Brain & Language 120 (2012) 321–331

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l

Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier–noun combinations

Yaxu Zhang [a,b,*], Jinlu Zhang [a,1], Baoquan Min [c]

[a] Department of Psychology, Peking University, Beijing 100871, China
[b] Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China
[c] Xuanwu Hospital, Capital Medical University, Beijing 100053, China

A R T I C L E I N F O

Article history:
Accepted 24 October 2011
Available online 17 November 2011

Keywords:
Sentence comprehension
Animacy
Classifier–noun combinations
N400
P600
Event-related brain potentials

A B S T R A C T

An event-related potential experiment was conducted to investigate the temporal neural dynamics of animacy processing in the interpretation of classifier–noun combinations. Participants read sentences that had a non-canonical structure, *object noun* + *subject noun* + *verb* + *numeral-classifier* + *adjective*. The object noun and its classifier were either (a) congruent, (b) incongruent, but matching in animacy, or (c) incongruent, mismatching in animacy. An N400 effect was observed for both incongruent conditions, but not for additional mismatch in animacy. When only data from participants who accepted the non-canonical structure were analyzed, the animacy mismatch elicited a P600 but still no N400. These findings suggest that animacy information is not used immediately for semantic integration of nouns and their classifiers, but is used in a later analysis reflected by P600. Thus, the temporal neural dynamics of animacy processing in sentence comprehension may be modulated by the relevance of animacy to thematic interpretation.

1. Introduction

Things in the world, such as humans, animals, and artifacts, and nouns denoting them, can differ in animacy. The human brain also appears to honor the property of animacy. For example, brain-damaged patients with category-specific semantic deficits can be disproportionately impaired for living things compared with non-living things, or the reverse: disproportionate impairment for non-living things compared with living things (for reviews and theoretical discussion, see Capitani, Laiacona, Mahon, & Caramazza, 2003; Caramazza & Mahon, 2003; Mahon & Caramazza, 2009). In addition, evidence from functional neuroimaging studies indicates that there are distinct regions of the brain involved to the response to animals and tools, the typical living and non-living things, respectively (e.g., Chao, Haxby, & Martin, 1999; see Martin (2007) for review).

During the past decades, numerous studies have investigated the role of noun animacy during on-line sentence comprehension, by examining both the processing of complex or syntactically ambiguous sentences and the processing of simple, syntactically unambiguous sentences. For example, some eye-tracking studies (e.g., Clifton et al., 2003; Ferreira & Clifton, 1986; Trueswell, Tanenhaus, & Garnsey, 1994) have demonstrated the influence of the animacy of sentence-initial noun phrase on the resolution of local syntactic ambiguity between main verb and reduced relative constructions, as in the sentence *The defendant/evidence examined by the lawyer turned out to be unreliable*, although the exact time course of the use of the animacy information remains a matter of debate (see Hsieh, Boland, Zhang, & Yan, 2009 for the use of noun animacy in revolving local ambiguity between two more complex syntactic interpretations).

By measuring eye movements, some other studies have demonstrated that syntactic complexity effects, that is, object-relative clauses being harder to process than subject-relative clauses, can be modulated by the animacy of the sentential subject and the noun within the relative clause (e.g., Mak, Vonk, & Schriefers, 2002, 2006; Traxler, Morris, & Seely, 2002; Traxler, Williams, Blozis, & Morris, 2005). Specifically, the difficulty associated with (complex) object relatives was reduced or even eliminated when the sentential subject was inanimate and the noun within the relative clause was animate, as in the sentence *The movie that the director watched received a prize at the film festival.*

Recently, two claims have been made about the role of animacy in sentence comprehension. One is that animacy information may be used in a heuristic way sometimes, resulting in a purely

* Corresponding author at: Department of Psychology, Peking University, Beijing 100871, China. Fax: +86 10 6276 1081.
E-mail address: yxzhang@pku.edu.cn (Y. Zhang).

[1] Present address: Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China.

0093-934X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.bandl.2011.10.007

animacy-based assignment of thematic roles (Hoeks, Stowe, & Doedens, 2004; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003). The evidence for this claim comes mainly from event-related brain potential (ERP) studies showing that animacy-related thematic role violations, in which a verb that required an animate agent was actually preceded by an inanimate subject noun phrase, as in the sentences *The meal was devouring*... (Kim & Osterhout, 2005) and *At breakfast the eggs would plant*... (Kuperberg et al., 2007), elicited P600 effects but no N400 effects (see Kuperberg (2007) for a review of studies observing such a pattern). The P600 effects were interpreted as probably reflecting the conflict between the output of the animacy heuristic and the output of the syntactic analysis (for discussion, see Kuperberg, 2007; Kuperberg et al., 2007).

The other claim is even stronger, in which animacy functions as a type of prominence information that influences the establishment of interpretive relations between agent/actor and patient/undergoer arguments even in simple, syntactically unambiguous, and both syntactically and semantically well-formed sentences (e.g., Philipp, Bornkessel-Schlesewsky, Bisang, & Schlesewsky, 2008; Roehm, Schlesewsky, Bornkessel-Schlesewsky, Frisch, & Haider, 2004; see Bornkessel-Schlesewsky and Schlesewsky (2009) for review). For example, Philipp et al. (2008) found a larger N400 for an inanimate agent argument compared to an animate agent argument when they were encountered after an animate patient argument during Chinese sentence comprehension (see Roehm et al., 2004 for the same N400 effects in German). The authors also provided further evidence suggesting that these N400 responses are not due to lexical difference between animate and inanimate nouns. Instead, they have been taken as reflecting the computation of thematic relationship between arguments (see Frisch & Schlesewsky, 2001 for N400 effects reflecting how animacy and case markings interact in the thematic interpretation of German embedded structures).

Note that the animacy N400 reviewed above has suggested a rapid use of animacy information in thematic processing, including both the assignment of thematic roles for nouns and the establishment of thematic relationships between verbs and their arguments (see Bornkessel-Schlesewsky & Schlesewsky, 2009; Kuperberg, 2007 for a brief review of how noun animacy influences case markings of a specific noun according to its thematic role in the grammar of some languages). For thematic processing, there is typically a close link between the animacy of a noun and the thematic role that the noun plays: the agent/actor tends to be high and the patient/undergoer low in animacy (see Bornkessel-Schlesewsky & Schlesewsky, 2009). This may be the reason why animacy information has been shown to be used very rapidly for sentence interpretation (determining "who does what to whom").

In a very recent ERP study in Polish, a language in which the animate/inanimate distinction is reflected in the inflectional morphology of nouns, Szewczyk and Schriefers (2011) directly compared the ERP effects of animacy and semantic violations. The animacy violations were realized by a conflict between the actual animacy value of the object noun and the expected animacy value based on the preceding context, as in the Polish version of *Although it was late autumn and bitter cold, little John was running in the backyard with his neck bare. His worried grandma prepared some wool and knitted* ***an employee*** *for her grandson*. For the semantic violations, the object noun was semantically incongruent with the preceding context, but there was no conflict of animacy, as in "... *knitted* ***a medicine*** *for her grandson*". Although the N400 effects did not differ between the two types of violation, a larger P600 was observed for animacy violations compared to semantic violations. It was therefore concluded that animacy and other "non-grammaticalized", semantic features are processed differently.[2] So far, however, little has been known about whether, as a basic semantic feature, animacy is used immediately in real time processing of word combinations in which one word modifies another, but without thematic processing being involved. It should be if animacy information is invariably prominent in real time combination of individual words in a sentence, regardless of the nature of the role it plays.

In the present study, we investigated the role of animacy in processing Chinese classifier–noun combinations, which do not involve thematic role assignments or the establishment of thematic relationships. In Chinese, a classifier language, noun classifiers conceptually classify the referent of the noun according to its semantic features, such as animacy, shape (length, roundness, etc.), and size. A noun classifier is obligatory when the noun is counted or is used in a demonstrative structure (Li & Thompson, 1981; also see Saalbach & Imai, 2007). In other words, nouns cannot be directly modified by numerals or demonstratives. Instead they are modified by a numeral-classifier combination, similarly to the way mass nouns are quantified in English, as in the Chinese numeral classifier–noun combination *san liang qiche* (three CL-liang [classifier indicating ground vehicles] cars, 'three cars') or *na liang qiche* (that CL-liang car, 'that car') (for a detailed description of Chinese classifier system, see Gao & Malt, 2009; Zhang, 2007).

