Unconscious effect of Chinese classifier during object categorization

A Thesis

Presented to

The Division of Philosophy, Religion, Psychology, and Linguistics

Reed College

In Partial Fulfillment

of the Requirements for the Degree

Bachelor of Arts

Charlotte Li

May 2022

Approved for the Division

(Psychology)

____

Enriqueta Canseco-Gonzalez

Acknowledgments

I would first like to thank Enriqueta for being the best advisor I could’ve hoped for. I was lucky enough to get involved in research early in my undergraduate career and I can’t imagine completing a thesis like this without the research experience I gained working for Enriqueta. Thank you for being patient, caring, and honest. Thank you for being so involved not only in my thesis but my entire research career. I will always remember the time you spent walking me through the decoding paper line by line in your office, and deeply miss the type of mentorship I was lucky enough to have as an undergraduate student.

Kevin and Michael, thank you for being helpful throughout the process of my thesis and thank you for all the time you dedicated to offering me suggestions on my research design and data analysis. It’s my honor to have you two on my thesis oral board.

Ariadna, thank you for making Hum 110 an exceptional time. I always viewed Hum with you as a fundamental moment of my Reed educational experience, and I cannot express to you enough how much you helped me in my very first year of Reed and how it positively impacted me for the rest of my Reed career. I knew I wanted to have you on my thesis board from freshman year, and I still feel the same after three years. Thank you for agreeing to be my fourth reader.

To my parents, I would not be able to do this without you. Thank you for supporting all of my decisions throughout the journey. I truly feel lucky and grateful to be

your daughter. Thank you for always being there, always listening, and always making me feel loved.

To my friends, Nirvana, Ella, and Clarissa. Thank you for making me so happy during my senior year. It was so great that we have one another during this challenging year and we've all accomplished so much. All the fun and emotional times we had together will be my fondest memory to look back to during my college years.

I also want to thank members of the SCALP lab for offering me great assistance and companionship during my thesis time, thank you, Abigail, Cole, Andries, and Angelica. Thank you, to all of my participants, for contributing good data.

List of Abbreviations

LAN	Left Anterior Negativity
ERP	Event Related Potentials
EEG	Electroencephalography
ICA	Independent Component Analysis

Table of Contents

List of Tables

List of Figures

Abstract

Does speaking a language influence how you perceive and categorize objects in your everyday life? In this thesis, I attempt to approach this question from the perspective of a grammatical feature commonly seen in East Asian languages, known as the classifier. In exp.#1 I asked Chinese-English bilinguals who are native Chinese speakers and English monolingual participants to rate pairs of pictures based on their similarity. I found that both bilingual speakers and monolingual speakers rated the objects as more similar when they belonged in the same semantic category, but also when their Chinese name shared the same classifier. In exp.#2, I also tested Chinese-English bilinguals and English monolinguals in a semantic categorization task using triads of pictures, while I measured their brain response (ERPs) to the third picture. Unbeknownst to the participants, in half of the trials, the third picture shared the same classifier in Chinese with the first two, and in the other half, it did not. No priming effect was found in the behavioral results of either semantic (category) relatedness or grammatical (classifier) congruency. However, electrophysiological data revealed a semantic priming effect in both groups, and a negativity in grammatically incongruent trials, only in the native Chinese speakers. Given that the study was entirely carried out in an English context, these results suggest that classifier information was unconsciously activated in the Chinese speakers during a perceptual categorization task. Findings from this study provide supporting evidence for the linguistic relativity hypothesis and in particular the label-feedback hypothesis in the grammatical domain.

Chapter 1: Introduction

Does speaking a language shape the way people think? Ever since the Sapir-Whorf hypothesis, also known as the linguistic relativity or Whorfian hypothesis was proposed (Whorf, 1956), there have been constant debates in the field regarding the relationship between language and human cognition. The hypothesis focuses on the structural difference between natural languages and explores whether these different structures and the classification of reality implicit within them affect our perception of reality. (Lucy, 2001).

The Whorfian hypothesis at its strongest and most extreme form is linguistic determinism, which suggests that language essentially dictates through and the influence of language is strong enough that it overrides pre-existing physical perceptions (Wolff & Holmes, 2010). While linguistic determinism had been rejected by most scholars, it cannot be said that language has absolutely no impact on cognitive processes. In the recent few decades, linguistic relativity, a milder take of the Whorfian hypothesis became more popular and there has been a reappearance of empirical research dedicated to exploring the interactions between language and thought more accurately (Wolff & Holmes, 2010).

