Cognition 166 (2017) 433–446

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Predicting semantic features in Chinese: Evidence from ERPs

Nayoung Kwon [a,*], Patrick Sturt [b], Pan Liu [c]

[a] Department of English, Konkuk University, Republic of Korea
[b] Department of Psychology, University of Edinburgh, UK
[c] Department of Psychology, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 22 June 2016
Revised 13 March 2017
Accepted 8 June 2017

Keywords:
Prediction
Classifiers
Chinese
ERPs
N400
Sentence processing

ABSTRACT

This article reports two ERP studies that exploited the classifier system of Mandarin Chinese to investigate semantic prediction. In Mandarin, in certain contexts, a noun has to be preceded by a classifier, which has to match the noun in semantically-defined features. In both experiments, an N400 effect was elicited in response to a classifier that mismatched an up-coming predictable noun, relative to a matching classifier. Among the mismatching classifiers, the N400 effect was graded, being smaller for classifiers that were semantically related to the predicted word, relative to classifiers that were semantically unrelated to the predicted word. Given that the classifier occurred before the predicted word, this result shows that fine-grained semantic features of nouns can be pre-activated in advance of bottom-up input. The studies thus extend previous findings based on a more restricted range of highly grammaticalized features such as gender or animacy in Indo-European languages (Szewczyk & Schriefers, 2013; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003).
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Prediction or anticipation refers to a mental process that generates information about future states based on what we know already. As one of the most fundamental principles of human cognition (Clark, 2013), prediction operates in various cognitive domains, such as visual processing, motor control and theory of mind (Friston & Stephan, 2007; Frith & Frith, 2006; Mehta & Schaal, 2002; Wolpert & Flanagan, 2001), and its importance has also been emphasized in studies of language processing (DeLong, Urbach, & Kutas, 2005; Federmeier, 2007; Pickering & Garrod, 2007; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). However, although there has been some compelling evidence for semantic prediction, most of the relevant studies were conducted based on unilateral semantic distinctions such as gender or animacy and in Indo-European languages, in which semantic features such as gender and/or animacy are highly grammaticalized, and correlate with overt morpho-syntactic markers. To evaluate the generality of semantic prediction, in this study we use the classifier system of Mandarin Chinese, in which agreement between a classifier and a noun is based on perceived similarity that is functionally or perceptually defined, representing various semantic relations.

1.1. Prediction in language processing

Language processing has been argued to be highly incremental and predictive. Previous studies have shown that language comprehenders do not delay the processing of incoming words until the end of a sentence, despite temporary structural ambiguity rampant in natural human language (Frazier & K. Rayner, 1988). Instead, the incoming words are parsed and interpreted immediately as they are perceived. This incremental nature of language processing is evidenced by garden-path effects, in which processing difficulty is elicited at the point of structural disambiguation when the initially built structure turns out to be incorrect and thus requires structural revision (Rayner & Frazier, 1987; Frazier & Rayner, 1982). On the other hand, the predictive nature of language processing has been mainly discussed in association with context effects. That is, it is argued that contextual information facilitates processing of a target item because that item has been predicted ahead of time based on the preceding context, as evidenced by shorter reaction times in a simple reading or lexical judgment task (Ehrlich & Rayner, 1981; Fischler & Bloom, 1979; Jordan & Thomas, 2002; Hess, Foss, & Carroll, 1995; Wright & Garrett, 1984) or by predictive eye-gaze patterns in the visual-world paradigm. For example, using the visual-world paradigm, Kamide, Altmann, and

* Corresponding author at: Department of English, School of Humanities, Konkuk University, Neungdong-ro 120, Gwangjin-gu, Seoul 143-701 Republic of Korea.
E-mail address: nayoung.kw@gmail.com (N. Kwon).

http://dx.doi.org/10.1016/j.cognition.2017.06.010
0010-0277/© 2017 Elsevier B.V. All rights reserved.

Haywood (2003) showed that sentence comprehenders moved their eyes towards a picture of an object that was appropriate given the context—for example, there were more fixations on a picture of a motorbike than a carousel given the context *The man will ride*, but the reversed pattern was observed given the context *The girl will ride*. Given that these effects were already significant at the verb position even before the participants perceived the object in the speech stream, Kamide at al. took these results to suggest that sentence comprehenders predict the semantic contents of an upcoming word based on the combinatory semantic constraints of an agent (*the man*) and a verb (*ride*) (for similar results, see also Kamide, Scheepers, & Altmann, 2003; cf. Kaiser & Trueswell, 2004; Weber, Grice, & Crocker, 2006).

Recent studies have suggested that ERPs also provide efficient means to examine the effects of prediction during on-line sentence processing (DeLong et al., 2005; Van Petten & Luka, 2012; Lau, Stroud, Plesch, & Phillips, 2006; and references therein). In particular, the N400 component has proven to be useful in the investigation of prediction effects in language processing. The N400 is a negative-going wave that peaks about 400 ms post-stimulus, with a centro-parietal maximum. It was first observed in response to semantic anomalies such as *He spread the warm bread with* <u>socks</u> (Kutas & Hillyard, 1980) and has been since shown to be sensitive to the processing of semantic information regardless of semantic anomaly (see Kutas & Federmeier, 2011 for a review). For example, the amplitude of N400 varies depending on an item's predictability in a given context (cf. Chow et al. 2014). Thus, N400 amplitude increases in response to a less probable sentential continuation (e.g. *dog* for *Don't touch the wet* ...) relative to the expected one (e.g. *paint*) (Kutas & Hillyard, 1984). In addition, the amplitude of the N400 also varies as a function of frequency of an item or as a function of semantic relations independent of the local context. Thus, a high frequency word elicits a smaller N400 compared with a low frequency a word, when they are presented in isolation (Rugg, 1990). Similarly, a high typicality category member (e.g. robin as a type of bird) elicits a smaller N400 compared to a low typicality category member (e.g. turkey as a type of bird), which in turn elicits a smaller N400 compared to an unrelated item (Kutas & Federmeier, 2000).

More relevant to the goal of this paper is the study of Federmeier and Kutas (1999). In this study, the authors compared ERP responses to two types of semantic category violation: 'within-category violation', with an unexpected item from the same semantic category as the predicted item, and 'between-category violation' with an equally unexpected item from a different semantic category. For example, when preceded by a lead-in sentence like *They wanted to make the hotel look more like a tropical resort*, a second target sentence (*So along the driveway, they planted rows of* ...) ended either with (i) the expected item (e.g. *palms*), (ii) a within-category violation (e.g. *pines*), or (iii) a between-category violation (e.g., *tulips*). The results showed that both the two unexpected endings elicited larger N400s than the predicted ending. Importantly, however, the between-category violation elicited still larger N400s than the within-category violation, despite their equally low cloze probability (see also Federmeier, McLennan, De Ochoa, & Kutas, 2002; Kutas & Hillyard, 1984; Kutas, 1993; Thornhill & Van Petten, 2012 for similar results). These results suggest that the neural representation of a word is based on a set of semantic features, and that an overlap in semantic features facilitates the processing of an unexpected item in on-line language processing. In the context of the current study, Federmeier and Kutas' (1999) results are compatible with the claim that some semantic content was predicted ahead of time. However, the results are also compatible with the integration hypothesis, namely that within-category violations were easier to integrate with the preceding context due to more overlap in semantic features than between-category violations (cf. Wlotko & Federmeier, 2015).

On the other hand, DeLong et al. (2005) provides stronger evidence for prediction effects. Using the phonological regularity of the English indefinite articles *a* and *an*, DeLong et al. examined ERP responses to sentence contexts that led to anticipation of specific words that started either with a consonant or a vowel. For example, *The day was breezy so the boy went outside to fly* ... was followed by *a kite*, a continuation that is highly expected, or *an airplane*, a continuation that is plausible but less likely. The study replicated the well-known correlation between N400 amplitude and cloze probability of target nouns (Kutas & Hillyard, 1984) with bigger N400 responses to *airplane* than to *kite*. More importantly, the strong correlation was also evident at the preceding article, with bigger N400 responses to *an* (an article preceding an unexpected noun *airplane*) compared to *a* (an article preceding an expected noun *kite*). Given that phonological variations of the two indefinite articles are based solely on the word forms of the predicted target nouns, the results at the article strongly suggest that sentence comprehenders predict specific words or specific phonological forms of words (for further evidence of form-based expectations, see DeLong, Groppe, Urbach, & Kutas, 2012; Ito, Corley, Pickering, Martin, & Nieuwland, 2016; Kim & Lai, 2012; Laszlo & Federmeier, 2009; Martin et al., 2013).

Similar results were also obtained from studies of Spanish, in which an article should agree with a noun in gender (e.g. *una canasta* 'a_$_{\text{feminine}}$ basket_$_{\text{feminine}}$'). Capitalizing on this grammatical gender system, Wicha, Bates, Moreno, and Kutas (2003) (see also Wicha, Moreno, & Kutas, 2003) showed that an unexpected article (e.g. *un* 'a_$_{\text{masculine}}$') elicited broadly distributed N400-like negativity compared to an expected article (e.g. *una* 'a_$_{\text{feminine}}$'). As the gender of an article is determined by the following noun (e.g. *canasta* 'basket_$_{\text{feminine}}$'), which is, in turn, predicted based on the prior sentence context (e.g. 'Red riding Hood carried the food for her grandmother in ~'), the authors took these N400 effects as evidence that readers predict gender-specific nouns (see also Szewczyk & Schriefers, 2013 for similar experimental results based on animacy manipulations in Polish).

A study in Dutch with a similar experimental design further supported the relevance of prediction in language processing, but the prediction effect in this study was observed in the form of positivity (Van Berkum et al., 2005). Similarly to Spanish, Dutch also has a grammatical gender system in which the gender of a preceding adjective has to agree with that of a noun, in indefinite noun phrases. Van Berkum et al. (2005) found that after a highly constraining discourse context (e.g. 'The burglar had no trouble locating the secret family safe. Of course, it was situated behind a...'), adjectives (e.g. *grote* 'big_$_{\text{common}}$') whose gender mismatched with that of the predicted noun (e.g. *schilderij* 'painting_$_{\text{neutral}}$') elicited a larger fronto-central positivity compared with their gender-matched counterparts (e.g. *groot* 'big_$_{\text{neutral}}$') (see also Wicha, Moreno, & Kutas, 2004 for positive-going effects with gender manipulations in Spanish).