In linguistic theory, there is a debate as to whether noun classifiers are functional elements (e.g., Cheng & Sybesma, 1999, 2005; Muromatsu, 1998) or semantic elements (e.g., Wu & Bodomo, 2009). According to the former view, noun classifiers in Chinese, a language that has no articles/determiners, carry out some of the functions of determiners that exist in other languages, including a deictic function. In contrast, according to the latter view, noun classifiers are contentful, rather than functional, morphemes indicating the semantic classes of nouns; hence, they impose selectional restrictions on the scope of the noun and help to disambiguate word meaning ambiguity (also see Saalbach & Imai, 2007). In other words, meaningfulness is one of the defining properties of noun classifiers, as suggested by Wu and Bodomo (2009).

Chinese have hundreds of noun classifiers, most of which are used with more than one noun. In addition, often several different classifiers can go with the same noun. More relevant to the present study, some classifiers are used with inanimate nouns only, such as *liang* [classifying ground vehicles] and *ben* [classifying objects that are bound into a book-like form], although others are used with both animate and inanimate nouns, such as *tiao*, which classifies long things and goes with both some animal nouns (snake, fish, dog, etc.) and some artifact nouns (rope, towel, trousers, etc.). In addition, the number of classifiers that can be used with animate (human and animal) nouns is about 10, which is much less than the number of classifiers that can go with inanimate nouns (see Gao & Malt, 2009).[3]

Although a few studies have investigated the nature of mental representation and/or non-linguistic cognitive consequences of Chinese noun classifiers, including the influence on categorization,

[2] Szewczyk and Schriefers (2011) argued that in their study the animate/inanimate distinction has no consequences for thematic processing. However, in the example for animacy violations just described, the object noun "*employee*" contained an animacy violation not only in terms of discourse context, but also in terms of thematic constraint, since the main verb "*knitted*" requires an inanimate noun as its object argument. Thus, thematic processing was actually involved, though the extent is not very clear.

[3] Among the quite small number of classifiers that can occur with animate nouns, about half of them (*pi*, *zhi*, *tiao*, *dai*, and *wei*) can also occur with inanimate nouns and thus are ambiguous in terms of animacy. For example, the classifier *pi* goes with both horse and cloth, *zhi* with both tiger and boat and *tiao* with both snake and trousers. For this reason, only unambiguous, inanimate classifiers were used in the present study.

Table 1
Design and sentence examples for all three critical conditions. Examples are given in Chinese, with English literal glosses and translations. The critical words are in bold and underlined.

Condition	Example
Congruent	(a) 汽车/赵庆丰/看见/<u>一辆</u>/黑色的。Car/Qingfeng Zhao/had seen/**<u>one CL-liang</u>** (classifying ground vehicle)/black. (Qingfeng Zhao had seen a black car.)
Incongruent, Animacy-Match	(b) 台灯/赵庆丰/看见/<u>一辆</u>/便宜的。Desk lamp/Qingfeng Zhao/had seen/**<u>one CL-liang</u>** (classifying ground vehicle)/cheap. (Qingfeng Zhao had seen a cheap desk lamp.)
Incongruent, Animacy-Mismatch	(c) 海豹/赵庆丰/看见/<u>一辆</u>/笨拙的。Seal/Qingfeng Zhao/had seen/**<u>one CL-liang</u>** (classifying ground vehicle)/clumsy. (Qingfeng Zhao had seen a clumsy seal.)

similarity judgment, inductive reasoning, and organization of information in memory (e.g., Gao & Malt, 2009; Saalbach & Imai, 2007), there have been very few studies examining how classifier–noun combinations are processed during on-line sentence comprehension (Zhou et al., 2010). In a very recent ERP study, Zhou et al. (2010) provided evidence for the rapid use of the semantic information embodied in noun classifiers during the processing of classifier–noun combination. They found that compared with the nouns that were congruent with their classifiers, as *changyi* ('chair') in *yi zhang changyi* (one CL-zhang [classifying flat objects or things with a flat surface] chair, 'a chair'), the nouns that were incongruent with their classifiers, as *changyi* ('chair') in *yi tai changyi* (one CL-tai [classifying machines, electric appliances, and performances] chair), elicited a larger negativity (N400) with a broad distribution in the 300–500 ms time window, followed by a late (550–800 ms) anterior negativity. The N400 was interpreted as reflecting semantic integration between nouns and their classifiers and the late anterior negativity as reflecting either a semantic reinterpretation or a heavy memory load caused by the semantic incongruence or both. However, note that Zhou et al. did not specifically manipulate animacy congruency between nouns and their classifiers: both the nouns and their classifiers were always inanimate.

The present study aimed at investigating whether animacy information is used immediately in real time combination of Chinese nouns and their classifiers. We used three types of sentences, like (1a)–(1c) in Table 1, as the critical items. These sentences contained a structure of *object noun* + *subject noun* + *verb* + *numeral classifier* + *adjective*, in which the object nouns were modified both by the numeral classifiers and by the adjectives, with the numeral classifiers serving as the critical words in the ERP experiment. Given the quite small number of the classifiers that go with animate nouns only (about 5; see above and Footnote 3), we used only unambiguous, inanimate classifiers. These classifiers can go with some artifact nouns, but not any animate nouns. For example, *liang*, a classier classifying ground vehicles, can modify *qiche* ('car'), but not artifact nouns like *taideng* ('desk lamp') or animate nouns like *haibao* ('seal').

As shown in Table 1, within each triplet, the three types of sentence differed in the object noun and sentence-final adjective only. In the Congruent condition, the object noun, which was inanimate, was congruent with the classifier. For the two incongruent conditions, the object noun was incongruent with the classifier. More importantly, whereas the object noun was an inanimate, artifact noun and thus matched with the classifier in animacy for the Incongruent, Animacy-Match condition, the object noun was an animate, animal noun and thus did not match with the classifier in animacy for the Incongruent, Animacy-Mismatch condition.[4]

The reason why the construction of *object noun* + *subject noun* + *verb* + *numeral-classifier* + *adjective*, a grammatical but non-canonical structure, was used is to have the numeral-classifier rather than the object noun as the critical word. The canonical counterpart of the structure we employed is *subject noun* + *verb* + *numeral-classifier* + *adjective* + *object noun*. If using such a canonical structure, one would have to use the object noun as the critical word. In that case, in order to avoid the influence of lexical difference between animate and inanimate nouns on the ERP effects of interest, one would have to manipulate the animacy of classifiers, which, however, is difficult for an ERP experiment given that there are only about five classifiers that can go with animate nouns, as mentioned earlier. While the findings from non-canonical structures have limited generalizability, they do provide valuable insight into how and when various types of information are used in the establishment of syntactic dependencies, as noted by many previous studies (e.g., Aoshima, Phillips, & Weinberg, 2004; Dickey & Thompson, 2004; Felser, Clahsen, & Münte, 2003; Phillips, Kazanina, & Abada, 2005; Schriefers, Friederici, & Kühn, 1995).

We expect a larger N400 at the numeral classifiers for the two incongruent conditions compared to the Congruent condition, based on the previous study that observed an N400 response for the classifier–noun mismatch (Zhou et al., 2010). Such an N400 would support the linguistic hypothesis that Chinese noun classifiers are semantic elements, imposing selectional restrictions on the scope of the noun (e.g., Wu & Bodomo, 2009). More importantly, the numeral classifiers would evoke a larger N400 in the Incongruent, Animacy-Mismatch condition than in the Incongruent, Animacy-Match condition, if the animacy information embodied in the classifiers and object nouns is used immediately during the semantic processing of classifier–noun combinations. The reason why such a prediction can be made is that N400 amplitude has been amply demonstrated to be highly sensitive to varying degrees of semantic relationships (e.g., Federmeier & Kutas, 1999a, 1999b; Li, Shu, Liu, & Li, 2006; for review, see Kutas & Federmeier, 2000). For example, using Chinese verb–noun pairs as stimulus, Li et al. (2006) manipulated semantic fitness between object nouns and their preceding verbs that required human nouns as their objects, as *guyong* ('hire'). In the Congruent condition, the object nouns were perfect objects of the verbs, as *baobiao* ('bodyguard'). For the incongruent conditions, the object nouns were incongruent with the verbs either in terms of specific semantic features only, as *yinger* ('baby'), or in terms of both specific and broader (humanness and animacy) semantic features, as *muji* ('hen') and *dianxian* ('wire'). Compared with the congruent object nouns, the incongruent nouns elicited a larger N400. More importantly, the additional incongruence in humanness or in both humanness and animacy resulted in a further increase of N400 amplitude (for findings of N400 amplitude varying with the fitness between words and their sentence context, see Federmeier & Kutas, 1999a, 1999b).

