Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Lexico-syntactic features are activated but not selected in bare noun production: Electrophysiological evidence from overt picture naming

Man Wang [a,b,c,*], *Yiya Chen* [b,c] and *Niels O. Schiller* [b,c]

[a] *School of Foreign Languages and Literature, Shandong University, China*
[b] *Leiden University Center for Linguistics, The Netherlands*
[c] *Leiden Institute for Brain and Cognition, The Netherlands*

ARTICLE INFO

Article history:
Received 18 October 2017
Reviewed 16 December 2017
Revised 10 February 2018
Accepted 22 May 2018
Published online 31 May 2018

Keywords:
Morphological processing
Overt speech production
Lexico-syntactic feature
Mandarin Chinese classifiers
ERP

ABSTRACT

To produce a word, speakers need to retrieve the lexico-syntactic representation of the word and encode the phonological form for articulation. It is not precisely known yet if a word's syntactic features (e.g., number, gender, etc.) are automatically activated and selected in bare noun production. Cubelli, Lotto, Paolieri, Girelli, and Job (2005) proposed that only in languages that have a complex morphological structure (e.g., Italian), the selection of grammatical gender is required. In languages with a relatively simpler morphological structure, the selection of grammatical gender is by-passed. Here, we investigated this issue further by employing a language with an extremely simple morphological structure, i.e., Mandarin Chinese. Using the picture-word interference paradigm, we manipulated the congruency of the lexico-syntactic *classifier* feature (comparable to grammatical gender) between the target picture and the superimposed distractor word. We measured participants' naming latencies and their electroencephalogram (EEG). As a result, relative to the classifier-congruent condition, classifier incongruency elicited a stronger N400 effect in the ERP analyses, suggesting the automatic activation of lexico-syntactic features in bare noun production. However, classifier congruency did not affect naming latencies, suggesting that the lexico-syntactic feature is not selected in bare noun naming when it is irrelevant for production.

* *Corresponding author*. School of Foreign Languages and Literature, Shandong University, 250100, Jinan, China.
E-mail address: wangm@sdu.edu.cn (M. Wang).

https://doi.org/10.1016/j.cortex.2018.05.014

1. Introduction

Words, together with their semantic, syntactic and phonological properties, are stored in our mental lexicon. When we speak, we access our mental lexicon at an amazingly high speed to select the to-be-produced words to express the meaning in their appropriate phonological forms within the syntactic constraints (Van Turennout, Hagoort, & Brown, 1998). Cognitive language production models predict *when* certain components of a to-be-produced word are activated, selected and encoded, *where* the activation is located in the brain, and *how* the activation flows. In terms of the temporal loci, most of these models agree on the main stages involved in word production: (a) conceptualization of the intended message, (b) retrieval of the semantic and grammatical representations of the to-be-produced words (hereafter lemma retrieval), (c) word-form encoding, and (d) articulation (e.g., 'Independent-Network model', Caramazza, 1997; 'interactive' spreading-activation model, Dell, 1986, 1988, 1990; Dell & O'Seaghdha, 1991, 1992; the WEAVER++ model, Levelt, 1992; Levelt, Roelofs, & Meyer, 1999a, 1999b; Roelofs, 1992, 1993; Roelofs & Meyer, 1998).

During lemma retrieval, a lemma is activated by the concept and selected for the next stage of phonological form encoding. The word's syntactic features (e.g., number, grammatical gender, etc.) receive activation from the lemma. Some syntactic features (e.g., number) may also receive activation from the concepts (e.g., MULTIPLE; Levelt et al., 1999a; see Nickels, Biedermann, Fieder, & Schiller, 2015 for an alternative account). For instance, in English, the *-s* affix needs to be selected for regular plural nouns (e.g., 'cats'). In Dutch, the determiner needs to be selected and to agree with the noun on its grammatical gender in noun phrase production ('de arm', *the arm*, common gender and 'het been', *the leg*, neuter gender). Empirical evidence has been reported to support the selection of syntactic features during word and phrase production (e.g., La Heij, Mak, Sander, & Willeboordse, 1998; Schriefers & Teruel, 2000; Schriefers, 1993; Van Berkum, 1997). Nevertheless, it is debated whether a word's syntactic features (e.g., grammatical gender) are always activated and whether consequently, they are also automatically selected, even when they are irrelevant for specific speech production tasks (e.g., 'cat' in English and 'been', *leg*, in Dutch).

Experimental studies have mostly made use of the picture-word interference paradigm (e.g., Glaser, 1992; see MacLeod, 1991 for a review) to examine the selection of syntactic features in speech production. For example, the selection of grammatical gender in noun phrase production in Dutch and German has been reported (e.g., La Heij et al., 1998; Schriefers & Teruel, 2000; Schriefers, 1993). Specifically, shorter naming latencies were observed when the grammatical gender of the distractor word (e.g., 'dak', *roof*, neuter gender) was congruent with that of the target picture name (e.g., 'boek', *book*, neuter gender) than in an incongruent condition (e.g., 'tafel', *table*, common gender). This has been observed in both article-adjective-noun (e.g., 'het groene boek', *the green book*) and plain adjective-noun (e.g., 'groen boek', *green book*) productions. The effect in naming latencies was called the "gender congruency effect" (La Heij et al., 1998; Schriefers & Teruel, 2000; Schriefers, 1993; Van Berkum, 1997). However, the gender congruency effect disappeared when the determiners are the same for common and neuter nouns, for instance, in the Dutch plural ('de boeken', *the books* – 'de tafels', *the tables*). Thus, the "gender congruency effect" was re-interpreted as determiner congruency effect related to the retrieval of determiners at the word-form level (e.g., Alario & Caramazza, 2002; Miozzo & Caramazza, 1999; Miozzo, Costa, & Caramazza, 2002; Schiller & Caramazza, 2003, 2006; see Caramazza, Miozzo, Costa, Schiller, & Alario, 2001 for a review).

However, no gender or determiner congruency effect was observed in bare noun production in Dutch (e.g., boek, *book*) by La Heij and colleagues (La Heij et al., 1998; see also; Starreveld & La Heij, 2004). By contrast, Cubelli and colleagues conducted a series of experiments using the picture-word interference paradigm and reported consistent effects of grammatical gender in bare noun naming in Italian (Cubelli et al., 2005). In their study, the grammatical gender congruency showed an inhibitory effect, compared to the incongruent condition, contradicting the facilitative effect observed in West-Germanic languages (such as in Dutch and German) when the determiner was included in the naming task. The inhibitory effect was interpreted as reflecting competition at the lemma level and the selection of grammatical gender is mandatory before accessing the morpho-phonological form of a given noun in word production (Cubelli et al., 2005).

So far, no agreement has been reached upon whether lexico-syntactic features such as grammatical gender are indeed automatically activated and selected in bare noun production. If they are, as suggested by Cubelli et al. (2005), it suggests that speakers select extra information such as task-irrelevant syntactic features in word production. If the lexico-syntactic features are not selected (e.g., La Heij et al., 1998; Starreveld & La Heij, 2004), there are still two possibilities for the theoretical account of the null effect in naming latencies. The null effect could be accounted for by speech production models (e.g., Caramazza, 1997; Levelt et al., 1999a) in various ways. One possibility is that the lexico-syntactic features are not activated in bare noun production. The other possibility is that they are always activated but not selected and consequently, do not affect the retrieval and production of the target word (La Heij et al., 1998).

As discussed in La Heij et al. (1998), even if the lexico-syntactic features are activated, there are still two possible explanations in alternative speech production models. It could be the case that the lexico-syntactic features receive spreading activation from the activated lemma (Levelt et al., 1999a). Since the lexico-syntactic features are activated after the retrieval of the lemma, they will not affect the production speed when irrelevant for production (La Heij et al., 1998). Alternatively, based on the assumptions derived from the model by Caramazza (1997), the syntactic layer is omitted. The lexico-syntactic information receives activation directly from the semantic representation or the phonological representation. Specifically, the lexico-syntactic features such as word class receive activation from the semantic representation and other features such as gender receive activation from the phonological representation (Caramazza, 1997; cf.; La Heij et al., 1998, p. 217).

