Processing Mandarin Chinese classifiers as a lexico-syntactic feature during noun phrase production

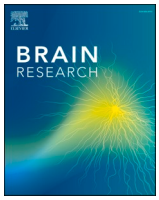
Wang, J.; Witteman, J.; Schiller, N.O.

Brain Research 1868 (2025) 149995

Contents lists available at ScienceDirect

Brain Research

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



Research paper

Processing Mandarin Chinese classifiers as a lexico-syntactic feature during noun phrase production

Jin Wang [a,*,1], Jurriaan Witteman [a,2], Niels O. Schiller [a,b,3]

[a] Leiden University Centre for Linguistics (LUCL), Leiden University, Leiden, The Netherlands
[b] Department of Linguistics and Translation, City University of Hong Kong, Kowloon Tong, Hong Kong

ARTICLE INFO

Keywords:
Picture naming
Picture-word interference
Lexico-syntactic feature
Classifier
N400

ABSTRACT

During speech production, lexico-syntactic features associated with nouns (e.g., grammatical gender, classifiers, number) are assumed to be automatically activated. Although previous studies have provided evidence for this assumption by examining *classifier congruency effects*, empirical validation of this mechanism in Mandarin Chinese remains limited. The present study investigated whether a *classifier congruency effect* can be reliably elicited during noun phrase production in Mandarin and explored how this effect relates to semantic processing. We employed a picture-word interference (PWI) paradigm, incorporating several methodological refinements. Both classifier congruency and semantic relatedness between the target and distractor words were manipulated. Behavioural results replicated the *semantic interference effect*, with longer naming latencies observed for semantically related distractors than semantically unrelated ones. Although no main effect of classifier congruency was found, a significant interaction with semantic relatedness emerged. Classifier incongruency led to delayed naming under semantically related conditions. ERP results further revealed that both the *semantic interference* and *classifier congruency effects* peaked within the N400 time window. These findings provide further evidence that classifier information is automatically activated as a lexico-syntactic feature during lemma access, and that this activation is influenced by semantic processing. The present results contribute both conceptually and methodologically to advancing our understanding of classifier processing in Mandarin Chinese.

1. Introduction

While natural speech unfolds linearly in time, the underlying structure of language is hierarchical. In many languages, content words in a sentence constrain the selection and morphological form of function words. For example, in German, all nouns are categorised into three grammatical genders (masculine, feminine, or neuter). The grammatical gender of a noun specifies the form of preceding determiners. In the phrase *das Wasser* ("the water"), the neuter noun *Wasser* requires the neuter determiner *das*, rather than the masculine *der* or the feminine *die*. A comparable system exists in Mandarin Chinese, wherein nouns are required to be paired with classifiers in quantifier-classifier phrases, subject to both syntactic and semantic constraints. In the phrase *yì běn zázhì* ("one + classifier + magazine"), the classifier *běn* must be used for book-like objects. Grammatical gender in Indo-European languages and classifiers in Mandarin Chinese constitute lexico-syntactic features hypothesised to be stored in the mental lexicon alongside lemmas (Levelt et al., 1999a). While extensive research has examined the processing of grammatical gender in Indo-European languages (for a review, see Wang & Schiller, 2019), the mechanisms underlying classifier processing in Mandarin Chinese remain notably limited. The present study builds on prior research by adopting well-established paradigms, introducing novel experimental materials and analytical methods to examine the cognitive processing of classifiers in Mandarin Chinese.

1.1. Retrieval of lexico-syntactic features during noun phrase production

Major language production models propose that speech involves

* Corresponding author at: Leiden University Centre for Linguistics, Reuvensplaats 3-4, 2311 BE Leiden, The Netherlands.
E-mail addresses: jin_wang_c@163.com (J. Wang), j.witteman@hum.leidenuniv.nl (J. Witteman), nschille@cityu.edu.hk (N.O. Schiller).

[1] Research interests: psycholinguistics, neurolinguistics, second language acquisition.
[2] Research interests: meta-analysis, neuroimaging, neurolinguistics, open science, replication.
[3] Research interests: psycho- and neurolinguistics. Syntactic, morphological, and phonological processes in language production and reading aloud. Articulatory-motor processes during speech production, language processing in neurologically impaired patients, and forensic phonetics.

https://doi.org/10.1016/j.brainres.2025.149995
Received 10 May 2025; Received in revised form 5 October 2025; Accepted 9 October 2025
Available online 10 October 2025

three stages: the conceptualisation stage, the formulation stage, and the articulation stage (e.g., Bock & Levelt, 1994; Caramazza, 1997; Dell, 1986, 1988; Garrett, 1975, 1980; Levelt, 1989, 1992, 1999; Levelt et al., 1999b; Oppenheim et al., 2010; Roelofs, 1997, 2000; Roelofs & Ferreira, 2019; for an overview, see Griffin & Ferreira, 2006). The *WEAVER++ model* (Levelt et al., 1999a) proposes that lexico-syntactic features are activated at the lemma stratum — an intermediate layer between the conceptual and word-form strata. During picture naming, activation spreads from the concept to the associated lemma and subsequently to the connected lexico-syntactic features (e.g., grammatical gender, number, or classifier). If these features are overtly produced, activation further spreads to their word form, enabling retrieval of appropriate determiners or classifiers. In contrast, Caramazza's (1997) *Independent Network model* suggests three distinct networks — lexical-semantic, syntactic, and phonological networks. In this model, the lexical-semantic network can activate the syntactic and phonological networks independently. The model argues for parallel activation of syntactic and phonological networks directly from semantics, positing that lexico-syntactic feature processing may be bypassed if not phonologically instantiated (e.g., when German nouns of different grammatical genders exhibit identical determiners in the plural form).

1.2. Gender congruency effect

Research into the processing mechanisms of lexico-syntactic features has initially concentrated on grammatical gender. Schriefers (1993) investigated the processing of grammatical gender using the *picture-word interference (PWI)* paradigm in Dutch. In this experimental paradigm, a distractor word is superimposed onto a target picture. Participants are required to name the target pictures verbally while ignoring the distractor words. A *gender congruency effect* was observed, whereby naming latencies increased when the target and distractor nouns differed in grammatical gender. This finding suggests that the grammatical gender information of distractors is automatically activated during the process of lemma retrieval, thereby competing with the grammatical gender nodes of the target nouns for selection. Schiller and Caramazza (2003) challenged this interpretation (i.e., *gender selection interference hypothesis*, GSIH), proposing a *determiner selection interference hypothesis* (DSIH): the observed effect arises not from grammatical gender activation per se but from competition between determiners. They found that the *gender congruency effect* diminished when the determiners of both the target and distractor stimuli were congruent, even in the presence of grammatical gender incongruency. Hence, they refer to this phenomenon as the *determiner congruency effect*. Although behavioural studies yield mixed interpretations, electrophysiological evidence supports the activation of grammatical gender during noun phrase production. Bürki et al. (2016) observed differences in ERP signals around 210 ms before articulation onset between gender-congruent and gender-incongruent conditions (*mean RT* = 798 ms). Together, these results suggest that grammatical gender information may be automatically activated and selected as a lexico-syntactic feature during noun phrase production.

1.3. Classifier congruency effect

The investigation of lexico-syntactic feature processing has also gained traction in research on Mandarin Chinese. Although Mandarin lacks rich morphological inflections and displays more flexible syntax, it features a classifier system similar in function to grammatical gender (Adams & Faires Conklin, 1973; Allan, 1977; Contini-Morava & Kilarski, 2013; Kilarski, 2013). In Mandarin Chinese, nouns are quantified through quantifier-classifier phrases (i.e., quantifier + classifier + noun), where the presence of a classifier is mandatory. The syntactic position of the classifier is usually fixed. Nouns may pair with various classifiers to further specify the quantity or form of the referent, enriching its meaning (Wang et al., 2025c; Zhang and Liu, 2009). For example, *yì zhī yáng* means "one sheep", while *yì qún yáng* means "a herd of sheep", indicating different quantities. The different classifiers in *yì dī shuǐ* ("a drop of water") and *yì tān shuǐ* ("a puddle of water") highlight the distinctions in the form of water.

Chinese classifiers can be categorised into several subtypes, such as individual classifiers (e.g., *duǒ* in *yì duǒ huā*, "one + classifier + flower"), group classifiers (e.g., *qún* in *yì qún rén*, "one + classifier + people"), and partition classifiers (e.g., *duàn* in *yí duàn cōng*, "one + classifier + green onion") (He, 2000). Individual classifiers denote a single unit of a person or object. Apart from the general classifier *gè*, which can be used with a wide range of nouns, most individual classifiers have specific collocational relationships with nouns. Although a noun may take different classifiers depending on context or pragmatic purpose, it generally pairs with a dominant individual classifier (Wang et al., 2025c). This study focuses on individual classifiers, which are referred to simply as classifiers in the following sections, unless otherwise specified.

Several studies have used the PWI paradigm to explore whether classifiers are activated during noun phrase production in a manner analogous to grammatical gender (Huang & Schiller, 2021; Li et al., 2006; Zhang & Liu, 2009). These studies revealed that naming latencies increased when the classifiers of the distractor and target noun were incongruent, demonstrating a *classifier congruency effect*. This suggests that classifier information is automatically activated during lemma access. Unlike grammatical gender processing, which typically occurs in the P600 window (e.g., Foucart & Frenck-Mestre, 2011; Gunter et al., 2000; Hagoort & Brown, 1999), the *classifier congruency effect* is often reflected in N400-like ERP responses (Huang & Schiller, 2021; Wang et al., 2019). The N400 component is generally associated with semantic processing (for a review, see Kutas & Federmeier, 2011), suggesting that classifier activation may be more semantically influenced than grammatical gender.

The extent to which classifier processing engages semantic or syntactic processing remains a matter of ongoing debate in both language comprehension and language production research (for reviews, see Qian, in press; Wang and Schiller, in press). In the studies by Wang et al. (2019) and Huang and Schiller (2021), both the congruency of the classifiers and the semantic relatedness between the target and distractor nouns were manipulated. The experiments also revealed a *semantic interference effect* (for a review, see Bürki et al., 2020), wherein naming latencies were longer when the distractor and the target noun belonged to the same semantic category, compared to semantically unrelated conditions. The behavioural effect was mirrored in the ERP data as an N400 component. This raises the question of whether the N400-like effects elicited by the *semantic interference effect* and the *classifier congruency effect* reflect similar underlying cognitive processes.

The categorisation of nouns by classifiers is primarily semantically driven. The pairing between a noun and a classifier must be consistent in semantic features such as the animacy, function, shape and size (Allan, 1977; Tai, 1994; Tai & Chao, 1994; Zhang & Schmitt, 1998). The classification of nouns according to these semantic features sometimes aligns with the semantic categories. For instance, the classifier *tái* is typically used for machines, whereas *liàng* is used for vehicles. The similar N400-like ERPs elicited by the *classifier congruency effect* and the *semantic interference effect* in previous studies (Huang & Schiller, 2021; Wang et al., 2019) seem to suggest that classifier processing is closely tied to the processing of semantic category information. However, the precise nature of this relationship and potential distinctions between the two ERP components remain underexplored in the literature.

1.4. Limitations of previous studies

Although several studies have reported evidence for a *classifier congruency effect* in Mandarin and have suggested that classifier information is automatically activated and selected during noun phrase production, these findings so far come from a relatively small number of investigations (Huang & Schiller, 2021; Li et al., 2006; Wang et al., 2019;

Table 1
Examples of distractors presented with the target noun "猴子 (*monkey*, classifier 只 zhi)" in all conditions.

	semantically related (S+)	**semantically unrelated (S−)**
classifier congruent (C+)	熊猫 (*panda*, 只 zhi)	袜子 (*socks*, 只 zhi)
classifier incongruent (C−)	马 (*horse*, 匹 pi)	门票 (*ticket*, 张 zhang)

Wang and Schiller, submitted; Zhang & Liu, 2009). Furthermore, some aspects of the experimental design and data analysis in these studies leave room for improvement. One issue concerns the use of the general classifier *gè*, which was included in the stimulus materials of some previous experiments. As a result of its extensive grammaticalization, *gè* has lost much of its original semantic content and is widely used with nearly all nouns in spoken Mandarin (Myers, 2000). This presents two potential problems: first, the processing of *gè* may rely primarily on syntactic routines, which distinguishes it from more semantically specific classifiers (Frankowsky et al., 2022; Qian & Garnsey, 2016). Second, stimuli in the classifier-incongruent condition may not have been truly incongruent, given *gè's* broad compatibility.