It seems reasonable to hypothesize a rapid evaluation of animacy congruency between nouns and their classifiers, resulting in an N400, if animacy indeed has an invariable prominence for real time combination of individual words in a sentence, especially for

[4] Since the subject noun was human, the object noun for the Animacy-Mismatch condition was designed to be an animal noun, rather than a human one, to induce animacy differences between the subject and object arguments and reduce the degree of processing difficulty.

Chinese, a language in which animacy is a relatively strong cue to the identification of the agent of a sentence (Li, Bates, & MacWhinney, 1993) and sentence comprehension has been assumed to rely on semantic analyses to a relatively large degree (e.g., Xu, 1997; for evidence supporting this conjecture, see Li, 1998; Yu & Zhang, 2008; Zhang, Yu, & Boland, 2010). Alternatively, instead of an N400, an ERP response reflecting a later analysis would occur, if the use of animacy information is merely restricted to the later stage of classifier–noun combinations.

2. Methods

2.1. Participants

Thirty healthy students gave informed consent to participate in the experiment (mean age 22 years, range: 19–25; 15 females). All participants were native speakers of Mandarin Chinese, right-handed and had normal or corrected-to-normal vision. They were paid a nominal sum for their participation. The experiment was approved by the Academic Committee of the Department of Psychology, Peking University.

2.2. Materials and normative measures

The critical materials consisted of 120 triplets of Chinese sentences (see Table 1 for examples). These sentences were assigned to three experimental lists using a Latin square procedure. In order to counterbalance the number of correct and incorrect sentences in the experiment and offset the sentence structure used in the critical items, in each list, the 120 items were mixed with 280 filler sentences. One-hundred and twenty of the fillers were critical items of an unrelated experiment and were of a structure of subject–verb–object (SVO) (two-third of them contained an anomaly of verb transitivity or a semantic incongruence between verb and its object). The other fillers consisted of 40 correct subject–verb sentences and 120 sentences having the same construction as the critical items (60 of them were correct and began with an animal noun, 20 of them were correct and began with an artifact noun, and 40 of them contained a semantic incongruence between verb and animal or artifact object noun).

As mentioned earlier, the numeral classifiers served as the critical words in the ERP experiment. Totally 59 classifiers were used, with most (45/59 or 76%) of them being from a classifier list developed by Gao and Malt (2009). We did not use any ambiguous classifiers that can modify both artifact and animal nouns, such as *tiao* [classifying slender, long-shape things].

The sentence-final adjectives were always congruent with the object nouns and were matched across conditions for written word frequencies and number of strokes (see Table 2). The ANOVA performed on the mean word frequencies did not reveal an effect of condition, $F(2,357) = 1.805$, $p = 0.166$, $MSE = 487.33$, nor did the ANOVA performed on the mean number of strokes, $F < 1$.

In addition, the sentence-initial object nouns were matched across conditions for written word frequencies, but not for number of strokes (see Table 2). The ANOVA performed on the mean word frequencies did not reveal an effect of condition, $F(2,357) = 1.88$, $p = 0.154$, $MSE = 2859.70$, although the ANOVA performed on the mean number of strokes did, $F(2,357) = 3.53$, $p = 0.03$, $MSE = 37.30$. The unmatched number of strokes was not considered a problem because the sentence-initial noun was the third segment before the critical word (numeral classifier, see Table 1).

In order to evaluate the degree of sentence acceptability, we conducted a sentence acceptability survey, in which a separate group of 30 participants was asked to judge the acceptability of each sentence on a 7-point scale, with '1' indicating that the sentence was completely unacceptable and '7' completely acceptable. One-hundred and twenty triplets of sentences were assigned to three experimental lists by using a Latin square procedure, with 120 sentences each. In each list, the 120 items were presented in a pseudo-random order, such that at most three trials with the same condition occurred consecutively. For each list, another version with a reverse order was formed to further counterbalance the order effects. Each participant received only one of the six experimental lists. Table 2 shows the rating results for the 120 triplets of items used in the ERP experiment. An analysis of variance (ANOVA) performed on the acceptability scores revealed a main effect of condition [by subjects, $F_1(2,58) = 147.47$, $p < 0.0005$, $MSE = 0.38$; by items, $F_2(2,357) = 519.48$, $p < 0.0005$, $MSE = 0.44$]. Post-hoc Newman–Keuls comparisons showed that the congruent sentences were rated more acceptable than both types of incongruent sentences ($ps < 0.01$). In addition, the sentences were more unacceptable in the Incongruent, Animacy-Mismatch than in the Incongruent, Animacy-Match condition ($ps < 0.01$).

Table 2
Mean frequencies (per million words) and number of strokes of the object nouns and adjectives, acceptability scores of the sentences, and congruence scores of the noun–numeral classifier pairs for all three critical conditions.

	Congruent	Incongruent, Animacy-Match	Incongruent, Animacy-Mismatch
Object noun			
Frequency	18.80 (88.85)	10.62 (20.29)	5.52 (16.53)
Number of strokes	17.62 (5.98)	15.58 (6.47)	19.41 (5.85)
Adjective			
Frequency	8.64 (16.47)	13.87 (25.95)	12.48 (22.75)
Number of strokes	23.98 (4.87)	24.18 (4.85)	24.17 (5.07)
Acceptability score of sentence	5.22 (1.06)	3.23 (1.22)	2.57 (1.15)
Congruence score of noun–numeral classifier pair	6.19 (0.57)	1.82 (0.52)	1.38 (0.37)

Note that the acceptability score was not very high even for the congruent sentences, due to the non-canonical structure (*object noun + subject noun + verb + numeral classifier + adjective*) used for these sentences (for a graded effect of word order on sentence acceptability ratings, see Friederici, Fiebach, Schlesewsky, Bornkessel-Schlesewsky, & von Cramon, 2006). In addition, although the sentences were more unacceptable in the Incongruent, Animacy-Mismatch than in the Incongruent, Animacy-Match condition, there are two alternative explanations for such a difference. First, the classifiers were more incongruent with the object nouns for the Incongruent, Animacy-Mismatch compared to the Incongruent, Animacy-Match condition, due to the additional mismatch in animacy. Second, there was a difference in the animacy hierarchy between the two conditions. For the Incongruent, Animacy-Match condition, the subject and object were animate and inanimate nouns, respectively. In contrast, for the Incongruent, Animacy-Mismatch condition, both the subject and object were animate nouns, resulting in the sentences being more unacceptable.

In order to justify the graded difference in the degree of meaning congruence of the object nouns and their classifiers between the three conditions, we conducted a meaning congruence survey, in which a separate group of 30 participants was asked to judge the meaning congruence for each noun–numeral classifier pair, as *qiche – yi-liang* (car – one CL-liang [classifying ground vehicles]), on a seven-point scale, with '1' indicating that the two words within a pair were completely incongruent and '7' completely congruent. One-hundred and twenty triplets of noun–numeral classifier pairs were assigned to three experimental lists by using a Latin square procedure, with 120 word pairs each. In each list, the 120 critical word pairs were pseudo-randomly mixed with 150 verb–

noun pairs (filler items), such that at most three critical trials occurred consecutively. For each list, another version with a reverse order was formed to further counterbalance the order effects. Each participant received only one of the six experimental lists.

Table 2 shows the rating results for the critical word pairs used in the ERP experiment. An analysis of variance (ANOVA) performed on the congruence scores revealed a main effect of condition [by subjects, $F_1(2,58) = 1173.23$, $p < 0.0005$, $MSE = 0.28$; by items, $F_2(2,357) = 2698.93$, $p < 0.0005$, $MSE = 0.31$]. Post-hoc Newman–Keuls comparisons showed that the noun–numeral classifier pairs were rated more congruent in the Congruent condition than in the incongruent conditions ($ps < 0.01$). In addition, the word pairs were more incongruent in the Incongruent, Animacy-Mismatch than in the Incongruent, Animacy-Match condition ($ps < 0.01$), suggesting that the participants' ratings were sensitive to whether or not there was a mismatch in animacy between nouns and their classifiers.

