Journal of Neurolinguistics 52 (2019) 100852

Contents lists available at ScienceDirect

Journal of Neurolinguistics

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

Research paper

An elephant needs a head but a horse does not: An ERP study of classifier-noun agreement in Mandarin

Shiao-hui Chan

Department of English, National Taiwan Normal University, Taipei, Taiwan

ARTICLE INFO

Keywords:
Classifier
Agreement
Morphosyntactic processing
N400
LAN
Mandarin Chinese

ABSTRACT

Classifiers are essential elements between numerals and nouns in Mandarin (e.g. "one-*tou*$_{CL}$-elephant"), but whether they serve a semantic or functional/morphosyntactic role in relation to the accompanying noun has been heatedly debated in linguistics. Previous ERP research consistently supported the semantic view with findings of N400; however, the apparent meaning clash in mismatched classifier-noun pairing in these studies might render morphosyntactic processing undetected. We created two violation conditions to explore classifier-noun agreement: incongruent GE-noun combinations (replacing a specific classifier with the meaning-devoid general classifier, GE) and outright grammatical mistakes (missing a required classifier). With congruent combinations as the baseline, GE-noun combinations elicited a negativity effect strikingly similar to that induced by the grammatical violation condition in phrases (Experiment 1) and sentences (Experiment 2), indicating the involvement of morphosyntactic processing in classifier-noun agreement. The finding suggests that there is a middle ground for the linguistic debate over the nature of classifier selection in relation to nouns.

1. Introduction

Quantity of objects can be expressed with language, and different languages do it differently. For example, an English speaker uses a numeral followed by a noun to describe the quantity of objects (e.g. "three books"), while a Mandarin speaker has to place a "classifier[1]" (CL hereafter in examples and experimental conditions) between a numeral and a noun (e.g. "three-*ben*$_{CL}$-book"). However, the selection of a proper classifier is not an easy task because there are 174 classifiers in Modern Mandarin (according to C. Huang, Chen, & Lai, 1997) and selectional constraints exist between a classifier and its accompanying noun (Allan, 1977; Gao & Malt, 2009; Tai, 1994; N. Zhang, 2013). For example, "an elephant needs a head but a horse does not", as stated in the article title, because the classifier *tou*, which literally means "head", can co-occur with "elephant" (e.g. 一頭大象 *yi-tou*$_{CL}$*-daxiang* 'one-CL$_{large\ animal}$-elephant') but generally not with "horse" (e.g. *一頭馬[2] *yi-tou*$_{CL}$*-ma* 'one-CL$_{large\ animal}$-horse'). Therefore, when using a classifier phrase (Numeral-Classifier-Noun, hereafter Num-CL-N), a Mandarin speaker is quantifying and classifying objects at the same time (Her & Hsieh, 2010; N. Zhang, 2013). Specifically, for quantifying, classifiers are there to name the unit already present in the semantic

E-mail address: shiaohui@ntnu.edu.tw.

[1] Mandarin classifiers can be divided into sortals/count-classifiers and mensurals/measure-classifiers from the perspectives of cognitive distinction (sortals refer to "inherent" property of entities while mensurals refer to "contingent" properties), syntactic patterns (*de* can be added between a mensural and the noun but not between a sortal and the noun) and language typology (sortals are specific to a few languages while mensurals can be found in all languages) (Her & Hsieh, 2010; Tai, 1994). This study focuses on sortal classifiers and the use of the word "classifier" henceforth refers to this type of classifiers.

[2] We follow the convention in linguistics by putting an asterisk (*) before an ungrammatical structure.

https://doi.org/10.1016/j.jneuroling.2019.100852
Received 9 January 2019; Received in revised form 11 May 2019; Accepted 19 June 2019
Available online 26 June 2019
0911-6044/

denotation of the noun and tell people what counts as "one" in the context (Cheng & Sybesma, 2005; Her, 2012); for classifying, classifiers group nouns into different classes according to their sematic dimensions, such as animacy, shape, consistency, or size (Gao & Malt, 2009; Tai, 1994). Linguists have long been intrigued to study the complex nature of the selectional constraints between classifiers and their accompanying nouns. When explaining the constraints from a cognitive perspective, linguists have found that the selection of a classifier is mainly determined by the cognitive categories (animacy, shape, size, consistency) or attributes (parts of an object, e.g. head, mouth) of the object, although there are still cases where the pairing relationship cannot be easily explained (e.g. dialectal variations) (Gao & Malt, 2009; Tai, 1994). However, when exploring the selectional restrictions from a formal/syntactic perspective, researchers' opinions differ: some believe that classifiers serve grammatical/functional functions without semantic contribution and that the lexical aspect of classifiers is about collocational restrictions since not all nouns can combine with the same classifiers (termed the "grammatical/functional" view hereafter) (e.g. Cheng & Sybesma, 1999, 2005), while others argue that classifiers are contentful morphemes used to indicate the semantic classes of nouns—they often carry information beyond that carried by their associated nouns (termed the "semantic" view hereafter) (e.g. Wu & Bodomo, 2009). The discrepancies between the semantic vs. grammatical view arise probably because of the result of the evolution of Mandarin classifiers, as explained below.

Mandarin classifiers are derived from nouns or verbs in older varieties of Mandarin. Along the course of evolution, classifiers started from sharing total (人十人 *ren-shi-ren* 'person-ten-person'), partial (三輛車 *san-liang*$_{CL}$*-che* 'three-CL$_{vehicle}$-car'), to no (一樁謀殺案 *yi-zhuang*$_{CL}$*-mousha-an* "'aCL$_{various\ categories}$-murder-case'", 十個人 *shi-ge*$_{CL}$*-ren* 'ten-GE-person') semantic feature with the accompanying nouns, and eventually only partial- and no-feature-sharing classifiers survive in Modern Mandarin (N. Zhang, 2013). As a result, it is not always intuitive for a contemporary native speaker to explain why a noun can co-occur with a certain classifier but not another. Again, take our article title, "*an elephant needs a head but a horse does not,*" as an example. The selectional constraint looks arbitrary because the classifier for "elephant" is *tou* (e.g. 一頭大象 *yi-tou*$_{CL}$*-daxiang* 'one-CL$_{large\ animal}$-elephant'), which literally means "head" and can still be used as a stand-alone noun, while the proper classifier for "horse" is *pi* (e.g. 一匹馬 *yi-pi*$_{CL}$*-ma* 'one-CL$_{horse}$-horse'), which literally meant "horse" in 1100 BCE but has lost its meaning and can only be used as a morpheme (Erbaugh, 1986). Even though a horse is large and conceptually should be able to co-occur with *tou*, most Mandarin speakers would find *一頭馬 *yi-tou*$_{CL}$*-ma* ('one-CL$_{large\ animal}$-horse') strange and are unable to explain why. Therefore, the loss of feature sharing between a classifier and a noun makes the selectional constraints not entirely transparent in the existent Mandarin classifier system, which in turn causes the discrepancies between the semantic vs. functional view of the nature of classifier-noun pairing mentioned above when linguists try to describe the restrictions from a formal/syntactic perspective.

Although mostly not designed to address this functional vs. semantic issue, event-related potential (ERP) studies using Mandarin classifier materials seem to overwhelmingly support the semantic view. These experiments manipulated incongruity between classifiers and nouns across classifier categories to investigate processing of classifier-noun pairing. For example, a classifier classifying electric appliance was paired with a non-appliance object (*一台信紙 *yi-tai*$_{CL}$*-xinzhi* 'one-CL$_{appliance}$-letter'), or a classifier normally paired with ground vehicle was paired with an appliance (*一輛檯燈 *yi-liang*$_{CL}$*-taideng* 'one-CL$_{ground\ vehicle}$-lamp') (e.g. Y. Zhang, Zhang, & Min, 2012; Zhou et al., 2010). By comparing incongruent classifier-noun sequences with congruent ones, all the researchers claimed that classifier-noun mismatches induced an N400 effect (although depending on the experimental design, the analyzed word could be the noun or the classifier), supporting the view that Mandarin classifiers are semantic elements (Hsu, Tsai, Yang, & Chen, 2014; Qian & Garnsey, 2016; Y.; Zhang et al., 2012; Zhou et al., 2010).

The case of functional vs. semantic status about classifiers would have been closed if it were not for the observation that almost all the N400 effects reported in these studies had an atypical, non-centro-parietal, scalp distribution (see Table 1 for a summary of the ERP findings). For example, the N400 effect induced by the nouns of classifier-noun mismatches in Zhou et al. (2010) and in one of the analyses in Qian and Garnsey (2016) was widespread, with the latter having a fronto-central concentration, and the N400 effect of the classifiers in noun-classifier mismatches in Y. Zhang et al. (2012) was also widespread. Interestingly, with such an atypical scalp distribution, these studies did not consider the possibility of such negativity being or at least containing a LAN (left anterior negativity, although the distribution is not always left lateralized), which shares similar latency with an N400 and is often associated with morphosyntactic/grammatical processing. In contrast, Hsu et al. (2014) did report an anterior negativity (AN) in processing classifier–noun pairs at varying dependency distances in object-gap relative clauses. They found that when the classifier was semantically incongruent with the adjacent noun (but semantically congruent or incongruent with the head noun further down, see Table 1), an effect with a midline anterior focus appeared on the adjacent noun. However, the authors argued against this component being a LAN because (1) the AN was not lateralized to the left, (2) their materials did not have syntactic violation, and (3) the adjacent noun of the mismatched pair did not induce a P600 to form a biphasic LAN-P600 pattern. Instead, the authors argued that the AN might reflect the metacognitive processes in resolving the classifier-noun conflict as well as the working memory demand for keeping the classifier active and anticipating some plausible sequence to appear. Note that typical N400 effects with centro-parietal distribution were indeed reported in two of the above studies in their additional analyses (Hsu et al.: on the head noun away from the classifier-noun mismatch; Qian & Garnsey: on the noun with a different baseline correction). Nevertheless, we believed that the atypical scalp distribution of the negativity merited further investigation to see if there was undetected morphosyntactic processing, which might have been eluded from the authors when interpreting the results with semantic incongruity of the materials in mind.