Van Petten and Luka (2012) broadly characterized the frontal positivity as indexing costs of failed prediction, while the N400 effect as related to "the other side of coin", an index of processing facilitation due to successful prediction. However, it is not clear at this point why highly analogous studies have elicited different ERP responses. That is, monophasic N400 and frontal positivity effects have each been independently elicited in studies of a similar design (negativity: Otten & Van Berkum, 2009; Wicha, Bates, et al., 2003; Wicha, Moreno, et al., 2003; positivity: Van Berkum et al., 2005; Wicha et al., 2004) while other studies have shown these two components simultaneously, in a bi-phasic effect (Delong, Urbach, Groppe, & Kutas, 2011; Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Kutas, 1993; Otten, Nieuwland, & Van Berkum, 2007; Thornhill & Van Petten, 2012). Thus, it has been difficult to pinpoint precise functional differences between the N400 and frontal positivity effects. Thornhill and Van Petten (2012) addressed this question, manipulating semantic relatedness as well as predictability. The results showed that while unexpected words (e.g., *mind* & *reputation*) in a given sentential context (e.g., *He was afraid that doing drugs would damage his* ...) elicited overall a larger N400 compared to an expected word (e.g., *brain*), the N400 effect was smaller in response to a semantically related word (e.g., *mind*) than to a semantically unrelated word (e.g., *reputation*). In contrast, frontal positivities elicited to unexpected words (e.g., *mind* & *reputation*) did not differ as a function of semantic relatedness. Based on the results, the authors argued that the N400 is related to conceptual expectations while the frontal positivity is related to specific lexical expectations.

In summary, previous studies have suggested that language processing can be predictive, at least under some circumstances. With highly constraining contextual information, studies have suggested that language comprehenders can make predictions that are fine-grained enough to include information about the phonological form, grammatical gender and/or animacy of an upcoming word. However, although evidence for semantic prediction has been compelling, all of the relevant studies have been conducted in Indo-European languages. Moreover, when prediction effects have been examined in a word position preceding the predicted word, the effects were always based on one-dimensional semantic features such as gender or animacy, which are highly grammaticalized and correlate with overt morpho-syntactic markers in the languages examined (Greenberg, 1978; Corbett, 1991). Given this, it is important to evaluate the generality of these findings in a typologically different language, using semantic features that are not limited to potentially salient grammaticalized features. To address this question, in this study we use the classifier system of the Chinese language, an agreement system based on various semantic classifications.

1.2. Classifiers in Chinese

In Chinese, a noun must be preceded by a *classifier* when the noun is modified by a numeral (e.g., *one, two*), a demonstrative (e.g., *this, that*) or a quantifier (e.g., *whole, a few*) (Li & Thompson, 1989). In addition, a classifier and its associated noun have to agree in semantic features such as animacy, shape and size (Allan, 1977; Tai, 1994). Accordingly, nouns grouped by a classifier reflect a linguistic categorization of perceived similarities, either physical or functional. For example in (1), the classifier张is used for纸 'paper', as 张 is the classifier for flat objects like paper and sheets, and in (2) the classifier只is used for小鸟 'bird', as 只 is the classifier for birds and some other animals.

(1)	三 three	张 classifier	纸 paper
	'three sheets of paper'		
(2)	七 seven	只 classifier	小鸟 bird
	'seven birds'		

However, the semantic classification represented by a classifier might not represent a homogeneous single group. For example, the classifier口is used for *people* and *pigs* as these form a semantic category relating to family members or domestic animals. The same classifier is also used for *language, tune* and *voice* as these form a semantic category relating to verbal behaviour. It is also used for *pots, bells, suitcases, coffins, ponds,* and *knives* as these form a semantic category of objects with an opening or edge (Song, 2009). In addition, perceived similarities among a group of nouns could be quite abstract or even seem arbitrary. For example, *fish*, *trousers*, *river*, *snake* and *road* all share the classifier条, as they are perceived to be "long and flexible". Likewise, *cloud* shares a classifier 朵 with *flower*, as *clouds* are perceived to be "flower-like". The classifier-noun agreement, however, is not simply based on a one-to-many relationship. Instead, a noun can take different classifiers as well, depending on how it was perceived. For example, *person* can take a general classifier 个 in a neutral context, and a classifier 位 when honorified, as well as 群 when referred to as a group. Similarly, *grape* can take a classifier 粒 when perceived as being grain-like, or a classifier 顆 when highlighted as being small and round. In addition, these characters used as classifiers can have completely different meanings, when they are not used as classifiers. For example, 张is a surname common in China, and 只 means 'only, merely, simply' when it is used alone or with another character, as in 只是.

However, a close relation between a classifier and its associated nouns has been attested in various experiments. For example, more fixations were made on classifier-match objects than on distractor objects when a classifier was present, suggesting that a classifier affects conceptual processing (Huettig, Chen, Bowerman, & Majid, 2010). In addition, noun-classifier mismatches elicit an N400. For example, Zhang, Zhang, and Min (2012) examined the ERP responses to a non-canonical construction in Chinese where an object NP precedes its agreeing classifier (i.e., object noun + subject noun + verb + classifier + adjective). Putting irrelevant details aside, the semantic relation between the fronted object and classifier was manipulated to be either (a) matching (e.g., *car* ...*classifier for vehicle*...) or (b) mismatching (e.g., *lamp/seal* ...*classifier for vehicle*...). The results showed that the group of participants who accepted the non-canonical target construction as grammatical elicited an N400 effect to the classifier mismatch condition in comparison to the match condition. The N400 effect to a classifier-noun mismatch has also been attested in Chinese when target words were presented using a canonical word order (Chou, Huang, Lee, & Lee, 2014; Jiang & Zhou, 2012; Qian & Garnsey, 2016; Zhou et al., 2010), as well as when a classifier and its associated noun were separated by an intervening linguistic element and thus formed a long-distance dependency (Hsu, Tsai, Yang, & Chen, 2014). These results suggest that semantic feature agreement between a classifier and its associated noun is obligatory in Chinese, and a mismatch results in increased processing difficulty.

In this study, we took advantage of classifier-noun semantic agreement to investigate cognitive mechanisms underlying the anticipation of semantic features of upcoming words. In particular, we used a pre-nominal relative clause as a constraining context and varied head nouns and their associated classifiers, as in (3).

(3) Expected:	张艺谋执导的	这	部	电影	提名了。
Related:	张艺谋执导的	这	场	演出	提名了。
Unrelated:	张艺谋执导的	这	座	大厦	提名了。
	Zhang Yimou directed	this	classifier	movie/ performance/ building	was nominated
	'This movie/performance/building that Zhang Yimou directed was nominated.'				

The relative clause (which always precedes the modified noun in Chinese) was always semantically highly constraining (e.g. ∼*which Zhang Yimou directed*) such that a specific head noun would be predicted to occur, as in the expected condition (e.g. *movie*). On the other hand, semantic features of the related and the unrelated conditions were manipulated such that the head nouns of the related condition (e.g. *performance*) were thematically or functionally related to those of the expected condition, while the head nouns of the unrelated condition (e.g. *building*) had little semantic overlap with the predicted head noun. In all of these three conditions, classifiers agreed with their associated head nouns, but neither the related condition nor the unrelated condition included classifiers which could be also compatible with the head nouns of the expected condition. Overall, the cloze probabilities of the classifiers and head nouns were 0.68 and 0.85 respectively in the expected condition, and all 0 in the related and unrelated conditions (see Experiment section for details).

Given the well-attested correlation between N400 amplitude and cloze probability (Kutas & Hillyard, 1984), we predict an N400 effect at the head noun of the related and unrelated conditions. Importantly, we predict that a comparable effect will already be visible at the preceding classifier position. All experimental sentences included a numeral, quantifier or demonstrative. Thus, as discussed above, they required the presence of a classifier, meaning that the following character had to be interpreted as a classifier and the classifier should semantically agree with its associated noun. However, the classifiers of the unrelated and the related conditions do not agree with the semantic features of the predicted head noun. Thus, if the semantic features of an upcoming word are activated ahead of time as argued by Kamide et al. (2003), a N400 effect will be observed as early as the classifier position for these two conditions. In addition, we predict that the N400 amplitude at the classifier position will be smaller in the related condition compared with the unrelated condition, despite their equally low cloze probability. This is because the related condition has more semantic overlap with the predicted item, and thus its processing is expected to be facilitated (cf. Federmeier & Kutas, 1999). Finally, the prediction effect may be also elicited in the form of a frontal positivity, as the effect has been claimed to be related to specific lexical expectations (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). However, it is not necessarily the case that any unexpected word elicits a frontal positivity. Instead, some researchers have found that unexpected but plausible continuations elicit the positivity, while incongruent continuations do not. This suggests that the frontal positivity is sensitive to both predictability and plausibility (Federmeier, Kutas, & Schul, 2010; DeLong, Quante, & Kutas, 2014). The current study did not overtly manipulate the plausibility of experimental sentences as it was not directly relevant for the main goal of the study. However, as we manipulated semantic distance of the target sentences, the events denoted by the unrelated sentences were less likely to occur than the events denoted by the related conditions, which are in turn less likely to occur than those of the expected conditions.[1] Given this, we predict that the frontal positivity will be more visible in response to the related condition than to the unrelated condition, even though both of these conditions involve an unfulfilled lexical expectation.

In summary, we predict that an N400 effect will be elicited to the related and the unrelated conditions, that the effect will be visible already at the classifier position before the predicted head noun occurs, and that the N400 amplitude will be smaller in the related condition compared with the unrelated condition. We also predict that a frontal positivity will be elicited to the related condition and that the effect will be visible at the predicted head noun position. If these predictions are borne out, they will provide solid evidence that semantic features of an upcoming word are predicted ahead of time.

2. Experiment 1

2.1. Methods

2.1.1. Materials

To create sentences with highly constraining sentential contexts, we first ran a norming study. 144 Mandarin Chinese sentence fragments were constructed, which were truncated after the relative clause marker *de*, followed by a numeral or a demonstrative, as shown in (4). Thirty native speakers of Mandarin at Dalian University of Technology in China received 18 RMB and participated in a questionnaire study. They were asked to complete the sentences however they saw fit. After obtaining head nouns and classifiers for all the sentences, we chose 102 sentences whose cloze probability for a given head noun was higher than or equal to 0.5. The mean cloze probability was 0.68 for classifiers and 0.85 for head nouns respectively. Classifiers had lower cloze probability than head nouns as sometimes different classifiers were used for the same head nouns.