2.3. Procedure

Participants were seated in a comfortable chair approximately 1 m from the computer screen in a dimly lit and sound-attenuated room. They read the sentences sequentially, as each word (or sometimes a short phrase) appeared in the center of the screen. Each trial started with a central fixation cross presented for 800 ms, followed by a 500 ms blank screen. Each word or short phrase was presented for 400 ms, with an additional 100-ms inter-stimulus interval.[5] After the presentation of the last segment of the sentence there was an 800-ms blank, followed by a row of question marks ('?????') reminding participants to judge the overall (syntactic **and** semantic) acceptability of each sentence by pressing one of two buttons (half of the participants pressed the "YES" button with their left index finger and the other half with their right index finger). This cue remained on the screen until the participant had responded or for maximum 3 s. Prior to the experimental blocks, participants received a practice block of 24 trials. The experimental session lasted about 1 h.

2.4. ERP recording

The EEG was recorded from 62 Ag/AgCl electrodes mounted in an elastic cap (Quik-Cap, NeuroScan Inc., Herndon, Virginia, USA). The electrodes were placed in the following sites: AF7, AF3, FP1, FPz, FP2, AF4, AF8, F7, F5, F3, F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, C3, C1, Cz, C2, C4, C6, T8, TP7, CP5, CP3, CP1, CPz, CP2, CP4, CP6, TP8, P7, P5, P3, P1, Pz, P2, P4, P6, P8, PO7, PO5, PO3, POz, PO4, PO6, PO8, O1, Oz, and O2. Recordings were referenced to the left mastoid, but re-referenced to linked mastoids offline. The horizontal electrooculogram (EOG) was recorded from electrodes placed at the outer canthus of each eye and the vertical EOG was recorded from electrodes placed above and below the participants' left eye. Electrode impedances were kept below 5 kΩ. The EEG and EOG were amplified with a band-pass from DC to 70 Hz and recorded continuously with a digitization rate of 500 Hz.[6] ERPs were additionally filtered off-line with 30 Hz low pass for the plots only. All statistical analyses were performed on the original data.

2.5. ERP data analysis

ERPs time-locked to the critical words (the numeral classifiers) were computed for each participant, condition, and electrode site. A detrending algorithm was applied to correct for a common linear component caused by the slow voltage shifts that are common for DC EEG recordings. We used a 15 s time interval, ranging from the onset of the critical words to 15 s after them, to estimate the linear component of any slow voltage shifts (for the use of a similar procedure, see Angrillia, Dobel, Rockstroh, Stegagno, & Elbert, 2000; Fiebach, Schlesewsky, & Friederici, 2002; Phillips et al., 2005; Zhang et al., 2010). The subsequent analyses were based on 1200 ms epochs, ranging from 200 ms before the onset of the critical words to 1000 ms after them. We used a 200-ms pre-stimulus baseline. All epochs were evaluated individually for EOG or other artifacts. Epochs with amplitudes exceeding ±100 μV were excluded from the averages through artifact rejection. The overall rejection rate was 5.7%, equal for all three conditions (Congruent: 5.5%; Incongruent, Animacy-Match: 6.1%; and Incongruent, Animacy-Mismatch: 5.6%).

Two time windows were chosen on the basis of visual inspection and earlier studies: 300–550 ms for N400 effects and 600–1000 ms for possible P600 effects. All statistical analyses were performed on the mean amplitudes in the selected time windows. ERPs were analyzed separately for midline and lateral electrodes. From the 62 electrodes, we first selected the three most often reported midline electrodes Fz, Cz, and Pz. Omnibus ANOVAs for midline electrodes included two within-subject factors: Electrode (Fz/Cz/Pz) and Condition (Congruent/Incongruent, Animacy-Match/Incongruent, Animacy-Mismatch). In order to cover distributional differences in both the left–right and anterior-to-posterior dimensions, we selected 36 lateral electrodes and formed six lateral regions of interest, with six electrodes each, by crossing Hemisphere (Left/Right) and Region (Anterior/Central/Posterior), similar to what has been done in some previous studies (e.g., Friederici & Frisch, 2000; Yu & Zhang, 2008; Zhang et al., 2010): left anterior (F3, F5, F7, FC3, FC5, and FT7); left central (C3, C5, T7, CP3, CP5, and TP7); left posterior (P3, P5, P7, PO3, PO7, and O1); right anterior (F4, F6, F8, FC4, FC6, and FT8); right central (C4, C6, T8, CP4, CP6, and TP8); and right posterior (P4, P6, P8, PO4, PO8, and O2). Omnibus ANOVAs for lateral electrodes included three within-subject factors: Hemisphere, Region, and Condition (Congruent/Incongruent, Animacy-Match/Incongruent, Animacy-Mismatch).

Only effects involving the factor Condition are reported. Post-hoc Newman–Keuls comparisons were performed only when the ANOVA revealed a significant effect of Condition. In addition, the Greenhouse–Geisser correction was applied when evaluating effects with more than one degree of freedom in the numerator. In

[5] In order to avoid some peculiar processing strategies being induced by a relatively slow presentation rate, we used the relatively fast presentation rate (500 ms/word). Note that reliable N400 and P600 effects were observed in ERP studies of sentence processing using a presentation rate of 500 ms/word (e.g., in German: Friederici & Frisch, 2000; Friederici, Steinhauer, & Frisch, 1999; Frisch, Hahne, & Friederici, 2004; in English: Ledoux, Gordon, Camblin, & Swaab, 2007; in Polish: Szewczyk & Schriefers, 2011; in Chinese: Yu & Zhang, 2008; Zhang et al., 2010). However, one might still be concerned about the overlap of ERPs between successive words and the possibility that the baseline for the ERPs of the critical words (the numeral classifiers) contained overlapping activity from the immediately preceding words (the verbs). Therefore, we analyzed the mean amplitudes of ERPs in the interval from 300 to 500 ms after the onset of the verbs. As expected, the results revealed neither a main effect of Condition nor any interaction involving Condition, suggesting that there were no systematic ERP activity differences across conditions in the baseline prior to the onset of the critical words. Thus, even if there was an overlap from the verbs, the overlap had the same influence on the ERPs to the critical words for each of the three conditions (see Luck, Woodman, & Vogel, 2000 for the subtraction procedure being used both to eliminate overlap and to isolate specific ERP components such as N400). In addition, we also analyzed the ERPs elicited by the adjectives, the words immediately following the critical words, to address the overlap from these words (for the results, see Section 3.2.2).

[6] In order to observe any possible slow potentials, the DC recording was used. A detrending algorithm was correspondingly used to correct for the possible slow drift caused by the DC recording (see Section 2.5). Reliable N400 and P600 effects have been observed in previous ERP studies employing the DC recording (e.g., Friederici & Frisch, 2000; Friederici et al., 1999; Ledoux et al., 2007; Phillips et al., 2005; Yu & Zhang, 2008; Zhang et al., 2010).

these cases, the original degrees of freedom and the corrected *MSE* and probability levels are reported.

3. Results

3.1. Behavioral data

As expected, compared with the two types of incongruent sentences, the congruent sentences were judged to be both syntactically and semantically acceptable more frequently (Congruent: 56.58%, SD = 44.95%; Incongruent, Animacy-Match: 17.50%, SD = 24.77%; Incongruent, Animacy-Mismatch: 9.98%, SD = 21.84%), yielding a main effect of Condition, $F(2,58) = 34.93$, $p < 0.0005$, $MSE = 0.05$. These results suggest that the congruence between the classifier and object noun resulted in the sentences in the Congruent condition being more acceptable. Note that, mirroring the findings in the off-line sentence acceptability survey, in the ERP experiment, the congruent sentences were not judged to be acceptable very frequently and there were large individual differences in the judgment, due to the grammatical but non-canonical structure used in these sentences.

In order to determine whether or not participants read the sentences attentively, we computed the average accuracy of the acceptability judgments for the 120 correct or incorrect filler sentences that were critical items of an unrelated experiment and had a canonical structure (subject–verb–object). For these sentences, a correct response was the judgment of "acceptable" for the correct sentences and "unacceptable" for the incorrect sentences. The average accuracy was 94.89% (SD = 4.69%), suggesting that participants read the sentences attentively.