In fact, some classifier studies, including some of those reviewed above, have observed LAN and/or P600 in mismatch conditions, suggesting possible involvement of morphosyntactic/grammatical processing. Mueller, Hahne, Fujii, and Friederici (2005) created a miniature version of Japanese and manipulated classifier–noun incongruity in the middle of a sentence. A late LAN (500–800 ms), but no N400, was found on the mismatched noun in native Japanese speakers. The authors suggested that the LAN might index an active backward search process when the unexpected noun was encountered in order to complete a "post hoc" agreement checking with the classifier in the front. Y. Zhang et al. (2012) explored processing of animacy information of classifiers in non-canonical sentences (i.e.

Table 1
ERP effects induced by classifier-noun mismatches in past research.

Authors	Materials	(L)AN (ms)	N400 (ms)	P600 (ms)	Reference
Mueller et al. (2005)	*two-**CL**$_{bird}$-GEN **cat**-ACC jump over take place	LAN (500–800)			Left mastoid
Zhou et al. (2010)	*Zhao/repaired/one/**CL**$_{appliance}$/ **chair**	AN (550–800)	Broad N400 (300–500)		Linked mastoids
Y. Zhang et al. (2012)	Congruent **Car**/Qingfeng Zhao/had seen/one **CL**$_{vehicle}$/black Incongruent, animacy matched *__Desk lamp__/Qingfeng Zhao/had seen/one **CL**$_{vehicle}$/cheap Incongruent, animacy unmatched ***Seal**/Qingfeng Zhao/had seen/one **CL**$_{vehicle}$/clumsy		Broad N400 (300–550)	Posterior P600 (600–1000) for animacy unmatched	Linked mastoids
Hsu et al. (2014)	At summer-end autumn-start time, Match-short one-**CL**$_{person}$/**neighbor**/planted/ DE/fruit-tree Match-long one **CL**$_{tree}$/neighbor/planted/DE/ **fruit-tree** Mismatch *one **CL**$_{flower}$/**neighbor**/planted/ DE/**fruit-tree** bear-LE many fruit.	AN (250–450) for adjacent nouns in Mismatch and Match-long	Centro-posterior N400 (250–450) for head nouns in Mismatch and Match-short	Posterior P600 (500–1000) for head noun in Mismatch and Match-long	Average frame
Qian and Garnsey (2016)	*On the table/DE/two **CL**$_{paper}$/ **coffee**/already cold		Broad, fronto-central focus N400 (350–550) with pre-number_classifier baseline Centro-parietal N400 (350–550) with pre-noun baseline		Not reported

a noun followed by a classifier). The noun and its classifier were either congruent, incongruent but matching in animacy, or incongruent but mismatching in animacy. The authors found that although the classifier in both incongruent conditions induced a widespread N400 effect, the mismatch in animacy did not induce a stronger N400, but a P600. The authors argued that the P600 might reflect a continued combinatorial analysis driven by a conflict between the output of the computation of the noun-classifier phrase structure and the output of the animacy processing.

A final evidence hinting at grammatical processing in classifier-noun pairing comes from a functional magnetic resonance imaging (fMRI) study where the comparison of inside-classifier violations (3 sub-conditions of classifier-noun mismatches) with congruent pairs produced greater activation in BA 45 of the left inferior frontal gyrus (Chou, Lee, Hung, & Chen, 2012). Although the authors attributed the activation solely to an increased demand on semantic processing, previous literature has shown that BA45 could also be involved in syntactic processing (A. D. Friederici, 2002; Kaan & Swaab, 2002; Santi & Grodzinsky, 2007).

Taken together, based on the findings in previous literature, including the atypical scalp distribution of the N400 effect, the occurrence of the late LAN and P600, and the activation of BA45 in an fMRI study, we speculated that morphosyntactic processing might be present in processing Mandarin classifier phrases but was ignored probably due to the use of semantic incongruity in the materials and the difficulty of distinguishing LAN from N400 because of their temporal proximity. Thus, the purpose of the current study was to examine whether Mandarin classifiers can play a grammatical role in relation to their accompanying nouns.

To this end, our study took a different approach in designing classifier-noun pairing: instead of adopting apparent semantic clash, we used two frequently found mistakes in young children and aphasic patients (e.g. Tzeng, Chen, & Hung, 1991; Uchida & Imai, 1999): the omission of classifiers and the inappropriate use of the general classifier “GE” (個). The omission mistake ignores the required presence of a classifier and thus is a clear grammatical violation, while the incongruity in the inappropriate use of GE is not as apparent. Although emerged quite late historically and is devoid of semantic content now, GE is the most frequently used classifier in Modern Mandarin, which can co-occur with nouns of various semantic types (e.g. 人 *ren* 'person', 漢堡 *hanbao* 'burger', 國家 *guojia* 'country'", 太陽 *taiyang* 'sun'") (Myers, 2000). Although most speakers know the appropriate specific classifier for a certain noun, they tend to use the general classifier instead in everyday conversation (Ahrens, 1994; Erbaugh, 1986; H.; Zhang, 2007). Interestingly, when children or aphasic patients are aware that a classifier is obligatory but do not know which one to use appropriately, they also often choose the general classifier GE (Ahrens, 1994; Erbaugh, 1986; Tzeng et al., 1991; Uchida & Imai, 1999). However, the replacement of a classifier by GE is not always possible: the more semantic features are shared between a classifier and a noun, the less likely the classifier can be replaced with GE (Ahrens, 1994). We thus made use of such “bad replacements” of GE to manipulate materials in the current study so that when compared with a congruent classifier-noun (CL-N) pair, an incongruent GE-noun (GE-N) sequence did not cause apparent semantic clash as in previous studies.

We conducted two ERP experiments with phrases and sentences as materials, respectively. Our hypothesis was straightforward. Missing a classifier in a classifier phrase is an outright violation of grammar, so the induced effect should be grammatical in nature. If replacing a proper classifier with GE also induces an effect similar to that triggered by the omission of a classifier, we can conclude that the brain response to an inappropriate GE is also grammatical in nature and further verify that successful classifier-noun processing requires morphosyntactic/grammatical agreement in addition to semantic processing found in previous research.

2. Experiment 1

2.1. Methods

2.1.1. Subjects

Twenty-one neurologically healthy, right-handed subjects (Mean age = 21.6, SD = 2.1, 14 Male) were recruited after giving informed written consent. Their right-handedness was determined by a simplified version of the Edinburgh handedness inventory (Oldfield, 1971). They were all native Mandarin speakers in Taiwan and were screened to rule out any history of neurological disorders that would interfere with language functioning. The research protocol was approved by the Research Ethics Office of National Taiwan University. Participants were paid as compensation for their time.

2.1.2. Materials

The stimuli included three types of classifier phrases: Congruent (Num-CL-N), Incongruent (Num-GE-N) and Missing (Num-Ø-N). First, classifiers were selected from *Mandarin Daily Dictionary of Chinese Classifiers* (C. Huang, , Chen, , & Lai, 1997), which was compiled based on the Academia Sinica Balanced Corpus of Modern Mandarin Chinese (http://asbc.iis.sinica.edu.tw/), a grammatically tagged and balanced corpus containing 5 million words collected from media in Taiwan. The dictionary divided 427 classifiers and measure words into 7 categories (individual classifiers, event classifiers, kind classifiers, proximation measure words, container measure words, standard measure words and activity measure words) and only individual classifiers, used to classify individual objects (Gao & Malt, 2009), were selected in the experiment. The candidate classifiers were then paired with 2-character nouns. Since the noun had to be congruent with a specific classifier but incongruent with GE, the congruity between classifiers and nouns and the incongruity between GE and nouns were evaluated via the collocation information provided by Chinese Word Sketch (C.-R. Huang et al., 2005), a web-based program with input of 1.4 billion words from Linguistic Data Consortium Chinese Gigaword (C.-R. Huang, 2009). The incongruent GE-N combinations were created based on the ratio of the GE-N collocation frequency to the selected CL-N collocation frequency (i.e. GE-N/CL-N). If the ratio was below 10%, the GE-N combination was considered incongruent. For example, the collocation frequency for “GE-ticket” and “CL_{zhang}-ticket (張 CL_{zhang}, being the most frequently used classifier with “ticket”, classifies two-dimensional objects)” was 3 and 924, respectively, and thus the ratio of GE-ticket/CL_{zhang}-ticket was 0.3% and “GE-ticket" was considered incongruent. Since the collocation information of the search engine was not always precise (e.g. sometimes a classifier does not pair with the searched noun, but a noun that is farther away), each CL/GE-N pairing was manually examined by five linguistics graduate students to ensure the accuracy of their collocation frequencies. In the end, 54 classifiers were chosen, which were repeated twice to pair with 108 nouns. The average frequency of the CL-N collocations (classifiers adjacent to the left of nouns) was 1435 (3–685795, SD = 3245) and the average GE-N/CL-N ratio (GE adjacent to the left of nouns) was 2% (0–10%, SD = 2.8%). It is worth noting that in real language use, the actual GE-N/CL-N ratio should be even lower (and thus, less acceptable for users) because many nouns could co-occur with more than one classifier (e.g. "candle" can co-occur with 根 *gen*$_{CL}$ or 支 *zhi*$_{CL}$), thus making the denominator of the ratio larger.

To complete the creation of grammatical classifier phrases, numerals ranging from one to nine (in Chinese characters) were added and repeated 12 times to pair with the established CL-N combinations to create 108 Num-CL-N phrases. The numerals were adjusted to ensure that the final classifier phrase sounds natural (e.g. 六段對話 "six-*duan*$_{CL}$-conversation" was a bit strange but一段對話 'one-*duan*$_{CL}$-conversation' was fine). While the Num-CL-N combinations were being constructed, the ungrammatical condition was also created by deleting the classifier from the Num-CL-N structure (i.e. Num-Ø-N). Only when the resulting Num-Ø-N structures did not form a legitimate compound word could the adjustment of numerals be finalized. After all the Num-CL-N and Num-Ø-N combinations were finalized, the same numerals were added to the corresponding GE-N combinations to form Num-GE-N phrases. In the end, there were 108 Num-CL-N (Congruent), 108 Num-GE-N (GE) and 108 Num-Ø-N (Missing) classifier phrases.

