are independent, then a P600 effect should be observed in all role-reversed and animacy-violated conditions, but the effect should be largest in the animacy-violated role-reversed condition in Experiment 1. However, if the P600 is fully attributable to the general implausibility of the surface meaning (e.g., Brouwer et al., 2012; Kuperberg, 2007; van de Meerendonk et al., 2009, 2010), then we should merely expect a significant P600 effect in all implausible conditions relative to the plausible control in both experiments.

Predictions for the N400 effects also differ in competing accounts, although the N400 effects alone are less theoretically decisive. Accounts that assume independent semantic composition mechanisms predict an N400 effect in the two non-combinable conditions in Experiment 1, but no N400 effects in role-reversed sentences in both experiments because of the presence of a plausible non-surface interpretation. An account that attributes N400 effects to lexical association differences makes very similar predictions, since the lexical association between the target verb and its preceding words was much stronger in the role-reversed condition in both experiments than in the non-combinable conditions in Experiment 1. Although these two rather different hypotheses cannot be distinguished based on the N400 results alone, only the former predicts that non-surface plausibility uniquely contribute to the P600. Meanwhile, Bornkessel-Schlesewsky and Schlesewsky's (2008) proposal predicts that the N400 should be larger in the animacy-violated conditions than in the animacy-congruous conditions, whereas Kuperberg et al.'s (2007) hypothesis predicts that the N400 should be attenuated or eliminated in the presence of an animacy violation.

2. Results

2.1. Acceptability judgments

Participants' average acceptability judgment accuracy and the target response in each condition are shown in Table 3. With an overall accuracy of 86.7%, participants reliably accepted canonical control sentences and rejected the semantically anomalous sentences, regardless of the presence or absence of thematic role-reversals. In Experiment 1 a repeated measures ANOVA revealed a marginal effect of animacy-congruity ($F(1,19)=3.03$, $p<0.10$), due to more accurate responses for animacy-violated sentences than animacy-congruous sentences. In Experiment 2 mean accuracy did not differ significantly between the canonical and role-reversed conditions ($t(18)=0.71$, $p=0.49$).

Table 3 – Target response and accuracy on acceptability judgment task.

Experimental condition	Target response	Percent accurate (sd)
Experiment 1		
Animacy-congruous, combinable	Yes	83.2 (11.6)
Animacy-violated, combinable (role-reversed)	No	89.5 (8.3)
Animacy-congruous, non-combinable	No	83.5 (12.2)
Animacy-violated, non-combinable	No	91.8 (7.5)
Experiment 2		
Canonical control	Yes	87.2 (7.4)
Role-reversed (animacy-congruous)	No	84.9 (10.2)

2.2. Event-related potentials

2.2.1. Experiment 1: effects of animacy-violations and non-combinability

Fig. 1 shows the grand average ERPs ($n=19$) at the target word in all four conditions in Experiment 1. The target words in all conditions elicited the pattern characteristic of ERPs to visual stimuli. These components include an initial positivity (P1) peaking at about 80 ms, followed by a negativity (N1) at 170 ms, and a positivity (P2) around 275 ms. These responses were followed by a centro-posterior negativity between about 300 and 500 ms (N400). In the conditions involving animacy violations, the N400 was followed by a large late positive-going wave starting from approximately 550 ms (P600).

Inspection of Fig. 1 reveals clear effects of both experimental factors. Combinability affected N400 amplitude and animacy-congruity affected P600 amplitude. Non-combinable target verbs elicited a larger N400 response compared to combinable target verbs. Starting at about 550 ms, animacy-incongruous target verbs elicited a larger posterior positivity (P600) than the animacy-congruous target verbs. These observations were confirmed by the statistical analyses. Results from the overall ANOVA and region of interest (ROI) analyses are presented in Table 4. The mean ERP values in the N400 and P600 intervals in the midline posterior region are presented in Fig. 2.

No significant differences were observed in the 0–300 ms interval. In the 300–500 ms interval the overall ANOVA revealed a significant main effect of combinability, reflecting that ERPs in the non-combinable conditions were more negative than in the combinable conditions across the entire scalp. The interaction between combinability and animacy, and a four-way interaction between combinability, animacy, anteriority and laterality was also significant, reflecting that the effect of combinability was slightly larger and more broadly distributed in the animacy-congruous condition than in the animacy-violated condition. ROI analyses revealed a significant main effect of combinability in all

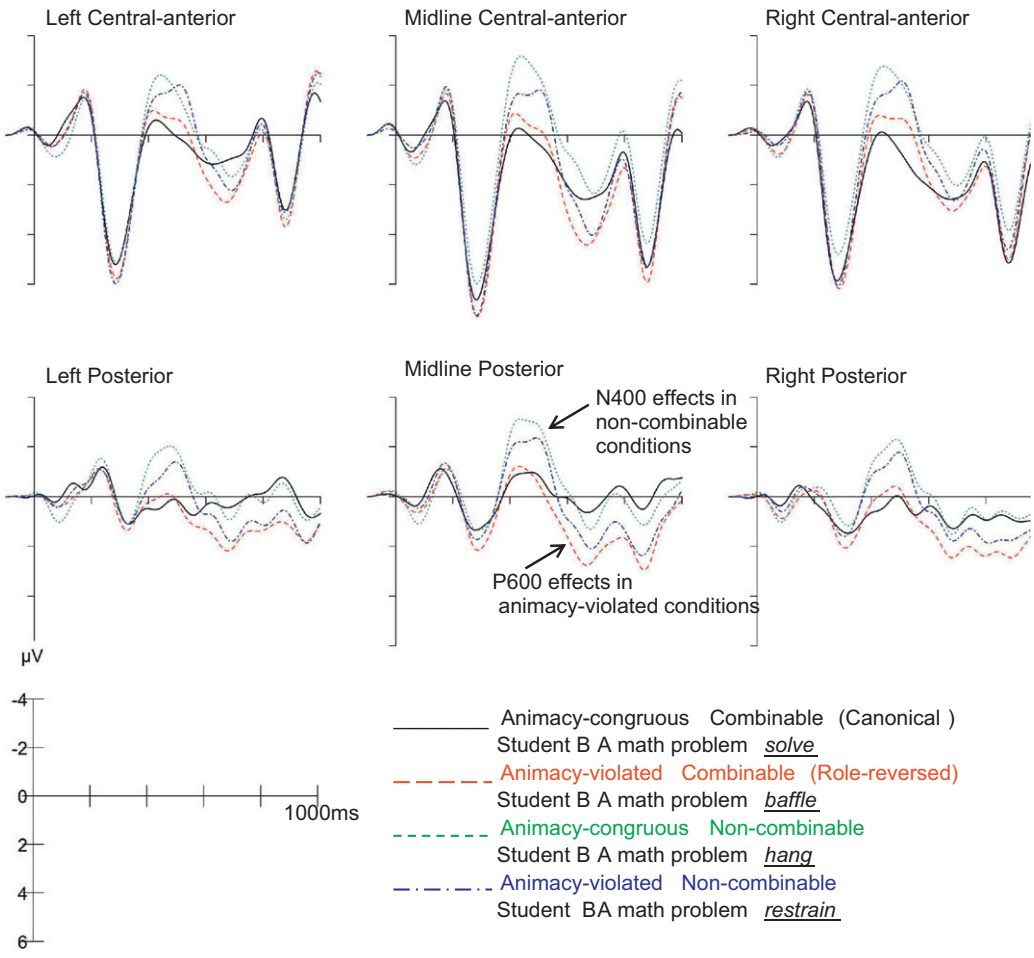

Fig. 1 – Grand average ERPs in six regions of interests in Experiment 1.