3.2. ERP data

Fig. 1 shows grand average ERPs elicited by the numeral classifiers for all three critical conditions at eleven representative electrodes.

As shown in Fig. 1, compared with the Congruent condition, both incongruent conditions elicited a larger negativity with a broad distribution in the 300–550 ms time window, with no difference between the two incongruent conditions. In addition, there were no very apparent effects in the 600–1000 ms window. These observations were statistically verified by ANOVAs performed on the mean amplitudes in the 300–550 ms and 600–1000 ms time windows, respectively. The results of the global ANOVAs are shown in Table 3.

3.2.1. The 300–550 ms time window

The global ANOVA revealed a main effect of Condition at both the midline and the lateral electrodes. Post-hoc Newman–Keuls comparisons revealed a larger negativity (N400) for the two incongruent conditions compared to the Congruent condition ($ps < 0.01$), with no difference between the two incongruent conditions ($ps > 0.10$).

In order to determine if the 300–550 ms ERP effects of Condition were modulated by whether or not participants accepted the non-canonical structure used in the critical sentences, we divided participants into two groups, with 15 participants each, according to the percentage of judging the congruent sentences as both syntactically and semantically acceptable. Table 4 shows the percentage of judgments of "syntactically and semantically acceptable" for the sentences in each of the three critical conditions for the Accepting and Not Accepting group, respectively.

The ANOVA performed on the percentage of "acceptable" judgments revealed a Condition × Group interaction, $F(2,56) = 36.16$, $p < 0.0005$, $MSE = 0.04$. This interaction was due to a larger effect of Condition for the Accepting group [$F(2,28) = 86.74$, $p < 0.0005$, $MSE = 0.05$] than for the Not Accepting group [$F(2,28) = 5.24$, $p = 0.034$, $MSE = 0.03$]. Post-hoc Newman–Keuls comparisons revealed a higher percentage of "acceptable" judgments for the Congruent condition compared to the two incongruent conditions for both the Accepting group ($ps < 0.01$) and Not Accepting group ($ps < 0.05$). In addition, the percentage was marginally higher for the Incongruent, Animacy-Match condition compared to the Incongruent, Animacy-Mismatch condition for the Accepting group ($0.05 < p < 0.10$), but it did not differ significantly between the two incongruent conditions for the Not Accepting group ($p > 0.10$).

Figs. 2 and 3 show grand average ERPs elicited by the numeral classifiers for all three critical conditions for the Accepting and Not Accepting group, respectively. As shown in Fig. 2, for the Accepting group, there was a larger negativity with a broad distribution in the 300–550 ms time window (N400) for both incongruent conditions compared to the Congruent condition, with no difference in N400

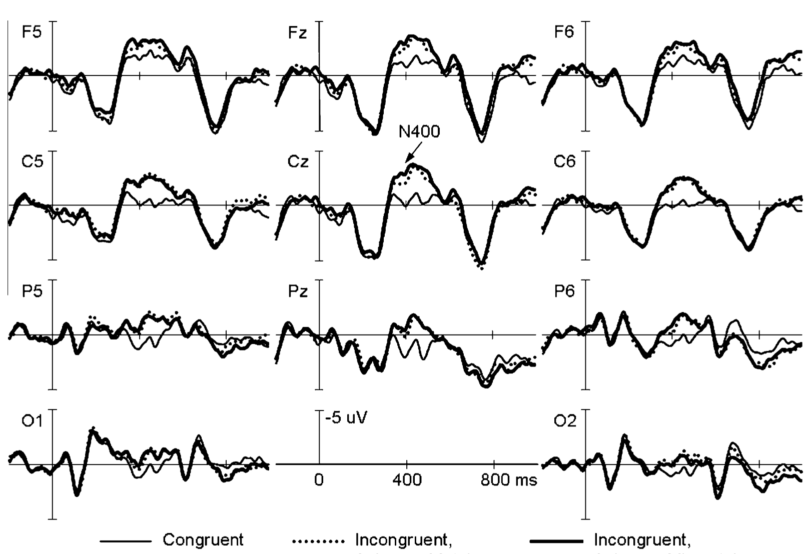

Fig. 1. Grand average ERPs time locked to the onset of the numeral classifiers for all participants ($n = 30$). This figure and the subsequent ones display ERPs over 11 representative electrodes for all three critical conditions.

Table 3
Overall analyses of variance in two time windows (in milliseconds) over midline and lateral electrodes, for amplitude data from all 30 participants.

Source	*dfs*	300–550			600–1000		
		F	*p*	*MSE*	*F*	*p*	*MSE*
Midline electrodes							
Condition	2, 58	14.73	<0.0005	5.88	<1		
Condition × Electrode	4, 116	<1			2.82	0.041	2.88
Lateral electrodes							
Condition	2, 58	15.68	<0.0005	6.21	<1		
Condition × Region	4, 116	<1			6.51	0.002	2.72
Condition × Hemisphere	2, 58	<1			<1		
Condition × Region × Hemisphere	4, 116	<1			2.11	0.096	0.39

Table 4
Mean percentage of judgments of "acceptable" for the sentences in all three critical conditions for the Accepting and Not Accepting group. Values in the parentheses indicate standard deviations.

Group	Condition		
	Congruent	Incongruent, Animacy-Match	Incongruent, Animacy-Mismatch
Accepting	97.17% (2.29%)	31.17% (28.56%)	18.80% (28.57%)
Not Accepting	16.00% (25.53%)	3.83% (7.43%)	1.17% (2.29%)

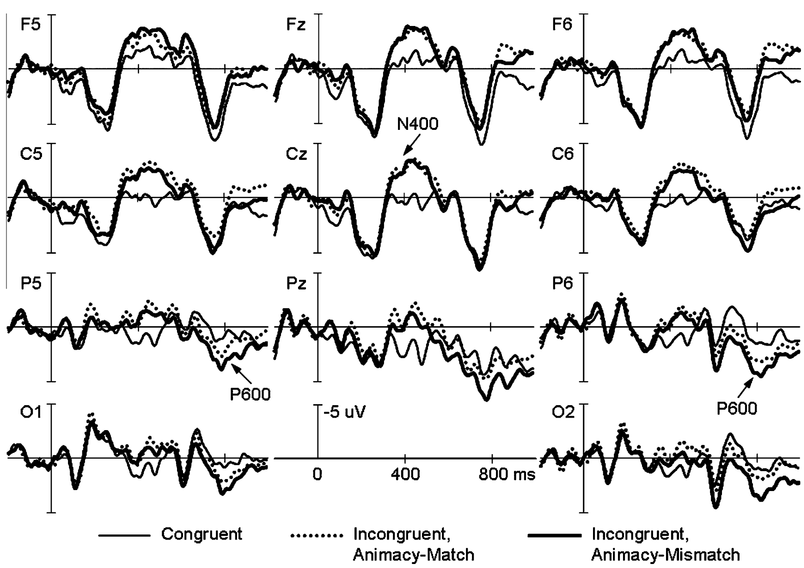

Fig. 2. Grand average ERPs time locked to the onset of the numeral classifiers for participants judging the congruent sentences as acceptable ($n = 15$).

amplitude between the two incongruent conditions. In contrast, as shown in Fig. 3, for the Not Accepting group, the N400 amplitudes were apparently larger for the Incongruent, Animacy-Mismatch condition, but not for the Incongruent, Animacy-Match condition, compared to the Congruent condition. These observations were statistically verified, as described below.

The ANOVA with the factors of Group, Condition, and Electrode (for midline sites) or Region and Hemisphere (for lateral sites) for the mean amplitudes revealed a Condition by Group interaction that was significant at the lateral sites, $F(2,56) = 3.43$, $p < 0.05$, $MSE = 5.59$, and almost reached significance at the midline sites, $F(2,56) = 2.94$, $p = 0.06$, $MSE = 5.41$.