(4)	张艺谋执导	的	这______________________。
	Zhang Yimou directed	DE	this
	'This _____ which Zhang Yimou directed ____________________. '		

Three experimental conditions were constructed based on these 102 sentences, as shown in (3).

We manipulated the target sentences such that the head nouns of the related condition were thematically or functionally related to those of the expected condition, while the head nouns of the unrelated condition had little semantic overlap with the predicted head nouns of the expected condition (see Table 1 for further example sentences). Importantly, classifiers were exclusively compatible with their associated head nouns; the head nouns of the related and the unrelated conditions were not compatible with the classifiers of the expected condition, and likewise, classifiers of the related and the unrelated conditions were not compatible with the head nouns of the expected condition. Overall, 60 distinct classifiers were used across the 102 sentences, with many of the classifiers reused several times. To evaluate whether a given classifier was equally likely to occur in the three experimental conditions, we first counted the number of times that a particular classifier was used for each condition and divided this number by the total number of times that it was used in the experiment. The probability that a classifier occurred in the three conditions, however, did not differ, suggesting that the use of specific classifiers did not trigger certain interpretational biases (Expected:

[1] While some related sentences are quite plausible (e.g., "draft which is in a mailbox∼"), some are less likely to occur (e.g., "cows which are swimming in a pond∼"). Likewise, some unrelated sentences are rather plausible (e.g., "a drop of blood which is in a mailbox∼") while some are quite anomalous (e.g., "flowers that are swimming in a pond").

Table 1
Sample experimental sentences.

Expected	樵夫砍伐的	这	棵	树	非常庞大
Related	樵夫砍伐的	这	束	花	非常庞大
Unrelated	樵夫砍伐的	这	本	书	非常庞大
Gloss	woodcutter.chop	this	classifier	tree/flower/book	extremely.huge
Trans.	'This tree/flower/book that the woodcutter cut is extremely huge.'				
Question	'Did the woodcutter cut something small?'				
Expected	点缀在夜空的	几	颗	星星	在闪烁
Related	点缀在夜空的	几	朵	云	在闪烁
Unrelated	点缀在夜空的	几	本	书	在闪烁
Gloss	adorn.night.sky	few	classifier	star/cloud/novel	flick
Trans.	'The few stars/cloud/novels that adorn the night sky are flickering'				
Question	'Is there something flickering on the night sky?'				
Expected	邮箱里的	那	封	信	是从国外
Related	邮箱里的	那	份	稿	是从国外
Unrelated	邮箱里的	那	滴	血	是从国外
Gloss	is.in.the.mailbox	that	classifier	letter/draft/blood	is.from.overseas
Trans.	'That letter/draft/blood which is in the mailbox is from overseas.'				
Question	'Is the mail box empty?'				
Expected	莎士比亚写的	这	篇	文章	是经典之作
Related	莎士比亚写的	这	封	信	是经典之作
Unrelated	莎士比亚写的	这	串	珍珠	是经典之作
Gloss	Shakespeare.wrote	this	classifier	writing/letter/pearl	is.master.piece
Trans.	'The writing/letter/pearl that was written by Shakespeare is a master piece.'				
Question	'Does Shakespeare have a master piece?'				
Expected	在树上歌唱的	几	只	小鸟	真是有趣
Related	在树上歌唱的	几	位	老师	真是有趣
Unrelated	在树上歌唱的	几	辆	车	真是有趣
Gloss	singing.on.tree	few	classifier	bird/teacher/car	is.very.interesting.
Trans.	'The few birds/teachers/cars singing on the tree are very interesting'				
Question	'Is the singing coming from grass?'				
Expected	在池塘 里游水的	几	条	鱼	得病了
Related	在池塘 里游水的	几	头	牛	得病了
Unrelated	在池塘 里游水的	几	束	花	得病了
Gloss	swimming.in.pond	few	classifier	fish/cow/flower	is.ill
Trans.	'The few fish/cows/flowers swimming in the pond are ill.'				
Question	'Is the thing in the pond sick?'				
Expected	用来上网的	这	台	电脑	是新款式
Related	用来上网的	这	根	电缆	是新款式
Unrelated	用来上网的	这	顿	饭	是新款式
Gloss	used.for.web.surfing	this	classifier	computer/cable/meal	is.new.model
Trans.	'This computer/coil cable/meal used for web surfing is a new model'				
Question	'Is something used for web surfing?'				
Expected	停泊在港口的	这	艘	船	体型庞大
Related	停泊在港口的	这	辆	车	体型庞大
Unrelated	停泊在港口的	这	头	牛	体型庞大
Gloss	docked.at.harbour	this	classifier	ship/car/cow	size.huge
Trans.	'This ship/car/cow docked at the harbour is huge.'				
Question	'Is there something docked at the harbour?'				
Expected	飞往印尼的	那	班	航班	出了意外
Related	飞往印尼的	那	列	火车	出了意外
Unrelated	飞往印尼的	那	根	草	出了意外
Gloss	fly.to.Indonesia	that	classifier	flight/train/grass	happen.accident
Trans.	'That flight/train/grass which flew to Indonesia got into an accident.'				
Question	'Did an accident happen?'				
Expected	收藏在相簿里的	那	张	照片	带来许多回忆
Related	收藏在相簿里的	那	卷	录像带	带来许多回忆
Unrelated	收藏在相簿里的	那	台	电脑	带来许多回忆
Gloss	kept.in.photo.album	that	classifier	photo/video/computer	bring.many.memory
Trans.	'That photo/video/computer kept in the photo album brings many memories'				
Question	'Did something bring many memories?'				
Expected	我从图书馆借了	两	本	书	很重
Related	我从图书馆借了	两	份	报纸	很重
Unrelated	我从图书馆借了	两	座	大厦	很重
Gloss	I.borrow.from.library	two	classifier	book/newspaper/building	very.heavy
Trans.	'Two books/newspapers/buildings that I borrowed from a library are very heavy'				
Question	'Did I borrow something from a library?'				
Expected	在草原上奔驰的	那	匹	马	很健壮
Related	在草原上奔驰的	那	头	牛	很健壮
Unrelated	在草原上奔驰的	那	张	床	很健壮
Gloss	grass.gallop	that	classifier	horse/cow/bed	very.strong
Trans.	'That horse/cow/bed that is galloping on the grassland is very strong'				
Question	'Is the grassland empty?'				
Expected	提供鲜奶的	五	头	牛	都很强壮
Related	提供鲜奶的	五	匹	马	都很强壮
Unrelated	提供鲜奶的	五	张	桌子	都很强壮

(*continued on next page*)

(*continued*)

Gloss	provide.fresh.milk	five	classifier	cow/horse/desk	all.very.strong
Trans.	'The five cows/horses/desks providing fresh milk are all very strong.'				
Question	"Was fresh milk provided?"				

mean = 0.39, *se* = 0.04; Related: mean = 0.34, *se* = 0.04; Unrelated: mean = 0.27, *se* = 0.04) [F(2, 118) = 1.39, n.s.]. In addition, classifiers of the three conditions did not differ from each other either in terms of (log transformed) frequency (Expected: mean = 4.1, *se* = 0.076; Related: mean = 4, *se* = 0.057; Unrelated: mean = 4.1, *se* = 0.067; F(2, 303) = 0.98, n.s.; Cai & Brysbaert, 2010) or complexity, measured by the number of strokes (Expected: mean = 7.8, *se* = 0.29; Related: mean = 8.1, *se* = 0.29; Unrelated: mean = 8.1, *se* = 0.28) [F(2, 303) = 0.338, n.s.]. Likewise, the head nouns of the three conditions did not differ from each other either in terms of (log transformed) frequency (Expected: mean = 3.2, *se* = 0.065; Related: mean = 3.1, *se* = 0.07; Unrelated: mean = 3.2, *se* = 0.068) [F(2, 303) = 2.3, p < 0.1] or complexity (Expected: mean = 11.5, *se* = 0.55; Related: mean = 11.8, *se* = 0.58; Unrelated: mean = 11.8, *se* = 0.64) [F(2, 303) = 0.11, n.s.].

The 102 experimental sentences were split into three lists along with 113 filler sentences of similar length and structural complexity, according to a Latin Square design. Each list contained 34 experimental sentences for each condition, such that no participant saw more than one sentence from any triplet, and equal number of experimental conditions were presented to each participant. Each list was further divided into four sub-lists, three of which contained 54 sentences and the last 53 sentences. The sentences in each list were pseudo-randomized for each participant such that sentences from the same condition did not appear consecutively and the stimuli were presented in a different random order for every participant.

2.1.2. Participants

21 native speakers of Mandarin from China were paid S$10/h for their participation in the ERP study (age range: 20–27; 10 females). At the time of the experiment, they were students enrolled at universities or graduate schools in Singapore. All the participants were right-handed with no neurological disorders and normal or corrected-to-normal vision.

2.1.3. Procedures

Participants were tested in a single session in a soundproofed electrically-shielded booth. They were seated in a comfortable chair approximately 120 cm in front of a monitor and instructed to read the stimulus for comprehension while minimizing blinks or any body movement. The experimental session started with a short practice block. Each trial began with a fixation cross in the center of a screen (500 ms). Sentences were then presented visually one phrase at a time (Zhang Yimou/ directed/ DE/ this/ classifier/ head noun/ ...). Each phrase was presented for 400 ms with an interstimulus interval of 400 ms (Zhou et al., 2010). The target classifier position varied from the 5th to the 9th phrase in a given sentence depending on the number of preceding phrases (mean: 5.8, range: 4–8 phrases). Sentence final words were followed by a blank screen for 1000 ms, after which Yes/No comprehension questions were presented. The comprehension questions probed the overall understanding of experimental sentences. For example, for the sentence (3) "Was what Zhang Yimou directed nominated?" was presented (see Table 1 for further examples). Participants were given a short break about every 10 min. The total experiment took about 1.5 h including preparation.

2.1.4. Electrophysiological recording and analysis

The electroencephalogram (EEG) was recorded using a 64-channel HydroCel Geodesic Sensor Nets (Electrical Geodesics Incorporated, Oregon, USA). Impedances were kept below 40 kΩ (Ferree, Luu, Russell, & Tucker, 2001). The EEG was amplified with an EGI EEG System 300 amplifier, with a bandpass of 0.01–100 Hz, digitized at a sampling rate of 250 Hz. The vertex electrode (Cz) was used as the reference during the recording and data were algebraically re-referenced off-line to the mean of the activity at the two mastoids. The offline EEG analysis was performed using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) toolboxes in MATLAB2013b (Mathworks, Natik, MA, USA). Trials contaminated by excessive muscle activity or eye movements were discarded offline before averaging, in an epoch starting 300 ms before and ending 1700 ms after stimulus onset. EEG data were low-pass filtered at 30 Hz. Six participants were removed due to a high artifact rejection rate, and one due to software failure. The average rejection rate was 15% across the remaining 14 participants. For visualization purposes, ERP waves were smoothed using a low pass filter with a cutoff frequency of 7 Hz.