In particular, Lupyan (2012) proposed the label-feedback hypothesis, suggesting that the assignment of linguistic labels to objects activates detections of existing perceptual features associated with that object and encourages categorization denoted by the labels assigned. According to Lupyan (2012), linguistic labels are actively modulating perceptual processing when an object is perceived. Although the original hypothesis referenced the direct naming of objects, Boutonnet et al. (2012) proposed that the feedback process outlined by the label-feedback hypothesis might apply to grammatical features as well.

In this chapter, I will first outline some studies that showed support for the linguistic relativity hypothesis and the label-feedback hypothesis. Specifically, I will first be focusing on studies that use traditional behavioral methods to study how grammatical features like grammatical gender and classifier affect cognitive processes. Next, I will explain the event-related potential (ERP) technique and examine some ERP studies in the field of linguistic relativity research and the advantages of the ERP approach. I will go into detail about the findings of Boutonnet et al. (2012) that suggested the unconscious activation of grammatical gender during an object categorization task, followed by the proposal of the present study inspired by Boutonnet et al. (2012), including the proposed hypothesis for the present study.

1.1 Label-feedback hypothesis and linguistic relativity

In this section, I will go into more detail about the label-feedback hypothesis proposed by Lupyan (2012) and the feedback process proposed in the hypothesis. After that, I will examine the existing empirical evidence of linguistic relativity using behavioral approaches.

1.1.1 Label-feedback hypothesis

Lupyan (2012) identified a paradox in the field of language and thought research: while there was a lot of evidence for cross-linguistic differences in color categorization and color perception, showing support for how language affects cognition and perception (Davies and Corbett, 1998; Davidoff et al., 1999; Roberson et al., 2005, 2008; Daoutis et al., 2006; Winawer et al., 2007; Thierry et al., 2009), it seems that these cross-linguistic differences were also easily removed when participants have less on-line access to language information (under verbal interference). Gleitman (2010) claimed this to be "anti-Whorf" as linguistic influences seem superficial if it's so easily erasable. In other words, if the cross-linguistic differences are only present when language is actively in use, that seems to suggest that fundamentally cognition and perception remain unaffected by language or that the influence of language is negligible (Gleitman and Papafragou, 2005; Dessalegn and Landau, 2008; Li et al., 2009).

However, Lupyan (2012) pointed out that this line of thinking is on the premise of two assumptions: a. Language is viewed as a "tool" (Gleitman et al., 2004), and b. There is a clear distinction between verbal and non-verbal ways of representation. Lupyan (2012) then proposed the label-feedback hypothesis as an attempt to offer alternative interpretations of such paradox: how can language be so evidently influencing perceptual processing, yet at the same time be so sensitive to manipulations like verbal interference?

Lupyan (2012) claimed that the paradox could be explained by the framework that language influences perception through augmenting perceptual processing as it's happening. Lupyan (2012) viewed language and thought as an interactive system in which language modulates thought automatically and rapidly “on-line”. He also argued that because language is used so ubiquitously, language, as part of this inherently interactive system, has the potential to be influential to a wide range of tasks.

The process of categorization requires recognition of the similarities and differences between objects (Lupyan, 2012). When perceptually processing an object in a meaningful way, the object needs to be recognized in larger groups (Lupyan, 2012). To be able to identify two objects both as “toys” requires the recognition of the two objects as members of the same group first, as Lupyan (2012) pointed out, naming is dependent on categorizing. At the same time, Lupyan (2012) claimed that the act of assigning verbal labels to objects also played a role in categorization, by activating the recognition of specific perceptual features that were shared by the members within the category that the label created. In this view, categorization is the process of recognizing perceptually different objects as the same class. (Lupyan, 2012). Although the same perceptual differences remain the same, objects that have been assigned the same name will be viewed as belonging in the same category, and the forming of such categories requires the detection of similarities between the objects.

The label-feedback hypothesis (Lupyan, 2012) proposed that language serves the role of transiently modulating perceptual processing while it is at work. Particularly, by assigning the same linguistic label to two different shades of green, the perceived “distance” between the two greens would momentarily be smaller, because after learning that the two colors are both named “green”, perceptual processing of the two colors now also activate the verbal label being assigned, creating a “hybrid visuo-linguistic experience” (Lupyan, 2012). Gliga et al. (2010), have found enhanced gamma-band oscillation in the visual cortex of 1-year-old children when labels have been assigned to objects, and the same effect was not found in conditions where no labels have been assigned. This shows that verbal labels modulate the process of visual perception, suggesting that verbal labeling affect object perception in a top-down process (Gliga et al., 2010), similar to what Lupyan (2012) described.