Separate analyses limited to each group revealed an effect of Condition at both the midline and lateral sites for both groups (Accepting group: midline, $F(2,28) = 14.11$, $p < 0.0005$, $MSE = 1.74$, lateral, $F(2,28) = 15.87$, $p < 0.0005$, $MSE = 0.99$; Not Accepting group: midline, $F(2,28) = 5.13$, $p < 0.05$, $MSE = 2.01$, lateral, $F(2,28) = 5.11$, $p < 0.05$, $MSE = 1.10$). Post-hoc Newman–Keuls comparisons for the Accepting group revealed a larger negativity (N400) at both the midline and lateral sites for both incongruent conditions than for the Congruent condition ($ps < 0.01$), with no difference between the two incongruent conditions ($ps > 0.10$). For the Not Accepting group, in contrast, at both the midline and lateral sites, compared with the Congruent condition, the Incongruent, Animacy-Mismatch condition elicited a larger negativity ($ps < 0.05$), but the Incongruent, Animacy-Match condition did not ($ps > 0.10$). In addition, the Incongruent, Animacy-Mismatch condition was marginally more negative than the Incongruent, Animacy-Match condition at both the midline and lateral electrodes ($0.05 < ps < 0.10$).

3.2.2. The 600–1000 ms time window

Although the global ANOVA revealed an interaction of Condition by Electrode and Condition by Region, separate analyses limited to each midline electrode or lateral region did not reveal any reliable effect of Condition (Fz, $F(2,58) = 3.13$, $p = 0.055$, $MSE = 3.26$; Cz and Pz, $Fs < 1$; anterior, $F(2,58) = 2.93$, $p = 0.063$,

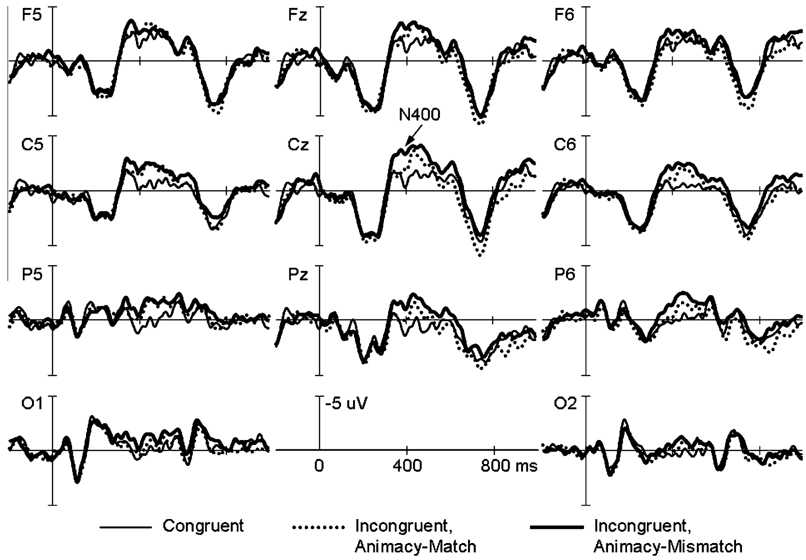

Fig. 3. Grand average ERP time locked to the onset of the numeral classifiers for participants judging the congruent sentences as unacceptable ($n = 15$).

$MSE = 1.61$; central, $F < 1$; posterior, $F(2,58) = 2.90$, $p = 0.07$, $MSE = 2.12$).

As shown in Fig. 2, for the Accepting group, in the 600–1000 ms time window, there was a larger anterior negativity for both incongruent conditions and an apparently larger posterior positivity (P600) for the Incongruent, Animacy-Mismatch condition only, both compared to the Congruent condition. In contrast, as shown in Fig. 3, for the Not Accepting group, there were no apparent effects in this late time window.

Again, in order to statistically determine if the ERP effects of Condition in this time window were modulated by whether or not participants accepted the non-canonical structure used in the critical sentences, we performed an ANOVA with Group, Condition, and Electrode (for midline sites) or Region and Hemisphere (for lateral sites) as factors. The results revealed a Condition × Electrode × Group interaction at the midline sites, $F(4,112) = 2.73$, $p < 0.05$, $MSE = 2.67$, and a Condition × Region × Group interaction at the lateral sites, $F(4,112) = 7.48$, $p < 0.0005$, $MSE = 1.89$.

At the midline sites, separate analyses restricted to each electrode and each group revealed an effect of Condition at Fz for the Accepting group (Accepting: Fz, $F(2,28) = 8.93$, $p < 0.01$, $MSE = 2.69$; Cz, $F < 1$; and Pz, $F(2,28) = 2.71$, $p = 0.09$, $MSE = 4.82$; Not Accepting: Fz, $F < 1$; Cz, $F(2,28) = 2.49$, $p = 0.11$, $MSE = 3.83$; and Pz, $F(2,28) = 2.00$, $p = 0.15$, $MSE = 1.89$). At the lateral sites, separate analyses restricted to each region and each group revealed an effect of Condition at anterior and posterior sites for the Accepting group (Accepting: anterior, $F(2,28) = 9.25$, $p < 0.01$, $MSE = 1.42$; central, $F(2,28) = 1.79$, $p = 0.20$, $MSE = 2.11$; and posterior, $F(2,28) = 5.83$, $p < 0.05$, $MSE = 2.50$; Not Accepting: anterior, $F(2,28) = 2.11$, $p = 0.14$, $MSE = 1.54$; central, $F(2,28) = 2.81$, $p = 0.08$, $MSE = 1.24$; and posterior, $F(2,28) = 1.40$, $p = 0.26$, $MSE = 1.24$). Post-hoc Newman–Keuls comparisons showed that the effects of condition observed in the Accepting group were mainly due to a larger anterior (including Fz) negativity for both incongruent conditions and a larger lateral posterior positivity (P600) for the Incongruent, Animacy-Mismatch condition only, both compared to the Congruent condition ($ps < 0.01$) (see Fig. 2). In addition, for the Accepting group, the Incongruent, Animacy-Mismatch condition was also marginally more positive than the Incongruent, Animacy-Match condition at the lateral posterior sites ($0.05 < p < 0.10$).

In order to ascertain whether the 600–1000 ms positivity specific to the Animacy-Mismatch condition was indeed a P600 elicited by the numeral classifiers or merely early ERP effects elicited by the upcoming adjectives, given the relatively fast presentation rate (500 ms/word), we analyzed the ERPs elicited by the adjectives. Fig. 4 shows the grand average ERPs for all three critical conditions at eleven representative electrodes.

As shown in Fig. 4, in the interval from 100 to 500 ms after the onset of the adjectives (corresponding to the interval from 600 to 1000 ms after the onset of the numeral classifiers), although both incongruent conditions elicited a positivity with a centro-parietal distribution, there were no differences in ERPs between the two incongruent conditions. These observations were statistically verified by ANOVAs performed on the mean amplitudes in this interval. The results revealed a Condition × Electrode interaction at the midline sites, $F(4,56) = 12.69$, $p < 0.0005$, $MSE = 1.29$, as well as a Condition × Region interaction at the lateral sites, $F(4,56) = 25.03$, $p < 0.0005$, $MSE = 1.24$.

At the midline sites, separate analyses restricted to each electrode revealed an effect of Condition at Cz and Pz (Fz, $F < 1$; Cz, $F(2,28) = 6.10$, $p = 0.01$, $MSE = 4.80$; and Pz, $F(2,28) = 18.93$, $p < 0.0005$, $MSE = 3.44$). At the lateral sites, separate analyses restricted to each region revealed an effect of Condition at central and posterior sites (anterior, $F(2,28) = 1.94$, $p = 0.17$, $MSE = 1.38$; central, $F(2,28) = 16.39$, $p < 0.0005$, $MSE = 1.92$; and posterior, $F(2,28) = 27.23$, $p < 0.0005$, $MSE = 2.01$). Post-hoc Newman–Keuls comparisons showed that these effects of condition were due to a larger positivity for both incongruent conditions compared to the Congruent condition (Cz: $ps < 0.05$; the other sites: $ps < 0.01$). The larger positivities most likely reflect more attention triggered by the anomalous classifiers, given that the adjectives were congruent with the object nouns. More importantly, there were no differences in ERPs between the two incongruent conditions ($ps > 0.10$). Thus, the 600–1000 ms positivity revealed by the analysis of the numeral classifiers, which was specific to the Incongruent, Animacy-Mismatch condition, cannot be accounted for by early ERP effects elicited by the adjectives. Instead, it must contain a P600 elicited by the numeral classifiers.