To examine effects at the classifier position, averages of ERP trials were calculated for each condition (expected, unrelated and related) after subtraction of the 100 ms pre-stimulus baseline. A latency window for mean amplitude analyses was decided based on a peak latency analysis (cf. Federmeier & Kutas, 1999). Mean negative peak latency between 350 and 450 ms was 397 ms for the expected condition, 394 ms for the unrelated condition and 400 ms for the related condition. Based on this, mean area voltage measurements were taken from a 100 ms window around 400 ms, between 350 and 450 ms post classifier onset.[2] For the head noun position, we analysed the data using a longer epoch starting from the preceding classifier, measuring against the 300 ms pre-classifier baseline. This was to avoid a potential baseline problem that could arise, as significant differences are predicted between the experiment conditions at the preceding classifier position (for relevant discussion, see Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004; Steinhauer & Drury, 2012). In doing so, we used the 1150–1250 ms latency window post classifier onset for mean amplitude analyses (i.e., the 350–450 ms post head noun onset) to use a comparable time window to that used for the classifier position.

The data were submitted to a full analysis, i.e. an overall ANOVA with repeated measures of experimental condition (expected, unrelated vs. related conditions) and electrode site (42 levels). In addition, a distributional analysis was conducted. For lateral electrodes, twelve regions of interests were defined by crossing three factors: hemisphere (left vs. right), laterality (lateral vs. medial), and anteriority (frontal vs. central vs. posterior) (Fig. 1). These regions included left frontal lateral (E13, E18, E19), left central lateral (E22, E25, E26), left posterior lateral (E27, E28, E30), left frontal medial (E9, E11, E12), left central medial (E16, E20, E21), left posterior medial (E31, E33, E35), right frontal lateral (E56, E58, E59), right central lateral (E46, E48, E49), right posterior lateral (E42, E44, E45), right frontal medial (E2, E3, E60), right central

[2] The 350 to 450 ms post critical word onset coincides with a latency window in DeLong et al. (2014; Fig. 2D), in which two unexpected sentential continuation types with varying congruity (i.e., somewhat plausible vs. anomalous) most differed from each other in successive statistical analyses covering 0 to 1200 ms post critical word.

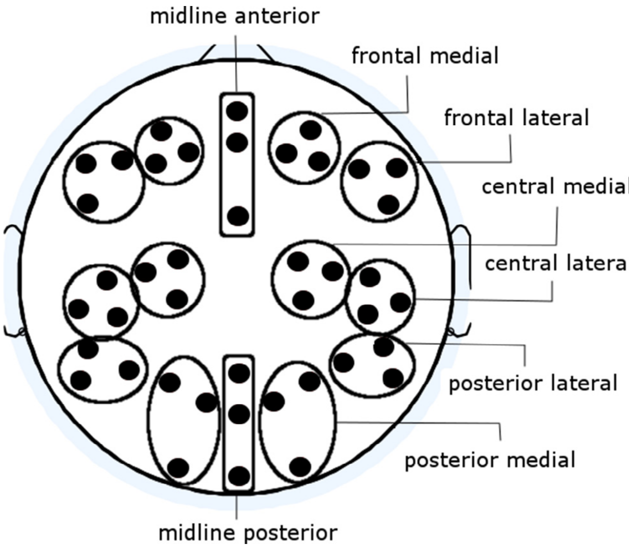

Fig. 1. Electrodes included in statistical analyses in Experiment 1.

medial (E41, E50, E51), and right posterior medial (E38, E39, E40). Analyses were conducted separately for midline electrodes with two levels of anteriority; anterior (E4, E6, E8) and posterior (E34, E36, E37). The Huynh and Feldt (1976) correction for lack of sphericity was applied, and corrected *p*-values are reported with the original degrees of freedom.

2.2. Results and discussion

The mean correct answer rates to comprehension questions following the expected, the unrelated and the related conditions were 95.8%, 92.6% and 93.8% respectively.

2.2.1. ERP results at the classifier position

Fig. 2 shows the averaged ERP responses elicited by the expected, unrelated and related conditions time-locked to classifiers. An omnibus ANOVA showed a main effect of sentence type ($F(2, 26) = 4.726$, $p < 0.018$). Planned pairwise comparisons were then conducted using ANOVAs to compare two levels of sentence type.

2.2.1.1. Unrelated vs. Expected classifiers. The unrelated condition was significantly more negative relative to the expected condition (lateral: $F(1, 13) = 5.382$, $p < 0.037$; midline: $F(1, 13) = 10.537$, $p < 0.006$).

2.2.1.2. Related vs. Expected classifiers. Similarly to the unrelated condition, the related condition elicited more negativity compared to the expected condition (lateral: $F(1, 13) = 5.079$, $p < 0.042$; cf: midline: all $Fs < 2.62$, n.s.).

2.2.1.3. Unrelated vs. Related classifiers. Finally, the pairwise comparison of the unrelated and the related conditions showed a significant main effect of sentence type in a midline analysis ($F(1, 13) = 5.568$, $p < 0.035$) and a marginal interaction of sentence type and hemisphere in a lateral analysis ($F(1, 13) = 3.612$, $p < 0.08$). The interaction was due to a stronger negativity to the unrelated condition compared to the related condition in the right hemisphere as shown in Fig. 2, E and in Fig. 3, which presents difference waves corresponding to the subtraction of expected from unrelated condition and the subtraction of expected from related condition. Indeed, a follow-up analysis only including the electrodes of the right hemisphere showed that the unrelated condition elicited significantly more negativity than the related condition at the classifier position ($F(1, 13) = 8.715$, $p < 0.011$; cf. left hemisphere: all $Fs < 1.6$, n.s.).

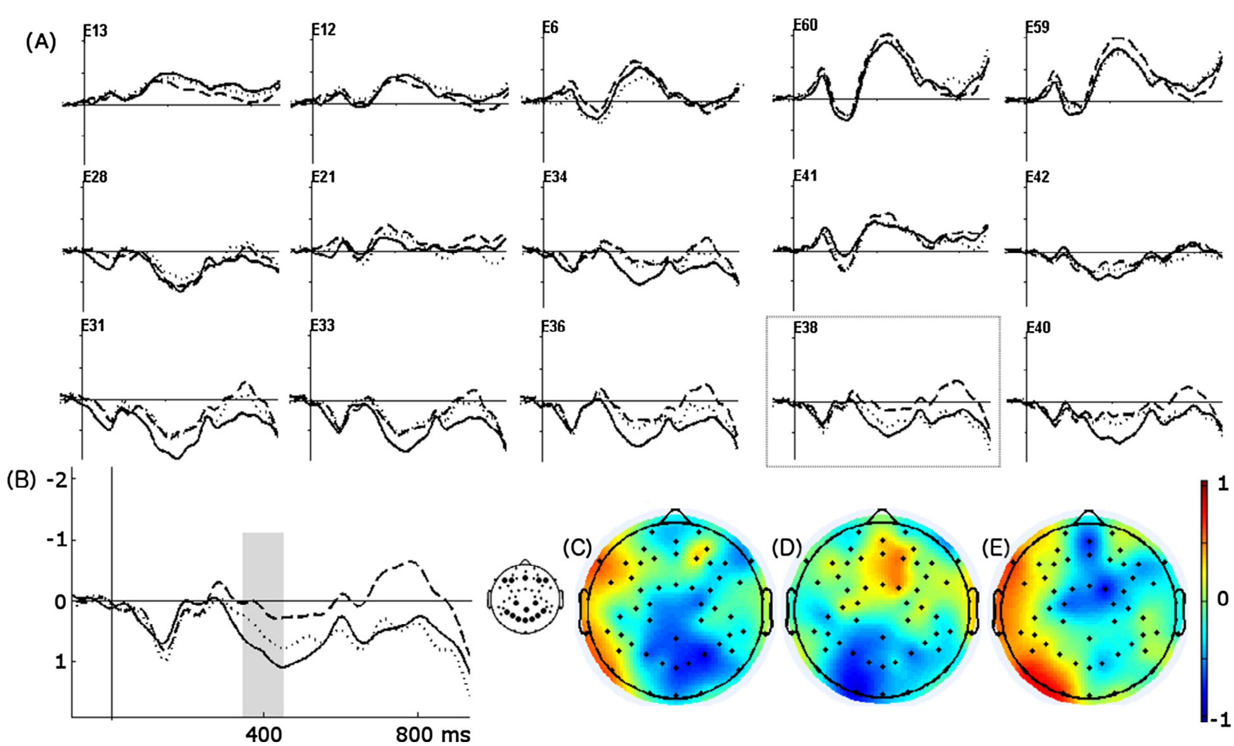

Fig. 2. (A) Experiment 1: Grand average ERP waveforms for expected (solid line), unrelated (dashed line) and related conditions (dotted line) at the classifier position. (B) Grand average ERP waveforms at E38 (the electrode in the dotted square in A). (C-E) Topographic scalp isovoltage map of the mean difference of unrelated and expected (C), of related and expected (D) and of the unrelated and related (E) (350–450 ms).

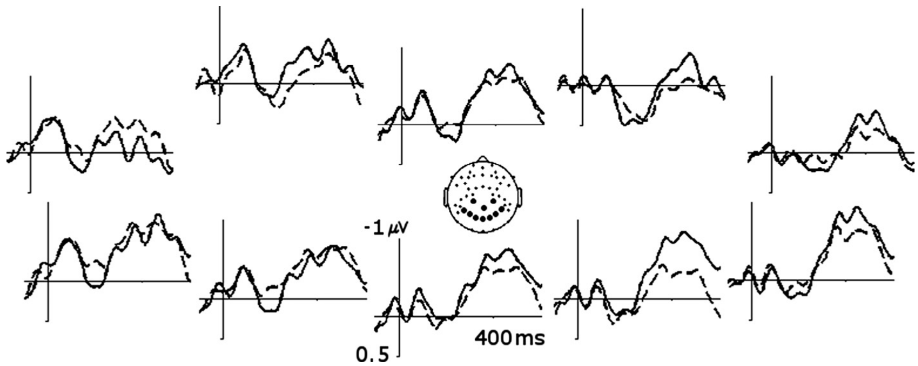

Fig. 3. Experiment 1: Grand average ERP difference waves at the classifier position corresponding to the subtraction of expected from unrelated condition (solid line) and the subtraction of expected from related condition (dashed line). The negativity to the unrelated condition is bigger than that to the related condition.