Another limitation relates to the analytical methods employed. Most prior studies used traditional average-based analyses, which are limited in their capacity to account for variation across participants and items or to handle unbalanced datasets. In comparison, (generalised) linear mixed-effects models (GLMMs/LMMs) have a better handle on missing data, larger statistical power, better control of the type I errors and allow for generalisation across items (Baayen et al., 2017; Barr, 2013; Frömer et al., 2018; Matuschek et al., 2017).

Finally, time windows and regions of interest (ROIs) in prior studies are typically chosen based on prior assumptions or findings. While this approach is widely used, it carries the risk of overlooking subtle effects that may occur in adjacent time windows or spatial areas. In contrast, a data-driven approach using permutation tests across all epochs may provide a clearer and more objective understanding of when and where effects emerge (Voeten, 2023a, 2023b).

Taken together, while the *classifier congruency effect* has been reported in Mandarin, the limited number of studies and methodological considerations suggest that further investigation is warranted. The present study seeks to contribute to this line of research by employing refined experimental materials and more detailed analytical methods to explore whether the *classifier congruency effect* can be reliably observed during noun phrase production in native Mandarin speakers.

1.5. The current study

The present study aims to replicate and extend the findings of Huang and Schiller (2021) by implementing three critical methodological updates (for detailed methodological differences between the two studies, see supplementary Table S.3). First, the experimental materials were revised to exclude the general classifier *gè*, ensuring that the classifier-incongruent condition remained valid and was not confounded by the pervasive use of *gè*. Classifiers were selected based on corpus-derived co-occurrence frequencies with target nouns, ensuring that they were dominant classifiers retaining semantic content and showing strong alignment with the semantic features of the nouns. This adjustment enables a more fine-grained examination of the potential interaction between the semantic features of classifiers and semantic categories of nouns. Second, we implemented a permutation-based and data-driven approach to identify temporally and spatially relevant EEG windows, enabling unbiased determination of time intervals and electrodes. By combining permutation testing, scalp topography analysis, and ERP component modelling, we systematically compared the neurophysiological signatures associated with classifier congruency and semantic relatedness, elucidating the relationship between classifier processing and semantic processing. Third, we used (generalised) linear mixed-effects models (GLMMs/LMMs) at the single-trial level. This approach retains more information while controlling for both subject-level and item-level variability, thereby increasing the interpretability and robustness of the analysis.

The experiment manipulated semantic relatedness and classifier congruency between target and distractor nouns using a PWI paradigm. We expect to observe both *semantic interference* and *classifier congruency effects* based on previous findings (Huang & Schiller, 2021; Wang and Schiller, submitted). Specifically, we predicted that semantically related distractors would lead to longer naming latencies than unrelated ones and that naming would be slower in classifier-incongruent conditions than classifier-congruent conditions. In the EEG data, we expect a more negative-going ERP component in the N400 time window for semantically unrelated distractors relative to related ones. Meanwhile, we expect classifier-incongruent trials to elicit more negative voltage amplitudes than classifier-congruent trials.

2. Methods

2.1. Participants

Thirty-five native Mandarin Chinese speakers (eight males and twenty-seven females) were recruited from the University of Münster in Germany. All participants were proficient in either English or German as a second language. Three participants also spoke Cantonese or Wu Chinese, but they primarily used Mandarin Chinese in their daily lives. The average age of the participants was 26.33 years (SD = 3.28), and six of them were left-handed. All participants reported having normal or corrected-to-normal vision, with no history of neurological, psychological, or language impairments. Informed consent was obtained prior to the experiment, and participants were provided with a debriefing form after completing the experiment, in accordance with the Ethics Code for Linguistic Research at the Faculty of Humanities. Participants received monetary compensation for their participation. Five participants were excluded due to insufficient valid data.

2.2. Materials

Twenty-five black-and-white line drawings representing objects used in daily life were selected from Liu's picture database (Liu et al., 2011) and used as target pictures in the picture naming task. The names of these pictures correspond to monosyllabic (36 %) or disyllabic (64 %) words in Mandarin Chinese. Four targets were identical to those in Huang and Schiller (2021), and five targets overlapped with those used by Wang et al. (2019). Each target picture was assigned four distractors. The distractors were paired with target words depending on whether they shared the same classifier as the target word or whether they belonged to the same semantic category as the target word, resulting in four experimental conditions (for the example stimuli, see Table 1; for the complete stimulus list, see Table S.2 in the supplementary materials), i.e., classifier-congruent and semantically-related (C+S+) condition, classifier-incongruent and semantically-related (C−S+) condition, classifier-congruent and semantically-unrelated (C+S−) condition, classifier-incongruent and semantically-unrelated (C−S−) condition. A proportion of 12 % of the distractor words overlapped with Huang and Schiller (2021), while 26 % overlapped with Wang et al. (2019).

Unlike Huang and Schiller (2021), who selected classifiers from a dictionary, the noun-classifier pairings in this study were retrieved from the BCC corpus (Xun et al., 2016). Since a noun may occur with multiple classifiers, the most frequent one for each noun was selected as the expected response in noun phrase production. On average, its collocation frequency (*mean collocation frequency* = 2,234.94, *SD* = 4,077.03) was 5.53 times higher than that of the second most frequent classifier for the same noun (*mean collocation frequency* = 536.01, *SD* = 9,918.19). All

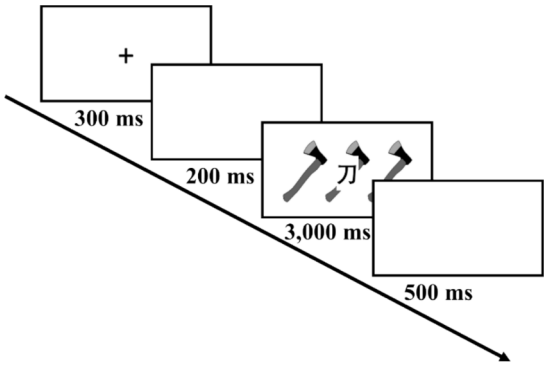


Fig. 1. Sequence of stimulus presentation.

noun-classifier pairs were reviewed by two native speakers of Mandarin Chinese with academic training in linguistics prior to the experiment. Based on feedback from a pilot study, any pairings deemed implausible or unnatural were excluded from the final set of experimental materials. Twenty-four native speakers of Mandarin Chinese (*Mean age* = 23.13 years, *SD* = 3.31; 10 male, 14 female), who did not participate in the other tasks of this study, evaluated the final set of experimental items. They rated the acceptability of each classifier–noun pairing on a 7-point scale, with 1 indicating least acceptable and 7 indicating most acceptable. The mean rating score was 6.11 (*SD* = 0.67).

The semantic relatedness of each pair of distractor and target words was assessed by fifteen native Chinese speakers who did not participate in the naming task. Ratings were made on a 7-point Likert scale, with higher scores indicating a stronger perception that the two words belong to the same semantic category. Statistical analysis of rating scores was conducted using the *clmm()* function from the *ordinal* package (Christensen, 2023) in RStudio Version 2024.04.2 + 764 (R Core Team, 2023). Fixed factors included classifier congruency and semantic relatedness, modelled using mixed-effects ordinal regression. Model selection followed a backward elimination procedure starting with the maximal random-effects structure. The final best-fitting model and its parameters are detailed in Table A.1 of Appendix A. The results demonstrated that there was a significant difference in rating scores between the semantically related ($M = 6.36$, $SD = 1.04$) and semantically unrelated ($M = 1.60$, $SD = 1.12$) conditions ($\beta = 3.831$, $SE = 0.359$, $z = 10.666$, *95 % CI* [3.127, 4.535], $p < 0.001$), whereas there was no significant difference between the classifier-congruent condition ($M = 4.07$, $SD = 2.59$) and the classifier-incongruent ($M = 3.92$, $SD = 2.64$) condition ($\beta = 0.070$, $SE = 0.059$, $z = 1.180$, *95 % CI* [−0.046, 0.185], $p = 0.238$).

To control for potential confounds, several lexical and visual features of the distractors were matched across conditions (for details, see Table S.1 in the supplementary materials). Results of *Kruskal-Wallis tests* indicated that distractors across the four conditions did not show significant differences in word frequency ($H(3) = 1.252$, $p = 0.741$, *95 % CI* [5515.998, 9584.582]), visual complexity determined by the number of strokes ($H(3) = 1.526$, $p = 0.676$, *95 % CI* [12.505, 14.715]), number of syllables ($H(3) = 2.095$, $p = 0.553$, *95 % CI* [1.651, 1.849]) and phrase frequency of noun-classifier collocations ($H(3) = 7.835$, $p = 0.050$, *95 % CI* [1105.028, 2274.792]). The word frequency data were obtained from the Chinese Lexical Database (Sun et al., 2018). In the present study, phrase frequency refers to how frequently the most used classifier for a given noun appears in quantifier-classifier phrases. The frequency of co-occurrence may affect the degree of classifier activation (Wang et al., 2025c). Therefore, we additionally controlled for this potential confounding factor, differing from previous research (Huang & Schiller, 2021; Wang et al., 2019). The phrase frequency data were retrieved and extracted from the BCC corpus (Xun et al., 2016). Last, distractors were not phonologically or orthographically related to the corresponding target nouns.

2.3. Design and procedure

This experiment followed a 2 × 2 within-subjects design, with classifier congruency (C) and semantic relatedness (S) as two fixed factors. Each of the four conditions (C+S+, C−S+, C+S−, C−S−) included 25 items. Participants were instructed to name all target pictures using noun phrases of the form "quantifier + classifier + noun" during the picture naming task. Either two or three identical target pictures were randomly presented for each target to minimise potential confounds from repeatedly naming the same number. Consequently, each participant completed a total of 200 experimental trials.[4] Eight additional trials were provided for warming up.

The presentation order of trials was pseudo-randomised using the Windows program Mix (Van Casteren & Davis, 2006) program, ensuring that trials with identical conditions, classifiers, or syllables were not presented consecutively. Trials with the same number of pictures could appear at most twice in a row. Additionally, the minimum distance between any two identical target words was ten trials. Also, the minimum distance between any two target words in the same semantic category was three trials. Each participant was presented with the stimuli in a different pseudo-random order.

The experiment was implemented in E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). The experimental procedure followed that of Huang and Schiller (2021), and comprised three sessions: a familiarisation session, a practice session, and an experimental session. During the familiarisation session, all target pictures were presented sequentially on the screen for 3000 ms, along with their corresponding names. Participants were instructed to indicate their familiarity with the pictures and target noun phrases by pressing a designated key. In the practice session, a string of letters ("XXXX") was superimposed on each target picture, and participants were instructed to ignore it while naming the picture using a noun phrase within 3000 ms. The experimenter provided corrections for any errors during this phase. The experimental session followed the same structure, except that distractor words replaced the letter strings (see Fig. 1). Each trial began with a fixation cross ("+") displayed for 300 ms, followed by a blank screen for 200 ms. The target picture, along with the distractor, was then shown for 3000 ms, followed by a final blank screen for 500 ms. Vocal responses were recorded automatically at the onset of each target picture using E-Prime 2.0 software. Throughout the experiment, EEG data were recorded simultaneously. In total, there were 200 trials distributed evenly across four blocks, with each block starting with two warm-up trials.

2.4. EEG recordings and data pre-processing

EEG data were recorded using a mobile Active-Two BioSemi system (BioSemi, Amsterdam) installed and configured in a controlled linguistic laboratory environment. The system and its setup were identical to those used by Huang and Schiller (2021). Thirty-two Ag/AgCl active electrodes were positioned on the EEG cap according to the standardised international 10/20 system (see Appendix C). In addition, six external electrodes were used: two were placed at the outer canthi of the eyes to record horizontal electrooculogram (HEOG), two were positioned above and below the left eye to record vertical electrooculogram (VEOG), and two were attached to the left and right mastoids to allow for offline re-referencing. The Common Mode Sense (CMS) and Driven Right Leg (DRL) electrodes served as the online reference and ground, respectively, to reduce noise and enhance signal quality. EEG signals were sampled at 512 Hz.

[4] Brysbaert and Stevens (2018) suggest that a well-powered reaction time experiment should have a minimum of 1,600 observations per condition. In the present study, each condition consisted of 1,750 observations, satisfying this requirement.