4. Discussion

The goal of the present study is to investigate whether animacy information is used immediately during the real-time semantic combination of Chinese nouns and their classifiers. For this purpose, we compared three types of sentences in which the classifiers

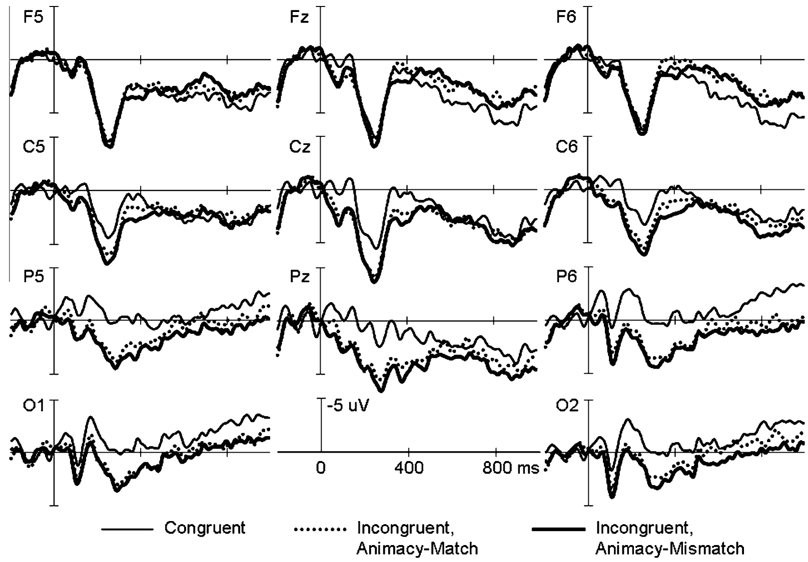

Fig. 4. Grand average ERP time locked to the onset of the adjectives for participants judging the congruent sentences as acceptable ($n = 15$).

and the sentence-initial object nouns were either (a) congruent, (b) incongruent, but matching in animacy, or (c) incongruent, mismatching in animacy. The analysis of data from all participants showed that compared with the congruent classifiers, the incongruent classifiers elicited a larger N400, mirroring the N400 effects evoked by object nouns that were incongruent with their preceding classifiers (Zhou et al., 2010). More importantly, by comparing the Incongruent, Animacy-Mismatch condition with the Incongruent, Animacy-Match condition, we found no increase in N400 amplitude for additional mismatch in animacy.

Considering the fact that a non-canonical structure (OSV) was used in the critical sentences and the effects of Condition may be modulated by whether or not participants accepted such a non-canonical structure, we divided participants into the Accepting and Not Accepting group. We found that the factor of Group did modulate the condition effects, both in the N400 and in the late (600–1000 ms) time window, suggesting that whether or not participants accepted the non-canonical structure is not simply a matter of individual differences in response criterion. For the Accepting group, both types of incongruent classifiers evoked a larger N400 than the congruent classifiers, reflecting difficulty in semantic integration (Kutas & Federmeier, 2000), or semantic implausibility that had essential consequences on the acceptability judgment (more NO responses made for the incongruent conditions compared to the congruent condition). More importantly, the additional mismatch in animacy did not result in an increase in N400 amplitude, which was the same as that observed when the data from all participants were analyzed, suggesting that animacy information is not used during semantic integration between the object nouns and their classifiers, no matter whether or not data from participants who did not accept the non-canonical structure were excluded.

In addition, the N400 effect observed for the Accepting group was followed by a P600 effect for the Incongruent, Animacy-Mismatch condition but not for the Incongruent, Animacy-Match condition. As that the amplitude of P600 has been shown to be sensitive to plausibility or acceptability (van de Meerendonk, Kolk, Vissers, & Chwilla, 2010), and that the Animacy-Mismatch condition was indeed less acceptable than the congruent condition, one might question whether the P600 reflects just the unacceptability of the sentences in the Animacy-Mismatch condition. This explanation, however, is highly unlikely because compared with the congruent condition, the Incongruent, Animacy-Match condition was also highly less acceptable, but no P600 response was observed for this condition.

The P600 may be partly related to the acceptability judgment task used and may reflect a continued combinatorial analysis driven by a conflict between the output of the computation of phrase structure based on syntactic category information (noun and classifier) and the output of the animacy processing that does not permit such a combination of noun and classifier (for a summary and discussion of the triggers of P600, see Bornkessel-Schlesewsky & Schlesewsky, 2008; Kuperberg, 2007). One might question whether the P600 reflects animacy violations per se. This possibility, however, is unlikely because no P600 effects were observed for animacy violations in a study that used a probe (word) verification instead of a sentence acceptability task (Li et al., 2006). We also observed a late (600–1000 ms) anterior negativity for the two incongruent conditions, which may reflect secondary semantic integration (for similar negativities, see Friederici et al., 1999; Zhang et al., 2010; Zhou et al., 2010).

Leaving aside the Animacy-Mismatch condition that was not included in the Zhou et al. (2010) study, the ERP (N400 and late anterior negativity) effects we observed among the participants who accepted the non-canonical structure were completely the same as those observed in the Zhou et al. study, in which the sentences had a canonical structure (SVO), with classifiers occurring before object nouns (the critical words). As that an N400 response is believed to reflect meaning processing (see Kutas and Federmeier (2011), for a recent review), the N400 effects observed both in the present study and in the Zhou et al. study provide evidence supporting that Chinese noun classifiers are semantic elements, imposing selectional restrictions on the scope of the noun, as suggested by some linguists (e.g., Wu & Bodomo, 2009).

For the participants who did not accept the non-canonical structure, we observed an N400 effect at the numeral classifiers for the Incongruent, Animacy-Mismatch condition but not for the Incongruent, Animacy-Match condition, although there was significant difference in the off-line judgment of overall (syntactic and semantic) acceptability of sentences between the latter condition and the Congruent condition. The discrepancy between the behavioral and ERP findings for the Incongruent, Animacy-Match condition may reflect the possibility that whereas the integration of more specific or concrete semantic information between the object nouns and classifiers did not occur when the numeral classifiers were read, resulting in the absence of N400 effects, it did occur later when the whole sentence had been read, as indicated by the higher percentage of "unacceptable" judgments for this condition compared to the Congruent condition. Such an absence of an immediate semantic integration may be either due to the possibility that the limited resource was completely occupied by the difficult word order processing, or due to a difficult or even failed word order processing that might temporally block the integration of specific semantic information (but see Yu & Zhang, 2008; Zhang et al., 2010 for evidence for the occurrence of semantic integration in the face of unsuccessful syntactic processing). Either way, when the word order cue appears not work well, such as in the case that the non-canonical structure is not acceptable, animacy cue becomes prominent, resulting in the N400 effects reflecting the integration of animacy information between the object nouns and their classifiers.

In sum, our results suggest that whereas more specific or concrete semantic information, such as functional semantic features of artifact nouns like *qiche* ('car') or *taideng* ('desk lamp'), is used during semantic integration of nouns and their classifiers, animacy, as a broader dimension of semantic information, is not used in this processing stage but is used in a later analysis reflected by P600, unless the word order is considered unacceptable and thus may not be an effective cue for sentence interpretation.

The time course of the use of animacy information in processing classifier–noun combinations in the case that the non-canonical structure is acceptable, as revealed in the present study, appears to significantly differ from that in the processing of thematic role or thematic relationships revealed by previous studies. As we mentioned at the outset, a previous study (Li et al., 2006) has demonstrated that in the case that there is already a mismatch in specific semantic features between Chinese verbs and their objects, an additional mismatch either in humanness only or in both humanness and animacy can result in a further increase of N400 amplitude, suggesting a rapid use of broader semantic information like humanness during semantic integration of verbs with their objects. In addition, there have been some ERP studies demonstrating the rapid use of animacy information in thematic role assignments in the N400 time window (e.g., Philipp et al., 2008; Roehm et al., 2004).

What, then, are we to make of the difference in the presence/absence of an N400 for animacy mismatch between the processing of classifier–noun combinations and the processing of thematic role or thematic relationships? No thematic processing being involved in classifier–noun combinations might be one property that is relevant. As we noted at the outset, for thematic processing, there is a typical connection between the animacy of a noun and the thematic role that the noun plays, such as agent/actor and patient/undergoer. Such a connection may be crucial for the immediate use of animacy information, resulting in an N400, for thematic processing (determining "who does what to whom"). For the processing of classifier–noun combinations, however, no such connection exists, which may be the reason why animacy information appears not to be used very rapidly, for example, in the N400 time window.