2.2.2. ERP results at the head noun position

Fig. 4 shows the two-word ERP averages time-locked to the onset of classifier. Visual inspection suggests that the unrelated and the related condition elicited greater negativity than the expected condition in the centro-posterior region. This observation was statistically confirmed. An omnibus ANOVA in the 1150–1250 ms latency range post classifier onset (i.e., 350–450 post head noun onset) showed a main effect of sentence type ($F(2, 26) = 4.064$, $p < 0.029$) and an interaction of sentence type × electrode ($F(82, 1066) = 1.837$, $p < 0.03$).

2.2.2.1. Unrelated vs. Expected head nouns. Planned pairwise comparisons were performed in the same time window and confirmed the difference between the unrelated and the expected condition with a main effect of the sentence type in a midline analysis ($F(1, 13) = 9.231$, $p < 0.01$; cf. lateral: $F(1, 13) = 2.55$, n.s.) with a stronger negativity to the unrelated condition than to the expected condition. In addition, there was a marginally significant interaction of sentence type × anteriority (lateral: $F(2, 26) = 3.849$, $p < 0.065$; midline: $F(1, 13) = 4.567$, $p < 0.052$) (Fig. 4C). This was because the negativity to the unrelated condition was particularly stronger in the centro-posterior region while in the frontal region, the same condition elicited more positive-going ERP responses (for similar observations, see also Thornhill & Van Petten, 2012). To confirm these apparently different effects, we analysed the data separately for the anterior region and centro-posterior regions. For the centro-posterior region (lateral: E16, E20, E21, E22, E25, E26, E27, E28, E30, E31, E33, E35, E38, E40, E39, E42, E45, E44, E49, E46, E48, E51, E41, E50; midline: E34, E36, E37), the unrelated condition elicited a significantly stronger negativity than the expected condition (midline: $F(1, 13) = 8.708$, $p < 0.011$; cf. lateral: $F(1, 13) = 3.818$, $p < 0.073$). On the other hand, in the anterior region (anterior electrodes in Fig. 1; lateral: E18, E13, E19, E11, E12, E9, E2, E3, E60, E59, E58, E56; midline: E8, E6, E4;), there was a marginally significant main effect of the sentence type with the unrelated condition eliciting more positivity than the expected condition in a lateral analysis ($F(1, 13) = 3.71$, $p < 0.076$; cf. midline: all Fs < 1).

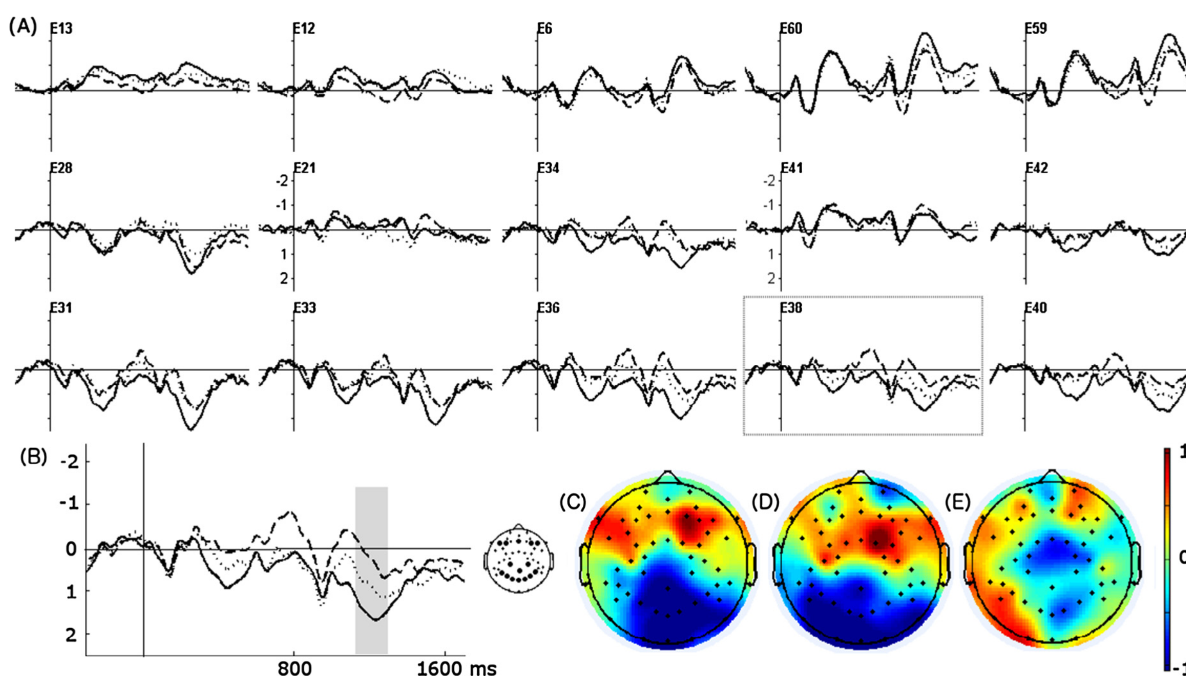

Fig. 4. Experiment 1: ERPs including head noun position: (A) Grand two-word average ERP waveforms for expected (solid line), unrelated (dashed line) and related conditions (dotted line) starting from the classifier. (B) Grand average ERP waveforms at E38 (the electrode in the dotted square in A). (C-E) Topographic scalp isovoltage map of the mean difference of unrelated and expected (C), of related and expected (D) and of unrelated and related (E) (1150–1250 ms post classifier onset; i.e., 350–450 ms post head noun onset).

2.2.2.2. Related vs. Expected head nouns. Pairwise comparison of the related and the expected conditions also confirmed the greater negativity of the related condition relative to the expected condition, with a marginal main effect of sentence type in lateral analyses ($F(1, 13) = 4.45$, $p < 0.055$; cf. midline: $F(1, 13) = 2.67$, n.s.). In addition, there was an interaction of sentence type × laterality × anteriority in a lateral analysis ($F(2, 26) = 6.001$, $p < 0.009$) and an interaction of sentence type × anteriority in a midline analysis ($F(1, 13) = 5.881$, $p < 0.031$). The interactions were due to a stronger negativity in the posterior region and an apparent positivity over the medial electrodes in the frontal region to the related condition compared to the expected condition (Fig. 4D). The positivity in the frontal region, however, was not significant in the latency range of 1150–1250 ms post classifier onset (i.e., 350–450 ms post head noun onset) or in any of later latency ranges either in a lateral and a midline analysis (i.e., every 200 ms interval from 400 to 900 ms post head noun onset; all Fs < 2). On the other hand, the posterior negativity was significant at midline ($F(1, 13) = 7.291$, $p < 0.018$), in the posterior medial region ($F(1, 13) = 6.497$, $p < 0.024$) and in the posterior lateral region ($F(1, 13) = 4.694$, $p < 0.049$).

2.2.2.3. Unrelated vs. Related head nouns. The pairwise comparison of the unrelated and the related conditions showed a significant interaction of sentence type × laterality × anteriority in a lateral analysis ($F(2, 26) = 3.996$, $p < 0.031$; cf. midline: $F(1, 13) = 2.85$, n.s.). This three-way interaction was due to a significant interaction of the sentence type with electrodes in the posterior-medial region ($F(5, 65) = 3.35$, $p < 0.024$) and a marginal main effect of the sentence type in the centro-medial region ($F(1, 13) = 3.968$, $p < 0.068$) with a stronger negativity to the unrelated condition (Fig. 4E).

In summary, predictions concerning the N400 were mostly confirmed (see Table 2 for a summary of the results). As predicted, both the unrelated and the related conditions elicited a negativity compared to the expected condition. Importantly, the elicited negativity was stronger for the unrelated condition than for the related condition. These effects were observed as early as the classifier position and were prominent in the centro-posterior region. These results suggest that head nouns were predicted ahead of time, and the semantic relatedness of unexpected words to predicted words eased processing difficulty. On the other hand, predictions concerning frontal positivity were not clearly supported. We predicted that the related condition (but not the unrelated condition) would elicit a frontal positivity. Unexpectedly, however, the positivity effect was marginally significant, in response to the unrelated condition but not to the related condition. Thus, the positivity effect in the current study does not seem overall compatible with the previous findings.

However, the number of participants in this study was on the low side, with only 14 subjects. Thus, it is possible that some of the crucial comparisons of the unrelated vs. the related conditions and of the related vs. the expected conditions were statistically marginal or did not even approach significance due to the small sample size. It is also important to replicate the overall results with a larger number of participants. To this goal, we conducted the experiment again, involving more number of participants but using the same sets of experimental sentences.

3. Experiment 2

Experiment 2 was a replication of Experiment 1, using the same experimental stimuli but with a larger number of participants. It is worth noting that Experiment 2 was run in a different lab, using a different ERP setup, so it is not possible to do a combined analysis of the data of the two experiments. We therefore report both studies as independent experiments. If a similar pattern of results is found using two different ERP setups, it will increase our confidence in the findings, and demonstrate the replicability of the findings.

3.1. Methods

3.1.1. Participants

30 native speakers of Mandarin from China participated in the ERP study, receiving 10,000 Korean Won per hour (age range: 18–28; 17 females). At the time of the experiment, they were students enrolled at Konkuk University in Korea. All the participants were right-handed, with no neurological disorders and normal or corrected-to-normal vision.

3.1.2. Procedures

The procedures were analogous to those described in Experiment 1.

3.1.3. Electrophysiological recording and analysis

The electroencephalogram (EEG) was recorded using a 19-electrode cap (Electro-Cap International, Inc), with electrodes located according to the 10–20 system (Jasper, 1958). The EEG was amplified using Neuroscan Synamps with a bandpass of 0.01–100 Hz, and continuously digitized at a sampling rate of 500 Hz. EEG channels were referenced online to the left mastoid and data were algebraically re-referenced off-line to the mean of the activity at the two mastoids. To monitor vertical and horizontal eye movements, external electrodes were placed on the outer canthi and under each eye, and were referenced to the left mastoid. Impedances were kept below 5 kΩ for each electrode. As in Experiment 1, trials affected by blinks, eye movements or muscle activity were excluded offline before averaging, in an epoch starting 300 ms before and ending 1700 ms after stimulus onset. EEG data were low-pass filtered at 30 Hz. Four participants were removed due to a high rejection rate, and one due to low comprehension accuracy rates (57.8%). The average rejection rate was 10.9% across the remaining 25 subjects. For visualization purposes only, ERP waves were smoothed using a low pass filter with a cutoff frequency of 7 Hz.