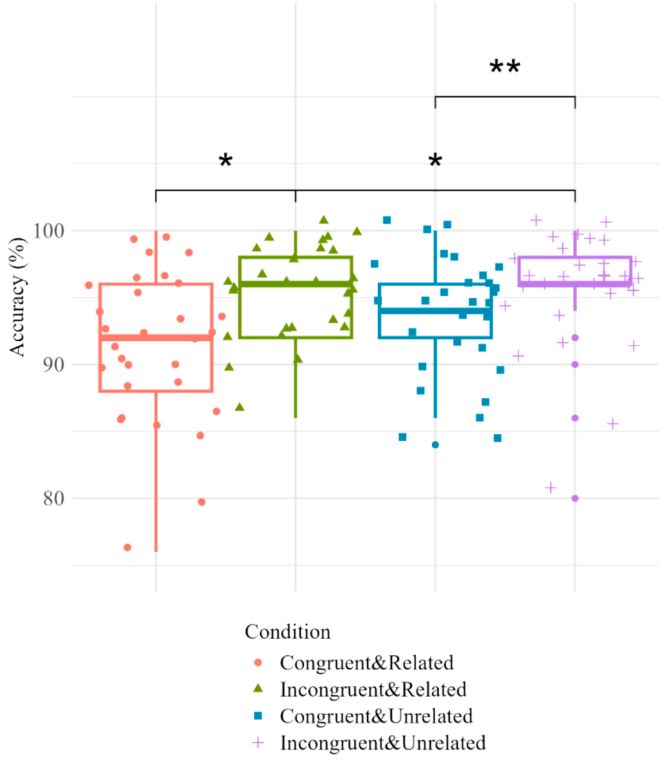


Fig. 2. Naming accuracy (%) for each condition.

EEG data pre-processing and ERP extraction were performed offline using Brain Vision Analyzer (Version 2.2.2, Brain Products GmbH, Gilching, Germany), following the procedure outlined in Huang and Schiller (2021) and Von Grebmer zu Wolfsthurn et al. (2021). Raw EEG signals were re-referenced to the average of the mastoid electrodes and band-pass filtered from 0.1 to 30 Hz. A notch filter at 50 Hz was applied to eliminate powerline interference. Noisy channels, constituting between 3.13 % and 12.5 % of electrodes per participant (mean = 6.92 %), were corrected using spherical spline interpolation. Ocular artefacts were identified through combined HEOG and VEOG channels using linear derivation and corrected via independent component analysis (ICA). Trials with voltage fluctuations exceeding ± 100 μV or containing other artefacts were excluded. Epochs were segmented for correctly named trials, from −200 ms to 800 ms relative to picture onset. Baseline correction was applied based on the mean voltage in the 200 ms pre-stimulus interval. Valid epochs were exported for statistical analysis. Five participants were excluded from further analysis due to insufficient valid trials (<60 %), resulting in a final dataset of thirty participants.

2.5. Data analysis

2.5.1. Behavioural data analysis

Audio recordings from each trial were annotated and manually reviewed offline using Praat 6.3.08 (Boersma, 2001) to extract the naming accuracy and latency (measured from stimulus onset to voice onset). Resulting behavioural data were then analysed using a single-trial modelling approach via the R package *lme4* (Bates et al., 2015b). Naming accuracy was modelled using generalised linear mixed models (GLMMs) via the *glmer()* function with a binomial distribution. Naming latencies, which showed positive skew, were analysed using *glmer()* with an inverse Gaussian distribution. Fixed-effect predictors in the models included Classifier Congruency and Semantic Relatedness, both sum-coded, with the classifier-incongruent and semantically unrelated conditions serving as reference levels, respectively. Random effects initially included random intercepts for participants and items and random slopes for each fixed effect by participants and by items. A backward elimination strategy was applied to refine the random-effects structure. Simplification was carried out when models failed to converge or additional random effects did not significantly improve model fit (Bates et al., 2015a). Model comparison and selection were performed using the *anova()* function, guided by a combination of Akaike's Information Criterion (AIC; Akaike, 1974), Bayesian Information Criterion (BIC; Neath & Cavanaugh, 2012), and log-likelihood ratio tests (Lewis et al., 2011). Model diagnostics included residual plots to assess homoscedasticity and normality. Post-hoc analyses of the interaction effects were conducted using the *emmeans* package (Lenth, 2024).

2.5.2. EEG data analysis

Different from Wang et al. (2019) and Huang and Schiller (2021), we did not predefine the time windows and electrodes for analysis. We first performed a permutation test on the ERP data before statistical modelling to examine the temporospatial distribution of *classifier congruency* and *semantic interference effects*. Using the *permutes* package (Voeten, 2023b), we computed F-values across all electrodes within the 0–700 ms time window relative to stimulus onset. Given that 700 ms post-stimulus onset is approaching the onset of articulation, as evidenced by the behavioural results, the time window for detecting the effects was restricted to before this time point. To assess spatial patterns in the effects, we introduced a factor of Anteriority and grouped electrodes into three regions: anterior (AF3, AF4, F7, F8, F3, F4, Fz), central (FC5, FC6, FC1, FC2, C3, C4, CP5, CP6, CP1, CP2, Cz) and posterior (P7, P8, P3, P4, PO3, PO4, O1, O2, Pz, Oz). Based on the results of the permutation analysis, we identified time windows and regions of interest (ROIs) and then conducted statistical modelling using single-trial linear mixed-effects models (LMMs) using the *lmer()* function (Amsel, 2011; Frömer et al., 2018). Unlike previous studies (Huang & Schiller, 2021; Wang et al., 2019) that analysed averaged ERPs using ANOVAs, this method accounts for both by-subject and by-item variance, providing greater explanatory power (Baayen et al., 2017; Barr, 2013; Frömer et al., 2018; Matuschek et al., 2017). Fixed effects included Classifier Congruency, Semantic Relatedness, and Anteriority (all sum-coded). The random-effects structure mirrored the approach used in the behavioural analysis, with backward elimination applied to determine the best-fitting model. Post-hoc tests were conducted to determine interaction effects. Previous studies (Huang & Schiller, 2021; Wang et al., 2019) have shown that both classifier congruency and semantic relatedness can elicit N400-like components. To further assess the similarity between these components, we analysed and compared the peak latencies of the two effects using a gamma-distributed *glmer()* model. The fixed factor *Effect* had two levels: the semantic interference effect and the classifier congruency effect. Subject and Electrode were included as random effects. The model selection procedure was identical to that used in the voltage amplitude analysis.

3. Results

3.1. Behavioural data exclusion

To maintain consistency with the EEG datasets, five participants were excluded from the behavioural data analysis, whereby a total of thirty datasets were retained. From a total of 6000 recorded trials collected from the thirty participants, we further excluded 1355 data points (22.58 %) when analysing the naming latencies. The exclusions were implemented in accordance with the following criteria: (1) 394 responses (6.57 %) were excluded due to the use of incorrect nouns or classifiers and the absence of responses; (2) 50 trials (0.89 %) were excluded for exhibiting naming latencies exceeding 2000 ms or falling below 200 ms; (3) 79 trials (1.32 %) were identified as outliers, given that their naming latencies exceeded three standard deviations from the mean latency for each participant and item; (4) further exclusions were made based on EEG data (as detailed in Section 3.3). As a result, a total

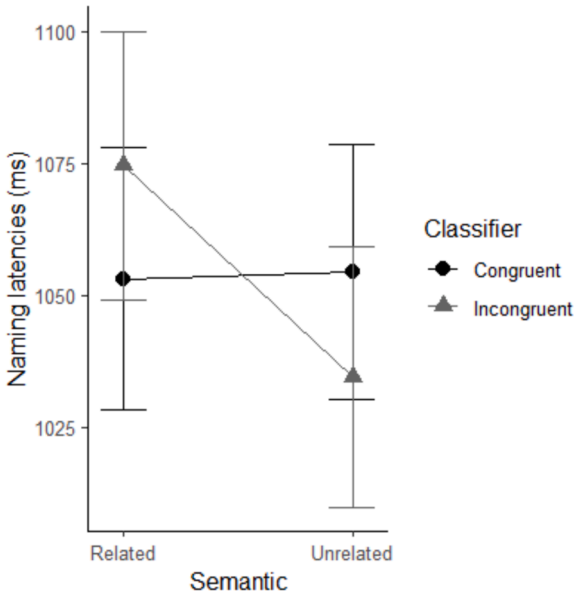


Fig. 3. Naming latencies (in ms) for each condition.

of 4645 trials (77.42 %) remained for subsequent analysis.

3.2. Behavioural data results

3.2.1. Naming accuracy

The best-fitting model (see Table A.2 of Appendix A) included the main effects of Classifier Congruency, Semantic Relatedness, and random intercepts for both subjects and items. The analysis revealed (see Fig. 2) a significant main effect of Classifier Congruency, with naming accuracy significantly lower for classifier-congruent conditions ($M = 0.921$, $SD = 0.269$) compared to incongruent ($M = 0.947$, $SD = 0.223$) conditions ($\beta = -0.240$, $SE = 0.055$, $z = -4.361$, *95 % CI* $[-0.348, -0.132]$, $p < 0.001$). A significant effect of Semantic Relatedness was also found, where naming accuracy was lower for semantically related items ($M = 0.925$, $SD = 0.263$) than for unrelated ($M = 0.943$, $SD = 0.23$) items ($\beta = -0.169$, SE = 0.055, $z = -3.068$, *95 % CI* $[-0.276, -0.061]$, $p = 0.002$). The interaction between Classifier Congruency and Semantic Relatedness did not reach statistical significance ($\beta = 0.036$, $SE = 0.055$, $z = 0.648$, *95 % CI* $[-0.072, 0.143]$, $p = 0.517$).

3.2.2. Naming latencies

The best-fitting model of naming latencies, as shown in Table A.3 of Appendix A and Fig. 3, indicated a significant main effect of Semantic Relatedness ($\beta = 9.388$, $SE = 4.325$, *95 % CI* $[0.909, 17.867]$, $p = 0.03$). Specifically, naming latencies were longer in the semantically related condition ($M = 1029$ ms, $SD = 248$) compared to the semantically

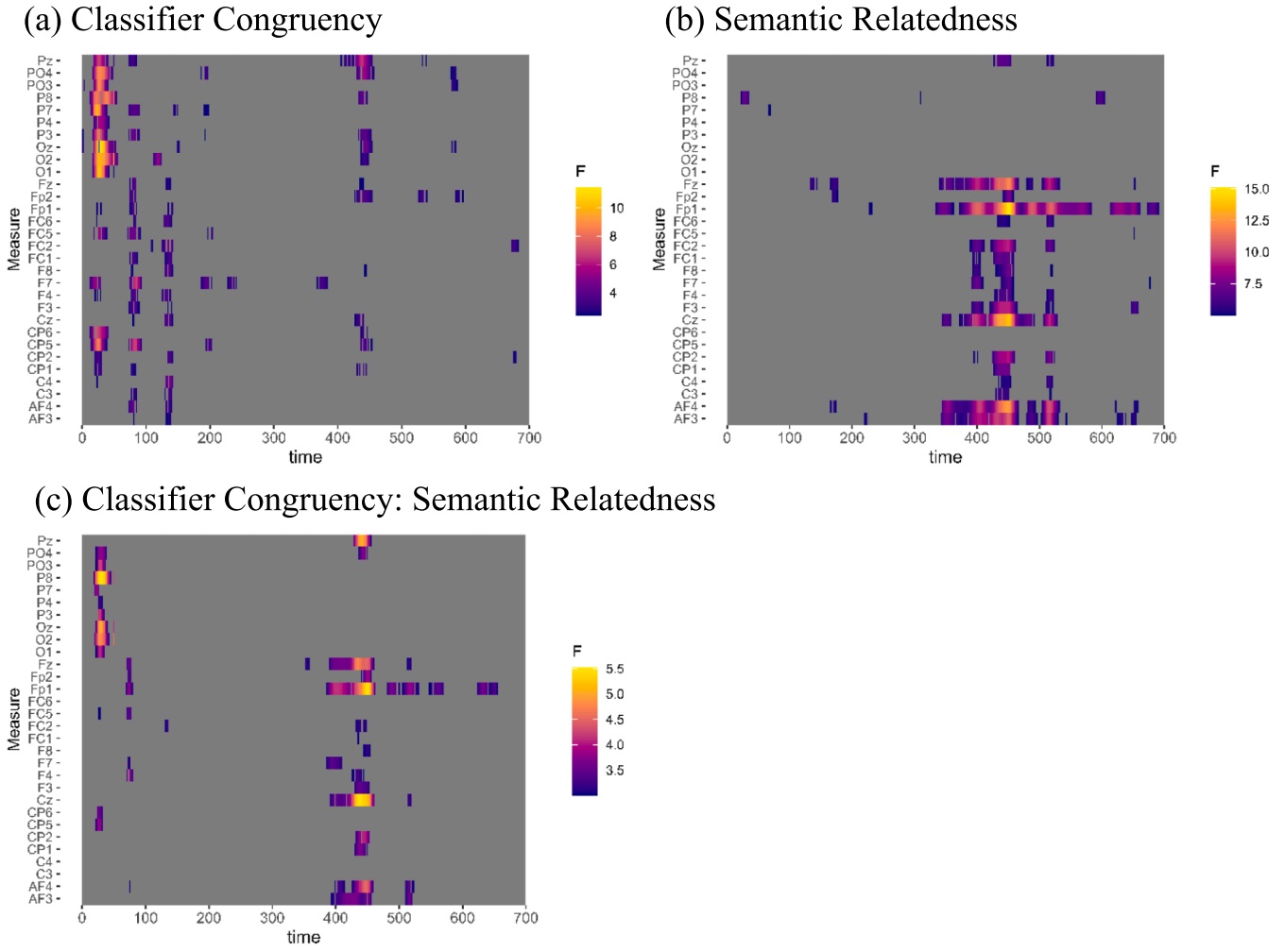


Fig. 4. Outcomes of the permutation test performed on all electrodes for the 0 – 700 ms time window relative to stimulus onset.