For stimulus presentation, each classifier phrase was divided into two frames. For Congruent and GE, the Num and CL/GE were presented together (i.e. 2 characters) followed by the 2-character N; for Missing, the Num (i.e. 1 character) was presented alone followed by the 2-character N. Each trial was thus 3–4 characters long. To prevent subjects from knowing the purpose of this experiment, 108 fillers were constructed, which were composed of 72 grammatical and 36 ungrammatical phrases, including 3- or 4-character compound nouns or verb phrases and 4-character idioms. In particular, to offset the tendency that all the experimental trials with GE were incongruent, 18 grammatical "Verb-GE-N" fillers, where GE had an adverbial use, were added (e.g. 買個 / 便當 *mai* GE/*biandang* ('buy GE/lunchbox', meaning buying a lunchbox in a casual or trivial manner)). Also, to prevent subjects from raising a red flag whenever a Num appeared alone in the ungrammatical "Num-Ø-N″ trials or from thinking that numerals were always followed by classifiers, 18 grammatical phrases/idioms with numeral initials were added (e.g. 四/伯父*si/bofu* ('four/uncle', meaning Father's fourth elder brother); 七上 / 八下 *qishang/baxia* ('7-up/8-down', meaning being agitated)) so that Nums could stand alone and were not always followed by a CL/GE.

In sum, the materials included 108 congruent phrases, 108 incongruent phrases with GE, 108 ungrammatical phrases with the omission of classifiers, and 108 fillers. All the experimental stimuli were identical across the three conditions except for the critical use/omission of CL/GE to ensure that the word frequency and concreteness of the nouns were perfectly matched across conditions. The experimental materials were split into three lists with 108 trials each so that each subject saw all the three conditions but only each classifier-noun pairing once (36 Congruent, 36 GE and 36 Missing trials). With the 108 fillers, each subject saw a total of 216 trials, with half of the trials grammatical/congruent and the other half ungrammatical/incongruent. See Table 2 for examples of the experimental materials.

2.1.3. Procedure

The experiment was carried out in a dimly lit and sound-attenuated chamber. Subjects were seated in a comfortable chair approximately 100 cm away from the computer screen. The stimulus presentation was controlled by E-Prime 2.0 (Psychology Software Tools, Inc.). Each trial started with a central fixation cross lasting for 500 ms, followed by a critical phrase. The critical phrase was divided into two frames, each containing 1–2 Chinese characters in the center of the screen (white text on black background; 2.1–2.3 degrees of horizontal and 0.6–0.9 degrees of vertical visual angle) and lasted 600 ms. The first frame displayed a Num and a CL/GE for the Congruent/GE condition and a Num only for the Missing condition, followed by the 2nd frame displaying a 2-character N. There was a 200-ms blank between the two frames. Following the 2nd frame was a response prompt (a question mark "?") reminding subjects to judge whether the presented stimuli were a correct combination by pressing one of two buttons with their

Table 2
Example materials in Experiments 1 and 2. Conditions include congruent use of classifiers (Congruent), incongruent use of GE (GE), and ungrammatical omission of classifiers (Missing). Slashes (/) indicate display frames on the computer screen. Num = numeral, CL = classifier, N = noun.

Condition	Experiment 1	Experiment 2
Congruent (Num-CL-N)	八張/門票 **ba-zhang$_{CL}$/menpiao** eight-CL/ticket '8 tickets'	導遊/到/售票亭/購買/八張/門票 Daoyou/dao/shoupiaoting/goumai/**ba-zhang$_{CL}$/menpiao** tour.guide/go/ticket.booth/buy/eight-CL/ticket 'The tour guide went to the ticket booth to buy 8 tickets.'
GE (Num-GE-N)	八個/門票 **ba-GE/menpiao** eight-GE/ticket	導遊/到/售票亭/購買/八個/門票 Daoyou/dao/shoupiaoting/goumai/**ba-GE/menpiao** tour.guide/go/ticket.booth/buy/eight-GE/ticket
Missing (Num-Ø-N)	八/門票 **ba/menpiao** eight/ticket	導遊/到/售票亭/購買/八/門票 Daoyou/dao/shoupiaoting/goumai/**ba/menpiao** tour.guide/go/ticket.booth/buy/eight/ticket

left or right index finger. The hand use was counterbalanced across subjects. Once a response was made, the question mark was erased and the screen was replaced by a 1200-ms blank. Prior to the experimental session, subjects completed a practice session until reliable performance was demonstrated. There were a total of 216 trials, and subjects took a break after every 54 trials, which amounted to 3 breaks during the experiment. The experimental session took about 40 min.

2.1.4. EEG recording and analysis

The electroencephalogram (EEG) was recorded from 32 Ag/AgCl electrodes mounted on an electrode cap according to the international 10–20 system. To detect eye movements and blinks, the electrooculogram (EOG) was recorded from electrodes placed below and above the left eye (VEOU and VEOL) and at the external canthi of the two eyes (HEOL and HEOR). Scalp EEGs were referenced to an average of the right and left mastoids for both online recording (NuAmps, Neuroscan, Inc.) and off-line analysis. The EEG was continuously recorded, filtered between DC to 100 Hz and digitized online with a sampling rate of 1000 Hz. The amplifier rate (Gain) was 19, corresponding to an input range of ± 131.5 mV. Electrode impedances remained below 5 kΩ.

Signal processing and analysis were performed in Matlab (Mathwork Inc.) using the EEGLAB toolbox (Delorme & Makeig, 2004) and ERPLAB toolbox (Lopez-Calderon & Luck, 2014). Data that contained large muscle artefacts or extreme movements with visual inspection were first manually removed and the remaining data were filtered with an Infinite impulse response (IIR) Butterworth high-pass filter (half-amplitude cutoff = 0.1 Hz, slope = 12 dB/octave). Independent component analysis (ICA) was performed on the continuous data to identify and remove components associated with eye blinks, eye movements, and excessive non-physiological environmental noise (Jung et al., 2000). To avoid a potential baseline problem caused by different brain response to the pre-noun component position of the three conditions (Congruent: Num-CL; GE: Num-GE; Missing: Num-Ø) (Kwon, Sturt, & Liu, 2017; Osterhout, McLaughlin, Kim, Greenwald, & Inoue, 2004; Steinhauer & Drury, 2012), the ICA-corrected EEG data were epoched from −1000 ms before and 600 ms after the head noun onset (i.e. starting from 200 ms before the pre-noun component position until the end of the head noun presentation) and baseline corrected by subtracting the average voltage of the 200 ms period before the pre-noun component position(i.e. −1000 – −800 ms). Four monopolar EOG channels were then transformed into 2 bipolar ones (VEOU and VEOL into VEOG; HEOL and HEOR into HEOG) to carry out the following automatic artefact rejection procedure. First, at each channel, a 200-ms window was moved across the data of the head noun (200 ms before and 600 ms after the stimulus) in 50-ms increments and any epoch where the peak-to-peak voltage exceeded 150 μV was rejected. A step function was then applied to the HEOG, with a window of 400 ms (200 ms before and 200 ms after the voltage transition) moving across the data of the head noun (200 ms before and 600 ms after the stimulus) in 10-ms steps, and epochs were rejected if they contained changes greater than 25 μV. The overall acceptance rate was 95.81 ± 3.91% (mean ± SD). The data were then averaged and filtered by an IIR Butterworth low-pass filter (half-amplitude cutoff = 30 Hz, slope = 12 dB/octave) and the grand average for each experimental condition was obtained by averaging all the participants' averaged ERP data.

To statistically test the results, we measured the mean amplitudes between 350 and 500 ms and analyzed the data separately for selected midline (Fz, Cz, Pz) and lateral (F3, F4, C3, C4, P3, P4) electrodes. For the midline electrodes, a 2-way repeated measures ANOVA with factors of Condition (Congruent, GE, Missing) and Electrode (Fz, Cz, Pz) was conducted. For the lateral electrodes, a 3-way repeated measures ANOVA with factors of Condition (Congruent, GE, Missing), Region (anterior, central, posterior) and Hemisphere (left, right) was performed. When the Mauchly's sphericity test was violated, the Greenhouse–Geisser correction was applied to compensate for inhomogeneous variances and covariances across treatment levels, with the original degrees of freedom and the adjusted probability levels reported. Follow-up comparisons were carried out and Bonferroni corrected only when an ANOVA revealed a significant effect involving the Condition factor.

2.2. Results

2.2.1. Behavioral data

The accuracy rates were 93.7 ± 5.0% (mean ± SD) for Congruent, 83.2 ± 11.7% for GE, and 98.6 ± 2.8% for Missing, showing that subjects were paying attention when performing the experimental task. To verify that the poor performance for the GE condition was not simply due to the appearance of GE, we also analyzed the 18 grammatical "Verb-GE-N" fillers, where GE had an adverbial use, and found out that the accuracy rate for the fillers was 93.7 ± 3.7%. A repeated measures analysis of variance (ANOVA) on the experimental trials with the factor of Condition (Congruent, GE, Missing) revealed a main effect [$F(2,40) = 21.08$, $p < .0001$]. Follow-up t-tests showed that the accuracy rate was highest in Missing, followed by Congruent, and then GE ($D_{(Missing\text{-}Congruent)} = .05$, $p < .01$; $D_{(Missing\text{-}GE)} = .15$, $p < .0001$; $D_{(Congruent\text{-}GE)} = .11$, $p < .01$).

The reaction times (RTs, note that subjects' response was delayed 600 ms after the onset of the noun) were 626.6 ± 311.2 (mean ± SD) ms for Congruent, 697.4 ± 377.6 ms for GE, and 508.1 ± 236.5 ms for Missing. A repeated measures ANOVA revealed a main effect of Condition [$F(2,40) = 11.67$, $p < .0005$]. Follow-up t-tests indicated that the RT for Missing was significantly shorter than that for Congruent and GE ($D_{(Missing\text{-}Congruent)} = -118.52$, $p < .005$; $D_{(Missing\text{-}GE)} = -189.35$, $p < .005$), and that the RTs for the latter two conditions did not differ ($D_{(Congruent\text{-}GE)} = -70.8$, $p = .34$).

In summary, the behavioral data showed that the Missing condition was the easiest for the subjects to evaluate (highest accuracy rate, shortest RT), while the GE condition was the hardest (lowest accuracy rate among the three conditions, although it did not take the subjects longer to respond compared with Congruent).