ROIs, and a significant interaction between animacy-congruity and combinability in three ROIs (midline central-anterior, right central-anterior, and right posterior regions). Follow-up comparisons revealed that the amplitude of the N400 was not different between the animacy-violated and animacy-congruent combinable conditions in any of these regions, whereas the N400 was less negative in the animacy-violated non-combinable condition than in the animacy-congruent non-combinable condition and this difference reached statistical significance in the midline central-anterior region ($t(1,18)=2.41$, $p<0.05$).

These results were also corroborated by pair-wise comparisons of the amplitude of the N400 between each of the anomalous conditions and the canonical control condition in each ROI. These comparisons revealed that the N400 never differed between the animacy-violated combinable condition and the control condition, and that the N400 was more negative in both non-combinable conditions compared to the control condition across the scalp. The N400 in the animacy-congruent non-combinable condition was significantly more negative than that in the canonical control condition across all ROIs (all $ps<0.02$); the effect in the animacy-violated non-combinable condition was marginally significant in the left central-anterior region ($p<0.06$) and significant in all other regions ($ps<0.02$).

Starting at around 550 ms ERPs in the animacy-violated condition were more positive than in the animacy-congruous condition and this effect persisted throughout the entire epoch. The effect was present across the entire scalp, but was largest at midline posterior sites. In the 600–800 ms interval the overall ANOVA revealed a marginally significant main effect of animacy-congruity ($p=0.054$) and significant interactions between animacy and anteriority and between animacy-congruity and laterality. Consistent with the typical distribution of P600 effects, ROI analyses confirmed a main effect of animacy-congruity that was significant in three regions (left posterior, midline posterior, and midline anterior) and marginally significant in the right posterior region ($p=0.06$). No significant interaction effects between animacy-congruity and combinability were observed. In the 800–1000 ms interval the overall ANOVA revealed a significant interaction between animacy-congruity and anteriority, along with a significant four-way interaction between combinability, animacy-congruity, anteriority, and laterality. ROI analyses revealed a main effect of animacy-congruity that was statistically significant in the midline posterior region and marginally significant in the left posterior region ($p=0.08$). No significant interaction effects between animacy-congruity and combinability were observed.

In summary, ERPs in Experiment 1 were significantly more negative in the non-combinable conditions than in the combinable conditions in the N400 interval, and significantly more positive in the animacy-violated conditions than in the animacy-congruous conditions in the P600 intervals. But with the exception of one ROI in the N400 analyses, the ERP effects of

Table 4 – Repeated measures ANOVA *F* values at the target word in Experiment 1.

	df	0–300 ms	300–500 ms	600–800 ms	800–100 ms
Omnibus ANOVA					
comb	1,18	<1	26.04**	<1	<1
anim	1,18	<1	<1	4.24***	<1
comb * anim	1,18	<1	11.04**	1.63	<1
comb * ant	1,18	<1	<1	<1	2.02
anim * ant	1,18	<1	1.23	7.54*	14.49**
comb * anim * ant	1,18	2.06	1.31	3.44***	5.66*
comb * lat	2,36	2.24	2.33	2.56	3.27***
anim * lat	2,36	<1	2.19	4.34*	2.02
comb * anim * lat	2,36	<1	<1	<1	1.57
comb * ant * lat	2,36	3.02***	2.4	2.33	3.37***
anim * ant * lat	2,36	<1	<1	<1	<1
comb * anim * ant * lat	2,36	<1	3.49*	3.12***	4.16*
ROI analyses					
Left central-anterior					
comb	1,18	<1	7.55*	<1	<1
anim	1,18	<1	<1	<1	<1
comb * anim	1,18	<1	2.04	1.77	<1
Midline central-anterior					
comb	1,18	<1	24.93**	<1	2.38
anim	1,18	1.09	<1	4.86*	<1
comb * anim	1,18	<1	7.56*	<1	2.78
Right central-anterior					
comb	1,18	2.84	25.61**	2.27	4.31***
anim	1,18	<1	<1	<1	<1
comb * anim	1,18	<1	14.18**	1.94	3.83***
Left posterior					
comb	1,18	<1	23.79**	<1	<1
anim	1,18	<1	<1	7.53*	3.43***
comb * anim	1,18	<1	2.43	1.18	<1
Midline posterior					
comb	1,18	<1	19.88**	<1	<1
anim	1,18	<1	1.37	9.77**	5.85*
comb * anim	1,18	<1	1.96	4.36***	2.37
Right posterior					
comb	1,18	1.28	18.81**	<1	<1
anim	1,18	<1	<1	3.87***	1.94
comb * anim	1,18	<1	6.69*	3.03***	<1

Factors: *comb*, combinability; *anim*, animacy; *ant*, anteriority; *lat*, laterality.
* $0.05 < p < 0.1$
** $p < 0.05$
*** $p < 0.01$.

animacy-congruity and combinability were independent of one another: the N400 amplitude was modulated by combinability but not by animacy-congruity, and the P600 amplitude was modulated by animacy-congruity but not by combinability.

2.2.2. *Experiment 2: effects of animacy-congruous role-reversal*

Fig. 3 shows the grand average ERPs ($n=19$) at the target verb in Experiment 2. Results from the overall ANOVA and ROI analyses are presented in Table 5. As shown in Fig. 3 the target words in both conditions elicited the pattern characteristic of ERPs to visual stimuli. The ERPs did not diverge early on, hence no significant effects of role-reversal were observed in the 0–300 ms and 300–500 ms intervals. Starting at around 550 ms the ERPs became more positive in the role-reversed condition than in the canonical condition and the effect persisted throughout the entire epoch. The effect was present across the entire scalp, but was most pronounced at posterior sites, showing a topographic distribution that is typical of P600 effects.

In the 600–800 ms interval the omnibus ANOVA revealed no significant main effect or interactions involving role-reversal ($p=0.13$), but ROI analyses confirmed a significant or near-significant effect of role-reversal in all posterior regions, consistent with the characteristically posterior distribution of P600 effects. In the 800–1000 ms interval the omnibus ANOVA revealed a marginal interaction between role-reversal and anteriority ($p=0.07$), and ROI analyses revealed a marginal effect of role-reversal in the right posterior region ($p=.06$).