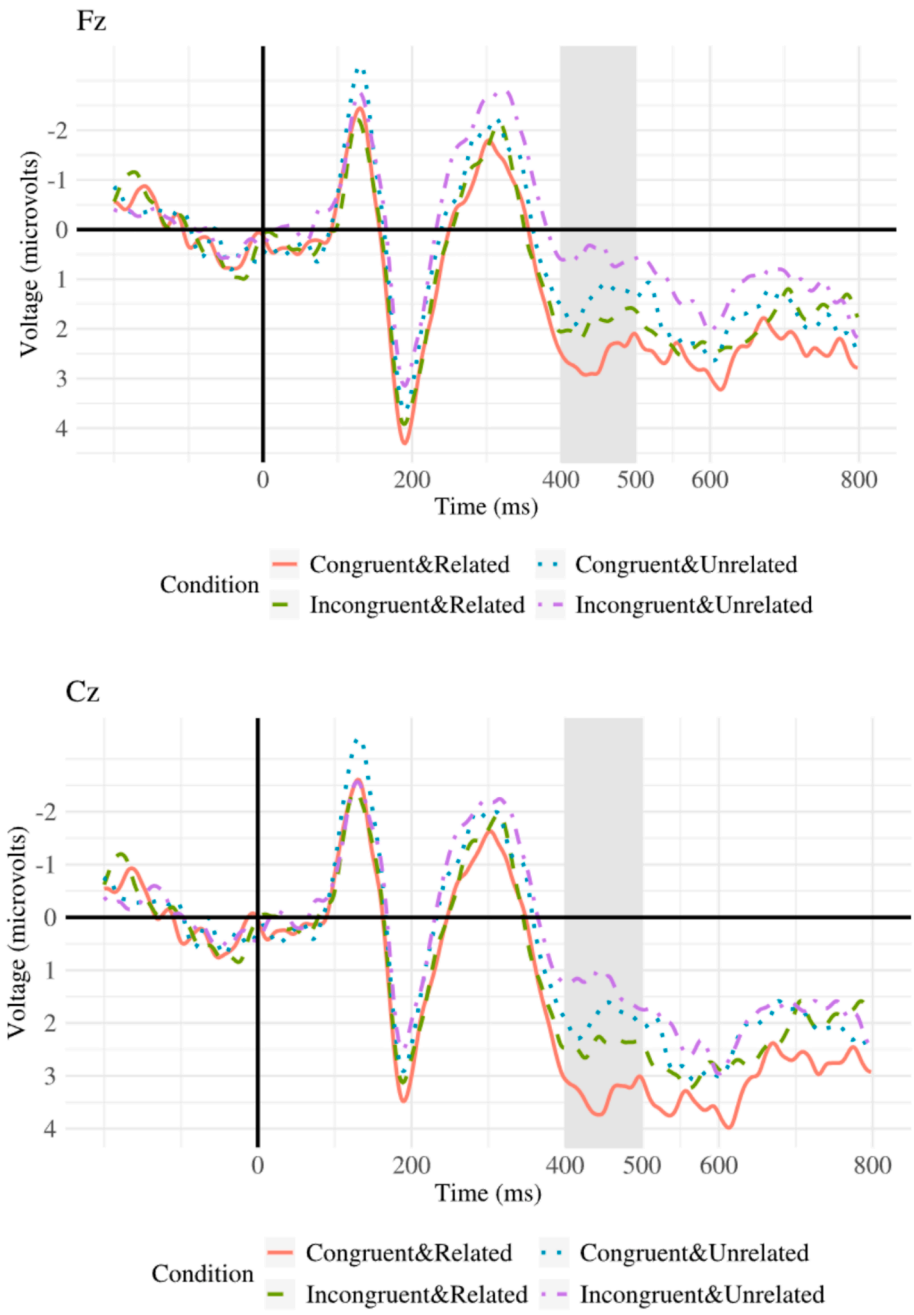


Fig. 5. Grand averages of ERPs from representative electrodes (Fz, Cz, Pz, Oz) for all conditions.

unrelated condition ($M = 1013$ ms, $SD = 241$). No main effect was found for Classifier Congruency — naming latencies did not differ significantly between classifier-congruent ($M = 1021$ ms, $SD = 248$) and classifier-incongruent ($M = 1021$ ms, $SD = 241$) conditions ($\beta = -0.161$, $SE = 4.153$, *95 % CI* [−8.303, 7.980], $p = 0.969$). A significant interaction between Classifier Congruency and Semantic Relatedness was observed ($\beta = -10.580$, $SE = 5.286$, *95 % CI* [−20.944, −0.217], $p = 0.045$). Post-hoc analyses showed that, for semantically related items, naming latencies were significantly shorter for classifier-congruent trials ($M = 1020$ ms, $SD = 249$) than incongruent ($M = 1038$ ms, $SD = 247$) trials ($\beta = -21.5$, $SE = 10.5$, $z = -2.043$, *95 % CI* [−42.1, −0.872], $p = 0.041$). For semantically unrelated pairs, however, there was no significant difference in naming latencies between classifier-congruent ($M = 1022$ ms, $SD = 247$) and incongruent ($M = 1004$ ms, $SD = 234$) trials ($\beta = 20.8$, $SE = 15.8$, $z = 1.315$, *95 % CI* [−10.2, 51.887], $p = 0.188$).

3.3. EEG data exclusion

The EEG data analysis was conducted following the same exclusion criteria as those applied to the behavioural analysis, including trials with incorrect responses and outliers. 15.02 % of the EEG data was contaminated by artefacts, which were removed during data pre-processing. Thirty datasets were analysed with the same fixed effects as in the behavioural analysis.

3.4. EEG data results

Visual inspection of the permutation test results indicated a potential modulatory effect of Classifier Congruency in the frontal region (Fz, F8) and centro-parietal region (Pz, PO4, P8, P3, Oz, O2, Cz, CP5) between 400 ms and 500 ms post-stimulus onset, as well as an effect of Semantic

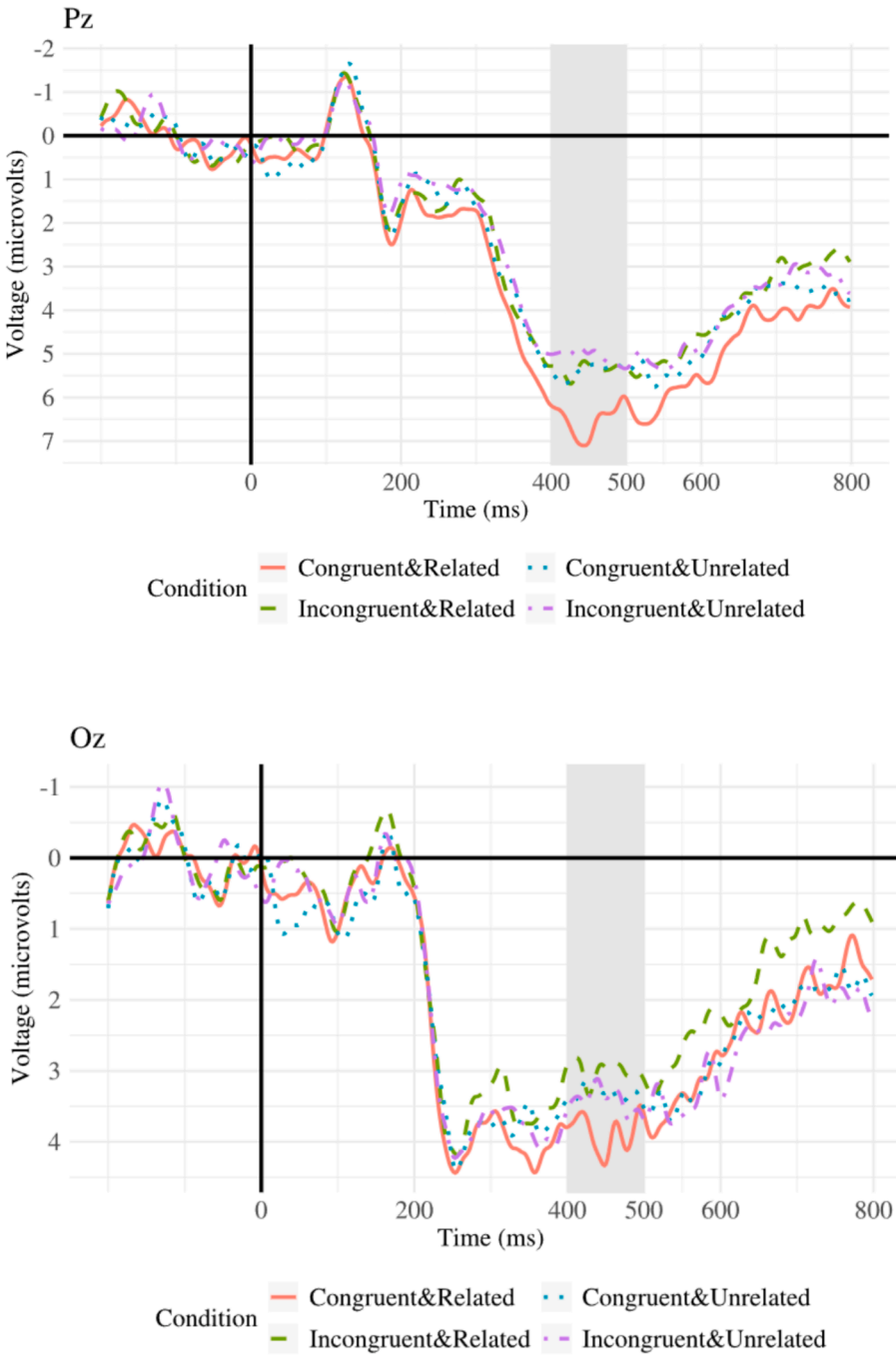


Fig. 5. (*continued*).

Relatedness in the fronto-central region (Fz, FC2, FC1, F8, F7, F4, F3, Cz) emerging between 350 ms and 550 ms post-stimulus onset (see Fig. 4).