Alternatively, Cubelli et al. (2005) proposed a two-layer architecture for language production: the lexico-semantic and lexico-syntactic representations. Both layers have to be activated and selected before accessing the phonological form of the target word. To explain the discrepancy between their finding and the null gender effect in Dutch, Cubelli et al. (2005) pointed out that only in languages that have a complex morphological structure (e.g., Italian), the selection of grammatical gender is required.

No study, to our knowledge, has direct evidence to tease apart these possibilities. Therefore, the following questions are empirically open: Are lexico-syntactic features always activated, even in singular bare noun production? If so, where do the lexico-syntactic features receive the activation from, i.e., via spreading activation from the activated lemma (as predicted in Levelt's model; Levelt et al., 1999a) or direct activation from the semantic or phonological representation (as predicted in Caramazza's IN model; Caramazza, 1997)? Furthermore, are they consequently selected in singular bare noun production? Are the cross-linguistic discrepancies attributed to the simplicity/complexity of the morphological structure?

Notably, most studies discussed above have drawn evidence from behavioral studies with reaction time data. It has been noted that reaction times only reflect the outcome of a number of cognitive processes involved in overt naming while electrophysiological data can provide a fine-grained measurement of the various ongoing cognitive processes stimulated by the experimental manipulation (Luck, 2005). Event-related potential (ERP) experiments have been carried out extensively in psycholinguistic research. However, the majority of the experiments investigate language perception processes and covert language production. This is mainly due to the concerns about muscle movements involved in language production that can distort the ERP signals and consequently make the acquired data unreliable. However, an increasing number of recent studies have investigated the functional characteristics of speech production with electrophysiological measurements and demonstrated that artifact-free ERP signals can be measured up to 400 msec post-stimulus presentation (Ganushchak, Christoffels, & Schiller, 2011). For instance, it has been proposed that the brain engages in lexical retrieval starting 200 msec after stimulus onset (Costa, Strijkers, Martin, & Thierry, 2009; Strijkers & Costa, 2011) and engages in syntactic processing 40 msec before phonological processing during speaking (Van Turennout et al., 1998). Semantic activation has been found to precede phonological encoding during picture naming (Schmitt, Münte, & Kutas, 2000; Van Turennout, Hagoort, & Brown, 1997) as reflected in both the lateralized readiness potentials (LRPs), an ERP component, and a response inhibition index, namely the N200. Morphological encoding has been observed around 400 msec after stimulus onset (Koester & Schiller, 2008), in line with the predictions of meta-analytic studies (Indefrey & Levelt, 2004; Indefrey, 2011). The reliability of electrophysiological measurement with overt speech production calls for more research to provide fine-grained data with high temporal resolution to reveal the underlying mechanisms of speech production.

These findings motivated us to seek electrophysiological evidence to tap into the issue of lexico-syntactic feature activation and selection in bare noun production. Our empirical base for this investigation is bare noun production in Mandarin Chinese, a language with a relatively simple morphological structure. As we will explain below, the nominal classifiers (hereafter classifiers) in Mandarin Chinese provide an interesting as well as important, but hitherto much ignored, test case for the debate.

In Mandarin Chinese, although gender or case is not overtly marked, it is compulsory to use a classifier between an article, a quantifier or another modifier and its associated noun. For instance, the common classifier for a piece of upper-body clothing (e.g., *coat*, *shirt*, etc.) is “jian^{4}”[1], and to refer to the noun “da^{4}yi^{1}” (*coat*) in a noun phrase using a numeral or an article, the classifier must occur between the modifier and the noun, i.e., “yi^{1} jian^{4} da^{4}yi^{1}” (*one classifier-jian^{4} coat*) or “zhe^{4} jian^{4} da^{4}yi^{1}” (*this classifier-jian^{4} coat*). Classifier choice is determined by the semantic-syntactic features (e.g., semantic category, number; see Wang, 1973). In some cases, an object's classifier is determined by its semantic category, e.g., the contrast between animal names that tend to be used with “zhi^{1}” and clothes names with “jian^{4}”. In some other cases, one lion is used with the classifier “tou^{2}” while a group of lions with the classifier “qun^{2}”. Sometimes, classifiers function as the grammatical marker, comparable to the number morphology in other languages (Cheng & Sybesma, 1999, 2005; Doetjes, 1997; Peyraube, 1998).

So far, we have only found two behavioral studies that manipulated classifier congruency (between the classifier and the noun) as well as semantic relatedness using the picture-word interference paradigm to investigate the role of classifiers in Mandarin Chinese speech production. Conflicting results, however, were reported regarding classifier effects in bare noun naming. Zhang and Liu (2009) found that a classifier-congruent distractor facilitated picture naming even in the bare noun production task where no classifier information was required. However, Wang, Guo, Bui, and Shu (2006) found contradictory results, and argued that only in noun phrase naming is classifier encoding required, but not in bare noun naming (Wang et al., 2006).

In psycholinguistic research, classifier information is considered comparable to grammatical gender information in some respects, as it is directly associated with the lexical item and regarded as a lexical property of nouns. It bears a transparent semantic relationship to the lexical item in some cases, but is arbitrary in others (Tzeng, Chen, & Hung, 1991). Given this similarity, the study of the effect of classifier in noun production is not only necessary but also provides an interesting line of comparison with regard to lexico-syntactic feature encoding between spoken word production in West-Germanic languages (where gender is a prominent feature) and that in East Asian languages (where classification is a prominent feature). In the current study, we used the picture-word interference paradigm and manipulated both semantic category and classifier congruency between target picture name and distractor word. This manipulation provides insights into the classifier choice as a function of semantic classes (e.g., Wu & Bodomo, 2009; but see; Cheng & Sybesma, 2005), which is necessary to tease apart.

We measure both naming latencies and electrophysiological activities. If classifiers are not automatically activated, we expect to see comparable naming latencies and electrophysiological activities between classifier congruent and

[1] As an example, “jian” indicates the phonetic notation of the lexical item, i.e., Pinyin of the word and the number 4 indicates the lexical tone.

incongruent conditions. If classifiers are automatically activated but not selected, we expect to see comparable naming latencies between classifier congruent and incongruent conditions but significant differences between the two conditions in electrophysiological activities. Specifically, as gender disagreement has been reported to elicit an "N400-type effect" (Barber & Carreiras, 2005), we expect to observe a reduced N400 effect for the classifier congruent trials, relative to incongruent trials. A cautionary note is that the experimental tasks in Barber and Carreiras (2005) were noun phrase and sentence (silent) reading, results of which therefore may not be optimal for us to base our predictions directly upon. Unfortunately, as far as we know, there has not been an ERP study to determine the ERP effect elicited by grammatical gender (dis)agreement. We will therefore build upon results reported in Barber and Carreiras (2005) while being aware of the different setups for our data interpretation.

Alternatively, if classifiers are activated as well as selected in bare noun naming, we expect to observe shorter naming latencies on the classifier congruent trials than on the incongruent ones (Zhang & Liu, 2009) as well as a stronger N400 effect elicited by the incongruent classifiers compared to the congruent ones. Moreover, we expect to see a general semantic interference effect as reflected in naming latencies, based on previous research using the picture-word interference paradigm (e.g., Glaser & Düngelhoff, 1984; La Heij, 1988; Zhu, Damian, & Zhang, 2015; see Spalek, Damian, & Bölte, 2013 for a review), as well as a negative effect around 400 msec as shown in previous electrophysiological studies using the picture-word interference paradigm or (in combination with) another paradigm (e.g., the cumulative semantic interference paradigm or the semantic blocking paradigm) (Aristei & Abdel Rahman, 2011; Zhu et al., 2015; but see Costa et al., 2009 for ERP effects obtained in the P2, N2, P3 and the N400 ranges). The dissociation of the N400-like effect and the semantic interference effect has also been discussed in Blackford, Holcomb, Grainger, and Kuperberg (2012).