2.2.2. ERP data

The grand average ERP responses to the classifier phrase, including the pre-noun component (Num-CL, Num-GE, Num-Ø) and the

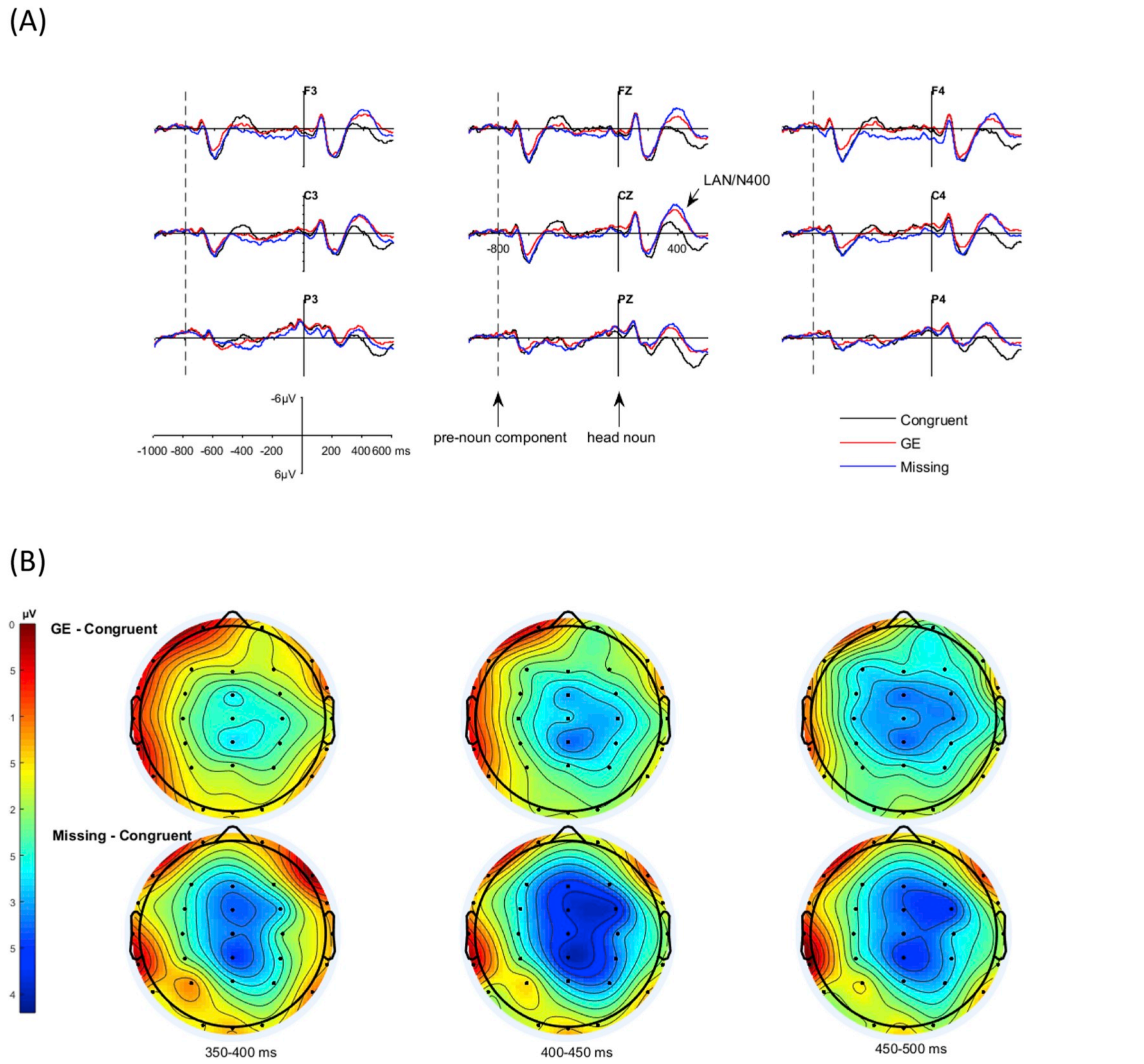

Fig. 1. Brain response in Experiment 1. (A) Grand average ERPs of the Congruent (black), GE (red) and Missing (blue) conditions in Experiment 1 at 9 representative electrode sites. Negative is plotted up. The epoch is synchronized to the onset of the head noun. The dashed line indicates the onset of the pre-noun component, including Num-CL (Congruent), Num-GE (GE) and Num-Ø (Missing). (B) Differential topographic maps of response to the head noun between 350 and 500 ms in Experiment 1, averaged over a 50-ms time window. Upper panel: GE – Congruent; lower panel: Missing – Congruent. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

head noun, and the topographic maps of the difference waves of "GE – Congruent" and "Missing – Congruent" for the head noun are presented in Fig. 1. As Fig. 1A shows, the brain responses to the three conditions at the pre-noun position are already different, confirming that using the 200 ms period before the pre-noun component as the baseline when analyzing the head noun data was legitimate. In response to the noun, there are visually-evoked N1–P2 components in all the three conditions, followed by a wide-spread negative-going brainwave, which seems stronger for Missing and GE than for Congruent. Fig. 1B further demonstrates that the negativity effects of Missing and GE have similar scalp distribution. To statistically test if the negativity effects for Missing and GE were significant and if their distributions were similar, we measured the mean amplitudes between 350 and 500 ms of the response to the head noun and analyzed the data separately for selected midline and lateral electrodes. A 2-way repeated measures ANOVA on the midline electrodes with factors of Condition (Congruent, GE, Missing) and Electrode (Fz, Cz, Pz) revealed a strong Condition effect (F (2,40) = 5.85, $p < .01$). Follow-up t-tests revealed that both Missing and GE induced stronger negativity than Congruent ($D_{(Missing\text{-}Congruent)} = -3.18$, $p < .05$; $D_{(GE\text{-}Congruent)} = -2.42$, $p < .05$) and the negativities were equally strong for Missing and GE ($D_{(Missing\text{-}GE)} = -.76$, $p = 1$). Similarly, a 3-way repeated measures ANOVA on the lateral electrodes with factors of Condition (Congruent, GE, Missing), Region (anterior, central, posterior) and Hemisphere (left, right) also revealed a strong Condition main effect (F (2,40) = 5.74, $p < .01$). Follow-up t-tests showed that both Missing and GE induced stronger negativity than Congruent ($D_{(Missing\text{-}Congruent)} = -2.40$, $p < .05$; $D_{(GE\text{-}Congruent)} = -2.31$, $p < .01$) and that there was no difference between Missing and GE ($D_{(Missing\text{-}GE)} = -.09$, $p = 1$). The Condition factor did not interact with other factors in the midline and lateral analyses. See Table 3

Table 3
ANOVAs on amplitudes over midline and lateral electrodes during 350–500 ms in Experiment 1.

Source	*dfs*	350–500	
		F	*p*
Midline electrodes			
Condition	2,40	5.85	< .01
Condition x Electrode	4,80	.55	ns
Lateral electrodes			
Condition	2,40	5.74	< .01
Condition x Region	4,80	.78	ns
Condition x Hemisphere	2,40	1.23	ns
Condition x Region x Hemisphere	4,80	.99	ns

for a summary of the ANOVA results.

In sum, both Missing and GE induced widespread negativity effects during 350–500 ms, which were equally strong in the two conditions and did not differ from each other in scalp distribution.

2.3. Discussion

We conducted Experiment 1 to explore whether morphosyntactic processing was involved in understanding classifier-noun pairing in Mandarin. Instead of using a semantic mismatch paradigm as in previous ERP studies (Hsu et al., 2014; Qian & Garnsey, 2016; Y.; Zhang et al., 2012; Zhou et al., 2010), we made use of the semantically devoid general classifier, GE, to create incongruent GE-N combinations (the GE condition) and compared their ERP pattern with that induced by ungrammatical classifier phrases missing a necessary classifier (the Missing condition). Our results revealed that, compared with a correct classifier-noun sequence (the Congruent condition), both GE and Missing induced a starkly similar negativity effect during 350–500 ms, both in amplitude and in scalp distribution. Although it is difficult to discern whether this negativity effect was an N400 or a LAN due to its widespread distribution, the similar responses in GE and Missing suggested that grammatical/morphosyntactic processing was involved in understanding classifier-noun pairing. Since previous classifier research used sentential materials, we conducted Experiment 2 to investigate if the finding of morphosyntactic agreement at the phrase level could be replicated in sentences.

3. Experiment 2

3.1. Methods

3.1.1. Subjects

The subject selection criteria were the same as those in Experiment 1. Twenty neurologically healthy, right-handed subjects (Mean age = 26.7, SD = 5.9, 10 Male) participated in this study after giving informed consent and none of them participated in Experiment 1. The experimental protocol was approved by the Research Ethics Office of National Taiwan University. Participants were paid as compensation for their time.

3.1.2. Materials

The 108 congruent classifier phrases in Experiment 1 were adopted to create 108 grammatical, declarative sentences in Experiment 2, with the classifier phrase appearing at the sentence-final position. Two minor modifications were made to the classifier phrases. First, some numerals were swapped between stimuli to make the sentences more plausible. For example, 九條河 'nine-*tiao*$_{CL}$-river' in Experiment 1 was changed to 三條河 'three-*tiao*$_{CL}$-river' in “The environmental protection group has been devoted to protecting the **three rivers** in the south (環保團體致力於保護南部三條河流)” in Experiment 2 since the number of large rivers was less than nine in the south of Taiwan, which is where the experiment took place. Also, while most of the CL-N pairings remained the same as those in Experiment 1, two of them were changed to make the sentence sound more natural. For example, the use of 位 *wei*$_{CL}$ in “*wei*$_{CL}$-government official” was perfectly fine in isolation, but it was more natural to use the classifier 名 *ming*$_{CL}$ in the following context, “The prosecutor interviewed the **three officials** who were involved in this case (檢方約談涉案的三名官員)” because the use of *wei*$_{CL}$ is usually associated with some degree of respect for the following noun and thus sound unnatural in this sentence where negative connation was given to the officials. The two changed pairings thus resulted in 2 of the 54 classifiers being repeated three times and 2 appearing only once (instead of repeating twice as in Experiment 1) and the frequency of the CL-N collocation (classifiers adjacent to the left of the nouns) of the material was 1401 (3–685795, SD = 3144) (the frequency was 1435 in Experiment 1).