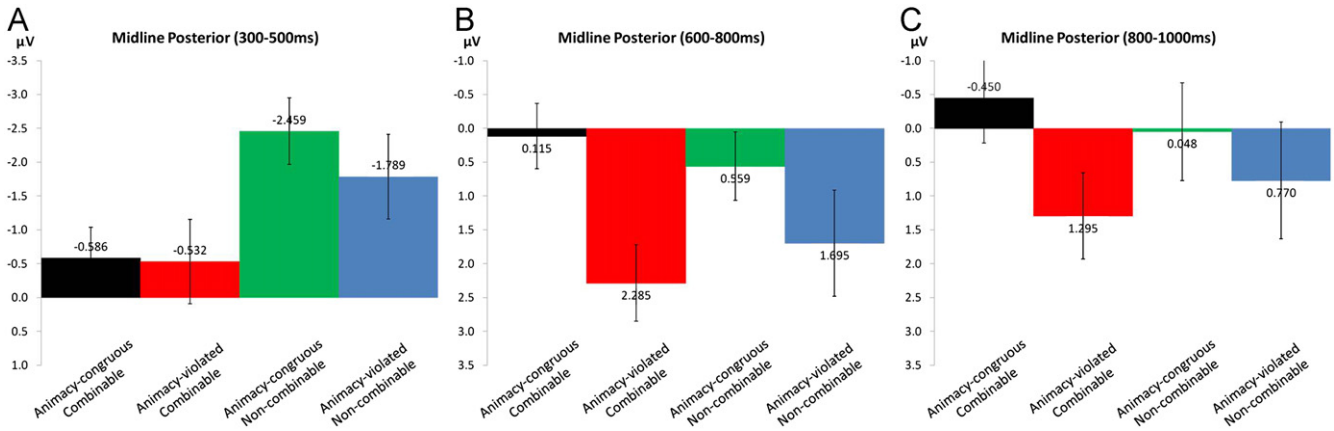

Fig. 2 – Grand average ERPs in midline posterior region in (A) 300–500 ms, (B) 600–800 ms, and (C) 800–1000 ms time intervals in Experiment 1.

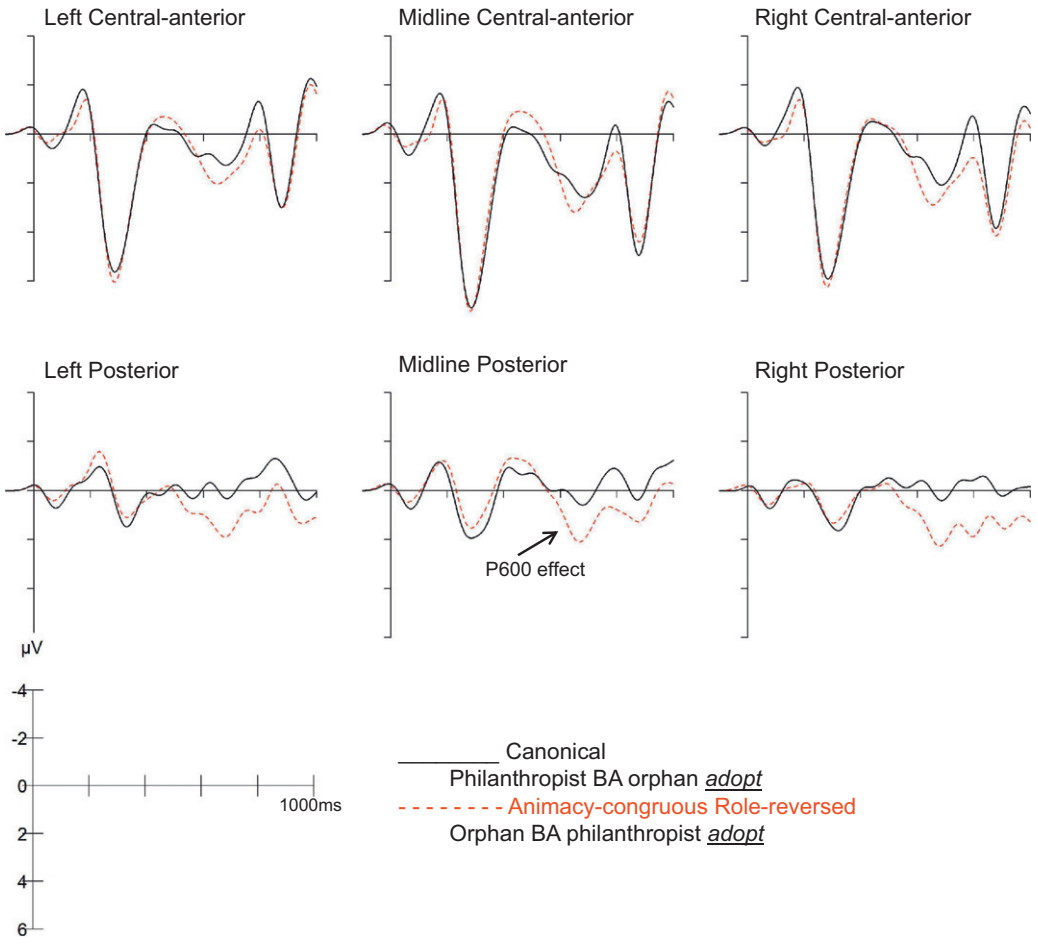

Fig. 3 – Grand average ERPs in six regions of interest in Experiment 2.

3. Discussion

The aim of the present study was to clarify the role of plausible non-surface interpretations ('semantic attraction'), lexical association/combinability, and animacy congruity in the ERP responses to role-reversal anomalies. Each of the individual results in the present study has precedents in previous studies, but the way in which they are combined here makes it possible to address architectural questions that were not so easily addressed before. First, the presence of a P600 effect and the absence of N400 effects in the role-reversed conditions of Experiments 1 and 2 are consistent with previous reports that role-reversed sentences, despite being syntactically well-formed and semantically incongruous, elicit a P600 effect and no N400 effects (e.g., Hoeks et al.,

Table 5 – Repeated measures ANOVA *F* values at the target word in Experiment 2.

	df	0–300 ms	300–500 ms	600–800 ms	800–100 ms
Omnibus ANOVA					
rev	1,18	<1	<1	2.48	<1
rev × ant	1,18	1.11	<1	2.73	3.81***
rev × lat	2,36	<1	2.13	2.2	2.11
rev × ant × lat	2,36	<1	<1	<1	<1
ROI analysis					
Left central-anterior	1,18	<1	<1	<1	<1
Midline central-anterior	1,18	<1	<1	<1	<1
Right central-anterior	1,18	<1	<1	1.13	<1
Left posterior	1,18	1.35	<1	4.57*	2.65
Midline posterior	1,18	<1	<1	3.12***	1.51
Right posterior	1,18	<1	<1	10.06**	3.86***

Factors: *rev*, reversal; *ant*, anteriority; *lat*, laterality.
* $0.05 < p < 0.1$
** $p < 0.05$
*** $p < 0.01$.

2004; Kim and Osterhout, 2005; Kolk et al., 2003, Kuperberg et al., 2003, 2007; Van Herten et al., 2005, 2006; Ye and Zhou, 2008). Further, the presence of a highly similar P600 effect across different semantically anomalous conditions in Experiments 1 and 2, independent of non-surface plausibility, suggests that the P600 is sensitive to the implausibility of the surface form of the sentence, but not to the availability of plausible non-surface interpretations. The current results suggest that the N400's disappearance in role-reversed sentences is likely due to strong lexical associations. Meanwhile, although the apparent reduction of the N400 in the animacy-violated non-combinable condition relative to the animacy-congruous non-combinable condition suggests that animacy-violations might attenuate the N400, we argue that such reduction is also attributable to component overlap.