Table 2
Summary of the results.

	Classifier (350–450 ms)		Head noun (350–450 ms)	
	Exp1	Exp2	Exp1	Exp2
Negativity effect				
3 conditions	Yes	Yes	Yes	Yes
Unrelated vs. Expected	Yes	Yes	Yes	Yes
Related vs. Expected	Yes	Yes	Yes	Yes ($p < 0.07$)
Unrelated vs. Related	Yes	Yes (300–600 ms)	Yes	Yes
Positivity effect				
Unrelated vs. Expected	n.a.		Yes ($p < 0.076$)	No
Related vs. Expected	n.a.		No	No

Peak latency analyses showed similar results to those in Experiment 1; mean negative peak latency for the expected condition, the unrelated condition and the related condition was 401 ms, 402 ms and 400 ms respectively. Accordingly, mean area voltage measurements were taken from a 100 ms window around the negativity peak, between a 350 and 450 ms latency window post classifier onset to examine effects at the classifier position. For the head noun position, a latency window of 1150–1250 ms post classifier onset (i.e., 350–450 ms post head noun onset) was used, similarly to Experiment 1.

The data were submitted to a full analysis, i.e. an overall ANOVA with repeated measures of experimental condition (expected, unrelated vs. related conditions) and electrode site (15 levels). In addition, a distributional analysis was conducted. For lateral electrodes, six regions of interests were defined by crossing two factors: hemisphere (left vs. right), and anteriority (frontal vs. central vs. posterior). These regions included left frontal (F7 & F3), left central (T3 & C3), left posterior (T5 & P3), right frontal (F4 & F8), right central (C4 & T4), and right posterior (P4 & T6). For midline electrodes, three regions of interest were defined; frontal (Fz), central (Cz), and posterior (Pz). The Huynh and Feldt (1976) correction for the violation of sphericity was applied, and corrected *p*-values are reported with the original degrees of freedom.

3.2. Results & discussion

The mean correct answer rates to comprehension questions following the expected, the unrelated and the related conditions were 94.5%, 91.7% and 94% respectively.

3.2.1. ERP results at the classifier position

Fig. 5 shows the averaged ERP responses elicited by the expected, unrelated and related conditions, time-locked to the classifier. Visual inspection suggests that, as in Experiment 1, both unrelated and related classifiers elicited an overall greater negativity relative to expected classifiers over centro-posterior regions. In addition, waveforms appear more negative-going for unrelated than for related classifiers. These observations were statistically confirmed. An omnibus ANOVA including the three conditions in the 350–450 ms latency range showed a main effect of sentence type ($F(2, 48) = 5.331$, $p < 0.008$) and an interaction of sentence type and electrodes ($F(28, 672) = 2.052$, $p < 0.001$). Planned pairwise comparisons were then conducted in the same latency window using ANOVAs to compare two levels of sentence type.

3.2.1.1. Unrelated vs. Expected classifiers. The pairwise comparison of the unrelated and the expected condition showed a main effect of sentence type (midline: $F(1, 24) = 11.321$, $p < 0.003$; lateral: $F(1, 24) = 4.259$, $p < 0.05$) with a stronger negativity to the unrelated condition compared to the expected condition.

3.2.1.2. Related vs. Expected classifiers. Similarly to the unrelated condition, the related condition elicited significantly more negativity compared to the expected condition (midline: $F(1, 24) = 6.298$, $p < 0.019$; cf. lateral: $F(1, 24) = 3.001$, $p < 0.096$). A lateral analysis also showed a marginal interaction of sentence type × anteriority ($F(2, 48) = 2.770$, $p < 0.095$). This was due to a stronger negativity to the related condition than to the expected condition in the centro-posterial region ($F(1, 24) = 4.997$, $p < 0.035$; Fig. 5, D).

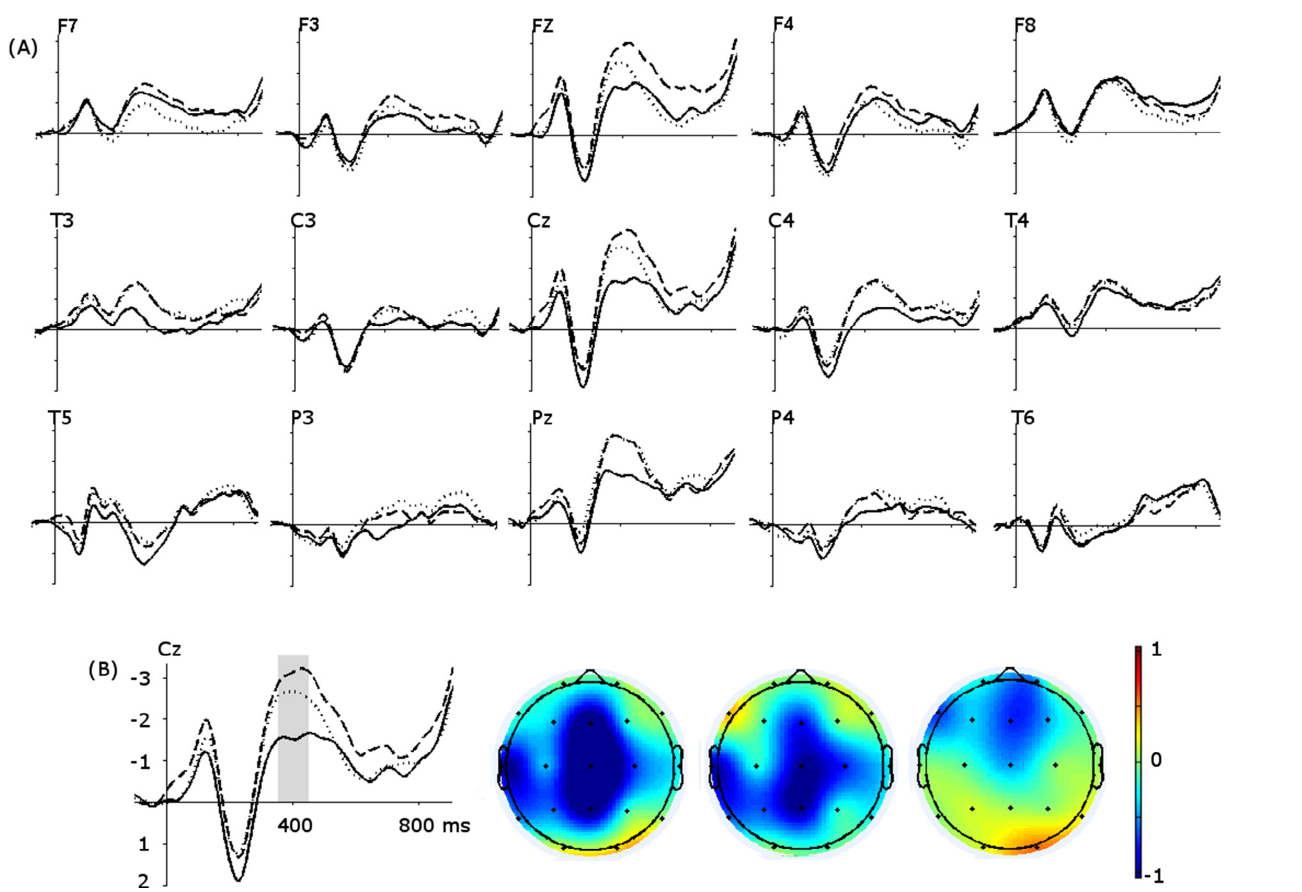

Fig. 5. (A) Experiment 2: Grand average ERP waveforms for expected (solid line), unrelated (dashed line) and related conditions (dotted line) at the classifier position. (B) Grand average ERP waveforms at Cz (the electrode in the dotted square in A). (C-E) Topographic scalp isovoltage map of the mean difference of unrelated and expected (350–450 ms) (C), of related and expected (350–450 ms) (D) and of unrelated and related (300–600 ms) (E).

3.2.1.3. Unrelated vs. Related classifiers. The pairwise comparison of the unrelated and the related conditions did not show any effect of the sentence type in the 350–450 ms latency range (all $Fs < 2$, n.s.). However, visual inspection suggested that negativity to the unrelated condition was stronger and thus sustained slightly longer than the negativity to the related condition (Fig. 5). Given this, we ran an additional analysis in a longer time window to confirm this observation. Indeed, the pairwise comparison of the unrelated and the related condition in the 300–600 ms latency range showed a marginal main effect of the sentence type ($F(1, 24) = 3.79$, $p < 0.063$) and a marginal interaction of sentence type × anteriority ($F(2, 48) = 2.662$, $p < 0.09$; cf. lateral: all $Fs < 2.3$) due to a stronger negativity to the unrelated condition compared to the related condition in the fronto-central region (Fig. 5, E). To corroborate these local effects, a follow-up analysis was conducted including midline electrodes in the fronto-central region (Fz, Cz). The analyses revealed a significant main effect of sentence type ($F(1, 24) = 5.886$, $p < 0.023$).

3.2.2. ERP results at the head noun position

For the head noun position, as in Experiment 1 we analysed the data using a longer epoch starting from the onset of the preceding classifier without re-baselining at the head noun position. Visual inspection suggests that the unrelated and the related conditions elicited greater negativity than the expected condition, and that negativity was stronger to the unrelated condition than to the related condition (Fig. 6). Statistical analyses were performed to examine these observations. An omnibus ANOVA in the 1150–1250 ms latency range post classifier onset (i.e., 350–450 post head noun onset) confirmed that there was a significant difference among the three conditions with a main effect of sentence type ($F(2, 48) = 3.452$, $p < 0.04$). Planned pairwise comparisons were then conducted in the same latency window to compare each pair of sentence type.

3.2.2.1. Unrelated vs. Expected head nouns. A pairwise comparison of the unrelated and the expected conditions showed a main effect of the sentence type with stronger negativity to the unrelated condition than to the expected condition (midline: $F(1, 24) = 10.692$, $p < 0.003$; lateral: all $Fs < 2.4$, n.s.). There was no other effect.