Previous studies in testing the linguistic relativity hypothesis involve both behavioral methods and electrophysiological techniques. EEG is commonly seen in psycholinguistic research. In the next section, I will introduce EEG and the ERP technique, as well as explain the rationale of choosing said technique for the design of the present study, and briefly introducing two ERP effects commonly seen in linguistic research.

1.2 EEG and ERPs

Electroencephalography (EEG) measures the postsynaptic electrical potentials generated by neurons that are active synchronously (Tivadar & Murray, 2019). Raw EEG waveforms are collected through electrodes placed on the scalp with electrolyte-gel (abralyt HiCl) applied to all the electrodes. The scalp is then gently abraded to enable a good connection between the electrodes and the scalp. Compared to other neurophysiological techniques commonly used like the fMRI, EEG is more temporally accurate and is most suitable for studies that focus on the time course of activity, rather than the localization of functions. (Luck, 2014)

Event-related potentials (ERPs) are time-locked around the appearance of the target object of particular interest and thus enable researchers to observe temporally accurate postsynaptic electrical activities in that period of interest. Usually, ERPs are the average of many trails of the same event of interest. Averaging the ERPs allows the noise in each individual trail to be canceled out and the more trials an ERP is averaged from, the clearer the effect being measured would be. An advantage of ERPs is that it reveals brain response to visual and/or audio stimuli without requiring the participants to give an overt behavioral response (Athanasopoulos & Casaponsa, 2020). This enables the experimenter to "hide" their actual intent by keeping the participants occupied with a task that does not directly reveal the interest of the experimenters, to avoid direct activation of the linguistic feature and demand characteristics (when participants realize what the experimenter was expects and adjust their performance to conform to that expectation). It then allows the researcher to measure unconscious cognitive processes. (Athanasopoulos & Casaponsa, 2020)

1.2.1 N400

One ERP component that has been especially useful for the study of language comprehension is the N400, a negative component peaking around 400ms after stimulus onset shown to vary systematically with the processing of semantic information. (Kutas & Federmeier., 2000). N400 was shown to be associated with word expectancy (Kutas & Federmeier., 2011). Particularly, when a sentence end with an unexpected word (e.g. *I took my coffee with cream and dog*.), a relatively broad distribution of negativity would peak around 400ms to the onset of the unexpected word *dog* (Kutas & Federmeier., 2011). According to Kutas and Federmeier (2011) N400 has been observed not only in spoken words, but in pseudo-words, American Sign Language, line drawing, and picture stimuli as well.

1.2.2 Left Anterior Negativity (LAN)

Another ERP marker relevant to linguistic relativity research is the left anterior negativity (LAN). The LAN has been identified as an electrophysiological marker of morphosyntactic violations (Molinaro et al., 2014, Tanner, 2015). LAN has been observed in sentences with grammatical mistakes, generally appearing as a negative wave around 300 to 500ms to the onset of the morphosyntactic unexpectancy (e.g. in *She run to the playground*. "*run*" would be the onset of LAN as it is a grammatical violation.)

1.3 Boutonnet et al. (2012), Unconscious effect of grammatical gender during object categorization

Boutonnet et al. (2012) looked at the unconscious effects of grammatical gender during an object categorization task, with a sample of Spanish-English bilingual speakers and monolingual English speakers. Grammatical gender is a feature present in many languages, which assigns all nouns to take a grammatical masculine, feminine, and/or neuter gender. Grammatical gender has been of particular interest to previous psycholinguistic research because the assignment is arbitrary and is not semantically based (Boutonnet et al., 2012). Participants were instructed to press a button indicating "yes" if the third picture they saw belonged to the same semantic category as the first two (e.g. Tomato, celery, asparagus), and press another button to indicate "no" if the third picture belonged in a different semantic category (e.g. Tomato, celery, truck). The participants were not informed whether the object depicted in the third picture had a different grammatical gender in Spanish to the first two (which happened in half of the trials), nor were they instructed to consider grammatical gender at all when pressing the button. According to the priming effect, refers to the phenomenon where exposure to a specific previous stimulus causes an unintentional influence on the response to the subsequent stimulus, it was expected that participants would take longer to respond to semantically incongruent trials than to semantically congruent trials. However, Boutonnet et al. (2012) found no priming effect of either semantic or (grammatical) gender congruency in the behavioral data. However, the ERP results supported the Whorfian hypothesis. In particular, a LAN was found only in the Spanish-English bilinguals and only in trials when the third picture was of a different gender from the first two. Boutonnet et al. (2012)'s findings suggest that during an object categorization task that only deals with semantic relatedness, information about the grammatical gender of the stimuli was activated in those who speak both Spanish and English, even if at the time of the study they used only English.