3.1. *P600 is not selectively sensitive to non-surface interpretations*

Table 6 summarizes some of the factors that have been proposed to account for the Semantic P600 phenomenon. Current accounts for the observation of P600 effects to semantic anomalies consider factors such as surface plausibility, non-surface plausibility, animacy congruity and competing representations.

The current results, along with some previous results discussed in the Introduction, are not compatible with accounts that assume independent semantic composition (Hagoort et al., 2009; Hoeks et al., 2004; Kim and Osterhout, 2005; Kolk et al., 2003; Van de Meerendonk et al., 2009, 2010; Van Herten et al., 2005, 2006). These accounts predict that semantic anomalies that have a (partially or wholly) plausible non-surface interpretation should make a unique contribution to the P600, and that a semantic P600 response should be conditioned by the absence of an N400 effect. In the current study, however, the animacy-violated non-combinable condition in Experiment 1 nonetheless elicited a P600 effect despite the absence of a plausible non-surface interpretation, and the size of this effect was almost identical to that elicited in the role-reversed condition. Further, the current findings are also not compatible with an account based on 'partial plausibility' (Van Herten et al., 2006), since the two non-combinable conditions had identical degrees of partial plausibility and only one of them elicited a P600 effect.

In fact, evidence for the P600's selective sensitivity to plausible non-surface interpretations has only been reported in the original study by Kim and Osterhout (2005). Other studies, including Stroud's (2008) replication study using Kim and Osterhout's (2005) experimental materials, have consistently found that semantic anomalies that have no plausible non-surface interpretations nonetheless elicit a P600 effect (see also Kuperberg et al., 2006; Kuperberg, 2007; Oishi and Sakamoto, 2010; Paczynski and Kuperberg, 2011, 2012; Stroud and Phillips, 2012). And in cases where there is a direct comparison between conditions with and without a plausible non-surface interpretation ('semantic attraction'), the P600 effects typically show identical amplitude (e.g., Stroud, 2008; Stroud and Phillips, 2012). Meanwhile, our observation that an N400 effect preceded this P600 effect shows that the presence of a P600 response to semantic anomalies is not conditioned by the absence of an N400 response. This is consistent with the observation that semantic anomalies frequently elicit both an N400 and a late positivity (e.g., Friederici et al., 1998; Kolk et al., 2003; Curran et al., 1993; Van den Brink et al., 2001; Van Herten et al., 2005). In fact, Van Petten and Luka (2012) noted that the N400 elicited by semantically incongruous words is followed by a posterior positivity in about one-third of the 64 published comparisons they reviewed. Taken together, our results provide convergent evidence that the P600 response to semantic anomalies is not modulated by the availability of (partially) plausible non-surface interpretations and therefore they undermine the original argument for independent semantic composition.

Meanwhile, however, we believe no existing accounts can straightforwardly capture the current results. A surface anomaly account that attributes the P600 to grammatical and animacy violations (e.g., Stroud, 2008) cannot capture the observations that fully grammatical and animacy-congruous role-reversed

Table 6 – Summary of some of the factors that have been proposed to account for the Semantic P600 phenomenon.

	N400				P600*			
Proposed contributors	Stored semantic relationships	Non-surface plausibility	Animacy (verb-dependent selectional restriction)	Animacy (linear order heuristic)	Surface plausibility	Non-surface plausibility	Competing representations	Animacy (verb-dependent selectional restriction)
Proposed generalizations / mechanisms	The N400 is reduced for words that are semantically/associatively related.	Semantic anomalies do not elicit an N400 effect when a plausible non-surface interpretation is computed.	Violations of a verb's animacy selectional restriction can "switch off" semantic processing and thus reduce the N400.	The N400 is modulated by the match between the animacy of a noun and its linear position.	Implausible interpretations elicit a larger P600.	The presence of a plausible non-surface interpretation elicits a larger P600.	Mismatches between a non-compositional representation and a syntax-based representation elicit a larger P600.	Animacy violations lead to a larger P600.
Accounts for the Semantic P600 phenomenon								
Kolk et al. (2003); van Herten et al. (2005, 2006); van de Meerendonk et al. (2009, 2010)	X	X			X	X	X	
Kim & Osterhout (2005)	X	X				X		
Hoeks et al. (2004)	X	X				X		
Kuperberg et al. (2007)	X	X	X		X	X	X	X
Paczynski & Kuperberg (2011)	X			X[a]	X		X	X
Bornkessel-Schlesewsky & Schlesewsky (2008); Bornkessel-Schlesewsky et al. (2011)	X	X		X[a]	X		X	
Hagoort et al. (2009)	X	X				X		
Stroud & Phillips (2012)	X							X
Brouwer et al. (2012)	X				X			

[a] Bornkessel-Schlesewsky and Schlesewsky (2008) proposed that violations of linear order animacy heuristics increase the N400, whereas Paczynski and Kuperberg (2011) proposed they reduce the N400.

* The role of context and task have also been discussed, although the generalization was not specified.

sentences (Experiment 2) nonetheless elicited a P600 effect. Meanwhile, accounts that attribute the P600 to the implausibility of the sentence (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Brouwer et al., 2012; Kuperberg, 2007; Van de Meerendonk et al., 2009, 2010) have yet to capture why certain kinds of implausible sentences fail to elicit a P600 effect. In the current study, even though the same group of participants had judged the sentences in all the anomalous conditions as unacceptable, only the animacy-congruous non-combinable condition failed to elicit a significant P600 effect. As 90% of the sentences in the animacy-congruous non-combinable condition (as opposed to 100% in the animacy-violated conditions and only 55% in the animacy-congruous role-reversed condition) involved a violation of the verb's selectional restriction (see Experimental procedures), sentences in the animacy-congruous non-combinable condition should not be less implausible than those in the other conditions. Therefore, we propose that neither the presence of animacy violations nor surface implausibility can straightforwardly account for the full set of current findings.

Previous studies have discussed the possibility that a P600 effect might be attenuated if it temporally overlaps with a large N400 effect, which has opposite polarity (e.g., Hagoort, 2003). One possibility is that a P600 effect was elicited in all semantically anomalous conditions, but that the response was fully masked in the non-combinable conditions due to an overlapping N400 effect. However, we regard this possibility as rather unlikely. Given that no apparent effects were observed in the animacy-congruous non-combinable condition in either of the P600 time intervals (600–800 ms; 800–1000 ms) in Experiment 1, the supposedly masked P600 effect would have to have been completely overlapping in time with the N400 effect. The N400 effects elicited by visually presented stimuli are typically confined to a well-defined time interval (e.g., 300–500 ms), during which the divergence between the conditions peaks at around 400 ms and gradually returns to baseline afterwards (for review see Kutas and Federmeier, 2011). P600 effects, meanwhile, tend to extend over a longer time interval (e.g., 500–1000 ms). Therefore, if a P600 effect were present in all the anomalous conditions in the present study, then its apparent absence in the animacy-congruous non-combinable condition must be attributable to (i) an N400 effect that extended well into the P600 time-window in that condition and/or (ii) a particularly short-lasting P600 effect in that condition.