2. Method

2.1. Participants

Thirty-three native Mandarin Chinese speakers (mean age = 26 years, SD = 3.05; 19 females) studying in the Netherlands (n = 28) or Beijing, China (n = 5) with comparable second language experience[2] gave informed consent for participation in the experiment. All participants were right-handed, had normal or corrected-to-normal vision, and no history of neurological impairments or language disorders. They were paid for their participation.

2.2. Materials

Thirty black-and-white line drawings from Severens' picture database (Severens, Van Lommel, Ratinckx, & Hartsuiker, 2005) or similarly drawn, corresponding to monosyllabic (20%), disyllabic (70%) or tri-syllabic (10%) names in Mandarin Chinese served as target pictures. Each picture was presented with four types of distractor words. The distractors were selected based on their congruency with the target picture names regarding two factors – classifier and semantic category (see Table 1). The distractors in the four conditions were matched in terms of word frequency, $F(3, 116) = .594$, $p = .620$, number of syllables, $F(3, 116) = 1.790$, $p = .153$, and visual complexity (number of strokes), $F(3, 116) = 1.437$, $p = .236$. Distractors were phonologically and orthographically unrelated to the target pictures.

2.3. Design and procedure

The experiment adopted a 2 by 2 factorial within-subject design, with classifier (C) and semantic category (S) as the two factors. Each factor had two levels: congruent (+) versus incongruent (–), resulting in four conditions: C+S+, C+S–, C–S+ and C–S–. On each trial, pictures were presented with a distractor (from one of the four conditions) superimposed on the center of the picture.

All participants saw each of the 30 pictures four times (once for each condition), resulting in 120 trials per participant, which were presented in a pseudo-random order such that the same picture did not occur within ten consecutive trials and no two consecutive trials were from the same condition or with the same corresponding classifier. The pseudo-randomized experimental lists were generated using the Windows program Mix (Van Casteren & Davis, 2006).

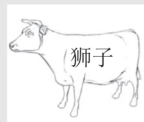
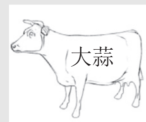
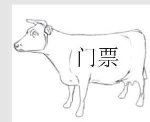
The experiment consisted of three sessions: a familiarization session, a practice session and an experimental session. In the familiarization session, each picture was presented once with its name underneath for 2 sec. Participants were requested to simply view the images and names. In the practice session, each picture was presented once with "XX" superimposed on it and participants were asked to name the pictures with the correct names while ignoring the "XX" on the pictures. Responses that deviated from the names given in the familiarization session were corrected by the experimenter.

In the experimental session, the 120 trials were divided equally into two blocks with a short break in between (length of the break was determined by the participant). On each trial, a fixation point ("+") was presented for 300 msec, followed by a blank screen (200 msec), the target picture with distractor (displayed until the participant initiated a vocal response, with a 2000 msec time-out), followed by another blank screen (500 msec) before the next trial began.

Participants sat in front of a computer in a dimly lit room and were asked to name the pictures using bare nouns as fast

[2] A Bartlett test for homogeneity of variance was performed on the behavioral data from the whole dataset, $p > .05$, indicating the homogeneity of the dataset, i.e., the variance does not differ across participant groups recruited in the two locations. We collected additional data from a second location to obtain sufficient statistical power, as the number of eligible participants was limited in the Netherlands. The participants we recruited in the Netherlands had lived in the Netherlands for less than half a year and those we recruited in Beijing had comparable language experience and proficiency. Nevertheless, we did re-run the analyses without the 5 participants and obtained the same patterns of results but with higher p-values (close to .1). Taken together, we decided to keep the additional 5 participants' data.

Table 1 – An example of a target picture presented with distractor in each condition. Distractors either match or mismatch the classifier (C) or semantic category (S) of target picture name.

	Condition			
Target picture name	C+S+	C+S-	C-S+	C-S-
“牛” *cow* /niu2/ classifier-“头” /tou2/				
distractor	*lion* /shi1zi0/	*garlic* /da4suan4/	*rat* /lao3shu3/	*ticket* /men2piao4/
classifier of distractor	“头” /tou2/	“头” /tou2/	“只” /zhi1/	“张” /zhang1/

and as accurately as possible. Vocal response times were measured by a voice-key and their electroencephalogram (EEG) was recorded simultaneously.

2.4. *Electrophysiological recording and data processing*

The EEG was recorded using 32 Ag/AgCI electrodes on the standard scalp sites of the extended international 10/20 system. Six flat electrodes were attached above and below the left eye to measure the eye blinks (2), at the external canthus of each eye to record horizontal eye movements (2) and at the mastoids for off-line re-referencing (2).

We used the Matlab toolbox FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) for the offline processing of the EEG data. The EEG signals were re-referenced to the average of both mastoids and band-pass filtered from .1 to 30 Hz. ERPs were time-locked to the onset of the target pictures. Epochs from −200 to 700 msec were computed, with baseline correction performed on the −200 to 0 msec prestimulus interval. Mean and linear trend were removed from the EEG data using a General Linear Modeling approach prior to resampling the EEG data acquired in two locations (sampled at 512 Hz in the Netherlands and 500 Hz in Beijing) to 256 Hz. We implemented the independent component analysis (ICA) function in FieldTrip (the codes are based on the function of EEGLAB; Delorme & Makeig, 2004) to remove the eye movement artifacts. At most two components per participant were identified as vertical or horizontal eye movements and removed from the EEG signal for further analysis.

Trials with amplitudes exceeding ±100 μV, or a 100 μV difference within a single trial, or exceeding 4 standard deviations of a participant's mean amplitude of all trials were considered as outliers and removed from the analysis. Data from six out of thirty-three participants were excluded from further analysis due to too many artifacts with available epochs below 50% after artifact rejection. The behavioral data from these six participants were excluded from analysis as well, leaving 27 effective datasets (mean age = 25 years, SD = 3.04; 18 females).

3. Results

3.1. *Behavioral data*

5.03% of all data points (3,240) were further removed from the behavioral data analysis, comprising: (a) incorrect responses; (b) voice-key failures (the first two types were counted as errors; the error rate was 3.58% and considered not informative enough for further analysis); (c) outliers (i.e., naming latencies exceeding 3 SDs above or below the participant's mean; 1.45%).

Repeated measures ANOVAs were performed on the participant means (F1) and item means (F2) with two within-subjects factors: classifier congruency (same classifier vs. different classifiers) and semantic relatedness (same semantic category vs. different semantic categories).

No significant effect of classifier congruency was obtained either in the by-participant analysis, $F1(1, 26) = .000$, $p = .994$, $\eta^2_P = .000$, or in the by-item analysis, $F2(1, 29) = .028$, $p = .867$, $\eta^2_P = .001$, indicating that classifiers are not selected in bare noun naming in Mandarin Chinese. There was a main effect of semantic relatedness in the by-participant analysis, $F1(1, 26) = 14.268$, $p = .001$, $\eta^2_P = .354$ and in the by-item analysis, $F2(1, 29) = 5.041$, $p = .033$. $\eta^2_P = .148$, with longer naming latencies on semantically related trials than semantically unrelated trials (Fig. 1). The interaction between the two factors was not significant either in the by-participant analysis, $F1(1, 26) = .008$, $p = .928$, $\eta^2_P = .000$, or in the by-item analysis, $F2(1, 29) = .000$, $p = .989$, $\eta^2_P = .000$.