However, grammatical gender is only one type of grammatical feature in languages, and as Boutonnet et al. (2012) mentioned, grammatical gender is very arbitrary. This is not the case for all the grammatical features, and therefore the findings of Boutonnet et al. (2012) are specific to grammatical gender only and cannot be applied to other grammatical features. To further explore whether grammatical information of a language not actively in use is activated during a task not involving said linguistic features and whether this effect is limited to something as arbitrary as grammatical gender, the present study replicates Boutonnet et al. (2012)'s design with another grammatical feature.

1.4 The Present Study

In Chinese, nouns are classified based on their features (including shape, size, function, etc.) and this classification determines the classifier they pair with the noun. In English, the nouns are modified by the number directly when described in quantity (e.g. *three chairs, seven books*). In Chinese and other East Asian languages, numerals modify the classifier that goes with the noun (Srinivasan, 2011). For example, in English we say *two tables*, the number directly modifies the object, but in Chinese, one would say something like *two [flat-thing] tables*, where the number modifies the classifier *[flat-thing]*, followed by the noun. Classifiers can be quite flexible in regard to the nouns they can apply to, and different classifiers can signify different features of the object that it is associated with. Although some classifiers are viewed as more arbitrary, it has been argued that even the more arbitrary classifiers could be explained in terms of "motivated extensions from central examples" (Gao & Malt., 2009). As cited by Gao and Malt (2009), Lakoff viewed the categories formed by classifiers as forms of conceptual categories.

Due to the fact that classifiers represent a system of categories based on pre-existing non-linguistic features like shape or functionality, classifiers are not as arbitrary as grammatical gender. For example, a fish and a dress both take the same classifier, tiao, which directly references a slender, long silhouette. Although fish and dress are not semantically related, they are visually similar. Therefore, it's possible that English speakers would also notice the perceptual similarity between objects that use the same classifier, although they do not speak a language containing classifier information. Speed, Chen, Huettig, and Majid (2020) found that objects sharing a classifier in Chinese were rated as more similar than objects that do not share a classifier, by both Mandarin and Dutch speakers. This could potentially be a confound if we replicate Boutonnet's (2012) design exactly. Therefore, I will first try to replicate Speed et al. (2020)'s finding on similarity rating with both Chinese-English bilinguals and monolingual English speakers.

I hypothesize that both Chinese-English bilingual speakers and English monolingual speakers will rate objects sharing the same classifier as more similar to one another than the objects that do not share the same classifier. I also expect to see both the bilingual and monolingual participants rating semantically related objects as more similar than semantically unrelated objects.

Chapter 2: Study 1

2.1 Method:

Participants:

Native English speakers (N = 31) and native Chinese speakers (N = 28) were recruited through CloudResearch and Prolific to complete a Qualtrics survey. Participants were compensated $1 for their participation upon completion of the survey. In order to make sure that participants (particularly native Chinese speakers) were proficient enough in English, they were asked to rate their language proficiency, on a scale of 0-10. Language proficiency was collected after the participants completed similarity ratings to avoid self-report of proficiency confounding their ratings. This was done in addition to the selection filtering implemented through the online recruiting platforms to make sure that we were getting a sample of participants as we aimed for. Participants who rated their language proficiency as lower than 9 (excellent) were excluded from the study, even if they indicated that they were native speakers.

Stimuli:

A hundred and ninety-two pairs of colored pictures were selected from The China Image Set (CIS), a set of colored pictures, normed by Ni et al. (2019) based on various Chinese psycholinguistic variables. In half of the 192 pairs the pictures were semantically related, and in the other half they were unrelated. Each half was further subdivided such that for half of the pairs, the Chinese names of the two objects use the same classifier and half of them use a different classifier. Only tangible objects that are easily recognizable and frequently seen in both English and Chinese contexts were chosen as stimuli.