Note, however, that even if the presence of an N400 effect did obscure a potential P600 effect in the current study, our results would still be incompatible with semantic illusion accounts of the P600. Since an N400 effect was present in the non-combinable conditions but not in the combinable conditions, resolving this overlap would yield larger P600 effects in the non-combinable conditions than in the combinable conditions. In particular, the P600 in the animacy-violated non-combinable condition would be larger than that in the role-reversed conditions. This is the opposite of the predictions of semantic illusion accounts, according to which the role-reversed condition should elicit a larger P600 than conditions in which no plausible non-surface interpretations are available, since a plausible non-surface interpretation is present in the former but not in the latter.

In sum, the presence of non-surface plausibility makes no unique contribution to the P600, and thus the P600's sensitivity to role-reversal anomalies does not constitute evidence for syntax-independent semantic composition. The present results suggest that the P600 is sensitive to both animacy violations and surface implausibility, but neither of these factors can fully account for the current findings in isolation. The current results add to the growing body of evidence that fully grammatical sentences with semantic anomalies do at times elicit a P600 effect. Future work is required to specify testable hypotheses about how these factors combine and/or interact in modulating the P600.

3.2. *Lexical relations, not non-surface plausibility or animacy-congruity, modulate the N400*

Current accounts for the absence of N400 effects to semantic anomalies consider factors such as lexical semantic association, non-surface plausibility and animacy congruity. Accounts that assume independent semantic composition (e.g., Kim and Osterhout, 2005) attribute the lack of N400 in role-reversed sentences to the presence of a plausible non-surface interpretation ('semantic attraction'). Meanwhile, others have proposed that the presence of animacy-violations makes a unique contribution to the modulation of the N400 (e.g., Bornkessel-Schlesewsky and Schlesewsky, 2008; Kuperberg, 2007). Alternatively, without positing any unique contribution of non-surface plausibility and/or animacy congruity, we outlined in the Introduction that the lack of N400 effects can be fully attributed to strong lexical association between the target verb and its arguments in role-reversed sentences. Our discussion about lexical association can be extended to include the semantic features, associative relationships and other types of semantic relationships between content words (e.g., the semantic memory-based mechanism in Kuperberg, 2007). It may also be extended to include event schemas (e.g. Bicknell et al., 2010; Paczynski and Kuperberg, 2012).

In the current study, the non-combinable conditions elicited a significantly larger N400 effect than the combinable conditions. Since the current study operationalized the manipulation of lexical association as the combinability between the verb and its arguments, these N400 findings alone do not allow us to determine whether the presence of a plausible non-surface interpretation makes a unique contribution to the N400 beyond that of lexical association. Meanwhile, we can better evaluate the merits of these competing accounts in the context of the broader array of findings. First, Kuperberg et al., 2006 observed that, when semantic relatedness is held constant, semantically anomalous words elicited the same N400 and P600 responses regardless of the presence of plausible non-surface interpretations ('thematic fit', in the authors' terms) and suggested that non-surface plausibility makes no unique contribution to their findings. Second, accounts that assume independent semantic composition predict that both the lack of an N400 effect and the presence of a P600 effect in role-reversed sentences are selectively conditioned by the presence of plausible non-surface interpretations. As we discussed above, however, most existing evidence shows that the P600 is less selective. Further, as discussed in the Introduction, instances of the N400's blindness to semantic incongruity have been reported outside of the "Semantic P600" literature and in different previous studies

(e.g., Fischler et al., 1983; Nieuwland and Kuperberg, 2008; Urbach and Kutas, 2010). Even though accounts that assume independent semantic composition are compatible with the current N400 findings, such accounts require that other instances of N400 blindness be given different interpretations. In contrast, our proposal attributes the N400's insensitivity to role-reversals to the (roughly) equal accessibility of the target word across conditions, due to a combination of strong lexical association and weak contextual constraints. This account does not give a special status to sentences with a plausible non-surface interpretation, and it can potentially provide a unified explanation for other instances of the N400's blindness to semantic anomalies.

Further, the current findings are consistent with the claim that animacy congruity makes no unique contribution to the N400. Specifically, even though role-reversals co-occurred with animacy-violations in Experiment 1 but not in Experiment 2, they did not elicit an N400 effect in either case. Meanwhile, the small reduction of the N400 in the animacy-violated non-combinable condition relative to the animacy-congruous non-combinable condition is compatible with the hypothesis that animacy-violations attenuate the N400 (Kuperberg et al., 2007). However, such N400 reduction can also be attributed to (i) potential component overlap, and/or (ii) semantic relatedness differences between the items used in the two non-combinable conditions. Since the animacy-violated condition elicited a much larger late positivity than the animacy-congruous condition, it is plausible that the N400 was reduced as a result of its overlap with the P600. Further, in the semantic relatedness judgment study the target verbs were judged to be slightly more related to the argument NPs in the animacy-violated non-combinable condition than in the animacy-congruous non-combinable condition (2.17 vs. 1.76 on a 7-point scale). This difference in semantic relatedness might have led to the N400 difference between the two non-combinable conditions. Therefore, we argue that the current results do not provide clear evidence that lexical relations and animacy-congruity interact to modulate the N400 effect.

We propose that in simple sentence contexts like "*The student BA the math problem…*", the processor might fail to differentially expect upcoming information, i.e., a congruous vs. an incongruous verb, based only on information about the subject and the object noun phrases. Under such circumstances the ease of lexical access, and by hypothesis the N400 amplitude, should only be modulated by lexical association between the target word and prior context. This, however, raises questions regarding the nature of expectation–generation mechanisms (e.g., DeLong et al., 2005; Federmeier 2007; Van Berkum et al., 2005; Wicha et al., 2004). Specifically, is the presumed difficulty in predicting an appropriate verb given its arguments reflective of a general property of the processor? Could the processor's apparent difficulty in predicting a verb be rectified? How do different word orders modulate the processor's success in predicting plausible thematic relations? Future work will need to address these questions by examining the effects of manipulations that are believed to facilitate predictions and by making carefully controlled comparisons across sentences with different word orders.

3.3. *Conclusion*

In this paper we investigated the theoretical implications of the 'Semantic P600' phenomenon. In previous studies both the presence of a P600 effect and the absence of an N400 effect in role-reversed sentences have been regarded as two central pieces of evidence for a syntax-independent semantic composition mechanism. We presented two ERP studies that tested competing explanations for these two pieces of evidence. We found that the P600's sensitivity to semantic anomalies is not restricted to cases in which plausible non-surface interpretations are available, and argued that the presence of a P600 effect in role-reversed sentences does not constitute evidence for independent semantic composition. We also showed that the N400's insensitivity to role-reversals cannot be attributed to the presence or the absence of animacy violations, and can instead be attributed to the lexical association between a verb and its arguments. We outlined a proposal in which the N400 reflects the ease of lexical access, and interpreted the lack of N400 effects in role-reversed sentences as reflecting the processor's temporary failure to generate specific lexical expectations in canonical vs. role-reversed sentences.