We modelled the voltage amplitudes from 400 to 500 ms post-stimulus onset based on the time window identified in the permutation test for the potential *classifier congruency effect*. The best-fitting model (see Table A.4 of Appendix A) showed a significant main effect of Semantic Relatedness ($\beta = 0.285$, $SE = 0.139$, $t = 2.053$, *95 % CI* [0.013, 0.557], $p = 0.040$) with less positive voltage amplitudes for semantically unrelated conditions ($M = 2.984$ μV, $SD = 12.771$) compared to related conditions ($M = 3.616$ μV, $SD = 13.015$); There was a trend for the *classifier congruency effect*, with less positive voltage amplitudes for classifier-incongruent conditions ($M = 3.020$ μV, $SD = 12.836$) than classifier-congruent ($M = 3.584$ μV, $SD = 12.953$) conditions ($\beta = 0.248$, $SE = 0.134$, $t = 1.850$, *95 % CI* [0.015, 0.511], $p = 0.064$). The interaction between Classifier Congruency and Anteriority was significant, $F(2, 3{,}622{,}948) = 4.114$, $p = 0.016$. Post-hoc analysis showed that the effect of Classifier Congruency was significant in the anterior region ($\beta = 0.533$, $SE = 0.268$, $z = 1.984$, *95 % CI* [0.007, 1.061], $p = 0.047$), but not in the others. The interaction between Semantic Relatedness and Anteriority was significant, $F(2, 3{,}622{,}948) = 298.997$, $p < 0.001$. Post-hoc analysis showed that the effect of Semantic Relatedness was significant in the anterior ($\beta = 0.834$, $SE = 0.278$, $z = 3.000$, *95 % CI* [0.289, 1.379], $p = 0.003$) and central regions ($\beta = 0.721$, $SE = 0.278$, $z = 2.592$, *95 % CI* [0.176, 1.267], $p = 0.010$). Post-hoc analyses of the three-way interaction between Classifier Congruency, Semantic Relatedness, and Anteriority revealed that, in the anterior region, the *classifier congruency effect* was significant in the semantically unrelated condition, $\beta = 0.903$, $SE = 0.322$, $z = 2.808$, *95*

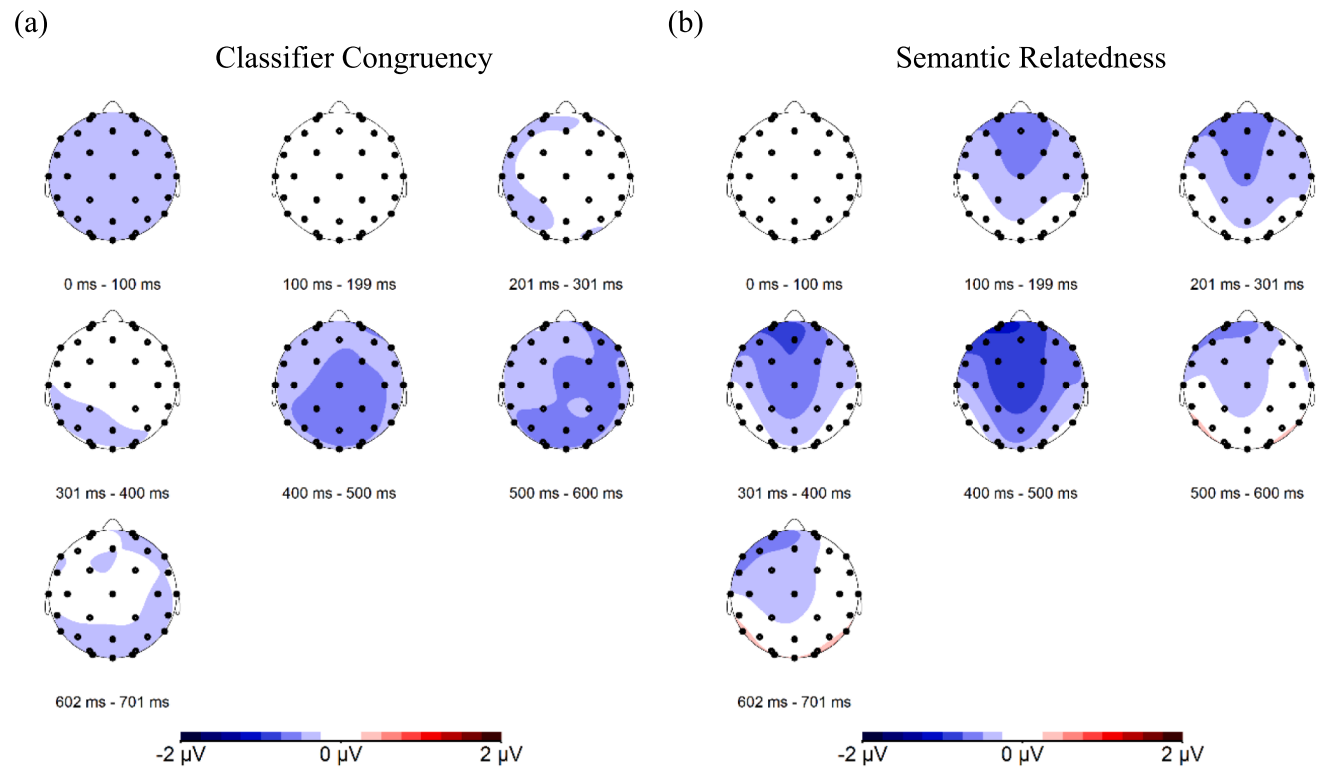


Fig. 6. Scalp topographies every 100 ms after stimulus onset. (a) The difference in voltage amplitudes between the classifier-incongruent condition and the classifier-congruent condition. (b) The difference in voltage amplitudes between the semantically unrelated condition and the semantically related condition.

% CI [0.273, 1.533], $p = 0.005$. The classifier-incongruent condition ($M = 0.977$ μ*V*, $SD = 13.553$) elicited more negative amplitudes than the classifier-congruent condition ($M = 1.869$ μ*V*, $SD = 13.415$). In the semantically related condition, the classifier-incongruent condition ($M = 2.198$ μ*V*, $SD = 13.489$) also elicited more negative amplitudes than the classifier-congruent condition ($M = 2.537$ μ*V*, $SD = 14.065$), but this difference was not significant, $\beta = 0.165$, $SE = 0.391$, $z = 0.421$, *95 % CI* [−0.602, 0.931], $p = 0.674$. Fig. 5 presents the grand mean of the ERPs on a selection of electrodes. Fig. 6 provides a visual presentation of the effect of Classifier Congruency and Semantic Relatedness using scalp topography.[5]

Since both the *classifier congruency effect* and the *semantic interference effect* elicited N400-like components, we further compared the peak latencies of these two effects (see Fig. 7). Specifically, within the 350–550 ms time window, we identified the peak latency of the voltage amplitude difference for the classifier congruency and semantic relatedness effects, respectively. The peak latency of the classifier congruency effect occurred at 455.16 ms after stimulus onset ($SD = 60.43$), slightly later than that of the semantic interference effect, which occurred at 446.71 ms ($SD = 60.51$). However, the model results (see Table A.5 in Appendix A) indicated no significant difference in peak latency between the two effects.

4. Discussion

The present study aimed to investigate the processing of classifier information during noun phrase production in Mandarin Chinese. It specifically examined the effects of classifier congruency and semantic relatedness utilising a picture-word interference (PWI) paradigm. We introduced several methodological refinements based on previous research (Huang & Schiller, 2021; Li et al., 2006; Wang et al., 2019; Zhang & Liu, 2009). These included the exclusion of highly grammaticalised general classifiers (e.g., *gè*), selecting the classifier that most frequently co-occurs with each noun from a corpus, applying single-trial linear mixed-effects modelling and adopting a permutation-based approach for defining EEG time windows in a data-driven manner. By implementing these improvements, we reassess the robustness of the *classifier congruency effect* and gain further insights into its relation to semantic processing.

Consistent with previous findings (e.g., Dell'Acqua et al., 2010; Huang & Schiller, 2021; Krott et al., 2019; Rose et al., 2019; Wang et al., 2019; Zhu et al., 2015), the current study replicated a reliable *semantic interference effect*: naming latencies were significantly longer when the target and distractor words belonged to the same semantic category compared to when they were unrelated. In the electrophysiological data, the semantically unrelated condition elicited more negative voltage amplitudes than the semantically related condition within the N400 time window in the fronto-central region, a component typically associated with semantic processing difficulty (Kutas & Federmeier, 2011). Regarding the *classifier congruency effect*, although no significant main effect was found on naming latencies across all conditions, a significant interaction between classifier congruency and semantic relatedness emerged. Specifically, classifier incongruency delayed naming significantly when the distractors were semantically related to the target nouns but not when the distractors were semantically unrelated. In the ERP data, although the main effect of classifier congruency only approached statistical significance across all regions within the 400–500 ms time window, further regional analyses revealed a significant *classifier congruency effect* in the anterior scalp region. In this region, classifier congruency interacted with semantic relatedness. *The classifier congruency effect* was significant in the semantically unrelated condition but did not

[5] Visual inspection of the ERP waveforms and scalp topography also revealed an N1–P1–N2 complex that may be influenced by semantic relatedness and classifier congruency. These ERP components are typically associated with the presentation of visual stimuli and are not relevant to the purpose of the present study and will therefore not be discussed further.

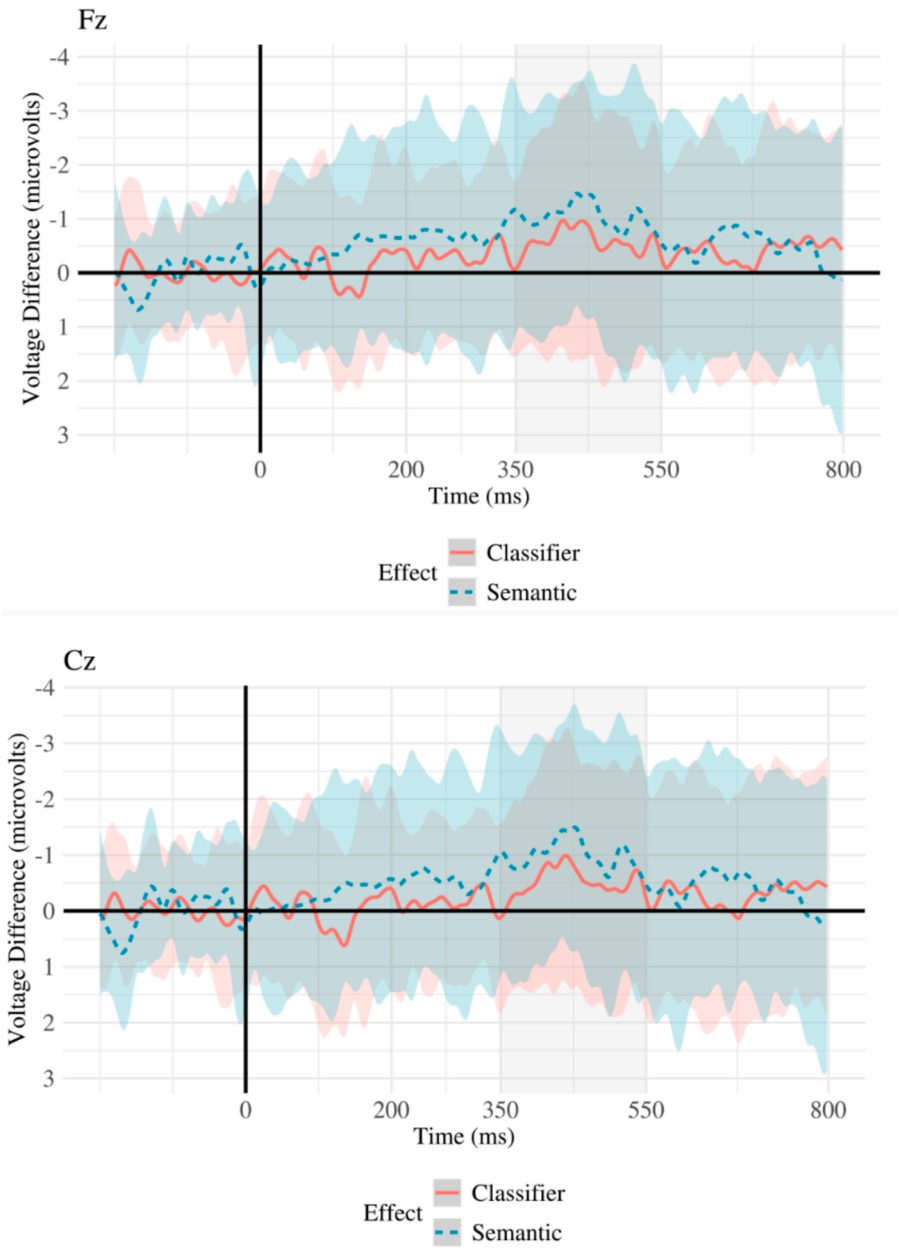


Fig. 7. Difference waves elicited by the classifier congruency effect and the semantic interference effect. The figure shows the four main electrodes (Fz, Cz, Pz, and Oz). Shaded areas indicate the standard deviations.

reach significance in the semantically related condition.