Participants viewed pictures in pairs and were asked to rate how similar they thought the two objects were on a scale of 1 - 7 (1 = not similar at all, 7 = very similar). Participants were instructed to try to use the whole scale, to give some 1s and 7s, and use also the numbers in between. Crucially, instructions were given in English to both groups to ensure consistency and avoid conscious activation of grammatical information unique to Chinese speakers. That is, there was no mention at all about the Chinese language or its classifier system.

2.2 Results

Similarity ratings were submitted to 2 x 2 x 2 Mixed Model ANOVA with congruency (congruent/incongruent) and type (semantic/grammatical) as our within-subjects variables, and group (English monolinguals/ Chinese-English bilinguals) as our between-subjects factor.

This analysis revealed a main effect of semantic relatedness, F (1, 57) = 302.099, $p < .001$, $\eta_p^2 = 0.841$ and a main effect of grammatical congruency, F (1, 57) = 327.547, $p < .001$, $\eta_p^2 = 0.852$. There was also a significant interaction between the effects of semantic relatedness and grammatical congruency, F (1, 57) = 50.732, $p < .001$, $\eta_p^2 = 0.471$. However, neither semantic relatedness (F (1, 57) = 0.103, $p = 0.750$, $\eta_p^2 = 0.002$) nor grammatical congruency (F (1, 57) = 0.034, p = 0.0.854, $\eta_p^2 < .001$) had a significant interaction with group (monolingual/bilingual). See Figure 1.

Planned comparison showed that English monolingual participants rated semantically related pairs (M = 4.560, SD = 0.980) as significantly more similar than the semantically unrelated pairs (M = 1.647, SD = 0.900), t (30) = 12.687, p < .001. The same effect was also significant for Chinese-English bilingual participants (semantically related: M = 4.302, SD = 0.263; semantically unrelated: M = 1.494, SD = 0.112), t (27) = 11.939, p < .001.

Similarly, for grammatical congruency, the effect is also significant for both monolingual and bilingual participants. Bilingual participants rated the pairs with same classifier (M = 3.097, SD = 0.861) as significantly more similar than the pairs that used different classifiers (M = 2.699, SD = 0.888), t (27) = 11.934, p < .001 . The same effect is also significant for monolingual participants (same classifier: M = 3.299, SD = 0.655; different classifier: M = 2.909, SD = 0.732), t (30) = 13.743, p < .001.

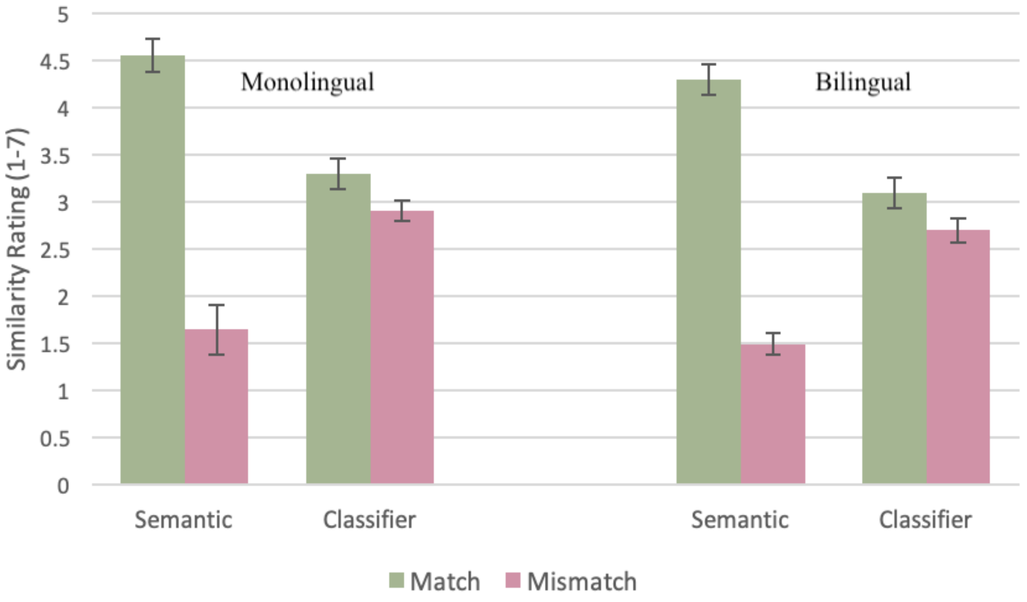


Figure 1. Plot of similarity ratings from Chinese-English bilinguals and English monolinguals, showing the main effect of semantical relatedness and grammatical congruency, and no effect of groups.