4. Experimental procedures

4.1. *Participants*

Nineteen students (11 female, mean age = 22 years, range 18–25 years) from Beijing Normal University participated in the current study. All participants were native speakers of Mandarin Chinese, were strongly right-handed based on the Edinburgh Handedness Inventory (Oldfield, 1971), and had normal or corrected-to-normal vision and no history of neurological disorder. Data from five additional participants were excluded due to excessive artifacts ($\geq$50% epochs rejected in one or more conditions). All participants gave informed consent and were paid 50 RMB/hour for their participation.

4.2. *Materials*

Each item set in Experiment 1 contained four sentence types (see Table 1). The materials preceding the verb in all sentences within each item set were identical, consisting of an animate subject, the coverb *Ba* and an inanimate object. Each condition had a different target verb, followed by materials that were identical within each item set. All the verbs were relatively common and three-characters long. Animacy-congruity was manipulated by using verbs that do or do not allow inanimate direct objects. Around 10% of verbs in the animacy-incongruous conditions showed a strong animacy bias rather than a strict animacy requirement. It should be noted that animacy-congruity does not entail the fulfillment of all the verb's selectional restrictions. For example, *The student hung the math problem* is considered animacy congruous because the verb 'hang' can take inanimate objects, even though 'math problem' does not fulfill the verb's requirement for a concrete object noun. Ninety percent of the sentences in the non-combinable animacy-congruous condition involved a violation of the verb's selectional restrictions. Combinability

was manipulated by using verbs that either do or do not yield a plausible interpretation when combined with the two arguments. A lexical association rating study was conducted to obtain objective measures of the lexical association between the verbs and their arguments (see below for more details). Role-reversal anomalies in this experiment were created by using a verb that does not allow inanimate direct objects but is combinable with the two arguments, i.e., the sentence would have a plausible interpretation had the two arguments been reversed. Therefore, within the current experimental design role-reversed sentences can be characterized as *animacy-violated* and *combinable*.

In order to avoid lexical confounds in the ERP data, each verb that was used in a combinable condition in one item set was used in a non-combinable condition in another item set. Specifically, the verbs in the canonical sentences were shuffled across item sets to create the *animacy-congruous non-combinable conditions*, and the verbs in the role-reversed sentences were shuffled to create the *animacy-violated non-combinable* sentences. Therefore, the two animacy congruous conditions and the two animacy-violated conditions used an identical set of verbs. Care was taken in the verb-shuffling procedure to ensure non-combinability in the resulting sentences.

Assessing the lexical frequency of the target verbs was made difficult by the fact that the target verbs often consisted of a main verb and an adjectival resultative, e.g., du-si, meaning 'poison-dead', which are considered separate lexical items in some Chinese corpora. We were able to obtain lexical frequency estimates for 66 of the 120 target verbs (after removing the aspectual marker *-le*) using SUBTLEX-CH (Cai and Brysbaert, 2010), which suggested that lexical frequencies were well-matched between the animacy-congruous verbs (mean=2.34; *SE*=0.11) and the animacy-violated verbs (mean=2.32; *SE*=0.14). Meanwhile, the average log character frequency of the first two characters of the target verbs (i.e., without the common aspectual marker *-le*) was numerically slightly higher for animacy-congruous verbs (mean=4.30; *SE*=0.09) than for animacy-violated verbs (mean=4.04; *SE*=0.08).

We asked 44 native Mandarin Chinese speakers who did not participate in the ERP study to rate, on a 7-point scale, the degree to which the verbs are considered 'related' to the corresponding pair of noun phrases used in the ERP study. Each participant only saw one of the target verbs for any given item. The results showed that the verbs were judged to be closely related to the arguments in the combinable conditions (animacy-congruous: *M*=6.55, *SE*=0.042; animacy-violated: *M*=5.15, *SE*=0.083), but not in the non-combinable conditions (animacy-congruous: *M*=1.76, *SE*=0.056; animacy-violated: *M*=2.17, *SE*=0.068).

As illustrated in Table 2, each item set in Experiment 2 contained two conditions: a canonical condition and a role-reversed condition. Role-reversed sentences were created by reversing the structural position of the arguments in the canonical sentences. Therefore, the two sentences within each set used an identical verb–argument triplet, and the two conditions differed only in the order of the arguments. Further, unlike Experiment 1 both preverbal arguments were animate and therefore the role-reversals in Experiment 2 never co-occurred with an animacy-violation.

All experimental sentences consisted of an adverbial phrase followed by a main clause. In order to avoid sentence-final wrap-up effects at the critical clause-final verb the SOV BA-construction was embedded in the adverbial phrase (*Zai… zhihou*, After…), followed by a grammatical main clause that was held constant across conditions within each item set. In all experimental sentences, no anomaly was evident before the critical verb.

Sixty sets of items were generated for each of the experiments and the sentences were distributed in two presentation lists, such that half of the participants read sentences from one presentation list and the remaining participants read sentences from the other list. Each list contained 180 experimental sentences (120 for Experiment 1 and 60 for Experiment 2) along with 180 unrelated fillers of similar length and structural complexity. Each list contained one sentence from each item set in Experiment 2, and two sentences from each item set in Experiment 1 (one combinable and one non-combinable). The sentences were presented in six blocks of 60 sentences each, and the order of the blocks was randomized across participants. The two conditions from the same item set never appeared within the same presentation block. Care was taken to ensure that the overall congruous-to-anomalous ratio in each presentation list was 1:1.

4.3. Procedure

Participants were comfortably seated in a testing room around 100 cm in front of a computer screen. Sentences were presented one word at a time in a white font (30 pt simplified Chinese characters) on a black background at the center of the screen. Each sentence was preceded by a fixation cross that appeared for 500 ms. Each word appeared on the screen for 400 ms, followed by 200 ms of blank screen. The last word of each sentence was marked with a period "○", followed 1000 ms later by a response cue "?". Participants were instructed to avoid eye blinks and movements during the presentation of the sentences, and they were asked to read each sentence attentively and to indicate whether the sentence was an acceptable sentence of Mandarin Chinese by pressing one of two buttons. The current study used this task because the phenomenon of interest has been observed in previous studies that used the same task. Prior to the experimental session, participants were presented with 12 practice trials to familiarize themselves with the task. The experimental session was divided into six blocks of 60 sentences each, with short pauses in between. Including set-up time, an experimental session lasted around 2.5 h on average.

4.4. EEG recording

EEG was recorded continuously from 30 AgCl electrodes mounted in an electrode cap (Electrocap International): midline: Fz, FCz, Cz, CPz, Pz, Oz; lateral: FP1/2, F3/4, F7/8, FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2. Recordings were referenced online to the left mastoid and re-referenced to linked mastoids offline. The electro-oculogram (EOG) was recorded at four electrode sites; vertical EOG was recorded from electrodes placed above and below the left eye and the

horizontal EOG was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5 kΩ. The EEG and EOG recordings were amplified (bandpass = 0.5–100 Hz) and digitized online at 1 kHz with a bandpass filter of 0.1-70 Hz.