For the stimulus presentation, each sentence was divided into 4–7 frames, with each frame containing a word with 1–3 characters. All the nouns were objects of the preceding verbs in the sentences, except for 5 sentences where the verb was an existential verb and the following noun was a theme. To ensure that the verb-noun co-occurrence (verb adjacent to the noun) was natural, the collocation frequency of the verb in the sentence context and the noun in the classifier phrase was at least 1 in the Chinese Word Sketch (mean = 398, range = 1–19182, SD = 911). After the materials for the Congruent condition were finalized, the GE and Missing conditions were derived by replacing the classifier with GE or by deleting the classifier in the grammatical sentences. Therefore, the ungrammaticality/incongruity of Missing and GE was determined by the final two frames completing the sentences (i.e. the sentence was grammatical before the final segments). To prevent subjects from knowing what this experiment was testing, 108 filler sentences—72 grammatical and 36 ungrammatical—were constructed. To parallel the ungrammaticality/incongruity in Missing and GE, the ungrammaticality of the filler sentences was also determined by the final two frames. The final two frames were fillers from Experiment 1, with six of them modified to avoid repetition with the sentence context of the experimental trials.

In sum, the materials included 108 grammatical sentences with the sentence-final phrase being a correct classifier phrase (Congruent), 108 incongruent sentences with GE in the classifier phrase (GE), 108 ungrammatical sentences with the classifier being omitted from the classifier phrase (Missing), and 108 filler sentences. All the experimental stimuli were identical across the three conditions except for the critical use/omission of CL/GE so that the word frequency or concreteness of the noun and the verb-noun collocation frequency were perfectly matched. All the nouns and verbs (i.e. contentful words) appeared only once in the materials. The experimental materials were split into three lists with 108 trials each so that each subject saw all the three conditions but only each CL-N pairing once (36 Congruent, 36 GE and 36 Missing trials). With the 108 fillers, each subject saw 216 trials in total, with half of the trials grammatical/congruent and the other half ungrammatical/incongruent. Example materials are shown in Table 2.

3.1.3. Procedure

The experiment setting was the same as Experiment 1 except for the following. Each trial started with a central fixation cross lasting for 700 ms. Following the fixation cross, sentence stimuli were presented sequentially in 4–7 frames, each containing 1–3 Chinese characters in the center of the screen (white text on black background; 2.1–2.3 degrees of horizontal and 0.6–0.9 degrees of vertical visual angle) and lasting 400 ms with a 200-ms inter-stimulus interval. The critical classifier phrase was presented in the last two frames: a Num with a CL/GE (Congruent and GE, respectively) or a Num alone (Missing) in the second last frame, and a 2-character N in the last. The final frame was followed by an 800-ms blank and then a response prompt (a downward arrow↓) appeared to remind subjects that it was the end of the sentence and that they had to judge whether the presented sentence was acceptable by pressing one of two buttons with their left or right index finger. The hand use was counterbalanced across subjects. Once a response was made, the arrow was erased and the screen was replaced by a 1500-ms blank. There were a total of 216 trials, and subjects took a break after every 54 trials, which amounted to 3 breaks during the experiment. Prior to the experimental session, subjects completed a practice session until reliable performance was demonstrated. The experimental session took about an hour.

3.1.4. EEG recording and analysis

The EEG recording setup was the same as that in Experiment 1. The signal processing steps were also identical to those in Experiment 1 except that the data in Experiment 2 were epoched from 800 ms before and 1000 ms after the head noun onset (i.e. starting from 200 ms before the pre-noun component position until 1000 ms after the head noun onset). The overall acceptance rate was 91.4 ± 9.2%.

As for data analysis, two sets of midline and lateral analyses for the sentence-final classifier phrases were analyzed: one for the three conditions and the other for the GE condition only. More detail about the analyses will be described in Section 3.2.2.

3.2. Results

3.2.1. Behavioral data

The accuracy rates were 95.6 ± 4.6% for Congruent, 61.2 ± 21.0% for GE, and 91.0 ± 14.3% for Missing. The high accuracy rates for Congruent and Missing showed that the participants were paying attention and understood the task. Note that the accuracy rate for the sentences containing 18 grammatical "Verb-GE-N" fillers, where GE had an adverbial use, was 96.7 ± 3.4%, indicating that subjects did not perform poorly on all the trials with GE, but only on trials in the GE condition. A repeated measures ANOVA on the experimental trials revealed that there was a main effect of Condition [F (2,38) = 30.63, $p < .0001$], with follow-up t-tests showing that the difference was due to the lower accuracy rate for GE than for the other two conditions ($D_{(GE\text{-}Congruent)} = -.34$, $p < .0001$; $D_{(GE\text{-}Missing)} = -.30$, $p < .0005$). There was no difference between Missing and Congruent ($D_{(Congruent\text{-}Missing)} = .04$, $p = .55$). The low accuracy rate and the large standard deviation for GE demonstrated that, although the replacement of a classifier by GE was inappropriate for our materials, such incongruity was more likely to be ignored in the sentential than in the phrasal context (note that the accuracy rate for the same GE-N combinations was 83.2 ± 11.7% in Experiment 1). To find out if the subjects' behavioral response was in line with our observation of the corpus data when constructing the experimental materials (see Section 2.1.2 for the creation of the GE-N combinations), we conducted a Pearson's correlation analysis on the 108 GE-N items (3 experimental lists with 36 GE trials each) and found a strong negative correlation between the error rates and the GE-N/CL-N ratios (i.e. the lower the frequency of the occurrence of GE-N in the corpus, the higher the error rate in the behavioral response, $r = .41$, $p < .0001$), showing that subjects' behavioral response indeed accorded with the corpus data.

The RTs (note that subjects' response was delayed 1200 ms after the onset of the critical noun) were 612 ± 369 ms for Congruent, 667 ± 484 ms for Missing, and 775 ± 487 ms for GE. A repeated measures ANOVA revealed a main effect of Condition (F (2,38) = 5.94, $p < .01$). Follow-up t-tests indicated that GE was responded to more slowly than Congruent and Missing ($D_{(Congruent\text{-}GE)} = -178.36$, $p < .05$; $D_{(Missing\text{-}GE)} = -133.01$, $p = .04$), and that there was no RT difference between Congruent and Missing ($D_{(Congruent\text{-}Missing)} = -47.36$, $p = 1$).

In sum, the GE condition was more difficult for the subjects to process than Missing and Congruent (lowest accuracy rate and longest RT for GE), while Missing and Congruent were equally easy.

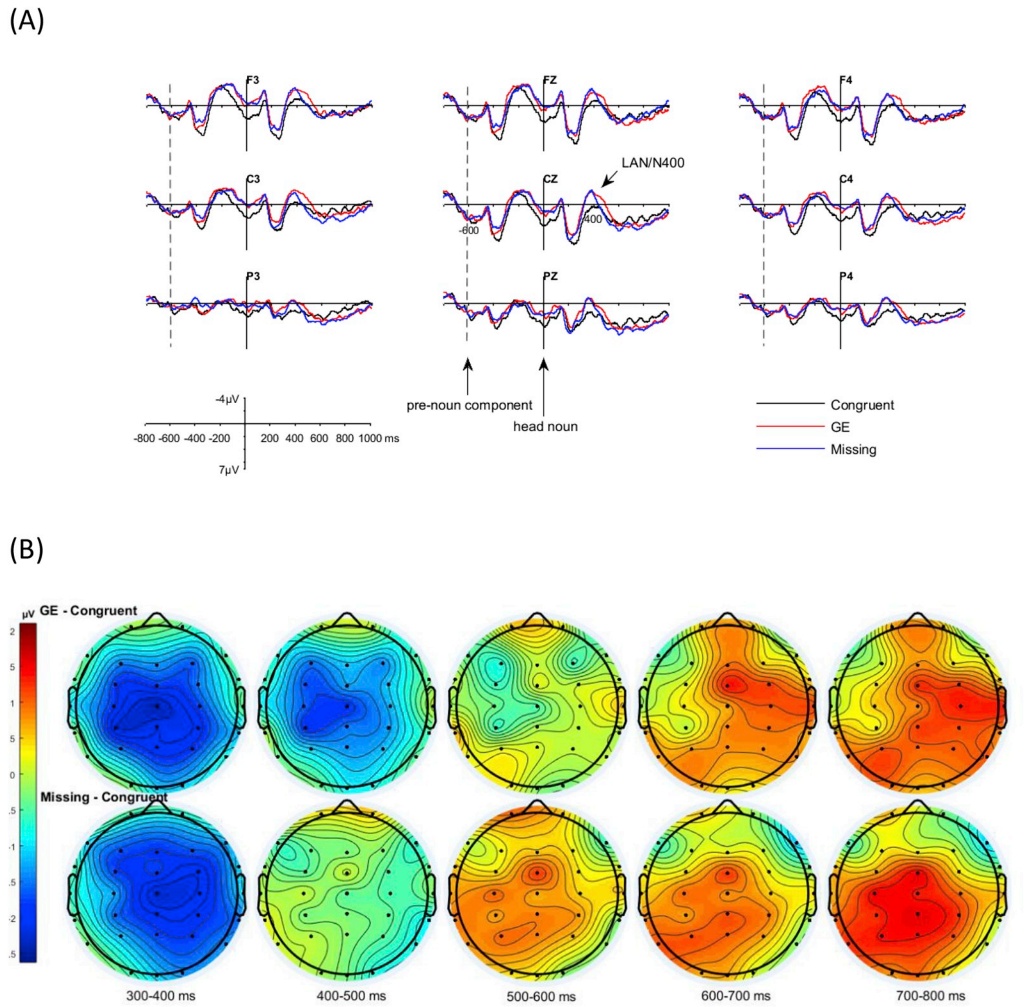

Fig. 2. Brain response in Experiment 2. (A) Grand average ERPs of the Congruent (black), GE (red) and Missing (blue) conditions in Experiment 2 at 9 representative electrode sites. Negative is plotted up. The epoch is synchronized to the onset of the head noun. The dashed line indicates the onset of the pre-noun component, including Num-CL (Congruent), Num-GE (GE) and Num-Ø (Missing). (B) Differential topographic maps of response to the head noun between 300 and 800 ms in Experiment 2, averaged over a 100-ms time window. Upper panel: GE – Congruent; lower panel: Missing – Congruent. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

3.2.2. ERP data

The grand average ERP responses to the sentence-final classifier phrase, including the pre-noun component (Num-CL, Num-GE, Num-Ø) and the head noun, and the topographic maps of the difference waves of "GE – Congruent" and "Missing – Congruent" for the head noun are presented in Fig. 2. As Fig. 2A shows, the brain responses to the pre-noun components are different, reassuring that baselining at the 200 ms period before the pre-noun component for the head noun analysis was appropriate. At the head noun position, there are visually-evoked N1–P2 components in all the three conditions, followed by a widespread negative-going brainwave, which seems to be stronger for Missing and GE than for Congruent. After passing the 400-ms mark, the brainwaves for Missing and GE start turning positive-going, but the transition for Missing seems to present a steeper slope than that for GE. At around 600 ms, the brainwaves for Missing and GE seem to become more positive than that for Congruent at the centroposterior sites (possibly P600) and the brainwaves for the three conditions start to merge toward the end of the epoch. Fig. 2B better visualizes the transition from negative to positive for Missing and GE: the two conditions present similar negativity effects between 300 and 400 ms, but the similarity starts to diminish during 400–600 ms and then reappears during 600–800 ms with the emergence of the positivity.