3.2.2.2. Related vs. Expected head nouns. A pairwise comparison of the related and the expected condition revealed a marginal main effect of the sentence type in a midline analysis ($F(1, 24) = 3.59$, $p < 0.07$) with a stronger negativity to the related condition compared to the expected condition. No significant effect was found in a lateral analysis (all $Fs < 2$).

3.2.2.3. Unrelated vs. Related head nouns. The pairwise comparison of the unrelated and the related condition revealed a marginal main effect of the sentence type in a midline analysis, with a stronger negativity to the unrelated condition than the related condition ($F(1, 24) = 4.23$, $p < 0.051$; cf. lateral: all $Fs < 1$, n.s.). There was also a significant interaction of the sentence type and anteriority in a midline analysis ($F(2, 48) = 3.44$, $p < 0.04$) due to a stronger congruency effect in the fronto-central region (Fz & Cz: $F(1, 24) = 5.314$, $p < 0.03$) with stronger negativity to the unrelated than to the related condition (Fig. 6E).

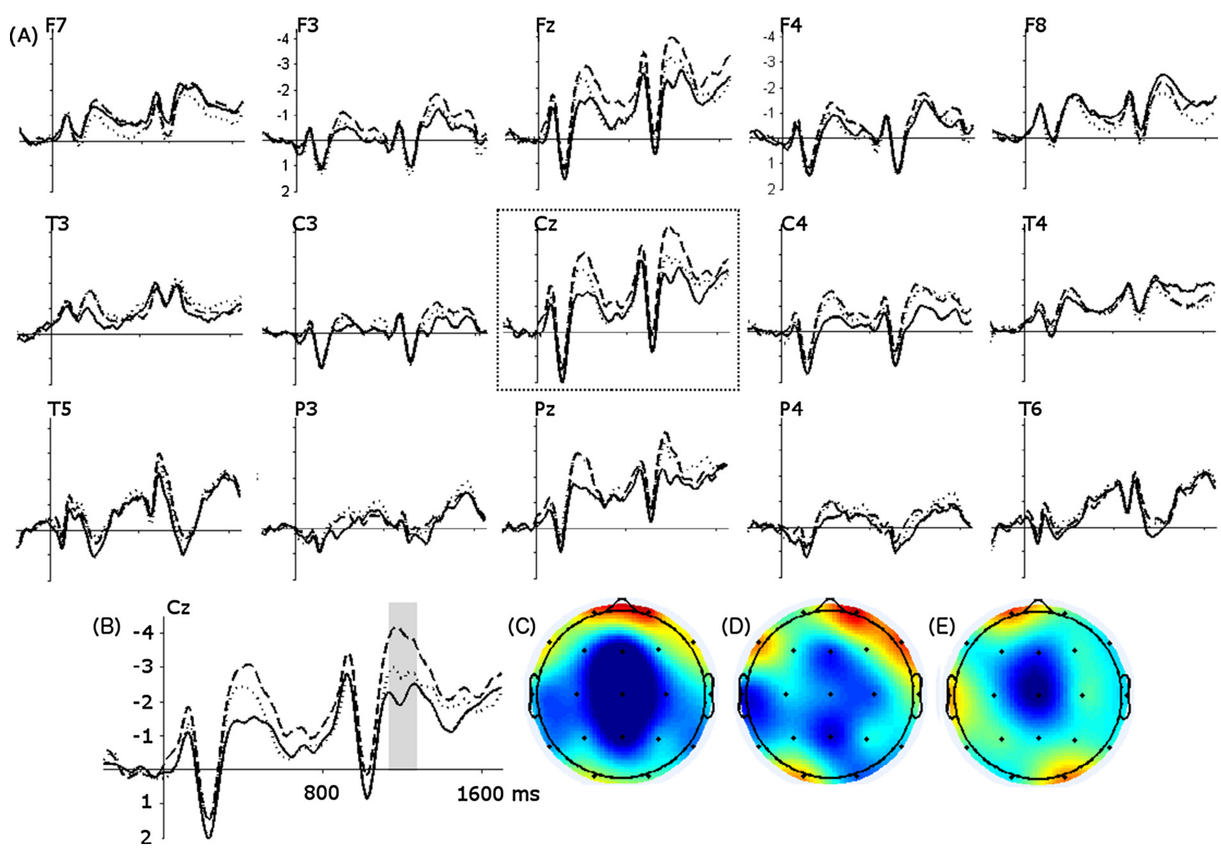

Fig. 6. (A) Experiment 2: ERPs showing the head-noun condition: Grand two-word average ERP waveforms for expected (solid line), unrelated (dashed line) and related conditions (dotted line) starting from the classifier position. (B) Grand average ERP waveforms at Cz (the electrode in the dotted square in A). (C-E) Topographic scalp isovoltage map of the mean difference of unrelated and expected (C), of related and expected (D) and of unrelated and related (E) conditions (1150–1250 ms post classifier onset; i.e., 350–450 ms post head noun onset).

To summarize, the critical findings of Experiment 1 were replicated in Experiment 2. Similarly to Experiment 1, brain responses to the unrelated and the related conditions were more negative-going relative to those to the expected condition at the classifier, as well as at the head noun position. Importantly, the negativity was stronger for the unrelated condition than for the related condition at these two critical word positions. On the other hand, a positivity effect was not visible in response either to the unrelated or to the related conditions at the head noun position, unlike in Experiment 1 (see Table 2 for a summary of the results).

4. General discussion

Two ERP experiments were conducted to investigate whether semantic features of an upcoming word are pre-activated based on prior contextual constraints. Below we discuss the results of Experiment 1 & 2 in detail in relation to the goal of the study.

Experimental sentences started with a relative clause with a highly constraining semantic context, such that a particular head noun would be predicted, but the design included three types of head nouns. Sentences with highly predicted head nouns (i.e., cloze probability 0.85) constituted the expected condition. The unrelated condition consisted of sentences with head nouns which were unlikely to be predicted (i.e., cloze probability 0.0) and were semantically unrelated to the predicted head noun. Finally, the related condition consisted of sentences with head nouns which were likewise unlikely to be predicted (i.e., cloze probability 0.0) but were semantically related to the predicted head noun. Crucially, all the experimental conditions included a classifier, which agreed with its associated head noun but was incompatible with the head nouns of the other conditions.

Given that previous studies have shown there is a strong correlation between N400 amplitude and the cloze probability of a target noun (Kutas & Hillyard, 1984), we first predicted that the unrelated and the related conditions would both elicit an N400 at the head noun position, in comparison to the expected condition, as the cloze probability was zero in both the unrelated and related conditions, but was high for the expected condition. Results confirmed the prediction. The unrelated head nouns elicited significantly stronger negativities compared to the expected head nouns both in Experiment 1 ($p < 0.01$) and Experiment 2 ($p < 0.003$). Likewise, the related condition elicited stronger negativities compared to the expected condition both in Experiment1 ($p < 0.018$) and, marginally, Experiment 2 ($p < 0.07$). These effects looked most pronounced over the centro-posterior sites of the brain, confirming the typical N400 distribution and morphology (Figs. 4 and 6).

Crucially, we also predicted that N400 effects would already be significant at the preceding classifier position. Although a classifier and its associated head noun should agree with each other in Chinese (Zhang et al., 2012), classifiers of the unrelated and the related conditions do not agree with the predicted head noun. Accordingly, if semantic features of the predicted head nouns are pre-activated, classifiers in the unrelated and the related conditions should elicit an N400 due to a semantic mismatch. The results confirmed the prediction; the negativity was already significant at the classifier position both in Experiment 1 (unrelated: $p < 0.006$; related: $p < 0.042$) and Experiment 2 (unrelated: $p < 0.003$; related: $p < 0.019$).

In addition, as semantic overlap with the predicted lexical item facilitates the processing of an unexpected item (Federmeier & Kutas, 1999), we predicted that the processing difficulty in the related condition would be reduced compared to the unrelated condition. The prediction was also confirmed in both experiments. The unrelated condition overall elicited significantly stronger N400s than the related condition at the head noun position (Experiment 1: $p < 0.031$; Experiment 2: $p < 0.03$) replicating Federmeier & Kutas. Importantly the effect was already visible as early as at the classifier position (Experiment 1: $p < 0.035$; Experiment 2: $p < 0.023$), extending the findings of Federmeier & Kutas in that the effect of the predicted word was found at a word position earlier (i.e., classifier position) than the expected word position itself (i.e. head noun position).

These results are consistent with the hypothesis that semantic features of upcoming words are pre-activated prior to the bottom-up input of the actual word and that these features facilitate the processing of unexpected items that share many semantic features in common with the predicted word (i.e., the related condition) relative to those with fewer shared features (i.e., the unrelated condition). Thus, these results can be taken as evidence to confirm the generality of predictive processing; at least with highly constraining contextual information, language processing is predictive even in a language without overt morpho-syntactic markers that correlate with salient semantic features (e.g., animacy or gender markers), and the predictions are fine-grained enough to include information about the broad range of semantic classifications that classifiers in Chinese are based on.

However, given that classifiers in Chinese have semantic content, it is possible that the N400s observed at the classifier position reflect semantic integration rather than prediction. Under this hypothesis, the N400 would have been the result of the difficulty of integrating the classifier with the content of the preceding relative clause. However, the experimental sentences in the current study were devised such that combinatory semantic constraints of an agent and a verb would lead to predictions of specific lexical items, just as in Kamide et al. (2003). For example, for (3) ('~which Zhang Yimou directed') 'a movie' would have been predicted, as the agent is 'Zhang Yimou' and the verb is 'direct'. If used with a different verb such as 'like' or 'talk about', all the experimental conditions (expected: 'movie'; related: 'performance'; unrelated: 'building') could have led to expected sentential continuations. Likewise, if used with a different agent such as *Britney Spears*, the related condition ('performance') would have made a better sentential continuation than the expected condition. Thus, given the nature of the sentences used, it is unlikely that the processing of a classifier is independent of the specific head nouns, which would be predicted on the basis of highly constraining sentential contexts.