The *semantic interference effect* observed in behavioural and ERP data aligns with a large body of literature demonstrating semantic processing during lexical retrieval (Bürki et al., 2020). According to the prominent model of lexical access (Levelt et al., 1999a; Roelofs, 1992), when participants attempt to name a target picture, activation initially spreads from the conceptual representation of the target to semantically related concepts within the network (see also Bloem & La Heij, 2003; Lupker, 1979; Roelofs, 1992; Schnur et al., 2006). This activation then spreads to the corresponding lemma nodes. Consequently, when distractors are semantically related to the target, their lexical nodes are activated to a higher level compared to unrelated distractors. During lemma selection, these co-activated lexical candidates compete with the target lemma, leading to increased selection difficulty and extended naming latencies.

The present study also observed a *classifier congruency effect*, which was weaker than the *semantic interference effect*. Under the semantically related condition, stimuli in the classifier-congruent condition were named significantly faster than those in the classifier-incongruent condition. This difference in naming latency is consistent with the findings of Huang and Schiller (2021). Wang et al. (2019) propose that the *classifier congruency effect* is similar to the *gender congruency effect* observed in some Indo-European languages. According to their account, the *classifier congruency effect* reflects the activation of classifiers as lexico-syntactic features during lemma access, thereby supporting the WEAVER++ model (Levelt et al., 1999a). In this framework, the associated classifier nodes are automatically activated when the target and the distractor are processed. When these classifiers are incongruent, this activation leads to competition for selection, which is behaviourally observed as prolonged naming latencies. Alternatively, Caramazza's (1997) Independent Network model suggests that lexico-syntactic features are activated when the corresponding lexical forms need to be specified. To produce classifier-noun phrases in the present study, the

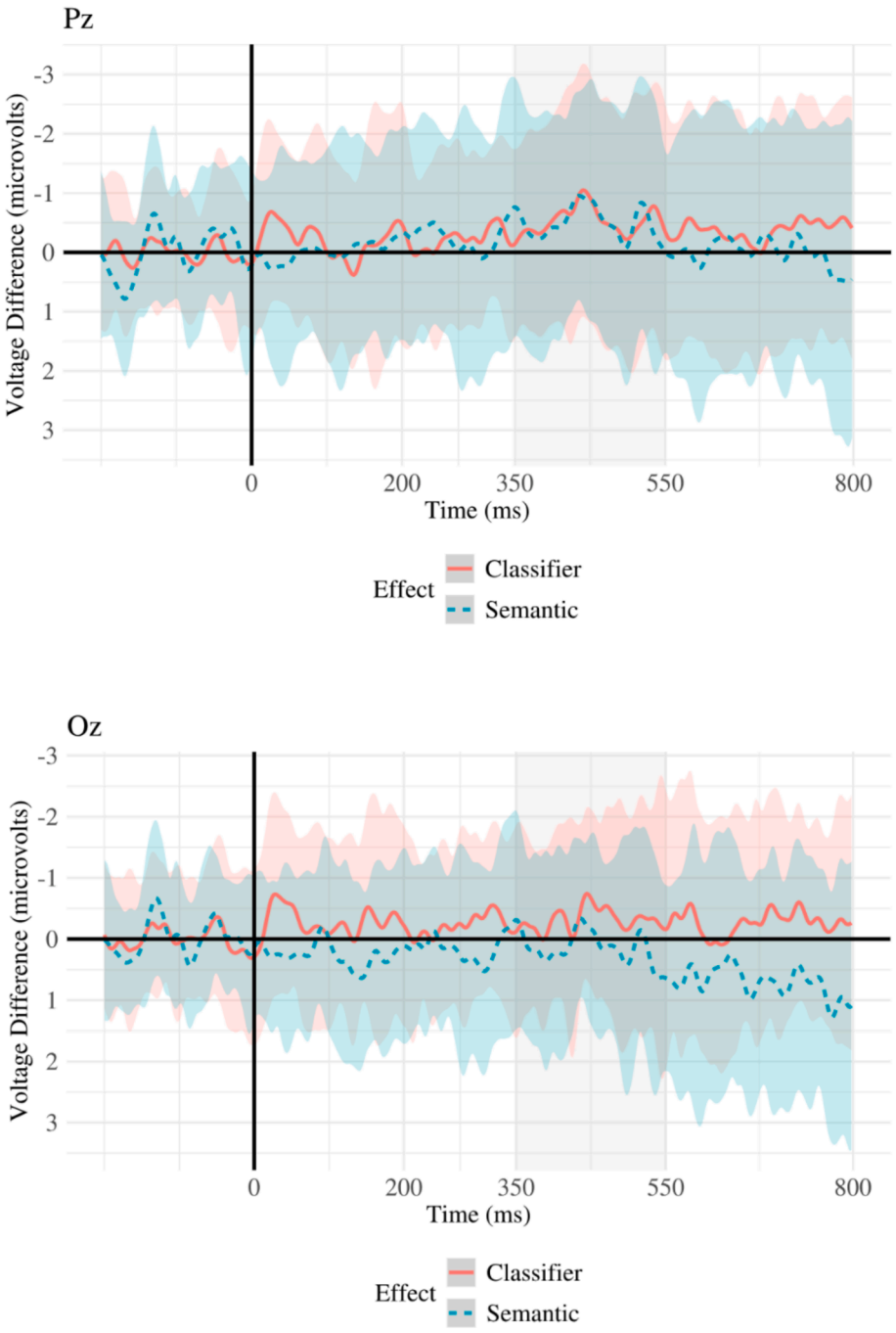


Fig. 7. (*continued*).

classifier must be retrieved and produced. The observed *classifier incongruency effect* suggests that the distractor's classifier was also automatically activated during noun phrase production, consistent with the *Independent Network* model. When the classifiers associated with the target and distractor words are incongruent, their activation may lead to competition during lexical form encoding, resulting in increased processing time.

For the EEG results, the *semantic interference effect* and the *classifier congruency effect* were generally reported within the N400 time window (Huang & Schiller, 2021; Wang et al., 2019). The ERP data from the present study corroborate these previous findings. The permutation-test results and scalp topography suggest that the *semantic interference effect* emerged slightly earlier than the *classifier congruency effect*. However, their peak latencies did not differ significantly. The N400-like component elicited by classifier congruency was mainly observed over frontal regions, while that elicited by semantic relatedness was found over both frontal and central regions. The two effects overlapped in onset window, peak latency, and scalp distribution. ERP components with such distributions are typically associated with semantic information processing (Kutas & Federmeier, 2011). The classifiers used in this study retained a certain degree of semantic content. They can categorise nouns according to the consistency between the semantic features of the nouns and the classifiers. This function is similar to the way nouns are grouped into different semantic categories according to their meanings. Therefore, the processing of classifiers might be related to the processing of semantic category information in the present study. The N400-like effect elicited by the semantic relatedness might arise during concept-lexical activation stages (Abdel Rahman & Melinger, 2009; Bloem &

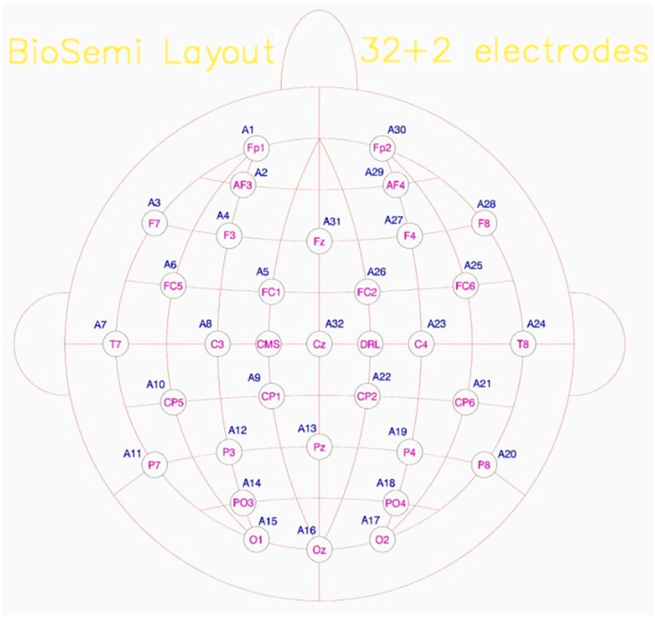


Fig. B1. 32-channels with 10/20 system layout including CMS and DRL (https://www.biosemi.com/headcap.htm).

La Heij, 2003; Zhang et al., 2016). When the semantic category of the target (e.g., fish) and distractor (e.g., road) words mismatched, this conflict imposed a cost at the stage of semantic information processing. On the other hand, the *classifier congruency effect* might emerge during lemma access and the retrieval of lexico-syntactic features (Huang & Schiller, 2021; Wang et al., 2019). Incongruency between the classifiers of the distractor and that of the target noun also involved conflicts in their semantic features, which might likewise cause difficulty in semantic processing. For example, the classifier for sugar ("*kē*") and the classifier for biscuits ("*kuài*") differ in the semantic feature of shape. Conflicts arising from a mismatch in semantic categories may be more difficult to resolve than conflicts based on specific semantic features of the classifier. We found that the semantic relatedness modulated the *classifier congruency effect* in the ERP data. When distractor and target belong to the same semantic category, category-level similarity may have masked or alleviated the difficulty of resolving the semantic conflict caused by classifier incongruency. As a result, although ERP differences were observed between classifier-congruent and incongruent conditions in the semantically related condition, this effect did not reach statistical significance. Together, the findings of the current study align with previous research showing the activation of classifiers during noun phrase production. For the classifiers retaining semantic information in this study, their processing may be related to the processing of semantic category information.

It is worth noting that the present study produced some findings that differ from previous research. The main difference is that we observed an interaction between classifier congruency and semantic relatedness. In naming latency, a significant *classifier congruency effect* was found only in the semantically related condition, while this effect was not significant in the semantically unrelated condition. This interaction pattern suggests that the *classifier congruency effect* was stronger in the semantically related condition than in the unrelated condition. Although Huang and Schiller (2021) did not report an interaction in their study on classifiers, a similar interaction has been found in a grammatical gender study using the same experimental paradigm. Schriefers (1993) reported a stronger *gender congruency effect* in the semantically related condition than in the semantically unrelated condition when examining the effects of gender congruency and semantic relatedness on naming latency. According to previous accounts of the *semantic interference effect* (Bürki et al., 2020), activation spreads from the target noun's concept node (e.g., *hóuzi*, "monkey") to semantically related concepts (e.g., *mǎ*, "horse"; *xióngmāo*, "panda"; etc.) and their corresponding lexical entries. The activation level of distractor words belonging to the same semantic category as the target may be enhanced. In contrast, distractors in the semantically unrelated condition may not receive such additional activation. Based on the assumption that activation can automatically spread from a lemma to its connected lexico-syntactic feature nodes, the activation levels of the classifier nodes for distractors in the semantically related condition might also be enhanced (e.g., classifier *pǐ* for *mǎ*; classifier *zhī* for *xióngmāo*). In other words, these distractors' classifiers might also receive extra activation. When the classifiers of the semantically related distractor and target were congruent, the highly activated classifier of the target noun could enter the form encoding stage more quickly than in the semantically unrelated condition, resulting in shorter naming latencies. When the classifiers were incongruent, the increased competition among highly activated but incongruent classifiers may have made it more difficult for the classifier of the target word to enter the lexical selection stage.

In the semantically unrelated condition, no significant effect of classifier congruency was observed on naming latency. Naming was slower in the classifier-congruent condition than in the classifier-incongruent condition. This result, which differs from previous findings, may be due to the selection of classifiers in the present study. Compared to the materials in previous studies (Huang & Schiller, 2021; Wang et al., 2019), we excluded the highly grammaticalised general classifier *gè*, which has largely lost its semantic content. In addition, noun-classifier pairs were carefully selected based on a corpus, ensuring the highest collocation frequencies. These adjustments ensured that the classifiers used in the experiment preserved semantic content and that the semantic associations between nouns and classifiers were relatively strong. In other words, there was substantial overlap in their semantic features. For example, *wéi jīn* ("scarf") is a long-shaped object, and its classifier *tiáo* has the semantic feature "long in shape." Classifiers categorise nouns based on such semantic features. In the semantically unrelated but classifier-congruent condition, the classifier-noun classification conflicted with the noun's semantic category. For instance, *yú* ("fish") and *wéi jīn* ("scarf") do not belong to the same semantic category, but both can take the classifier *tiáo*. The classifier *tiáo* groups them as long, strip-shaped objects. The EEG results in the present study showed a significant *classifier congruency effect* in the semantically unrelated condition, indicating that the classifiers of nouns were activated. Previous studies have suggested that the semantic features (e.g., shape, animacy) carried by classifiers may also be involved in classifier processing (Bi et al., 2010; Wang et al., 2025a,b). Therefore, during processing, the semantic information in classifiers that categorise nouns based on semantic features might conflict with the semantic category information of the nouns. This conflict might require additional processing time, leading to longer naming latencies in the semantically unrelated but classifier-congruent condition. Together, these findings suggested that semantic category information might modulate classifier processing. However, interaction effects often require larger sample sizes for validation. The conclusions drawn from the interaction effects in the present study should be further examined and explored in future experiments.