2.3 Discussion:

Overall, both Chinese-English bilingual speakers and English monolingual speakers rated objects in the same semantic category as more similar to each other, which is consistent with my hypothesis; both groups also rated objects sharing the same classifier as more similar than the objects that do not share the same classifier, which is also consistent with the hypothesis. The purpose of study 1 was to investigate whether speaking a language with classifier information leads people to recognize the objects that use the same classifier as more similar. Our results show that *both* English monolingual speakers and Chinese-English bilinguals rate object pairs as more similar when they use the same classifier. This suggests that objects that use the same classifier in Chinese share some physical or functional similarity that is evident even to people who do not have access to grammatical information in Chinese.

This suggests that, as expected, the classifier is a less arbitrary grammatical feature than grammatical gender. As previously mentioned, classifier categorization is based on various dimensions, including animacy, shape, and function (Adams & Conklin, 1973). All these factors could contribute to objects using the same classifier being viewed as more similar as they share shape or function or other features, even when they might not belong in the same semantic category. It is important to mention that in study 1, participants were not given instructions on what exactly they should base their similarity ratings on, so their ratings could reflect similarities on all kinds of dimensions, including semantic category and visual similarity.

With that in mind, results from study 1 still inform us that speaking a language containing classifier information does not aid speakers of that language to pick up on these nonlinguistic similarities more than people who do speak that language. In the case of classifiers in Chinese, at least based on the similarity ratings, we failed to find support for the label-feedback theory and the Whorfian hypothesis. However, as Boutonnet et al. (2012) findings suggested, Electrophysical data reveals what could be overlooked just by analyzing the behavioral data. Additionally, based on similarity rating, we cannot conclude anything about the undergoing cognitive processes, which analyzing ERPs would allow me to do. The participants will be occupied by a semantic categorization task that ask them to engage with the stimuli as they are without actively mentioning any grammatical information. However, by comparing the ERPs in four different conditions, we will be able to observe whether the participants are detecting the grammatical manipulation. Neuroimaging techniques would reveal more unconscious activation of classifier information when participants are occupied with a task completely unrelated to grammatical information. Therefore, I carried out study 2 to investigate if what Boutonnet et al. (2012) found with grammatical gender in Spanish could be replicated with a purportedly less arbitrary syntactic feature such as the classifier grammatical system of Chinese.

In study 2. For behavioral data in the EEG experiment, I hypothesize that we will replicate Boutonnet et al. (2012) and find no priming effect of either semantic relatedness or grammatical congruency in terms of reaction time.

For the ERP analysis, I expect to replicate Boutonnet et al. (2012) and find a significant N400 effect in semantically unrelated trials than in semantically related trials, in both English monolinguals and Chinese-English bilinguals. For the grammatical congruency manipulation, I do not expect to see left-anterior negativity (LAN) like Boutonnet et al. (2012) did in grammatically mismatch trials for the Spanish-English bilinguals. However, I do expect to see a significant difference between bilingual participants and monolingual participants, in trials where the third picture does not share the same classifier as the first two.

Chapter 3: Study 2

3.1 Method:

3.1.1 Participants:

Chinese-English Bilinguals who were native Chinese speakers (N = 18) and monolingual native English speakers (N = 18) participated in this experiment. The study was conducted at Reed College, in Portland, Oregon. Native Chinese speakers were international students from China who currently attend Reed College. All participants had normal or corrected-to-normal vision and no history of brain injury. Participants were compensated $20 for their participation. All procedures were approved by the Reed College Institutional Review Board.

3.1.2 Stimuli

The same set of stimuli as study 1, a total of 288 pictures, selected from The CIS (Ni et al., 2019), were presented in 96 triads. In each triad, the Chinese name of the first two pictures uses the same classifier and belongs to the same semantic category, but the third picture was either from the same or a different semantic category, and its Chinese name used either, the same or a different classifier. This creates four conditions (see Table 1 below for examples), with 24 triads in each experimental condition, for a total of 96.

Table 1. Example of experimental conditions. The positive signs indicate whether the third picture matches the semantic category and/or the classifier of the two prior pictures. The minus sign indicates the corresponding mismatching conditions.