Since the accuracy rate for the GE condition was only 61.2 ± 21.0%, it would be informative to find out if the brainwave response varied with the subjects' rejection or acceptance of the incongruent GE-N combinations. To investigate this, we split the trials in the GE condition into categories of "correct rejection" (i.e. the subjects correctly rejected the combinations) and "incorrect acceptance" (i.e. the subjects incorrectly accepted the combinations) and plotted their respective ERP response in Fig. 3. As Fig. 3 shows, the head nouns of the two categories elicit brainwaves with a similar profile: N1–P2 components followed by a widespread negative-going brainwave, which seems slightly stronger in the correct than in the incorrect trials. Both brainwaves then become

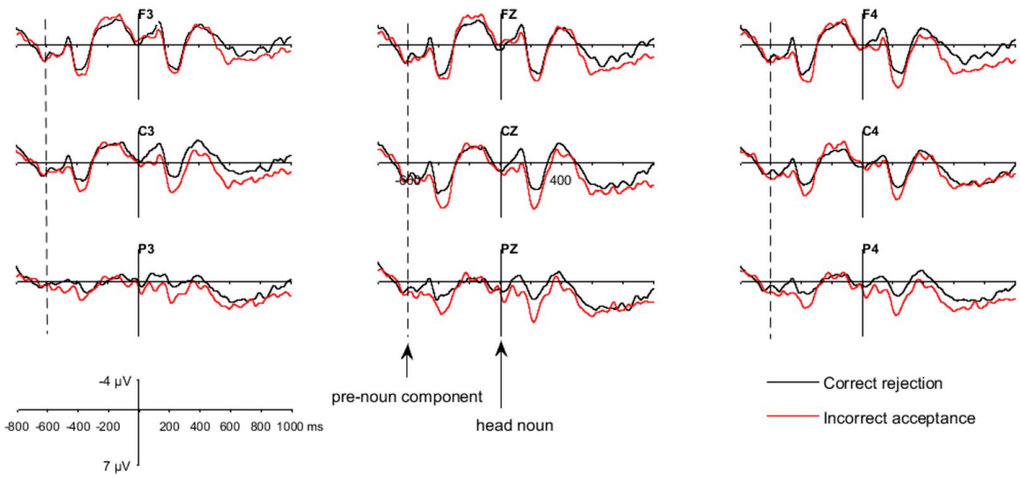

Fig. 3. Grand average ERPs of the correct rejection (black) and incorrect acceptance (red) of the GE condition in Experiment 2. Data are retrieved from 9 representative electrode sites, low-pass filtered at 25 Hz for display. Negative is plotted up. The epoch is synchronized to the onset of the head noun. The dashed line indicates the onset of the pre-noun component, Num-GE. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

positive-going and merge at later time points in posterior electrode sites.

To statistically verify the findings of visual inspection described above, we conducted two sets of midline and lateral analyses for the sentence-final classifier phrases. The first set of analysis was to examine the observation of and the transition from the negativity to the positivity in response to the sentence-final classifier phrase. We measured the mean amplitudes of the critical head noun during 350–500 ms for the negativity and during 600–800 ms for the positivity. However, as visual inspection revealed that the negativity patterns for Missing and GE seem to be more similar during 300–400 ms before their transition to positivity starts, we also shifted the analysis time window to 50 ms earlier for another analysis. As a result, we measured the amplitudes between 350 and 500 ms and 300–450, in that order, to capture the negativity effect for Missing and GE. As in Experiment 1, we submitted the ERP data, including those from the time windows of 300–450, 350–500 and 600–800 ms, to repeated measures ANOVAs for selected midline and lateral electrodes (see Section 2.1.4 for detail of the ANOVAs). The second set of analysis was to verify the similarity between the brainwaves induced by the "correct rejection" and "incorrect acceptance" trials in the GE condition. We measured the mean amplitudes of the ERP response to the critical head noun during 350–500 ms for a negativity and during 600–800 ms for a positivity on selected midline and lateral electrodes and submitted the data for further analyses. For the midline electrodes, a 2-way repeated measures ANOVA with factors of Acceptability (Correct rejection, Incorrect acceptance) and Electrode (Fz, Cz, Pz) was conducted. For the lateral electrodes, a 3-way repeated measures ANOVA with factors of Acceptability (Correct rejection, Incorrect acceptance), Region (anterior, central, posterior) and Hemisphere (left, right) was performed.

3.2.2.1. The classifier phrase: congruent, GE and missing. Negativity. To test the observed negativity effect of the critical head noun, we first analyzed the mean amplitudes between 350 and 500 ms for the Congruent, GE and Missing conditions and submitted the data for midline and lateral analyses. The midline analysis showed a main effect of Condition (F (2,38) = 3.73, $p < .05$). Follow-up t-tests indicated that the main effect was mainly driven by the GE condition ($D_{(GE\text{-}Congruent)} = -1.66$, $p < .05$), not Missing ($D_{(Missing\text{-}Congruent)} = -.94$, $p = .46$). Similarly, the lateral analysis revealed a Condition main effect (F (2,38) = 4.27, $p < .05$) and follow-up comparisons also showed that the effect was attributed to GE only ($D_{(GE\text{-}Congruent)} = -1.47$, $p < .01$; $D_{(Missing\text{-}Congruent)} = -.78$, $p = .49$). We then shifted the analysis time window into an earlier latency range, 300–450 ms. Both the midline and lateral analyses at this latency range revealed a strong Condition effect (Midline: F (2,38) = 6.06, $p < .01$; Lateral: F (2,38) = 6.83, $p < .005$) and follow-up comparisons showed that both Missing and GE induced a negativity effect and that both effects were equally strong (Midline: $D_{(Missing\text{-}Congruent)} = -1.58$, $p < .05$; $D_{(GE\text{-}Congruent)} = -1.79$, $p < .01$; $D_{(Missing\text{-}GE)} = .21$, $p = 1$; Lateral: $D_{(Missing\text{-}Congruent)} = -1.31$, $p < .05$; $D_{(GE\text{-}Congruent)} = -1.57$, $p < .005$; $D_{(Missing\text{-}GE)} = .26$, $p = 1$). Taken together, the results indicated that Missing and GE induced equally strong negativity effects during 300–450 ms, but the effect for GE was more sustained than that for Missing, continuing into the 350–500 ms period.

Positivity. The analysis of the mean amplitudes between 600 and 800 ms was to test if the observed positivity (possibly P600) for Missing and GE at the centroposterior sites could be statistically verified. Although visual inspection revealed such a tendency, both the midline and lateral analyses failed to show a main effect of Condition (Midline: F (2,38) = 1.63, $p = .21$; Lateral: F (2,38) = 1.58, $p = .22$) or any interaction involving the Condition factor (Midline: Condition x Electrode: F (4,76) = 1.72, $p = .16$; Lateral: Condition x Region: F (4,76) = 1.74, $p = .17$; Condition x Hemisphere: F (2,38) = 2.45, $p = .10$; Condition x Region x Hemisphere: F (4,76) = .84, $p = .51$). Therefore, the observed positivity was not strong enough to be statistically significant. See Table 4 for a summary of the statistical results of the Congruent, GE and Missing conditions.

Table 4
ANOVAs on amplitudes over midline and lateral electrodes in 3 time windows (in milliseconds) for the Congruent, GE and Missing conditions in Experiment 2.

Source	*dfs*	300–450		350–500		600–800	
		F	*p*	*F*	*p*	*F*	*p*
Midline electrodes							
Condition	2,38	6.06	< .01	3.73	< .05	1.63	ns
Condition x Electrode	4,76	1.27	ns	.82	ns	1.72	ns
Lateral electrodes							
Condition	2,38	6.83	< .005	4.27	< .05	1.58	ns
Condition x Region	4,76	.11	ns	.09	ns	1.74	ns
Condition x Hemisphere	2,38	1.10	ns	.73	ns	2.45	ns
Condition x Region x Hemisphere	4,76	1.05	ns	.90	ns	.84	ns

3.2.2.2. Further analysis of the GE condition. To statistically test if the observed activities induced by the correct rejection and incorrect acceptance of GE trials were indeed similar, we analyzed the mean amplitudes of the critical head noun between 350-500 ms and 600–800 ms and submitted the data to midline and lateral analyses, respectively. For the 350–500 ms time window, neither the midline nor the lateral analysis revealed a main effect of Acceptability (Midline: F (1,19) = .13, p = .72; Lateral: F (1,19) = .48, p = .50) or an interaction involving Acceptability (Midline: Acceptability x Electrode: F (2,38) = 2.06, p = .14; Lateral: Acceptability x Region: F (2,38) = .94, p = .40; Acceptability x Hemisphere: F (1,19) = .04, p = .85; Acceptability x Region x Hemisphere: F (2,38) = 2.91, p = .07). Similarly, for the 600–800 ms time range, there was no main effect of Acceptability (Midline: F (1,19) = .43, p = .52; Lateral: F (1,19) = .80, p = .38) or an interaction involving Acceptability (Midline: Acceptability x Electrode: F (2,38) = .80, p = .46; Lateral: Acceptability x Region: F (2,38) = 1.00, p = .38; Acceptability x Hemisphere: F (1,19) = .03, p = .86; Acceptability x Region x Hemisphere: F (2,38) = 3.12, p = .06). The insignificant results from both time windows suggested that the brainwaves induced by the correct and incorrect GE trials were similar both in amplitude and scalp distribution. See Table 5 for a summary of the ANOVA results.

3.3. Discussion

We conducted Experiment 2 to examine whether the involvement of morphosyntactic processing in understanding isolated classifier phrases in Experiment 1 could also be detected in the sentence context. Indeed, the finding of the morphosyntactic agreement was replicated, as indexed by the negativity effect in Missing and GE, and the statistical results further revealed that the effect for GE was more sustained than that for Missing.