5. Conclusions

This study investigated whether the *classifier congruency effect* could be reliably elicited during noun phrase production in Mandarin Chinese and how this effect relates to semantic processing. By employing refined experimental materials and advanced analytical approaches, the present study provides further evidence for the automatic activation of classifiers during lemma access. Furthermore, the results suggest that semantic processing may modulate classifier activation. The processing of classifiers that retain semantic content may be related to the processing of semantic category information. Future research could build on these

findings by exploring how different classifiers (e.g., specific classifiers versus general classifiers) differentially engage semantic and syntactic processing streams.

CRediT authorship contribution statement

Jin Wang: Writing – review & editing, Writing – original draft, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jurriaan Witteman:** Writing – review & editing, Supervision, Formal analysis. **Niels O. Schiller:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research received partial funding from the China Scholarship Council (CSC) under grant number 202006200006. Niels. O. Schiller. was supported by a start-up grant of the City University of Hong Kong [grant number: 9380177]. We would like to express our gratitude to all the members of the Experimental Linguistics Laboratories at Leiden University and Münster University and our participants. Lastly, we extend our gratitude to the anonymous reviewers for their valuable contributions.

Appendix

A. Model parameters

Table A1
Model of best fit for semantic relatedness rating scores, including log-odds ratios, standard errors, confidence intervals, z-values and p-values (n = 15).

Formula: rating score ~ semantic relatedness + classifier congruency + (1 + semantic relatedness | subject) + (1 | item)

Predictors	log-*odds*	*std. error*	*95 % CI*	*z-value*	*Pr(>|z|)*
1|2	−2.677	0.263	−3.192 – −2.162	−10.191	**<0.001**
2|3	−1.158	0.259	−1.666 – −0.649	−4.463	**<0.001**
3|4	−0.185	0.258	−0.692 – 0.321	−0.718	0.473
4|5	0.049	0.258	−0.457 – 0.556	0.192	0.848
5|6	1.495	0.260	0.985 – 2.004	5.750	**<0.001**
6|7	3.218	0.264	2.700 – 3.736	12.180	**<0.001**
Semantic relatedness [related]	3.831	0.359	3.127 – 4.535	10.666	**<0.001**
Classifier congruency [congruent]	0.070	0.059	−0.046 – 0.185	1.180	0.238
Random Effects					
σ^2	3.29				
$\tau_{00\ Item}$	0.33				
$\tau_{00\ Subject}$	1.58				
$\tau_{11\ Subject.Semantic\ relatednessUnrelated}$	6.57				
$\rho_{01\ Subject}$	−0.84				
ICC	0.37				
$N_{Subject}$	15				
N_{Item}	25				
Observations	1,453				
Marginal R^2 / Conditional R^2	0.739 / 0.835				

Table A2
Model of best fit for naming accuracy, including log-odds ratios, standard errors, confidence intervals, z-values and p-values (n = 30).

Formula: naming accuracy ~ 1 + classifier congruency × semantic relatedness + (1 | subject) + (1 | item)

Predictors	log-*odds*	*std. error*	*95 %CI*	*z-value*	*Pr(>|z|)*
(Intercept)	3.198	0.234	2.739 – 3.657	13.653	**<0.001**
Classifier[Congruent]	−0.240	0.055	−0.348 – −0.132	−4.361	**<0.001**
Semantic[Related]	−0.169	0.055	−0.276 – −0.061	−3.068	**0.002**
Classifier[Congruent] × Semantic[Related]	0.036	0.055	−0.072 – 0.143	0.648	0.517
Random Effects					
σ^2	3.29				
$\tau_{00\ Subject}$	0.55				
$\tau_{00\ Item}$	0.75				
ICC	0.28				
$N_{Subject}$	30				
N_{Item}	25				
Observations	6,000				
Marginal R^2 / Conditional R^2	0.019 / 0.298				

Table A3
Model of best fit for naming latencies, including estimated means, standard errors, confidence intervals and p-values (n = 30).

Formula: naming latencies ~ 1 + classifier congruency × semantic relatedness + (1 + classifier congruency × semantic relatedness | subject) + (1 | item)

Predictors	*Estimates*	*std. error*	*95 %CI*	*Pr(>|z|)*
(Intercept)	1054.762	21.866	1,011.895 – 1,097.630	**<0.001**
Classifier[congruent]	−0.161	4.153	−8.303 – 7.980	0.969
Semantically[related]	9.388	4.325	0.909 – 17.867	**0.030**
Classifier[congruent] × Semantically[related]	−10.580	5.286	−20.944 – −0.217	**0.045**
Random Effects				
σ^2	0.01			
$\tau_{00\ Subject}$	2,106.02			
$\tau_{00\ Item}$	882.14			
$\tau_{11\ Subject.ClassifierCongruent}$	101.85			
$\tau_{11\ Subject.SemanticallyRelated}$	135.46			
$\tau_{11\ Subject.ClassifierCongruent:SemanticallyRelated}$	230.52			
$\rho_{01\ Subject.ClassifierCongruent}$	−0.06			
$\rho_{01\ Subject.SemanticallyRelated}$	0.18			
$\rho_{01\ Subject.ClassifierCongruent:SemanticallyRelated}$	−0.09			
ICC	1.00			
$N_{Subject}$	30			
N_{Item}	25			
Observations	4,645			
Marginal R^2 / Conditional R^2	0.055 / 1.000			

Table A4
The best-fitting model for voltage amplitudes in the 400–500 ms time window post-stimulus onset, including estimated means, standard errors, confidence intervals, t-values and p-values (n = 30).

Formula: voltage amplitudes ~ 1 + classifier congruency × semantic relatedness × anteriority + (1 + classifier congruency × semantic relatedness | subject) + (1 | item)

Predictors	*Estimates*	*std. error*	*95 %CI*	*t-value*	*Pr(>|t|)*
(Intercept)	3.139	0.738	1.693 – 4.586	4.255	**<0.001**
Classifier[Congruent]	0.248	0.134	−0.015 – 0.511	1.850	0.064
Semantic[Related]	0.285	0.139	0.013 – 0.557	2.053	**0.040**
Anteriority[Posterior]	1.670	0.009	1.653 – 1.687	192.010	**<0.001**
Anteriority[Central]	−0.400	0.010	−0.419 – −0.381	−41.560	**<0.001**
Classifier[Congruent] × Semantic[Related]	−0.036	0.118	−0.267 – 0.195	−0.305	0.760
Classifier[Congruent] × Anteriority[Posterior]	−0.023	0.009	−0.040 – −0.006	−2.677	**0.007**
Classifier[Congruent] × Anteriority[Central]	0.005	0.010	−0.014 – 0.024	0.489	0.625
Semantic[Related] × Anteriority[Posterior]	−0.208	0.009	−0.225 – −0.191	–23.914	**<0.001**
Semantic[Related] × Anteriority[Central]	0.076	0.010	0.057 – 0.095	7.874	**<0.001**
Classifier[Congruent] × Semantic[Related] × Anteriority[Posterior]	0.130	0.009	0.113 – 0.147	14.933	**<0.001**
Classifier[Congruent] × Semantic[Related] × Anteriority[Central]	0.019	0.010	−0.000 – 0.038	1.948	0.051
Random Effects					
σ^2			147.14		
$\tau_{00\ Subject}$			15.40		
$\tau_{00\ Item}$			0.78		
$\tau_{11\ Subject.ClassifierCongruent}$			0.54		
$\tau_{11\ Subject.SemanticRelated}$			0.58		
$\tau_{11\ Subject.ClassifierCongruent:SemanticRelated}$			0.42		
$\rho_{01\ Subject.ClassifierCongruent}$			0.24		
$\rho_{01\ Subject.SemanticRelated}$			0.05		
$\rho_{01\ Subject.ClassifierCongruent:SemanticRelated}$			0.55		
ICC			0.11		
$N_{Subject}$			30		
N_{Item}			25		
Observations			3,623,100		
Marginal R^2 / Conditional R^2			0.011 / 0.117		

Table A5
The best-fitting model for peak latencies in the 350–550 ms time window post-stimulus onset, including estimated means, standard errors, confidence intervals, and p-values (n = 30).

Formula: peak latencies ~ 1 + effect + (1 + effect | subject) + (1 + effect | electrode)

Predictors	*Estimates*	*std. Error*	*95 %CI*	*Pr(>|z|)*
(Intercept)	454.857	6.872	441.369 – 468.345	**<0.001**
Effect [Semantic]	−3.993	8.597	−20.865 – 12.879	0.642
Random Effects				
σ^2	0.01			

(*continued on next page*)

Table A5 (*continued*)

Formula: peak latencies ~ 1 + effect + (1 + effect | subject) + (1 + effect | electrode)				
Predictors	*Estimates*	*std. Error*	*95 %CI*	*Pr(>|z|)*
$\tau_{00\ Subject}$	281.59			
$\tau_{00\ Electrode}$	2.84			
$\tau_{11\ Subject.EffectSemantic}$	356.58			
$\tau_{11\ Electrode.EffectSemantic}$	22.22			
$\rho_{01\ Subject}$	0.03			
$\rho_{01\ Electrode}$	−1.00			
ICC	1.00			
$N_{Subject}$	30			
$N_{Electrode}$	15			
Observations	900			
Marginal R^2 / Conditional R^2	0.023 / 1.000			

B. EEG montage

See Fig. B1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brainres.2025.149995.

References

Abdel Rahman, R., Melinger, A., 2009. Semantic context effects in language production: a swinging lexical network proposal and a review. Lang. Cogn. Process. 24 (5), 713–734. https://doi.org/10.1080/01690960802597250.

Adams, K.L., Faires Conklin, N., 1973. Toward a theory of natural classification. Proc. Annu. Meet. Chicago Linguist. Soc. 9 (1), 1–10.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723. https://doi.org/10.1109/TAC.1974.1100705.

Allan, K., 1977. Classifiers. Language 53 (2), 285–311. https://doi.org/10.1353/lan.1977.0043.

Amsel, B.D., 2011. Tracking real-time neural activation of conceptual knowledge using single-trial event-related potentials. Neuropsychologia 49 (5), 970–983. https://doi.org/10.1016/j.neuropsychologia.2011.01.003.

Baayen, H., Vasishth, S., Kliegl, R., Bates, D., 2017. The cave of shadows: addressing the human factor with generalized additive mixed models. J. Mem. Lang. 94, 206–234. https://doi.org/10.1016/j.jml.2016.11.006.

Barr, D.J., 2013. Random effects structure for testing interactions in linear mixed-effects models. Front. Psychol. 4. https://doi.org/10.3389/fpsyg.2013.00328.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. ArXiv Preprint ArXiv:1506.04967.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015b. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67 (1), 1–48. https://doi.org/10.18637/jss.v067.i01.

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 (1), 101–109. https://doi.org/10.1016/j.cognition.2010.04.004.

Bloem, I., La Heij, W., 2003. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. J. Mem. Lang. 48 (3), 468–488. https://doi.org/10.1016/S0749-596X(02)00503-X.

Bock, K., Levelt, W., 1994. Language production: Grammatical encoding. In: Gernsbacher, M.A. (Ed.), Handbook of Psycholinguistics, Vol. 5. Academic Press, San Diego, pp. 945–984.

Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot Int. 5 (9/10), 341–347.

Brain Vision Analyzer (Version 2.2.2). (2021). Gilching, Germany: Brain Products GmbH.

Brysbaert, M., Stevens, M., 2018. Power Analysis and effect size in mixed Effects Models: a Tutorial. J. Cogn. 1 (1). https://doi.org/10.5334/joc.10.

Bürki, A., Elbuy, S., Madec, S., Vasishth, S., 2020. What did we learn from forty years of research on semantic interference? a Bayesian meta-analysis. J. Mem. Lang. 114, 104125. https://doi.org/10.1016/j.jml.2020.104125.