Classifier information	Semantic information	Stimuli 1	Stimuli 2	Stimuli 3 (target)
+	+	Pig	Donkey	**Lion**
+	-	Pig	Donkey	**Onion**
-	+	Pig	Donkey	**Horse**
-	-	Pig	Donkey	**Blanket**

3.1.3 Procedure:

Participants expressing interest in participating in the study went through a brief online screening on Qualtrics where they answered brief questions about their native language and foreign language proficiency. Chinese-English bilinguals whose native language is Chinese and English speakers who do not speak a language that uses classifiers (e.g. Chinese, Japanese, Korean, Southeast Asian languages, Bengali, Assamese, Persian, Austronesian languages, Mayan languages), were considered eligible and contacted through email. If the participants asked why they were being recruited to participate, they would be informed that it is because we are studying categorization processes in bilingual speakers. In addition, participants will NOT be informed about the classifier manipulation in Chinese.

During the experiment, participants were instructed to press a button if the three pictures of a triad belonged to the same semantic category and another button if they didn't. We recorded the accuracy and reaction time of these responses.

On each trial, a fixation cross was presented for 1000 ms, followed immediately by the first picture for a duration of 600 ms, then a blank screen for 250 ms, followed by the second picture for 600 ms. Finally, after a random interval between 300 and 500 ms, the target (third picture) appeared and remained on the screen until the participant pressed a button to respond (See Figure 2). All 96 trials (24 trials per experimental condition) were presented in a randomized order. Participants were offered one break halfway through the study, and they were instructed to take as long as they felt necessary before continuing with the study. Although the main experimenter is a native Chinese speaker, all instructions were given in English to both native English speakers and native Chinese speakers to ensure consistency between the two groups, and to avoid any activation of grammatical information (classifier).

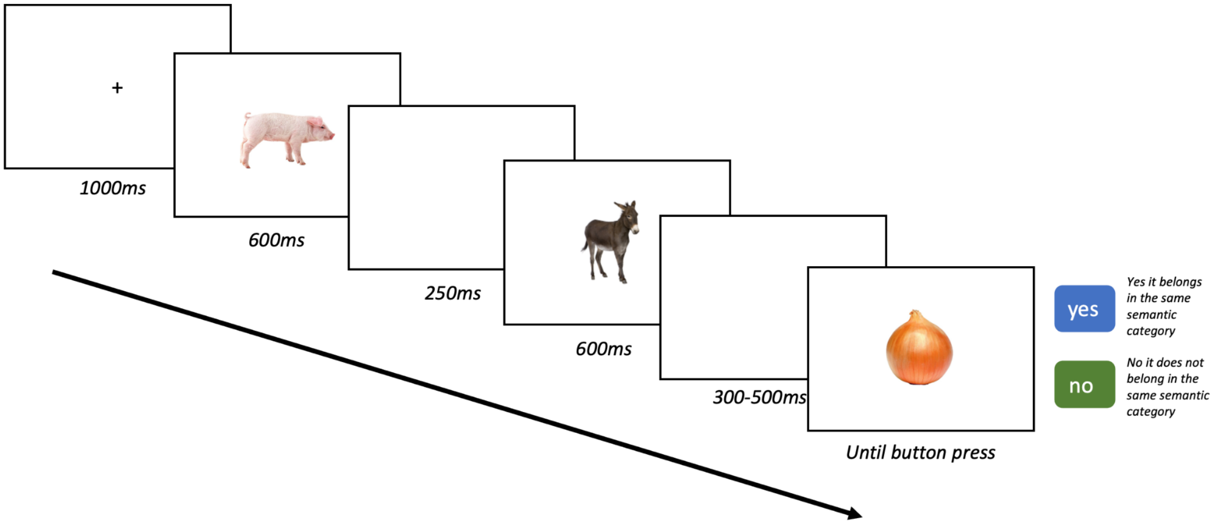


Figure 2. Diagram of what a typical trial looks like from the participant's point of view.

3.1.4 EEG Recording:

The EEG was recorded from 64 electrodes using the EasyCap system (Brain Products, Garching, Germany) with an on-line right mastoid electrode ("Ref") used as a reference. The EEG was later re-referenced offline to an average reference. Two electrodes were placed adjacent to the left and right eyes (HEOG: electrodes 62 and 63) to detect horizontal eye movement artifacts. One electrode was placed below the left eye (VEOG: electrode 61) to detect eye blink artifacts.