We believe the longer duration of the negativity effect for GE than for Missing may reflect the subjects' latency variability in responding to some, if not all, of the GE-N combinations; after all, Missing is an outright grammatical mistake and was easier to detect/process but GE is not so clear-cut. As stated in the Introduction, GE can co-occur with a variety of nouns and is used much more often than a specific classifier in everyday conversation (Ahrens, 1994; Erbaugh, 1986; H.; Zhang, 2007). As a result, even though the GE-N combinations in our study had low frequency of collocations and were considered incongruent when constructed (see Section 2.1.2 for material creation in Experiment 1), some of the incongruent pairing might not look as "bad" as the omission of a classifier at first glance, especially in the sentence context.

In fact, the finding about the difficulty of the GE condition was also corroborated by its lower accuracy rate than Missing (GE: 61.2 ± 21.0% vs. Missing: 91.0 ± 14.3%). We conducted a correlation analysis for the GE trials and discovered that the error rates were negatively correlated with the frequency of the occurrence of GE-N in the corpus, indicating that the subjects' behavioral response was sensitive to the actual use of GE in real life. However, while splitting the GE trials into "correct rejection" and "incorrect

Table 5
ANOVAs on amplitudes over midline and lateral electrodes for the "correct rejection" and "incorrect acceptance" of trials in the GE condition in Experiment 2.

Source	*dfs*	350–500		600–800	
		F	*p*	*F*	*p*
Midline electrodes					
Acceptability	1,19	.13	ns	.43	ns
Acceptability x Electrode	2,38	2.06	ns	.80	ns
Lateral electrodes					
Acceptability	1,19	.48	ns	.80	ns
Acceptability x Region	2,38	.94	ns	1.00	ns
Acceptability x Hemisphere	1,19	.04	ns	.03	ns
Acceptability x Region x Hemisphere	2,38	2.91	ns	3.12	ns

acceptance" to examine if the ERP response differed between these two categories, we were surprised to discover that there was virtually no amplitude or scalp distribution difference between them, despite the polarized behavioral response (rejection vs. acceptance). This interesting contrast revealed that the behavioral response, which was delayed 1200 ms after the critical noun onset and might have involved controlled, strategic processes, was possibly too metalinguistic to be indexed by the immediate brain response. While the behavioral judgment was sensitive to the actual GE-N ratio in everyday life and demonstrated the subjects' tolerance/awareness of linguistic variations in the speech community, the automatic brain response actually left little room for the acceptance of incongruent combinations.

In sum, the analysis of the negativity effect seemed to suggest that, although not as "sharp" as detecting trials in the Missing condition (as shown in the more sustained brainwaves for GE), the subjects' automatic ERP response was still more "categorical" than their metalinguistic judgment in detecting the incongruent GE-N combinations. Importantly, the divergence between the ERP and the behavioral responses in the GE condition demonstrated how transient and elusive the morphosyntactic processing could be in studying classifier-noun agreement.

In addition to the negativity effect, Fig. 2 also showed a slight centroposterior positivity for Missing and GE, but the pattern did not reach statistical significance. In fact, even after splitting the GE trials into correct rejection and incorrect acceptance groups, we still failed to observe a positivity effect for the correct rejection trials. We suspected that the absence of a P600 might be because not all subjects reanalyzed the sentence when encountering the ungrammatical/incongruent classifier phrase. In fact, the P600 effect was not consistently observed in previous classifier studies using semantic clash materials. We believe the inconsistency of the P600 effect may have to do with the complexity of experimental materials in different studies. Research reporting a P600 effect used complex structures: Y. Zhang et al. (2012) placed noun-classifier mismatches in a non-canonical sentence structure and Hsu et al. (2014) inserted an object-gap relative clause between a classifier and a noun. In contrast, Zhou et al. (2010), which did not report a P600 for classifier-noun mismatch, manipulated incongruity to various degree and thus the single mismatch of classifier-noun pairs were easier to process than double (verb-noun mismatch plus classifier-noun mismatch) and triple mismatches (classifier–noun, verb–noun and verb-classifier mismatch). Since posterior P600 is often related to reanalysis (see Molinaro, Barber, & Carreiras, 2011 for a review), we think that the presence of P600 in previous classifier studies is related to the more complex materials because subjects may tend to reanalyze the processed sentence so that the problematic stimulus can be properly integrated. Since our sentential materials were not as complex as those in Y. Zhang et al. (2012) and Hsu et al. (2014) (i.e. non-canonical word order, relative clauses), reanalysis might not invariably occur for all the subjects, resulting in the absence of a P600 effect.

Taken together, the equal strength of the amplitude as well as the similar scalp distribution of the negativity effects induced by GE and Missing in Experiment 2 demonstrated that processing GE-N incongruity was similar to processing ungrammaticality of Missing, suggesting that grammatical/morphosyntactic agreement was involved in understanding classifier phrases in the sentence context.

4. General discussion

The current study set out to examine whether Mandarin classifier-noun agreement can be grammatical in nature, hoping to contribute to the long semantic vs. grammatical debate in linguistics (Cheng & Sybesma, 1999, 2005; Wu & Bodomo, 2009). Previous ERP research, although not specifically designed to test this issue, consistently supported the semantic view (Hsu et al., 2014; Qian & Garnsey, 2016; Y.; Zhang et al., 2012; Zhou et al., 2010). However, the mismatched classifier-noun pairing used in these studies contained apparent meaning clash, which might render morphosyntactic processing undetected, if any. The current study avoided meaning clash by creating two kinds of violations often seen in aphasic patients and young children: omitting the classifier in a classifier phrase (the Missing condition, which is an outright grammatical violation) and inappropriately replacing a specific classifier with the meaning-devoid general classifier GE (the GE condition). Through the stark similarity of the negativity effects induced by these two types of violations, both in their amplitude and broad scalp distribution, our study revealed that morphosyntactic agreement was involved in processing classifier phrases in isolation (Experiment 1) as well as in a sentence (Experiment 2).

While we associated the negativity effect with morphosyntactic processing, one could alternatively argue that such an effect might reflect increased task difficulty in linguistic information processing (e.g. Zhang, Eppes, & Diaz, 2019) or controlled semantic retrieval and selection (e.g. Lau, Phillips, & Poeppel, 2008). In fact, similar interpretations were given in previous classifier studies. Mueller et al. (2005) found a delayed LAN (500–800 ms) in classifier violation in the middle of a sentence and suggested that the LAN effect might reflect greater working memory demand either for an active "post hoc" agreement checking between a noun and an earlier classifier or for retention of a noun in working memory in order to wait for the occurrence of a correct classifier. Zhou et al. (2010) also found a late AN (550–800 ms) for a mismatched noun at the sentence final position and proposed that it reflected semantic reinterpretation and/or heavy memory load caused by semantic incongruence. Hsu et al. (2014) observed an early AN (250–450 ms) in processing mismatched classifier–noun pairs and argued that the effect might reflect the metacognitive processes in resolving the conflict as well as the working memory demand for keeping the classifier active and anticipating some plausible sequence to appear later. Nevertheless, we believe that the negativity effect in our study is unlikely to be attributed to working memory load. Based on the behavioral results, Missing was easier than Congruent in Experiment 1 (higher accuracy rate and shorter RT) and as easy as Congruent in Experiment 2 (similar accuracy rate and RT), but the ERP response to Missing was consistently stronger than that to Congruent. In the same vein, GE was more difficult than Missing in both experiments (i.e. lower accuracy rates and longer RTs), yet the negativity effect was equally strong in both conditions. Therefore, task difficulty is unlikely to be tracked by the negativity effect found in this study.

Even though a task difficulty account may not be plausible, we cannot completely rule out the possibility that some kind of lexical process might be involved in processing classifier-noun agreement. Indeed, researchers have shown that morphosyntactically-based agreement operations are processed on a syntactic basis, but lexical representations could be recruited depending on the agreement pattern that has to be resolved (Molinaro et al., 2011). Since classifier-noun agreement in Mandarin is not morphophonolgically realized as those agreement patterns in inflectional languages, (at least part of) the words need to be recognized to make the computation of the agreement relationship possible. However, such "required" recognition process should be similar across the three conditions, yet GE and Missing induced a stronger effect than did Congruent. One might ascribe the effects evoked by GE and Missing to controlled, higher-order retrieval or selection of lexical representations (Lau et al., 2008). However, we do not think this controlled view can be valid: the negativity effects for Missing and GE were similar in amplitude and scalp distribution, suggesting that the underlying process should be identical for these two conditions, but if the effects were indeed related to controlled semantic processes, they should differ in the two conditions. This is because trials in the Missing condition looked like Mandarin compound nouns (Num-Ø-N vs. Num-N) (although efforts were made to ensure that none of these Missing combinations constituted an existing noun) and the subjects might be searching the mental lexicon for matched items when evaluating the correctness (as in Experiment 1) or acceptability (as in Experiment 2) of the trials, whereas trials in the GE condition (Num-GE-N) could not form a compound noun, so a lexicon search for matched items was not necessary to determine whether the consecutively presented stimuli were correct/acceptable and thus a smaller effect should have been observed. Therefore, we believe that the equally strong negativity effect observed in GE and Missing should index a similar process and is thus unlikely to be associated with controlled semantic processes, which were different in these two conditions.

Although our conclusion about the morphosyntactic agreement in classifier processing does not require the confirmation of the observed negativity to be a LAN, it is still informative to discuss what ERP component(s) the negativity could be. In fact, with the atypical scalp distribution of the N400 in previous classifier research and the negativity effects in our study, it is almost impossible to clearly discern whether the observed effect is an N400 or a LAN. After all, the negativities in the present study do not have a typical (right) centro-parietal concentration, so it is inappropriate to identify it as an N400 and entirely associate it with semantic processing, nor can they be treated as a LAN because their distribution (in our study and in previous classifier research) does not resemble a typical LAN, which usually, although not always, has a left-lateralized distribution. In fact, there has been a debate over whether the LAN in past literature is a byproduct of the averaging process: some participants may show a P600 and others may present an N400, and thus the brainwaves cancel each other out at midline electrodes and the profile of a left lateralized LAN followed by a P600 is found (Osterhout, 1997; Molinaro et al., 2011; Tanner, D., & Van Hell, 2014; Tanner, 2015). However, the negativity effect in our study (and also in Zou et al., 2010; Y. Zhang et al., 2012, and Qian & Garnsey, 2016) was widespread and not lateralized to the left, and there was no sign of a positivity in Experiment 1 and the visually detected weak positivity in Experiment 2 was also not significant (note that there was also no P600 in Zou et al., 2010 and Qian & Garnsey, 2016). Therefore, even if we do identify the negativity as a LAN, its presence cannot possibly be attributed to an averaging problem.