3.2. *ERP data*

21.02% of the experimental trials were removed from the ERP data analysis including error trials (3.83%) and epochs

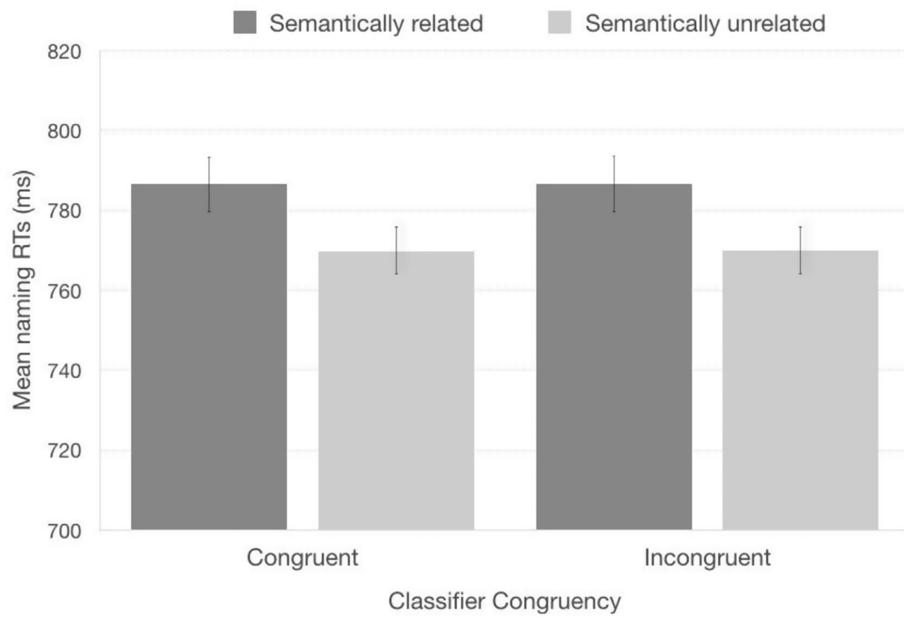


Fig. 1 – There was no significant difference between the classifier congruent and incongruent conditions. Naming latencies for the semantically related condition were significantly longer than the unrelated condition. There was no interaction between semantic relatedness and classifier congruency.

removed during artifact rejection (17.19%). For each condition, on average, there were 24 remaining epochs ($1.9 < SD$ per condition < 2.3). The following analyses were performed on four ROIs (left fronto-central: F3, FC1, FC5, C3; left centro-parietal: CP1, CP5, P3, PO3; right fronto-central: F4, FC2, FC6, C4; right centro-parietal: CP2, CP6, P4, PO4). Three consecutive time windows (0–275 msec, 275–575 msec, 575–650 msec) were chosen based on visual inspection of the data without taking the putative effects into account to avoid circularity (Luck & Gaspelin, 2017, Figs. 2 and 3; see Zhu et al., 2015 for a similar approach). The mean amplitudes in the above-mentioned time windows across all remaining channels were submitted to repeated measures ANOVA analysis in R (Team, 2014) using the *car* package (Fox & Weisberg, 2011), with classifier congruency (2 levels), semantic relatedness (2 levels) and ROI (4 levels) as three factors.

In the time window of 275–575 msec, there was a main effect of classifier congruency, $F(1, 26) = 6.11$, $p = .020$, $\eta^2_P = .190$, a main effect of semantic relatedness, $F(1, 26) = 4.67$, $p = .040$, $\eta^2_P = .152$, and a main effect of ROI, $F(3, 78) = 40.78$, $p < .001$, $\eta^2_P = .611$. No significant two-way interactions between any two of the three factors found, *p*-values $> .05$. No significant main effect or interaction was found in the other two time windows.

Next, to confirm the results of ANOVA analyses and to further explore the topographic distributions of classifier and semantic effects, two cluster-based permutation tests were performed considering data at all time points (about every 4 msec; see Zhu et al., 2015 for a similar approach). The permutation tests (Maris & Oostenveld, 2007) based on t-statistics were performed in FieldTrip (Oostenveld et al., 2011) on the participants' mean amplitudes within the time window 275–575 msec where significant semantic and classifier effects were statistically confirmed by the ANOVA analysis. This nonparametric randomization test was selected to control for the false alarm rate due to the multiple comparison problem with EEG data. This test first collects the trials into one single set regardless of experimental conditions. A random partition procedure is then performed on the data set 1,000 times and a histogram is constructed of the Monte Carlo approximation of the permutation distribution. The resulting *p*-value reflects the proportion of randomizations that result in a larger test statistic than the observed one. If this *p*-value is smaller than the critical alpha level of .05, then it is concluded that the data between the two experimental conditions are significantly different (see Maris & Oostenveld, 2007 for a detailed description of the method and see e.g., Wang, Bastiaansen, & Yang, 2015 for similar applications of the permutation tests).

Two pairs of comparisons were performed on the amplitudes in the time windows 275–575 msec. We then used the cluster-based permutation test based on t-statistics. First the classifier-congruent condition (C+) was compared with the classifier incongruent condition (C–) (both semantically unrelated), and then the semantically-related condition (S+) was compared with the semantically-unrelated condition (S–) (both classifier unrelated). The classifier-congruent and semantically-related condition was omitted (for a similar approach see Zhu et al., 2015).

A significant cluster (*p*-value smaller than .05) associated with the comparison between the classifier congruent and incongruent conditions was found from around 370 to 430 msec. The ERP amplitudes were more negative for the incongruent condition than for the congruent condition (Fig. 4). Similarly, a significant cluster (*p*-value smaller than .05) associated with the comparison between the semantically