If the explanation above is correct, there should be no N400 effects to be evoked by animacy mismatch that occurs in other types of combinations not involving thematic processing, such as adjective–noun combinations. We look to future research to assess this admittedly speculative hypothesis. In addition, it is not clear whether the results of the present study can be generalized to sentences with canonical word order. Further studies are clearly needed, probably using other classifier languages.

Finally, in the present study, animacy level was manipulated between animal and artifact domains. Interestingly, no N400 effects were observed for this manipulation, although evidence from brain-damaged patients with category-specific semantic deficits suggests a clear dissociation between animal and artifact domains (for reviews, see Capitani et al., 2003; Caramazza & Mahon, 2003; Mahon & Caramazza, 2009). However, it should be noted that compared with animal, human is higher on the animacy hierarchy, although both are animate. Future studies are necessary to determine whether the same results can be observed when animacy level is manipulated in other ways, for example, between human and artifact categories.

What can be concluded from the present study is that the animacy information encoded in animal and artifact nouns and their classifiers is normally not used very rapidly, say, in the N400 time window, during the computation of classifier–noun combinations that do not involve thematic processing, at least in non-canonical sentences. Instead, the animacy information is merely used at the later stage of classifier–noun combinations, resulting in both the P600 effects in the ERP experiment and the difference in the rating scores between the two incongruent conditions in the survey of meaning congruence of noun–numeral classifier pairs. Thus, the temporal neural dynamics of animacy processing in sentence comprehension may be modulated by whether or not animacy is relevant to thematic interpretation.

The conclusion above also has important implications for understanding/limiting the extent to which semantic cues like animacy are prominent for Chinese sentence comprehension, especially when considering that in Chinese, animacy has been shown to be a relatively strong cue to sentence comprehension, such as the identification of the agent of a sentence (Li et al., 1993), and sentence comprehension has been assumed to rely on semantic analyses to a relatively large degree (e.g., Xu, 1997; see Li, 1998; Yu & Zhang, 2008; Zhang et al., 2010).

Acknowledgments

This research was supported by grants from the National Natural Science Foundation of China (31070898) and the Natural Science Foundation of Beijing (7112069) to Yaxu Zhang. We thank two anonymous reviewers for their extremely helpful comments on an earlier version of this article.

References

Angrillia, A., Dobel, C., Rockstroh, B., Stegagno, L., & Elbert, T. (2000). EEG brain mapping of phonological and semantic tasks in Italian and German languages. *Clinical Neurophysiology, 111*, 706–716.

Aoshima, S., Phillips, C., & Weinberg, A. (2004). Processing filer-gap dependencies in a head-final language. *Journal of Memory and Language, 51*, 23–54.

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

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2009). The role of prominence information in the real-time comprehension of transitive constructions: A cross-linguistic approach. *Language and Linguistics Compass, 3*, 19–58.

Capitani, E., Laiacona, M., Mahon, B. Z., & Caramazza, A. (2003). What are the facts of category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology, 20*, 213–261.

Caramazza, A., & Mahon, B. Z. (2003). The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences, 7*, 354–361.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2*, 913–919.

Cheng, L., & Sybesma, R. (1999). Bare and not so bare nouns and the structure of NP. *Linguistic Inquiry, 30*, 509–542.

Cheng, L., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. In G. Cinque & R. Kayne (Eds.), *The oxford handbook of comparative syntax* (pp. 259–292). Oxford: Oxford University Press.

Clifton, C., Jr., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 47*, 571–588.

Dickey, M. W., & Thompson, C. K. (2004). The resolution and recovery of filler-gap dependencies in aphasia: Evidence from on-line anomaly detection. *Brain and Language, 88*, 108–127.

Federmeier, K. D., & Kutas, M. (1999a). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469–495.

Federmeier, K. D., & Kutas, M. (1999b). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research, 8*, 373–392.

Felser, C., Clahsen, H., & Münte, T. F. (2003). Storage and integration in the processing of filler-gap dependencies: An ERP study of topicalization and wh-movement in German. *Brain and Language, 87*, 345–354.

Ferreira, F., & Clifton, C. Jr., (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language, 47*, 250–272.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel-Schlesewsky, I., & von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex, 16*, 1709–1717.

Friederici, A. D., & Frisch, S. (2000). Verb argument structure processing: The role of verb-specific and argument-specific information. *Journal of Memory and Language, 43*, 476–507.

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

Frisch, S., Hahne, A., & Friederici, A. D. (2004). Word category and verb-argument structure information in the dynamics of parsing. *Cognition, 91*, 191–219.

Frisch, S., & Schlesewsky, M. (2001). The N400 indicates problems of thematic hierarchizing. *NeuroReport, 12*, 3391–3394.

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes, 24*, 1124–1179.

Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*, 59–73.

Hsieh, Y., Boland, J. E., Zhang, Y., & Yan, M. (2009). Limited syntactic parallelism in Chinese ambiguity resolution. *Language and Cognitive Processes, 24*, 1227–1264.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205–225.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenge to syntax. *Brain Research, 1146*, 23–49.

Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D. N., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentence: Evidence from event-related potentials. *Brain and Language, 100*, 223–237.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*, 117–129.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science, 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). *Annual Review of Psychology, 62*, 621–647.

Ledoux, K., Gordon, P., Camblin, C., & Swaab, T. (2007). Coreference and lexical repetition: Mechanisms of discourse integration. *Memory & Cognition, 35*, 801–815.

Li, P. (1998). Crosslinguistic variation and sentence processing: The case of Chinese. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective* (pp. 33–51). San Diego: Academic Press.

Li, P., Bates, E., & MacWhinney, B. (1993). Processing a language without inflections: A reaction time study of sentence interpretation in Chinese. *Journal of Memory and Language, 32*, 169–192.

Li, X., Shu, H., Liu, Y., & Li, P. (2006). Mental representation of verb meaning: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience, 18*, 1774–1787.

Li, C., & Thompson, S. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences, 4*, 432–440.

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology, 60*, 27–51.

Mak, W. M., Vonk, W., & Schriefers, H. (2002). The influence of animacy on relative clause processing. *Journal of Memory and Language, 47*, 50–68.

Mak, W. M., Vonk, W., & Schriefers, H. (2006). Animacy in processing relative clauses: The hikers that rocks crush. *Journal of Memory and Language, 54*, 466–490.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58*, 25–45.

Muromatsu, K. (1998). *On the syntax of classifiers*. Unpublished doctoral dissertation, University of Maryland.

Philipp, M., Bornkessel-Schlesewsky, I., Bisang, W., & Schlesewsky, M. (2008). The role of animacy in the real time comprehension of Mandarin Chinese: Evidence from auditory event-related brain potentials. *Brain and Language, 105*, 112–133.

Phillips, C., Kazanina, N., & Abada, S. H. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*, 407–428.

Roehm, D., Schlesewsky, M., Bornkessel-Schlesewsky, I., Frisch, S., & Haider, H. (2004). Fractionating language comprehension via frequency characteristics of the human EEG. *NeuroReport, 15*, 409–412.

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General, 136*, 485–501.

Schriefers, H., Friederici, A., & Kühn, K. (1995). The processing of locally ambiguous relative clauses in German. *Journal of Memory and Language, 34*, 499–520.

Szewczyk, J. M., & Schriefers, H. (2011). Is animacy special? ERP correlates of semantic violations and animacy violations in sentence processing. *Brain Research, 1368*, 208–221.

Traxler, M., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*, 69–90.

Traxler, M., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language, 53*, 204–224.

Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33*, 285–318.

van de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. J. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience, 22*, 67–82.

Wu, Y., & Bodomo, A. (2009). Classifiers ≠ determiners. *Linguistic Inquiry, 40*, 487–503.

Xu, T. (1997). *Yuyan lun (On Language)*. Changchun, China: Northeast Normal University Press.

Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *NeuroReport, 19*, 745–749.

Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics, 16*, 43–59.

Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 765–781.

Zhou, X., Jiang, X., Ye, Z., Zhang, Y., Lou, K., & Zhan, W. (2010). Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An ERP study. *Neuropsychologia, 48*, 1551–1562.