4.5. *ERP data analysis*

All trials were evaluated individually for EOG or other artifacts. Trials contaminated by eye blinks, excessive muscle artifact, or amplifier blocking were excluded from the averaging procedure. This affected 10.6% of experimental trials, equally distributed across conditions (ranging between 9.6 and 11.6% across conditions). Event-related potentials were computed separately for each participant and each condition for the 1000 ms after the onset of the critical verb relative to a 100 ms baseline preceding the critical verb. Averaged waveforms were filtered offline using a 10 Hz low-pass filter for presentation purposes only. All statistical analyses were performed using the original data.

Statistical analyses on average voltage amplitudes were conducted separately for four time windows chosen based on previous literature and on visual inspection of the data: 0–300 ms for possible early differences, 300–500 ms for the N400, and 600–800 ms and 800–1000 ms for the P600. Separate analyses were conducted for mean amplitudes in each time window. While the 600–800 ms interval is most commonly used to analyze P600 effects in previous studies, the additional 800–1000 ms interval was included in the current study to examine possible component overlap between the N400 and the P600 effects. We reasoned that, if the P600 partially overlapped in time with the N400, the amplitude of the P600 might be affected by the N400 more strongly in an earlier interval (e.g., 600–800 ms) than in a later interval (e.g., 800–1000 ms). Although similarity between the ERPs in these intervals does not exclude the possibility that the P600 did overlap with the N400, systematic differences in these P600 intervals could be informative about the extent to which our findings about the P600 is attributed to potential overlap with the N400.

Data from the two experiments were analyzed separately and in two ways. A traditional omnibus repeated measures ANOVA was conducted using anteriority (anterior vs. posterior) and laterality (left vs. midline vs. right) as topographic factors. Since the current study was designed to modulate two ERP components with well-established topographic distributions, a repeated measures ANOVA was conducted in each of the six regions of interest (ROI) to test for the predicted differences (left-anterior: F3, FC3, C3; midline-anterior: FZ, FCZ, CZ; right-anterior: F4, FC4, C4; left-posterior: CP3, P3, O1; midline-posterior: CPZ, PZ, OZ; right-posterior: CP4, P4, O2). Data from Experiment 1 were analyzed using combinability (combinable vs. non-combinable) and animacy-congruity (animacy-congruous vs. animacy-violated) as within-subjects factors. Follow-up comparisons were carried out only when the interaction between animacy-congruity and combinability reached statistical significance. Data from Experiment 2 were analyzed using role-reversal (control vs. role-reversed) as a within-subjects factor.

Acknowledgments

We would like to thank Brian Dillon, Taomei Guo, Fengqin Liu and Peiyao Chen for their valuable help in carrying out these studies. We would like to thank Gina Kuperberg for valuable discussion of earlier versions of this paper. This work was supported in part by NSF Grant BCS-0848554 to CP.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.brainres.2013.02.016.

REFERENCES

Bever, T., 1970. The cognitive basis for linguistic structures. In: Hayes, R. (Ed.), Cognition and Language Development. Wiley, New York, pp. 279–362.

Bicknell, K., Elman, J.L., Hare, M., McRae, K., Kutas, M., 2010. Effects of event knowledge in processing verbal arguments. J. Mem. Lang. 63, 489–505.

Bornkessel-Schlesewsky, I., Kretzschmar, F., Tune, S., Wang, L., Genç, S., Philipp, M., Roehm, D., Schlesewsky, M., 2011. Think globally: cross-linguistic variation in electrophysiological activity during sentence comprehension. Brain Lang. 117, 133–152.

Bornkessel-Schlesewsky, I., Schlesewsky, M., 2008. An alternative perspective on "semantic P600" effects in language comprehension. Brain Res. Rev. 59, 55–73.

Brown, C., Hagoort, P., 1993. The processing nature of the N400: evidence from masked priming. J. Cogn. Neurosci. 5, 34–44.

Brouwer, H., Fitz, H., Hoeks, J., 2012. Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. Brain Res. 1446, 127–143.

Cai, Q., Brysbaert, M., 2010. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. Plos ONE 5 (6), e10729.

Camblin, C.C., Gordon, C.G., Swaab, T.Y., 2007. The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. J. Mem. Lang. 56, 103–128.

Caramazza, A., Zurif, E.B., 1976. Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. Brain Lang. 3, 572–582.

Carpenter, P.A., Just, M.A., 1975. Sentence comprehension: a psycholinguistic processing model of verification. Psychol. Rev. 82, 45–73.

Curran, T., Tucker, D.M., Kutas, M., Posner, M.T., 1993. Topography of the N400: brain electrical activity reflecting semantic expectancy. Electroencephalogr. Clin. Neurophysiol. 88, 188–209.

Deacon, D., Hewitt, S., Yang, C.M., Nagata, M., 2000. Event-related potential indices of semantic priming using masked and unmasked words: evidence that the N400 does not reflect a post-lexical process. Cogn. Brain Res. 9, 137–146.

Delong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat. Neurosci. 8, 1117–1121.

Ditman, T., Holcomb, P.J., Kuperberg, G.R., 2007. The contributions of lexico-semantic and discourse information to the