cortex 116 (2019) 294–307

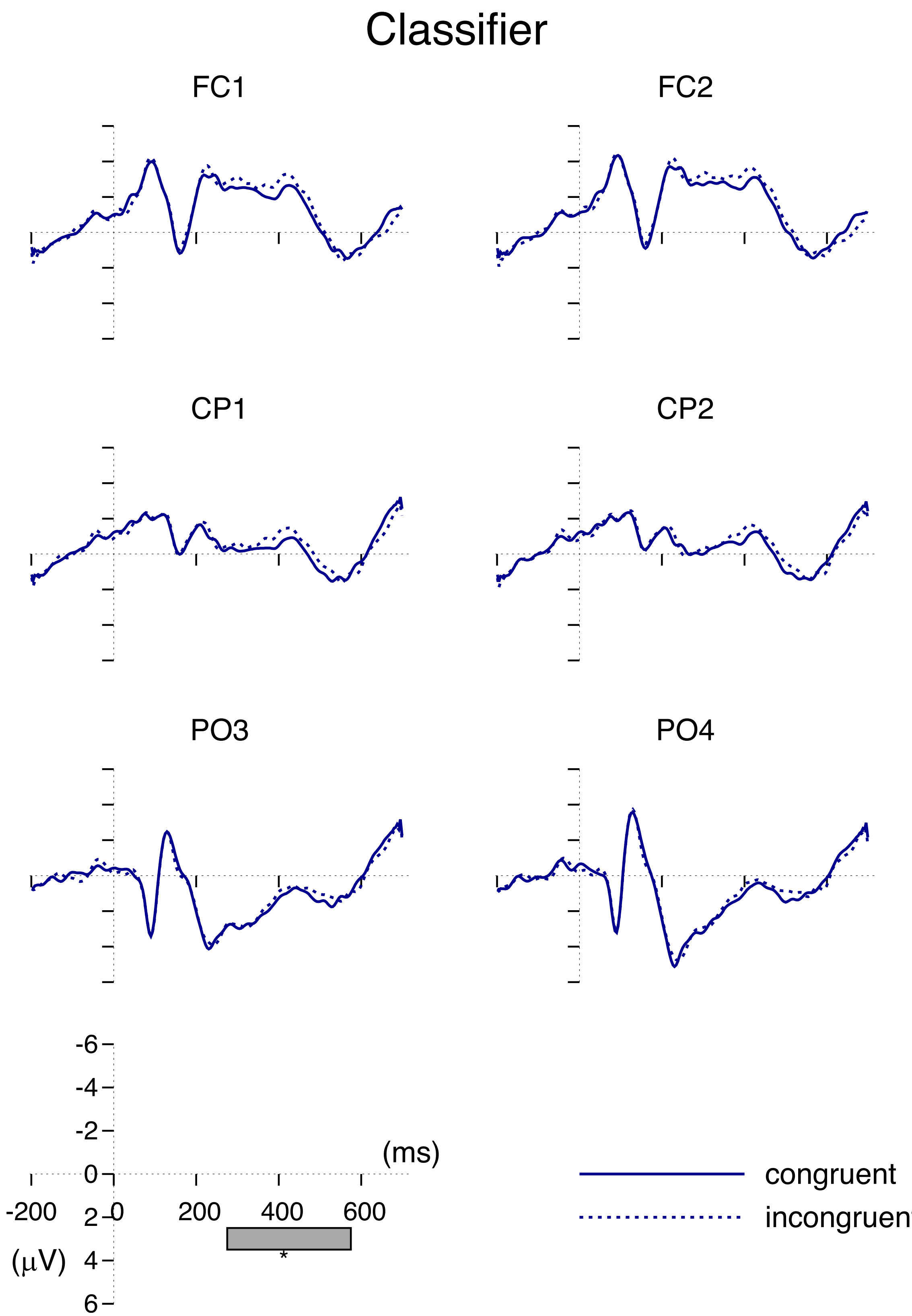


Fig. 2 – Grand averages of ERPs from six representative electrodes (FC1, FC2, CP1, CP2, PO3, PO4) for classifier congruent (C+) and incongruent (C−) conditions.

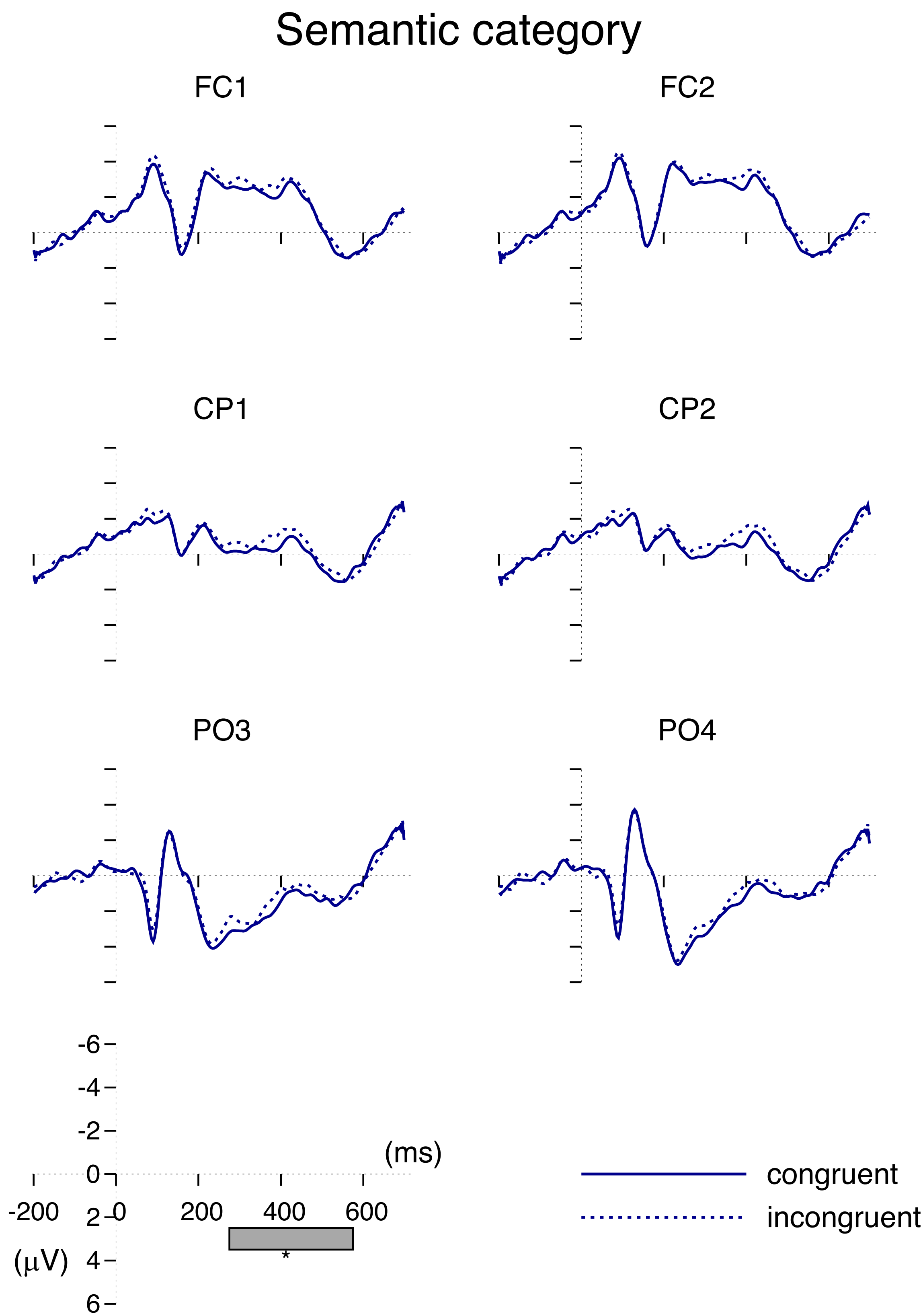


Fig. 3 – Grand averages of ERPs from six representative electrodes (FC1, FC2, CP1, CP2, PO3, PO4) for semantically related (S+) and unrelated (S−) conditions.

related and unrelated conditions was found within the window of around 370–430 msec. The amplitudes were more negative for the unrelated condition than for the related condition (Fig. 5).

4. Discussion

Using the picture-word interference paradigm, we manipulated the classifier congruency and semantic category congruency between the distractor word and the target picture. By measuring the participants' naming latencies and EEG activities, we investigated if lexico-syntactic features are activated and selected in bare noun production. We will first discuss the semantic effect and then the classifier effect.

The results obtained from manipulating the semantic category were in line with our predictions. The semantic interference effect (e.g., Glaser, 1992; MacLeod, 1991) was revealed by longer naming latencies when pictures were presented with a distractor word from the same semantic category relative to different semantic categories. This is consistent with previous studies (e.g., Glaser & Düngelhoff, 1984; La Heij, 1988; Zhu et al., 2015). The semantic interference effect can be interpreted as reflecting competition during lexical selection (see, e.g., Levelt et al., 1999a; but see, e.g., Mahon, Costa, Peterson, Vargas, & Caramazza, 2007; see Spalek et al., 2013 for a review).

In the ERP analyses, a larger negative ERP wave was observed for the semantically-unrelated condition compared to the related condition in the time window of 275–575 msec (Fig. 3). The effect was most robust in the parietal and central regions from about 370 to 430 msec according to a more conservative statistical analysis (Fig. 5). The ERP modulation by semantic category congruency is consistent with previous studies in Indo-European languages (e.g., Dell'Acqua et al., 2010; Janssen, Carreiras, & Barber, 2011; Jescheniak, Hahne, & Schriefers, 2003; Jescheniak, Schriefers, Garrett, & Friederici, 2002; but see Costa et al., 2009 who did not find significant correlations between naming latencies and ordinal positions in the N400 range) and Mandarin Chinese (e.g., Zhu et al., 2015), which also reported greater ERP negativities for the semantically-unrelated condition compared to the related condition. This negative effect at the parietal and central regions and peaking around 400 msec after stimulus presentation resembles a classic N400 effect. It is worth noting that Blackford and colleagues (Blackford et al., 2012) dissociated the behavioral semantic interference effect and the electrophysiological N400-like effect. While the N400-like effect is also possibly elicited by semantic priming (e.g., Blackford et al., 2012; Kreher, Holcomb, & Kuperberg, 2006), a cautionary note is that further research is needed to understand the electrophysiological effect that correlates with the semantic interference effect as shown in RTs.