More importantly, as already discussed in the introduction, the semantic content of classifiers is often fairly abstract, and can even be "empty". For example, for (3) ('~which Zhang Yimou directed'), the expected condition involved a classifier 部, which is a classifier for 'movie', 'novels', 'cars' and 'telephones' but means 'department' when used as a noun. However, the preceding sentential context would lead to the prediction of 'movie' not 'literature' and certainly not 'cars' or 'telephones'. There may be at least two semantic categories represented by a classifier 部, with 'movie' and 'literature' forming one and 'cars' and 'telephones' the other. However, even so, it is not easy to exclusively characterize the semantic content of the classifier 部 for these two categories in a way that could be distinguished from other related concepts such as 'performance', which was the related condition in (3) and requires a different classifier, and 'train', which also requires a different classifier, 辆. In addition, some classifiers do not even have any meaning. For example, 匹lost its meaning and is only used as a classifier for a particular noun (i.e., horse) in modern Chinese.

Given this, it is unlikely that classifiers are interpreted without consideration of the predicted head nouns in our materials. Alternatively, the abstract semantic content of a classifier could be evaluated against semantic features of the predicted head noun. For example, the classifier for 'performance' in the related condition of (3), 场 is used for 'events' and 'recreational activities' and thus is not compatible with the predicted head noun (i.e., 'movie'). However, 场 has more semantic feature overlap with the predicted head noun than the classifier of the unrelated condition, 座which is a classifier for 'buildings' or immovable objects. Accordingly, more semantic feature overlap in the related condition could have relatively eased the processing difficulty of the unexpected sentential continuation. Therefore, the N400s observed at the classifier position would be better accounted for in terms of pre-activation of semantic features of the predicted head noun.

Finally, we also predicted that a frontal positivity would be elicited in response to the head noun in the related condition. The frontal positivity has been characterized as an index of the cost of failed specific lexical prediction (Thornhill & Van Petten, 2012; Van Petten & Luka, 2012) or as being related to inhibition of words which were predicted but not actually presented (Kutas, 1993). However, the prediction for positivities was not confirmed in the study. In Experiment 1, visual inspection suggested frontal positivities at the head noun position both in the related and the unrelated condition in comparison to the expected condition. However, unexpectedly the effect was only marginally significant in response to the unrelated condition, and the apparent positivity to the related condition was not significant. In Experiment 2, the positivity effect was not significant at all, either in response to the unrelated condition or the related condition.

The frontal positivity is contingent on both predictability and plausibility (Federmeier et al., 2010; DeLong et al., 2014). As we did not overtly manipulate the plausibility of the target sentences, relative plausibility of the experimental conditions is not clear. Nonetheless, given semantic distance manipulations to the expected sentences, the related condition was more plausible than the unrelated condition although both are unexpected. Given this, a stronger effect to the unrelated condition than to the related condition found in the current study is not compatible with the results of the previous studies. However, the prediction-related positivity effect has been less frequently elicited and/or is relatively fragile compared to the N400 effect (Van Petten & Luka, 2012). Thus, it is not completely unlikely that the frontal positivities observed in Experiment 1 are related to lexical expectations.[3] Nonetheless, as the predictions concerning the frontal positivity were not clearly confirmed both in Experiment 1 and 2, it is difficult to further conjecture the nature of the positivity based on the results in this study. We believe that future research is needed in this regard.

In summary, N400s were elicited in response to the unrelated and the related condition in comparison to the expected condition. Importantly, the N400 effects were visible as early as the classifier position preceding the predicted head noun position and were stronger for the unrelated condition than for the related condition. The overall results extend previous findings of semantic anticipation by showing that semantic activation is not limited to languages with grammaticalized semantic features (e.g., Polish with animacy marking or Spanish/Dutch with gender marking) but can extend to a language without overt morpho-syntactic markers for semantic feature agreement. In addition, the gradient semantic prediction effects (i.e., graded N400 amplitudes for expected vs. related vs. unrelated conditions) suggest that a context may lead to the prediction of semantic features that are fine-grained enough to include information about broad types of semantic classification that form the basis of Chinese classifiers.

[3] We do not think that the positivity effect observed in Experiment 1 is a P600 effect. While a P600 typically has a parietal distribution, the effect in Experiment 1 shows a frontal distribution. Also, importantly, as discussed in the introduction, noun-classifier mismatches in Chinese elicit an N400 not a P600 (Zhang et al., 2012). Given this, it is unlikely that the observed effect is a P600.

Acknowledgements

This research was supported by the faculty research fund of Konkuk University in 2015. We would like to thank Qiang Li and Aili Zhang for their help with data collection and Huimin Tang, Huifang Or, Dan Ho, Eshley Gao, Yubin So, and Rachel Chen for their help with stimulus generation. We are also immensely grateful to reviewers and the editor for their constructive and insightful comments, and to Matthew Pontifex for his EEGLAB data import toolbox (loadcurry v2.0).

References

Allan, K. (1977). Classifiers. *Language*, 285–311.
Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE, 5*(6), e10729.
Chou, C.-J., Huang, H.-W., Lee, C.-L., & Lee, C.-Y. (2014). Effects of semantic constraint and cloze probability on Chinese classifier-noun agreement. *Journal of Neurolinguistics, 31*, 42–54.
Clark, An (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*, 181–204.
Chow, W-Y., Lago, S., Barrios, S., Parker, D., Morini, G., & Lau, E. (2014). Additive effects of repetition and predictability during comprehension: Evidence from event-related potentials. *PLoS One*, 9.
Corbett, G. (1991). Gender. Cambridge textbooks in linguistics. *Cambridge: Cambridge*.
DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language, 121*(3), 226–239.
DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia, 61*, 150–162.
Delong, K. A., Urbach, T. P., Groppe, D. M., & Kutas, M. (2011). Overlapping dual ERP responses to low cloze probability sentence continuations. *Psychophysiology, 48*, 1203–1207.
DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.
Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.
Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 641–655.
Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491–505.
Federmeier, K. D., & Kutas, M. (1999). 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., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language, 115*(3), 149–161.
Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*, 133–146.
Ferree, T. C., Luu, P., Russell, G. S., & Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clinical Neurophysiology, 112*, 536–544.
Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*(2007), 75–84.
Fischler, I., & Bloom, P. A. (1979). Automatic and attentional processes in the effects of sentence contexts on word recognition. *Journal of Verbal Learning and Verbal Behavior, 18*, 1–20.
Frazier, L. & K. Rayner (1988). Parameterizing the language processing system: Left- vs. right-branching within and across languages. In J. A. Hawkins (Ed.), *Explaining Language Universals* (pp. 247–279). Basil Blackwell.
Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.
Friston, K., & Stephan, K. (2007). Free energy and the brain. *Synthese, 159*, 417–458.
Frith, C. D., & Frith, U. (2006). How we predict what other people are going to do. *Brain Research, 1079*, 36–46.
Greenberg, J. H. (1978). How does a language acquire gender markers. *Universals of Human Language, 3*, 47–82.
Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124*, 62.
Hsu, C. C., Tsai, S. H., Yang, C. L., & Chen, J. Y. (2014). Processing classifier–noun agreement in a long distance: An ERP study on Mandarin Chinese. *Brain and Language, 137*, 14–28.

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behaviour, but only during linguistic processing. *Journal of Cognition and Culture, 10*, 39–58.

Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational and Behavioral Statistics, 1*, 69–82.

Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157–171.

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 10*, 371–375.

Jiang, X., & Zhou, X. (2012). Multiple semantic processes at different levels of syntactic hierarchy: Does the higher-level process proceed in face of a lower-level failure? *Neuropsychologia, 50*, 1918–1928.

Jordan, T. R., & Thomas, S. M. (2002). In search of perceptual influences of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 34.

Kaiser, E., & Trueswell, J. C. (2004). The role of discourse context in the processing of a flexible word-order language. *Cognition, 94*, 113–147.

Kamide, Y., Altmann, G., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*, 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research, 32*, 37–55.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: Evidence from ERPs. *Journal of Cognitive Neuroscience, 24*, 1104–1112.

Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes, 8*, 533–572.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 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.

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. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationship and prediction in context. *Journal of Memory and Language, 61*, 326–338.

Lau, E. F., Stroud, C., Plesch, S., & Phillips, C. (2006). The role of structural prediction in rapid syntactic analysis. *Brain and Language, 98*, 74–88.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*.

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69*, 574–588.

Mehta, B., & Schaal, S. (2002). Forward models in visuo-motor control. *Journal of Neurophysiology, 88*, 942–953.

Osterhout, L., McLaughlin, J., Kim, A., Greenwald, R., & Inoue, K. (2004). Sentences in the brain: Event-related potentials as real-time reflections of sentence comprehension and language learning. In M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Psychology Press.

Otten, M., Nieuwland, M. S., & Van Berkum, J. J. A. (2007). Great expectations: Specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8*, 89.

Otten, M., & Van Berkum, J. J. A. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92–101.

Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences, 11*, 105–110.

Qian, Z., & Garnsey, S. M. (2016). A sheet of coffee: An event-related brain potential study of the processing of classifier-noun sequences in English and Mandarin. *Language, Cognition and Neuroscience, 31*, 761–784.

Rayner, K., & Frazier, L. (1987). Parsing temporarily ambiguous complements. *Quarterly Journal of Experimental Psychology, 39*, 657–673.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory and Cognition, 18*, 367–379.

Song, J. (2009). *The semantic structure of Chinese classifiers and its implications for linguistic relativity* (Unpublished doctoral dissertation). University of Hawaii, Manoa.

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language, 120*, 135–162.

Szewczyk, J., & Schriefers, H. (2013). Prediction in language comprehension beyond specific words: An ERP study on sentence comprehension in Polish. *Journal of Memory and Language, 68*, 297–314.

Tai, J. H. (1994). Chinese classifier systems and human categorization. In honor of William S.-Y. Wang (Ed.), *Interdisciplinary studies on language and language change* (pp. 479–494).

Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*, 382–392.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 176–190.

Weber, A., Grice, M., & Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities: A study of anticipatory eye movements. *Cognition, 99*, 63–72.

Wicha, N. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165–168.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex, 39*, 483–508.

Wicha, N. 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. *Journal of Cognitive Neuroscience, 16*, 1272–1288.

Wlotko, E. W., & Federmeier, K. D. (2015). Time for prediction? The effect of presentation rate on predictive sentence comprehension during word-by-word reading. *Cortex, 68*, 20–32.

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology, 11*, R729–R732.

Wright, B., & Garrett, M. (1984). Lexical decision in sentences: Effects of syntactic structure. *Memory and Cognition, 12*, 31–45.

Zhang, Y., Zhang, J., & Min, B. (2012). Neural dynamics of animacy processing in language comprehension: ERP evidence from the interpretation of classifier–noun combinations. *Brain and Language, 120*, 321–331.

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.