resolution of ambiguous categorical anaphors. Lang. Cogn. Process. 22, 793–827.
Federmeier, K.D., Kutas, M., 1999. A rose by any other name: long-term memory structure and sentence processing. J. Mem. Lang. 41, 469–495.
Federmeier, K.D., 2007. Thinking ahead: the role and roots of prediction in language comprehension. Psychophysiology 44, 491–505.
Ferreira, F., Bailey, K.G.D., Ferraro, V., 2002. Good-enough representations in language comprehension. Curr. Direct. Psychol. Sci. 11, 11–15.
Ferreira, F., Clifton, C., 1986. The independence of syntactic processing. J. Mem. Lang. 25, 348–368.
Fischler, I., Bloom, P., Childers, D., Roucos, S., Perry, N., 1983. Brain potentials related to stages of sentence verification. Psychophysiology 20, 400–409.
Ford, M., Bresnan, J., Kaplan, R., 1982. A competence based theory of syntactic closure. In: Bresnan, J. (Ed.), The Mental Representation of Grammatical Relations. The MIT Press, Cambridge, MA.
Frazier, L., 1987. Sentence processing: a tutorial review. In: Coltheart, M. (Ed.), Attention and Performance XII: The Psychology of Reading. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 559–586.
Friederici, A.D., 1995. The time course of syntactic activation during language processing: a model based on neuropsychological and neurophysiological data. Brain Lang. 50, 259–281.
Friederici, A.D., Hahne, A., von Cramon, D.Y., 1998. First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: electrophysiological evidence for a double dissociation. Brain Lang. 62 (3), 311–341.
Friederici, A.D., Steinhauer, K., Frisch, S., 1999. Lexical integration: sequential effects of syntactic and semantic information. Mem. Cogn. 27, 438–453.
Hagoort, P., 2003. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. J. Cogn. Neurosci. 15 (6), 883–899.
Hagoort, P., Baggio, G., Willems, R., 2009. Semantic unification. In: Gazzaniga, M.S. (Ed.), The Cognitive Neurosciences fourth ed. MIT Press, Cambridge, MA, pp. 819–836.
Hagoort, P., Brown, C.M., Groothusen, J., 1993. The syntactic positive shift (SPS) as an ERP measure of syntactic processing. Lang. Cogn. Process. 8, 439–484.
Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304, 438–441.
Hahne, A., Friederici, A.D., 2002. Differential task effects on semantic and syntactic processes as revealed by ERPs. Cogn. Brain Res. 13, 339–356.
Hoeks, J.C.J., Stowe, L.A., Doedens, G., 2004. Seeing words in context: the interaction of lexical and sentence level information during reading. Cogn. Brain Res. 19, 59–73.
Jackendoff, R., 2002. Foundations of Language. Oxford University Press, Oxford.
Kim, A., Osterhout, L., 2005. The independence of combinatory semantic processing: evidence from event-related potentials. J. Mem. Lang. 52, 205–225.
Kim, A., Sikos, L., 2011. Conflict and surrender during sentence processing: an ERP study of syntax-semantics interaction. Brain Lang. 118, 15–22.
Kolk, H.H.J., Chwilla, D.J., van Herten, M., Oor, P., 2003. Structure and limited capacity in verbal working memory: a study with event-related potentials. Brain Lang. 85, 1–36.
Kuperberg, G.K., 2007. Neural mechanisms of language comprehension: challenges to syntax. Brain Res. 1146, 23–49.
Kuperberg, G.R., Sitnikova, T., Caplan, D., Holcomb, P.J., 2003. Electrophysiological distinctions in processing conceptual relationships within simple sentences. Cogn. Brain Res. 217, 117–129.
Kuperberg, G.R., Caplan, D., Sitnikova, T., Eddy, M., Holcomb, P.J., 2006. Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. Lang. Cogn. Process. 21 (5), 489–530.
Kuperberg, G., Kreher, D.A., Sitnikova, T., Caplan, D., Holcomb, P., 2007. The role of animacy and thematic relationships in processing active English sentences: evidence from event-related potentials. Brain Lang. 100, 223–238.
Kutas, M., Federmeier, K.D., 2000. Electrophysiology reveals semantic memory use in language comprehension. Trends Cogn. Sci. 4, 463–470.
Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62, 621–647.
Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207, 203–205.
Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307, 161–163.
Laszlo, S., Federmeier, K., 2009. A beautiful day in the neighborhood: an event-related potential study of lexical relationships and prediction in context. J. Mem. Lang. 61, 326–338.
Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9, 920–933.
MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S., 1994. The lexical nature of syntactic ambiguity resolution. Psychol. Rev. 101, 676–703.
McElree, B., Griffith, T., 1995. Syntactic and thematic processing in sentence comprehension: evidence for a temporal dissociation. J. Exp. Psychol.: Learn. Mem. Cogni. 21, 134–157.
Nieuwland, M.S., Kuperberg, G.R., 2008. When the truth isn't too hard to handle: an event-related potential study on the pragmatics of negation. Psychol. Sci. 19, 1213–1218.
Oishi, H., Sakamoto, T., 2010. Immediate Interaction Between Syntactic and Semantic Outputs: Evidence from Event-Related Potentials in Japanese Sentence Processing. Poster Presented at the 22nd Annual CUNY Human Sentence Processing Conference, Davis, CA.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.
Osterhout, L., Holcomb, P.J., 1992. Event-related potentials elicited by syntactic anomaly. J. Mem. Lang. 31, 785–806.
Paczynski, M., Kuperberg, G.R., 2011. Electrophysiological evidence for use of the animacy hierarchy, but not thematic role assignment, during verb argument processing. Lang. Cogn. Process. 26 (9), 1402–1456.
Paczynski, M., Kuperberg, G.R., 2012. Multiple influences of semantic memory on sentence processing: distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. J. Mem. Lang. 67 (4), 426–448.
Pickering, M.J., Traxler, M.J., Crocker, M.W., 2000. Ambiguity resolution in sentence processing: evidence against frequency-based accounts. J. Mem. Lang. 43, 447–475.
Rugg, M.D., 1985. The effects of semantic priming and word repetition on event-related potentials. Psychophysiology 22, 642–647.
Slobin, D., 1966. Grammatical transformations and sentence comprehension in childhood and adulthood. J. Verbal Learn. Verbal Behav. 5, 219–227.
Stroud, C., 2008. Structural and Semantic Selectivity in the Electrophysiology of Sentence Comprehension. Ph.D. Dissertation, University of Maryland College Park.

Stroud, C., Phillips, C., 2012. Examining the evidence for an independent semantic analyzer: an ERP study in Spanish. Brain Lang. 120 (2), 108–126.

Townsend, D.J., Bever, T.G., 2001. Sentence Comprehension: The Integration of Habits and Rules, Cambridge, MA. MIT Press.

Trueswell, J.C., Tanenhaus, M.K., Kello, C., 1993. Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. J. Exp. Psychol.: Learn. Mem. Cogn. 19 (3), 528–553.

Trueswell, J.C., Tanenhaus, M.K., Garnsey, S.M., 1994. Semantic influences on parsing: use of thematic role information in syntactic disambiguation. J. Mem. Lang. 33, 285–318.

Urbach, T.F., Kutas, M., 2010. Quantifiers more or less quantify online: ERP evidence for partial incremental interpretation. J. Mem. Lang. 63, 158–179.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol.: Learn. Mem. Cogni. 31, 443–467.

Van Berkum, J.J.A., 2009. The neuropragmatics of 'simple' utterance comprehension: an ERP review. In: Sauerland, U., Yatsushiro, K. (Eds.), Semantics and Pragmatics: From Experiment to Theory. Palgrave Macmillan, Basingstoke, pp. 276–316.

Van Berkum, J.J.A., Van den Brink, D., Tesink, C.M.J.Y., Kos, M., Hagoort, P., 2008. The neural integration of speaker and message. J. Cogn. Neurosci. 20 (4), 580–591.

Van de Meerendonk, N., Kolk, H.H.J., Chwilla, D.J., Vissers, W.M., 2009. Monitoring in language perception. Lang. Linguist. Compass, 1211–1224.

Van de Meerendonk, N., Kolk, H.H.J., Vissers, C.Th.W.M., Chwilla, D.J., 2010. Monitoring in language perception: mild and strong conflicts elicit different ERP patterns. J. Cogn. Neurosci. 22, 67–82.

Van den Brink, D., Brown, C.M., Hagoort, P., 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. J. Cogn. Neurosci. 13 (7), 967–985.

Van Herten, M., Chwilla, D.J., Kolk, H.H., 2006. When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. J. Cogn. Neurosci. 18, 1181–1197.

Van Herten, M., Kolk, H.H.J., Chwilla, D., 2005. An ERP study of P600 effects elicited by semantic anomalies. Cogn. Brain Res. 22, 241–255.

Van Petten, C., Luka, B.J., 2012. Prediction during language comprehension: benefits, costs, and ERP components. Int. J. Psychophysiol. 83, 176–190.

Wicha, N.Y.Y., Moreno, E.M., Kutas, M., 2004. Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. J. Cogni. Neurosci. 16, 1272–1288.

Ye, Z., Zhou, X., 2008. Involvement of cognitive control in sentence comprehension: evidence from ERPs. Brain Res. 1203, 103–115.