No significant classifier effect, however, was observed in the naming latencies of the bare-noun naming task, which is in line with the classifier null effect in bare noun naming reported by Wang et al. (2006) but contradicts the finding of Zhang and Liu (2009). This null effect is at odds with the grammatical gender effect observed in Italian (Cubelli et al., 2005) but compatible with the result that no gender/determiner effect is observed in Dutch bare noun naming (e.g., La Heij et al., 1998; Starreveld & La Heij, 2004). Cubelli et al. (2005) have proposed that only in languages with a complex morphological structure, the selection of grammatical gender is required. The null effect of classifier in Mandarin Chinese, a language with a rather simple morphological structure, can be taken as another case for the by-passing of the selection of the lexico-syntactic features in bare noun production, in line with the predictions by Cubelli et al. (2005). As discussed in the *Introduction*, the null effect in naming latencies has left open the question of whether the lexico-syntactic features are always activated, even when they are irrelevant for production.

A statistically significant effect of classifier incongruency, however, was found between 370 and 430 msec after the target picture onset (Fig. 4), albeit in the absence of any significant effect of classifier incongruency in naming latencies. Classifier encoding is not required in bare noun naming, but by manipulating the congruency of classifiers between target pictures and distractors, we observed a stronger N400 effect with the classifier incongruent trials compared to congruent ones. This resembles the effect elicited by morphological priming in speech production (Koester & Schiller, 2008) and gender disagreement (Barber & Carreiras, 2005). The existence of the electrophysiological effect of classifier congruency lends support for the automatic activation of classifier features even in bare noun naming.

However, different from the condition where explicit morphological primes are used to elicit a morphological priming effect (Koester & Schiller, 2008), the present task does not require classifier feature (form) encoding. Therefore, the automatic activation of classifiers is at odds to take place at the form-encoding level when the morphological priming was obtained (Koester & Schiller, 2008).

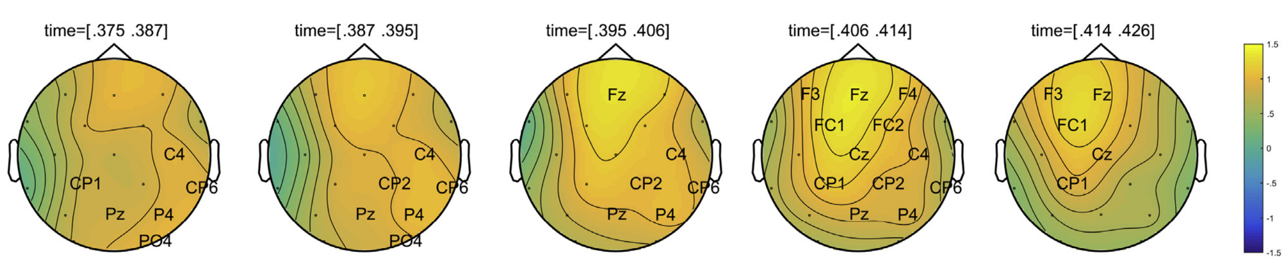


Fig. 4 – A significant positive cluster (C+ minus C−) was found for the classifier effect, ranging from around 370 to 430 msec. Electrodes with significant effects were highlighted with channel labels and asterisks. Due to limited space, the graphs were plotted with longer intervals instead of every 4 msec.

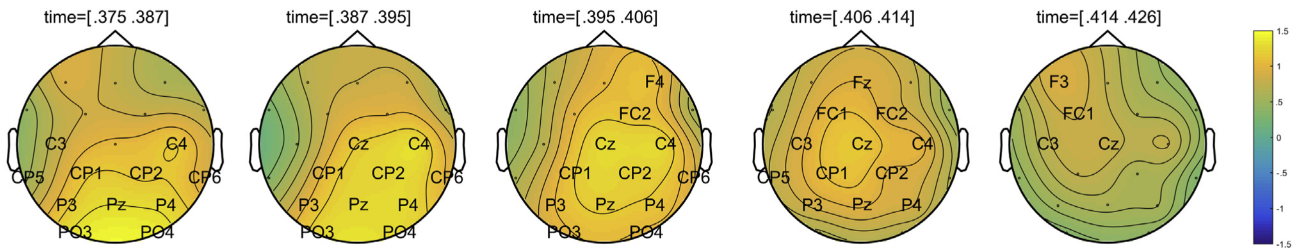


Fig. 5 – A significant positive cluster (S+ minus S−) was found for the semantic effect, ranging from around 370 to 430 msec. Electrodes with significant effects were highlighted with channel labels and asterisks. Due to limited space, the graphs were plotted with longer intervals instead of every 4 msec.

The remaining question then is how the classifier feature is activated in bare noun naming. There are two possible accounts. Based upon the Levelt et al. (1999a)'s model, one possibility is that classifier receives activation from the activated lemma, as a lexico-syntactic feature. Since this process happens after the lemma retrieval, we then would not expect the activation to affect the naming latency. Alternatively, based upon Caramazza's (1997) model, the other possibility is that the classifier, as a lexico-syntactic feature, receives activation directly from semantic representations or phonological representations. We know that classifiers in Mandarin Chinese can be independent from both the semantic representation and the phonological representation. For instance, native speakers of Mandarin Chinese acquire the classifier–noun combinations around four and five years old (e.g., Erbaugh, 1986; Fang, 1985) and 'there is no transparent or unequivocal mapping between conceptual properties and classifiers' (cf. Bi, Yu, Geng, & Alario, 2010, p. 103). As a consequence, the correct classifier–noun combinations have to be memorized. In other words, the connection between the classifier and its corresponding lemma is relatively fixed and reliable, while the connection between the classifier and the semantic representation is rather opaque and unreliable. Therefore, it is more likely that it is the activated lemma that spreads activation to the classifier feature, rather than that the classifier feature receives activation directly from semantic representation. Another possibility is that the classifier feature receives activation from the activated phonological representation. However, if this were the case, the ERP effect elicited by classifier incongruency would have been localized at a later point in time, following the activation of the phonological representation.

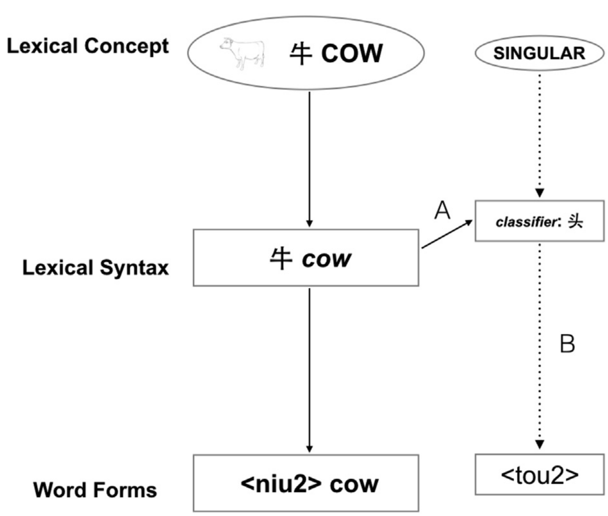


Fig. 6 – The automatic activation of the lexico-syntactic representation of classifiers in word production of Mandarin, adapted from Levelt et al. (1999a). The phonological form encoding of classifiers is not necessary in bare noun naming so Link B is only present when the production of classifier is required. Other lexico-syntactic features such as number and case that require more on-line processing rather than retrieval from long-term memory are not included in this model.