If the observed negativity was not an N400 or a LAN effect, what could it be? We believe that the broad negativity might be a result of overlapping LAN and N400, signaling the coexistence of morphosyntactic and semantic agreement in processing classifier phrases, because there is a good reason to connect the concurrence of the two types of agreement with the diversity of classifiers. As mentioned in the Introduction, Mandarin classifiers were originally derived from verbs or nouns, and they underwent a gradual loss of semantic information sharing with the nouns throughout the way of becoming a classifier. Such gradual loss of semantic content has resulted in the co-existence of contentful and meaningless classifiers in Modern Mandarin. Gao and Malt (2009) divided existent Mandarin classifiers into three types depending on the knowledge evoked by the classifier alone and on the knowledge when the classifier is used with a noun: (1) "well-defined" (all the objects associated with a given classifier share one or more defining features, e.g. "ben_{CL}" (本) for objects that are bound into a book-like form, such as books, notebooks or dictionaries), (2) "prototype" (the associated objects can often be deduced from the non-classifier use of the noun, but can also be distant from the prototype evoked by the classifier, e.g. "ke_{CL}" (顆) for small, three dimensional, roundish objects, such as pearls or grapes, or for something more distant, such as bombs and stars), and (3) "arbitrary" (no set of defining features nor any prototype can give a sense of coherence to the set of objects associated with the classifier; e.g. "dao_{CL}" (道) can be used with nouns, such as wall, door, procedure, or sunlight). It is thus likely that different kinds of processing are triggered by different types of classifiers (e.g. more semantically oriented processing for well-defined and prototype classifiers, while more morphosyntactically oriented processing for arbitrary ones); it is also possible that some classifiers induce both kinds of processing. Future research is needed to find out whether the semantic and morphosyntactic processing are separately induced by different classifier-noun pairings or are simultaneously triggered by the same classifier-noun combinations.

We believe that Mandarin classifiers possibly play both a semantic and a grammatical role in relation to its accompanying noun. It is likely that morphosyntactic agreement might have existed in previous classifier research even though grammatical violation was not included in the materials, and that semantic agreement might be present in our study, even though semantic clash of the materials was avoided. This argument actually recognizes that the proponents of the semantic or grammatical side of the debate are both on the right track—it is just that they are looking at only one side of the coin. Therefore, the new question to ask in linguistics is probably not *whether* classifier plays a grammatical or semantic role, but *why* these two types of agreement co-occur.

The discussion of the current study would not be complete without relating our results to findings in past agreement research. The negativity effect reported in our study is associated with morphosyntactic processing, which is reminiscent of the LAN effect found in previous literature on grammatical gender agreement in languages, such as Spanish or Dutch (see Molinaro et al., 2011, for a review). In fact, linguists have proposed a continuum of linguistic devices in nominal classification across languages from purely grammatical (e.g. gender agreement systems) to purely lexical (e.g. measure terms such as "a pile of books"), with classifier systems being located midway on this continuum (Grinevald, 2000). Therefore, the finding of similar brainwave patterns in response to classifier and gender systems (especially those studies on opaque gender, which induced early negativity extending to central and posterior brain regions, see Molinaro et al., 2011) is not completely surprising. What is surprising, however, is that morphosyntactic agreement is not expected to be present in Mandarin, a language without rich inflectional morphology. After all, there are more than a hundred classifiers in Mandarin, but usually the number of gender agreement devices, such as determiners or morphological markers on the noun/verb, is much less in languages with a grammatical gender system. Also, classifier use is variable: a noun can be paired with more than one classifier, as mentioned in Section 2.1.2, while usually a single value is carried by a noun in a gender system (although there are cases where a noun can have more than one gender, Dixon, 1968). Nevertheless, the violation of classifier-noun pairings in Mandarin induces an ERP effect similar to that found in agreement violation in languages with rich inflections, suggesting that the underlying neural mechanisms for computing dependency relations may bypass the diversity of morphological structure in different languages. We believe this is an important finding and future cross-linguistic study is needed to find out how similar the underlying processing can be in computing dependency relations in languages with or without rich inflectional morphology.

5. Conclusions

Our study reveals the existence of morphosyntactic agreement in Mandarin classifier phrases, which adds to previous findings about semantic involvement and elucidates that there should be a middle ground for the linguistic debate over whether Mandarin classifiers are functional or semantic elements. The concurrence of semantic and morphosyntactic agreement may be related to the co-existence of contentful and meaningless classifiers. Importantly, morphosyntactic agreement is usually considered absent in language without rich inflections. Its existence in Mandarin may suggest comparable computation for dependency relations in general, and agreement relationship in particular, in human language.

Acknowledgements

The author thanks the two anonymous reviewers for their constructive comments that significantly enhanced the quality of the manuscript. The author also sincerely appreciates the help from the research assistants in the Neurolinguistics Laboratory at National Taiwan Normal University for experimental material preparation and data collection.

Funding

This work was supported by the Ministry of Science and Technology (formerly known as National Science Council) of Taiwan under Grant [NSC 100-2410-H-003-088] and Grant [MOST 106-2410-H-003-137-MY2].

Disclosure of interest

The author reports no conflict of interest.

References

Ahrens, K. (1994). Classifiers production in normals and aphasics. *Journal of Chinese Linguistics, 22*(2), 203–248.
Allan, K. (1977). Classifiers. *Language, 53*(2), 285–311.
Cheng, L.-S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry, 30*(4), 509–542.
Cheng, L.-S., & Sybesma, R. (2005). Classifiers in four varieties of Chinese. In G. Cinque, & R. S. Kayne (Eds.). *The oxford handbook of comparative syntax* (pp. 259–292). Oxford: Oxford University Press.
Chou, T.-L., Lee, S.-H., Hung, S.-M., & Chen, H.-C. (2012). The role of inferior frontal gyrus in processing Chinese classifiers. *Neuropsychologia, 50*(7), 1408–1415.
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*(1), 9–21.
Dixon, R. M. W. (1968). Noun classes. *Lingua, 21*, 104–125.
Erbaugh, M. S. (1986). Taking stock: The development of Chinese noun classifiers historically and in young children. In C. Craig (Ed.). *Noun classes and categorization.* Amsterdam/Philadelphia: J. Benjamins.
Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84.
Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language & Cognitive Processes, 24*(7–8), 1124–1179.
Grinevald, C. (2000). A morphosyntactic typology of classifiers. In G. Senft (Ed.). *Systems of nominal classification* (pp. 50–92). Cambridge: Cambridge University Press.
Her, O.-S. (2012). Distinguishing classifiers and measure words: A mathematical perspective and implications. *Lingua, 122*(14), 1668–1691.
Her, O.-S., & Hsieh, C.-T. (2010). On the semantic distinction between classifiers and measure words in Chinese. *Language and Linguistics, 11*(3), 527–551.
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.
Huang, C. R. (2009). *Tagged Chinese Gigaword version 2.0.* Philadelphia, PA: Lexical Data Consortium.
Huang, C., Chen, K., & Lai, C. (Eds.). (1997). 國語日報量詞典 *[Mandarin daily dictionary of Chinese classifiers]*. Taipei: Mandarin Daily Press.
Huang, C.-R., Kilgarriff, A., Wu, Y., Chiu, C.-M., Smith, S., Rychly, P., ... Chen, K.-J. (2005). Chinese sketch engine and the extraction of collocations. *Proceedings of the fourth SigHan workshop on Chinese language processing. Jeju, korea*.
Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*(2), 163–178.
Kaan, E., & Swaab, T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences, 6*(8), 350–356.
Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition, 166*, 433–446.
Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience, 9*(12), 920–933.
Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213.
Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex, 47*(8), 908–930.
Mueller, J. L., Hahne, A., Fujii, Y., & Friederici, A. D. (2005). Native and nonnative speakers' processing of a miniature version of Japanese as revealed by ERPs. *Journal*

of Cognitive Neuroscience, 17(8), 1229–1244.
Myers, J. (2000). Rules vs. analogy in Mandarin Classifier selection. *Language and Linguistics, 1*(2), 187–209.
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113.
Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language, 59*(3), 494–522.
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, & J. C. Clifton (Eds.). *The on-line study of sentence comprehension: Eyetracking, ERP, and beyond*. Psychology Press.
Qian, Z., & Garnsey, S. (2016). An ERP study of the processing of Mandarin classifiers. In H. Tao (Ed.). *Integrating Chinese linguistic research and language teaching and learning* (pp. 59–80). Amsterdam Philadelphia: J. Benjamins.
Santi, A., & Grodzinsky, Y. (2007). Working memory and syntax interact in Broca's area. *NeuroImage, 37*(1), 8–17.
Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language, 120*(2), 135–162.
Tai, J. H.-Y. (1994). Chinese classifier systems and human categorization. In M. Chen, & O. Tseng (Eds.). *In honor of professor william S-Y. Wang: Interdisciplinary studies on language and language change* (pp. 479–494). Taipei: Pyramid Publishing Company.
Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A commentary on "grammatical agreement processing in reading: ERP findings and future directions" by Molinaro et al., 2014. *Cortex, 66*, 149–155.
Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia, 56*, 289–301.
Tzeng, O. J. L., Chen, S., & Hung, D. L. (1991). The classifier problem in Chinese aphasia. *Brain and Language, 41*(2), 184–202.
Uchida, N., & Imai, M. (1999). Heuristics in learning classifiers: The acquisition of the classifier system and its implications for the nature of lexical acquisition. *Japanese Psychological Research, 41*(1), 50–69.
Wu, Y., & Bodomo, A. (2009). Classifiers ≠ determiners. *Linguistic Inquiry, 40*(3), 487–503.
Zhang, H. (2007). Numeral classifiers in Mandarin Chinese. *Journal of East Asian Linguistics, 16*(1), 43–59.
Zhang, N. (2013). *Classifier structures in Mandarin Chinese.* Berlin: Mouton de Gruyter.
Zhang, H., Eppes, A., & Diaz, M. T. (2019). Task difficulty modulates age-related differences in the behavioral and neural bases of language production. *Neuropsychologia, 124*, 254–273.
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*(3), 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*(6), 1551–1562.