Bürki, A., Sadat, J., Dubarry, A.S., Alario, F.X., 2016. Sequential processing during noun phrase production. Cognition 146, 90–99. https://doi.org/10.1016/j.cognition.2015.09.002.

Caramazza, A., 1997. How many levels of processing are there in lexical access? Cogn. Neuropsychol. 14 (1), 177–208. https://doi.org/10.1080/026432997381664.

Christensen, R. H. B. (2023). ordinal—Regression models for ordinal data (R package version 2023.12-4). https://cran.r-project.org/package=ordinal.

Contini-Morava, E., Kilarski, M., 2013. Functions of nominal classification. Lang. Sci. 40, 263–299. https://doi.org/10.1016/j.langsci.2013.03.002.

Dell, G.S., 1986. A spreading-activation theory of retrieval in sentence production. Psychol. Rev. 93 (3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283.

Dell, G.S., 1988. The retrieval of phonological forms in production: Tests of predictions from a connectionist model. J. Mem. Lang. 27 (2), 124–142. https://doi.org/10.1016/0749-596X(88)90070-8.

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. Front. Psychol. 1, 1–10. https://doi.org/10.3389/fpsyg.2010.00177.

Foucart, A., Frenck-Mestre, C., 2011. Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. Biling. Lang. Congn. 14 (3), 379–399. https://doi.org/10.1017/S136672891000012X.

Frankowsky, M., Ke, D., Zwitserlood, P., Michel, R., Bölte, J., 2022. The interplay between classifier choice and animacy in Mandarin-chinese noun phrases: an ERP study. Lang. Cogn. Neurosci. 1–17. https://doi.org/10.1080/23273798.2022.2026420.

Frömer, R., Maier, M., Abdel Rahman, R., 2018. Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. Front. Neurosci. 12. https://doi.org/10.3389/fnins.2018.00048.

Garrett, M.F., 1975. The analysis of sentence production. In: Bower, G.H. (Ed.), Psychology of Learning and Motivation, Vol. 9. Academic Press, New York, pp. 133–177, 10.1016/S0079-7421(08)60270-4.

Garrett, M.F., 1980. Levels of processing in sentence production. In: Butterworth, B. (Ed.), Language Production, Vol. 1. Academic Press, Orlando/London, pp. 177–220.

Griffin, Z.M., Ferreira, V.S., 2006. Properties of spoken language production. In: Traxler, M.J., Gernsbacher, M.A. (Eds.), Handbook of Psycholinguistics. Elsevier, Amsterdam, pp. 21–59, 10.1016/B978-012369374-7/50003-1.

Gunter, T.C., Friederici, A.D., Schriefers, H., 2000. Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. J. Cogn. Neurosci. 12 (4), 556–568. https://doi.org/10.1162/089892900562336.

Hagoort, P., Brown, C.M., 1999. Gender electrified: ERP evidence on the syntactic nature of gender processing. J. Psycholinguist. Res. 28 (6), 715–728. https://doi.org/10.1023/A:1023277213129.

He, J. (2000). A study of classifiers in modern Chinese [现代汉语量词研究] (F. Zhou, Ed.; 1st ed.). The Ethnic Publishing House.

Huang, S., Schiller, N.O., 2021. Classifiers in Mandarin chinese: Behavioral and electrophysiological evidence regarding their representation and processing. Brain Lang. 214, 104889. https://doi.org/10.1016/j.bandl.2020.104889.

Kilarski, M., 2013. Nominal Classification, Vol. 121. John Benjamins Publishing Company, 10.1075/sihols.121.

Krott, A., Medaglia, M.T., Porcaro, C., 2019. Early and late effects of semantic distractors on electroencephalographic responses during overt picture naming. Front. Psychol. 10. https://doi.org/10.3389/fpsyg.2019.00696.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Annu. Rev. Psychol. 62 (1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123.

Lenth, R. V. (2024). emmeans: Estimated marginal means, aka least-squares means (R package version 1.10.4.900001). https://rvlenth.github.io/emmeans/.

Levelt, W.J.M., 1989. Speaking: from intention to articulation. The MIT Press, Cambridge, MA, 10.7551/mitpress/6393.001.0001.

Levelt, W.J.M., 1992. Accessing words in speech production: Stages, processes and representations. Cognition 42 (1–3), 1–22. https://doi.org/10.1016/0010-0277(92)90038-J.

Levelt, W.J.M., 1999. Models of word production. Trends Cogn. Sci. 3 (6), 223–232. https://doi.org/10.1016/S1364-6613(99)01319-4.

Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999a. A theory of lexical access in speech production. Behav. Brain Sci. 22 (1), 1–38. https://doi.org/10.1017/S0140525X99001776.

Levelt, W.J.M., Roelofs, A., Meyer, A.S., 1999b. Multiple perspectives on word production. Behav. Brain Sci. 22 (1). https://doi.org/10.1017/S0140525X99451775.

Lewis, F., Butler, A., Gilbert, L., 2011. A unified approach to model selection using the likelihood ratio test. Methods Ecol. Evol. 2 (2), 155–162. https://doi.org/10.1111/j.2041-210X.2010.00063.x.

Li, W., Jia, G., Yanchao, B., Hua, S., 2006. Classifier congruency effect in the production of noun phrases. Stud. Psychol. Behav. 4 (1), 34. https://psybeh.tjnu.edu.cn/EN/Y2006/V4/I1/34.
Liu, Y., Hao, M., Li, P., Shu, H., 2011. Timed picture naming norms for Mandarin chinese. PLoS One 6 (1). https://doi.org/10.1371/journal.pone.0016505.
Lupker, S.J., 1979. The semantic nature of response competition in the picture-word interference task. Mem. Cognit. 7 (6), 485–495. https://doi.org/10.3758/BF03198265.
Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D., 2017. Balancing Type I error and power in linear mixed models. J. Mem. Lang. 94, 305–315. https://doi.org/10.1016/j.jml.2017.01.001.
Myers, J., 2000. Rules vs. analogy in Mandarin classifier selection. Lang. Linguistics 1 (2), 187–209.
Neath, A.A., Cavanaugh, J.E., 2012. The Bayesian information criterion: Background, derivation, and applications. WIREs Comput. Stat. 4 (2), 199–203. https://doi.org/10.1002/wics.199.
Oppenheim, G.M., Dell, G.S., Schwartz, M.F., 2010. The dark side of incremental learning: a model of cumulative semantic interference during lexical access in speech production. Cognition 114 (2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007.
Qian, Z. (in press). The psycholinguistics of classifiers. In N. O. Schiller & T. Kupisch (Eds.), The Oxford Handbook of Gender and Classifiers. New York: Oxford University Press.
Qian, Z., Garnsey, S.M., 2016. A sheet of coffee: an event-related brain potential study of the processing of classifier-noun sequences in English and Mandarin. Lang. Cogn. Neurosci. 31 (6), 761–784. https://doi.org/10.1080/23273798.2016.1153116.
R Core Team, 2023. R: a language and environment for statistical computing (2024.4.1.748). R Foundation for Statistical Computing https://www.r-project.org/.
Roelofs, A., 1992. A spreading-activation theory of lemma retrieval in speaking. Cognition 42 (1–3), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F.
Roelofs, A., 1997. The WEAVER model of word-form encoding in speech production. Cognition 64 (3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9.
Roelofs, A., 2000. WEAVER++ and other computational models of lemma retrieval and word-form encoding. In: Wheeldon, L. (Ed.), Aspects of Language Production. Psychology Press, London, United Kingdom, pp. 71–114.
Roelofs, A., Ferreira, V.S., 2019. The architecture of speaking. In: Hagoort, P. (Ed.), Human Language: from Genes and Brains to Behavior. MIT Press, Cambridge, MA, pp. 35–50.
Rose, S.B., Aristei, S., Melinger, A., Abdel Rahman, R., 2019. The closer they are, the more they interfere: Semantic similarity of word distractors increases competition in language production. J. Exp. Psychol. Learn. Mem. Cogn. 45 (4), 753–763. https://doi.org/10.1037/xlm0000592.
Schiller, N.O., Caramazza, A., 2003. Grammatical feature selection in noun phrase production: evidence from German and dutch. J. Mem. Lang. 48 (1), 169–194. https://doi.org/10.1016/S0749-596X(02)00508-9.
Schnur, T.T., Schwartz, M.F., Brecher, A., Hodgson, C., 2006. Semantic interference during blocked-cyclic naming: evidence from aphasia. J. Mem. Lang. 54 (2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002.
Schriefers, H., 1993. Syntactic processes in the production of noun phrases. J. Exp. Psychol. Learn. Mem. Cogn. 19 (4), 841–850. https://doi.org/10.1037/0278-7393.19.4.841.
Sun, C.C., Hendrix, P., Ma, J., Baayen, R.H., 2018. Chinese lexical database (CLD): a large-scale lexical database for simplified Mandarin chinese. Behav. Res. Methods 50 (6), 2606–2629. https://doi.org/10.3758/s13428-018-1038-3.
Tai, J. H. Y. (1994). Chinese classifier systems and human categorization. In Honor of William S. Y. Wang: Interdisciplinary Studies on Language and Language Change, 479–494.
Tai, J.H.Y., Chao, F., 1994. A semantic study of the classifier zhang. J. Chin. Lang. Teach. Assoc. 29 (3), 67–78.
Van Casteren, M., Davis, M.H., 2006. Mix, a program for pseudorandomization. Behav. Res. Methods 38 (4), 584–589. https://doi.org/10.3758/BF03193889.
Voeten, C. C. (2023a). Analyzing time series data using permu.test.
Voeten, C. C. (2023b). Permutation tests for time series data. In Journal of Neuroscience Methods (Version 2.8; Vol. 164, Issue 1). https://CRAN.R-project.org/package=permutes.
Von Grebmer zu Wolfsthurn, S., Pablos Robles, L., Schiller, N.O., 2021. Noun-phrase production as a window to language selection: an ERP study. Neuropsychologia 162, 108055. https://doi.org/10.1016/j.neuropsychologia.2021.108055.
Wang, J., Witteman, J., Schiller, N.O., 2025a. Processing of visual shape information in chinese classifier-noun phrases. Cogn. Neuropsychol. https://doi.org/10.1080/02643294.2025.2485974.
Wang, M., Chen, Y., Schiller, N.O., 2019. Lexico-syntactic features are activated but not selected in bare noun production: Electrophysiological evidence from overt picture naming. Cortex 116, 294–307. https://doi.org/10.1016/j.cortex.2018.05.014.
Wang, M., & Schiller, N. O. (in press). The neurolinguistics of classifiers. In N. O. Schiller & T. Kupisch (Eds.), The Oxford Handbook of Gender and Classifiers. New York: Oxford University Press.
Wang, M., Schiller, N.O., 2019. A review on grammatical gender agreement in speech production. Front. Psychol. 9. https://doi.org/10.3389/fpsyg.2018.02754.
Wang, S., & Schiller, N. O. (submitted). Classifier processing in L1 and L2 noun phrase production: behavioral and electrophysiological evidence.
Wang, Y., Witteman, J., Schiller, N.O., 2025b. The role of animacy in language production: evidence from bare noun naming. Q. J. Exp. Psychol. 78 (7), 1461–1473. https://doi.org/10.1177/17470218241281868.
Wang, Y., Witteman, J., Schiller, N.O., 2025c. Activation of classifiers in word production: insights from lexico-syntactic probability distributions. Lang. Cogn. Neurosci. 1–15. https://doi.org/10.1080/23273798.2025.2498017.
Xun, E., Rao, G., Xiao, X., Zang, J., 2016. The construction of the BCC Corpus in the age of big Data. Corpus Linguistics 3 (1), 93–109.
Zhang, J., Liu, H., 2009. The lexical access of individual classifiers in language production and comprehension. Acta Psychol. Sin. 41 (7), 580–593. https://doi.org/10.3724/SP.J.1041.2009.00580.
Zhang, Q., Feng, C., Zhu, X., Wang, C., 2016. Transforming semantic interference into facilitation in a picture–word interference task. Appl. Psycholinguist. 37 (5), 1025–1049. https://doi.org/10.1017/S014271641500034X.
Zhang, S., Schmitt, B., 1998. Language-dependent classification: the mental representation of classifiers in cognition, memory, and ad evaluations. J. Exp. Psychol. Appl. 4 (4), 375–385. https://doi.org/10.1037/1076-898X.4.4.375.
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 Lang. 144, 16–25. https://doi.org/10.1016/j.bandl.2015.03.007.