In Fig. 6, extending the speech production model from Levelt et al.'s (1999a), we show that for the lexical concept COW, the consequently activated target lemma (e.g., 牛 *cow*) automatically spreads the activation to the classifier feature (e.g., *classifier* 头) of this target lemma via link A. When we have a distractor word (e.g., 门票 *ticket*), which also activates its lemma and automatically its classifier (e.g., *classifier* 张) that differs from that of the target (头), it elicits a stronger N400 effect, relative to the condition where a distractor (e.g., 大蒜 *garlic*) has the same classifier as that of the target (e.g., *classifier* 头). However, in bare noun naming where the classifier information is not required for production, the incongruency between different classifier features does not affect the naming latencies.

To conclude, our behavioral and electrophysiological results jointly suggest that the Mandarin classifier feature is automatically activated by its associated target lemma but it is not selected in bare noun naming. Future research can be beneficial to further investigate to what extent automatic activation of lexico-syntactic features is language universal.

Acknowledgments

This research was supported by the Fundamental Research Funds of Shandong University (11050078614048) and grants from "Talent and Training China–Netherlands" program. We thank Gareth O'Neill, Jenny Doetjes and Niels Janssen for their comments on this study. We thank Frank Mertz for help with comprising the Matlab script.

REFERENCES

Alario, F.-X., & Caramazza, A. (2002). The production of determiners: Evidence from French. *Cognition, 82*, 179–223.
Aristei, S., Melinger, A., & Abdel Rahman, R. (2011). Electrophysiological chronometry of semantic context effects in language production. *Journal of Cognitive Neuroscience, 23*, 1567–1586.
Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience, 17*, 137–153.
Bi, Y., Yu, X., Geng, J., & Alario, F.-X. (2010). The role of visual form in lexical access: Evidence from Chinese classifier production. *Cognition, 116*, 101–109.
Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition, 123*, 84–99.
Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*, 177–208.
Caramazza, A., Miozzo, M., Costa, A., Schiller, N. O., & Alario, F.-X. (2001). A cross-linguistic investigation of determiner production. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 209–226). Boston: MIT Press.
Cheng, L. L.-S., & Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry, 30*, 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). New York, NY: Oxford University Press.
Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences, 106*, 21442–21446.
Cubelli, R., Lotto, L., Paolieri, D., Girelli, M., & Job, R. (2005). Grammatical gender is selected in bare noun production: Evidence from the picture–word interference paradigm. *Journal of Memory and Language, 53*, 42–59.
Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*, 283–321.
Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124–142.
Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*, 313–349.
Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*, 604–614.
Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287–314.
Dell'Acqua, R., Sessa, P., Peressotti, F., Mulatti, C., Navarrete, E., & Grainger, J. (2010). ERP evidence for ultra-fast semantic processing in the picture–word interference paradigm. *Frontiers in Psychology, 1*, 177.
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.
Doetjes, J. S. (1997). *Quantifiers and selection. On the distribution of quantifying expressions in French, Dutch and English* (Doctoral dissertation). Leiden University.
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, Eugene, Oregon (pp. 399–436)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Fang, F. (1985). An experiment on the use of classifiers by 4-to 6-year-olds. *Acta Psychologica Sinica, 17*, 384–392.
Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.
Ganushchak, L., Christoffels, I., & Schiller, N. O. (2011). The use of electroencephalography in language production research: A review. *Frontiers in Psychology, 2*, 208.
Glaser, W. R. (1992). Picture naming. *Cognition, 42*, 61–105.
Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology – Human Perception and Performance, 10*, 640–654.
Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology, 2*, 255.
Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.
Janssen, N., Carreiras, M., & Barber, H. A. (2011). Electrophysiological effects of semantic context in picture and word naming. *Neuroimage, 57*, 1243–1250.
Jescheniak, J. D., Hahne, A., & Schriefers, H. (2003). Information flow in the mental lexicon during speech planning: Evidence from event-related brain potentials. *Cognitive Brain Research, 15*, 261–276.
Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience, 14*, 951–964.
Koester, D., & Schiller, N. O. (2008). Morphological priming in overt language production: Electrophysiological evidence from Dutch. *Neuroimage, 42*, 1622–1630.
Kreher, D. A., Holcomb, P. J., & Kuperberg, G. R. (2006). An electrophysiological investigation of indirect semantic priming. *Psychophysiology, 43*, 550–563.
La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory and Cognition, 16*, 400–410.
La Heij, W., Mak, P., Sander, J., & Willeboordse, E. (1998). The gender-congruency effect in picture-word tasks. *Psychological Research, 61*, 209–219.
Levelt, W. J. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1–22.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999a). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999b). Multiple perspectives on word production. *Behavioral and Brain Sciences, 22*, 61–69.
Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge: MIT Press.
Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54*, 146–157.
MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.
Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology – Learning, Memory, and Cognition, 33*, 503–535.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.
Miozzo, M., & Caramazza, A. (1999). The selection of determiners in noun phrase production. *Journal of Experimental Psychology – Learning, Memory, and Cognition, 25*, 907–922.
Miozzo, M., Costa, A., & Caramazza, A. (2002). The absence of a gender congruency effect in romance languages: A matter of

stimulus onset asynchrony? *Journal of Experimental Psychology — Learning, Memory, and Cognition, 28*, 388–391.

Nickels, L., Biedermann, B., Fieder, N., & Schiller, N. O. (2015). The lexical-syntactic representation of number. *Language, Cognition and Neuroscience, 30*, 287–304.

Oostenveld, R., Fries, R., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.

Peyraube, A. (1998). On the history of classifiers in archaic and medieval Chinese. In B. T'sou (Ed.), *Studia linguistica serica* (pp. 39–68). Hong Kong: City University of Hong Kong.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition, 47*, 59–87.

Roelofs, A., & Meyer, A. S. (1998). Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology — Learning, Memory, and Cognition, 24*, 922–939.

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language, 48*, 169–194.

Schiller, N. O., & Caramazza, A. (2006). Grammatical gender selection and the representation of morphemes: The production of Dutch diminutives. *Language and Cognitive Processes, 21*, 945–973.

Schmitt, B. M., Münte, T. F., & Kutas, M. (2000). Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology, 37*, 473–484.

Schriefers, H. (1993). Syntactic processes in the production of noun phrases. *Journal of Experimental Psychology — Learning, Memory, and Cognition, 19*, 841–850.

Schriefers, H., & Teruel, E. (2000). Grammatical gender in noun phrase production: The gender interference effect in German. *Journal of Experimental Psychology — Learning, Memory, and Cognition, 26*, 1368–1377.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R. J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica, 119*, 159–187.

Spalek, K., Damian, M. F., & Bölte, J. (2013). Is lexical selection in spoken word production competitive? Introduction to the special issue on lexical competition in language production. *Language and Cognitive Processes, 28*, 597–614.

Starreveld, P., & La Heij, W. (2004). Phonological facilitation of grammatical gender retrieval. *Language and Cognitive Processes, 19*, 677–711.

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology, 2*, 356.

Team, R. C. (2014). *R: A language and environment for statistical computing*. Vienna, Austria, 2014. URL: http://www.R-project.org.

Tzeng, O. J., Chen, S., & Hung, D. L. (1991). The classifier problem in Chinese aphasia. *Brain and Language, 41*, 184–202.

Van Berkum, J. J. (1997). Syntactic processes in speech production: The retrieval of grammatical gender. *Cognition, 64*, 115–152.

Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology — Learning, Memory, and Cognition, 23*, 787–806.

Van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during Speaking: From syntax to phonology in 40 milliseconds. *Science, 280*, 572–574.

Wang, L., Bastiaansen, M., & Yang, Y. (2015). ERP responses to person names as a measure of trait inference in person perception. *Social Neuroscience, 10*, 89–99.

Wang, L., Guo, J., Bui, Y., & Shu, H. (2006). Classifier congruency effect in the production of noun phrases. *Studies of Psychology and Behavior, 4*, 34–38.

Wang, S.-Y. (1973). The Chinese language. *Scientific American, 228*, 51–60.

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

Zhang, J.-J., & Liu, H.-Y. (2009). The lexical access of individual classifiers in language production and comprehension. *Acta Psychologica Sinica, 41*, 580–593.

Zhu, X., Damian, M. F., & Zhang, Q. (2015). Seriality of semantic and phonological processes during overt speech in Mandarin as revealed by event-related brain potentials. *Brain and Language, 144*, 16–25.

Appendix. Stimuli used in the experiment.

Target picture	Classifier	Distractor type			
		Semantically related		Semantically unrelated	
		Classifier congruent	Classifier incongruent	Classifier congruent	Classifier incongruent
兔子 tu4zi0 *rabbit*	只 *zhi1*	企鹅 qi3e2 *penguin*	马 ma3 *horse*	袖子 xiu4zi0 *sleeve*	雨伞 yu3san3 *umbrella*
刀 dao1 *knife*	把 *ba3*	叉子 cha1zi0 *fork*	碗 wan3 *bowl*	扇子 shan4zi0 *hand fan*	雪茄 xue3jia2 *cigar*
裤子 ku4zi0 *pants*	条 *tiao2*	围巾 wei2jin1 *scarf*	雨衣 yu3yi1 *raincoat*	路 lu4 *road*	白菜 bai2cai4 *Chinese cabbage*
古琴 gu3qin2 *traditional Chinese musical instrument*	把 *ba3*	琵琶 pi2pa2 *traditional Chinese musical instrument*	大鼓 da4gu3 *traditional Chinese musical instrument*	火 huo3 *fire*	箭头 jian4tou2 *arrowhead*
叶子 ye4zi0 *leaf*	片 *pian4*	花瓣 hua1ban4 *petal*	树枝 shu4zhi1 *branch*	废墟 fei4xu1 *ruins*	夫妻 fu1qi1 *couple*
吉他 ji2ta1 *guitar*	把 *ba3*	二胡 er4hu2 *traditional Chinese musical instrument*	鼓 gu3 *drum*	斧子 fu3zi0 *axe*	毛笔 mao2bi3 *writing brush*
壁虎 bi4hu3 *lizard*	只 *zhi1*	章鱼 zhang1yu2 *octopus*	公牛 gong1niu2 *bull*	梨 li2 *pear*	词典 ci2dian3 *dictionary*
大衣 da4yi1 *coat*	件 *jian4*	毛衣 mao2yi1 *sweater*	帽子 mao4zi0 *hat*	行李 xing2li3 *luggage*	拖把 tuo1ba3 *mop*
小提琴 xiao3ti2qin2 *violin*	把 *ba3*	木琴 mu4qin2 *Xylophone*	钢琴 gang1qin2 *piano*	锁 suo3 *lock*	飞机 fei1ji1 *airplane*
手 shou3 *hand*	只 *zhi1*	脚 jiao3 *foot*	头 tou2 *head*	鸭子 ya1zi0 *duck*	书桌 shu1zhuo1 *desk*
手指 shou3zhi3 *finger*	根 *gen1*	脚趾 jiao3zhi3 toe	指甲 zhi3jia3 *nail*	木头 mu4tou2 *wood*	奶酪 nai3lao4 *cheese*
支票 zhi1piao4 *check*	张 *zhang1*	钞票 chao1piao4 *bank note*	硬币 ying4bi4 *coin*	嘴 zui3 *mouth*	椅子 yi3zi0 *chair*
教堂 jiao4tang2 *church*	座 *zuo4*	寺庙 si4miao4 *temple*	银行 yin2hang2 *bank*	山 shan1 *mountain*	彩虹 cai3hong2 *rainbow*
松鼠 song1shu3 *squirrel*	只 *zhi1*	猴子 hou2zi0 *monkey*	驴 lu:2 *donkey*	股票 gu3piao4 *stock*	眼镜 yan3jing4 *glasses*
河 he2 *river*	条 *tiao2*	小溪 xiao3xi1 *brook*	海 hai3 *sea*	毛巾 mao2jin1 *towel*	蚊子 wen2zi0 *mosquito*
耳朵 er3duo3 *ear*	只 *zhi1*	眼睛 yan3jing1 *eye*	头发 tou2fa4 *hair*	天鹅 tian1e2 *swan*	火车 huo3che1 *train*
蛇 she2 *snake*	条 *tiao2*	龙 long2 *dragon*	猪 zhu1 *pig*	街 jie1 *street*	牙齿 ya2chi3 *teeth*
照片 zhao4pian4 *photo*	张 *zhang1*	相纸 xiang4zhi3 *photographic paper*	相机 xiang4ji1 *camera*	门票 men2piao4 ticket	足球 zu2qiu2 *football*

(continued)

Target picture	Classifier	Distractor type			
		Semantically related		Semantically unrelated	
		Classifier congruent	Classifier incongruent	Classifier congruent	Classifier incongruent
牛 niu2 *cow*	头 *tou2*	狮子 shi1zi0 *lion*	鳗鱼 man4yu2 *eel*	大蒜 da4suan4 *garlic*	马车 ma3che1 *carriage*
石头 shi2tou0 *stone*	块 *kuai4*	玉 yu4 *jade gemstone*	沙 sha1 *sand*	肉 rou4 *meat*	手套 shou3tao4 *glove*
纸 zhi3 *paper*	张 *zhang1*	地图 di4tu2 *map*	笔 bi3 *pen*	床 chuang2 *bed*	汤 tang1 *soup*
老鼠 lao3shu3 *mouse*	只 *zhi1*	猫 mao1 *cat*	猛兽 meng3shou4 *beast*	靴子 xue1zi0 *boot*	火柴 huo3chai2 *match*
蛋糕 dan4gao1 *cake*	块 *kuai4*	饼干 bing3gan1 *cookie*	冰淇淋 bing1qi1lin2 *ice cream*	肌肉 ji1rou4 *muscle*	电脑 dian4nao3 *computer*
卫生纸 wei4sheng1zhi3 *toilet paper*	卷 *juan3*	画纸 hua4zhi3 *drawing paper*	餐巾纸 can1jin1zhi3 *paper napkin*	胶卷 jiao1juan3 *camera film*	萝卜 luo2bo0 *radish*
螃蟹 pang2xie4 *crab*	只 *zhi1*	虾 xia1 *shrimp*	鲤鱼 li3yu2 *common carp (type of fish)*	耳环 er3huan2 *earring*	镜子 jing4zi0 *mirror*
西红柿 xi1hong2shi4 *tomato*	个 *ge4*	柠檬 ning2meng2 *lemon*	葱 cong1 *Welsh onion*	包 bao1 *bag*	墙 qiang2 *wall*
钢笔 gang1bi3 *fountain pen*	支 *zhi1*	铅笔 qian1bi3 *pencil*	尺子 chi3zi0 *ruler*	箭 jian4 *arrow*	钥匙 yao4shi0 *key*
香蕉 xiang1jiao1 *banana*	根, *gen1*	甘蔗 gan1zhe4 *sugar cane*	葡萄 pu2tao2 *grape*	汗毛 han4mao2 *fine hair*	灯塔 deng1ta3 *lighthouse*
衬衫 chen4shan1 *shirt*	件 *jian4*	衣服 yi1fu2 *clothes*	裙子 qun2zi0 *dress*	艺术品 yi4shu4pin3 *art piecework*	珠子 zhu1zi0 *bead*
袜子 wa4zi0 *sock*	只 *zhi1*	鞋 xie2 *shoe*	上衣 shang4yi1 *top (clothing)*	青蛙 qing1wa1 *frog*	礼物 li3wu4 *gift*