Only trials where participants responded correctly were included in the analysis. The data was segmented into four groups according to conditions, see Table 1. An average reference was performed using all the electrodes. Artifacts (blinks and eye movements) were mathematically corrected using Ocular Correction Independent Component Analysis. Performing artifact correction is especially beneficial in studies that contains fewer trials, as it allows the experimenter to keep the trials that contain artifacts as oppose to traditional artifact rejection technique. All individual segments of each condition were averaged together. The segmentation of each ERP was time-locked from -100ms to 1000ms to the onset of the third stimuli.

The EEG was segmented into epochs from -100 ms to 1000 ms relative to the onset of the target (3rd) picture. The 100ms interval before the onset of the picture was used later for baseline correction. The EEG was acquired at a 500 Hz sampling rate, and impedance levels were kept below 10 kΩ. The raw EEG was filtered with a 25 Hz high cutoff filter. All EEG data were processed using BrainVision Analyzer Software (Brain Products, Germany).

3.2 Results:

3.2.1 Behavioral data:

Accuracy:

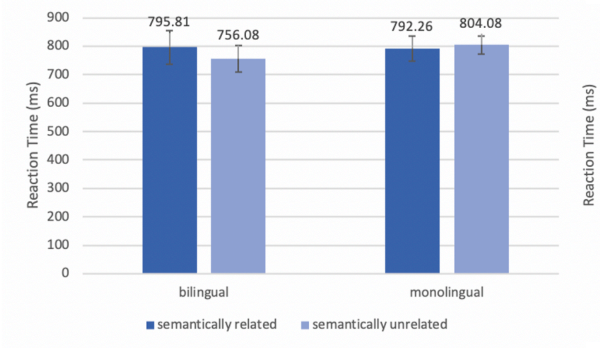
Both monolingual English speakers (96.35%) and Chinese-English bilingual participants (96.64%) responded at a very high accuracy level in the object categorization task, with an average accuracy of 96.50% and no significant difference between groups.

Reaction time:

Trials with incorrect responses and trials where the reaction time was shorter than 250 ms or larger than 2.5 SDs of the mean for that individual in that condition, were excluded from the analysis.

Figure. 3a and 3b, show the mean reaction times for each condition, type of information, and group. Reaction times were submitted to a mixed-design ANOVA with condition (match/mismatch), type of information (semantic/classifier) as our within-subjects variables, and Group (monolingual/bilingual) as the between-subjects variable. Overall, there were no significant main effects or interactions in reaction times. In particular, there was no main effect of condition based on semantic information ($F(1, 34) = 0.551$, $p = 0.463$, $\eta_p^2 = 0.016$), nor based on grammatical information ($F(1, 34) = 0.961$, $p = 0.334$, $\eta_p^2 = 0.027$). There was also no significant main effect of group, $F(1, 34) = 1.882$, $p = 0.179$, $\eta_p^2 = 0.052$, nor any interaction.

a.



b.

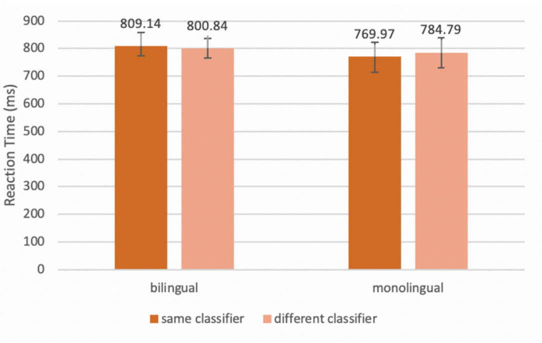


Figure 3a. Reaction time in trails where the object being depicted in the third picture belongs in the same semantic category as the first two (semantically related) and when it does not belong in the same semantic category (semantically unrelated), from English monolingual and Chinese-English bilingual participants.

Figure 3b. Reaction time in trails where the object being depicted in the third picture uses the same classifier as the first two (grammatically congruent) and when it does not use the same classifier (grammatically incongruent), from English monolingual and Chinese-English bilingual participants.

3.2.2 Electrophysiological data:

ERPs elicited by pictures matching/mismatching semantic:

ERPs elicited by pictures matching/mismatching the semantic category. Based on prior N400 studies (including Boutonnet et al.), and on visual inspection of our waveforms, we pooled seven electrodes (1, 2, 3, 4, 5, 6, 7) from the central area of the scalp to create individual averages for each condition (semantic match and mismatch) that focus on the N400 effect. Figure. 4 shows grand averages across all individuals for each condition and for each group.

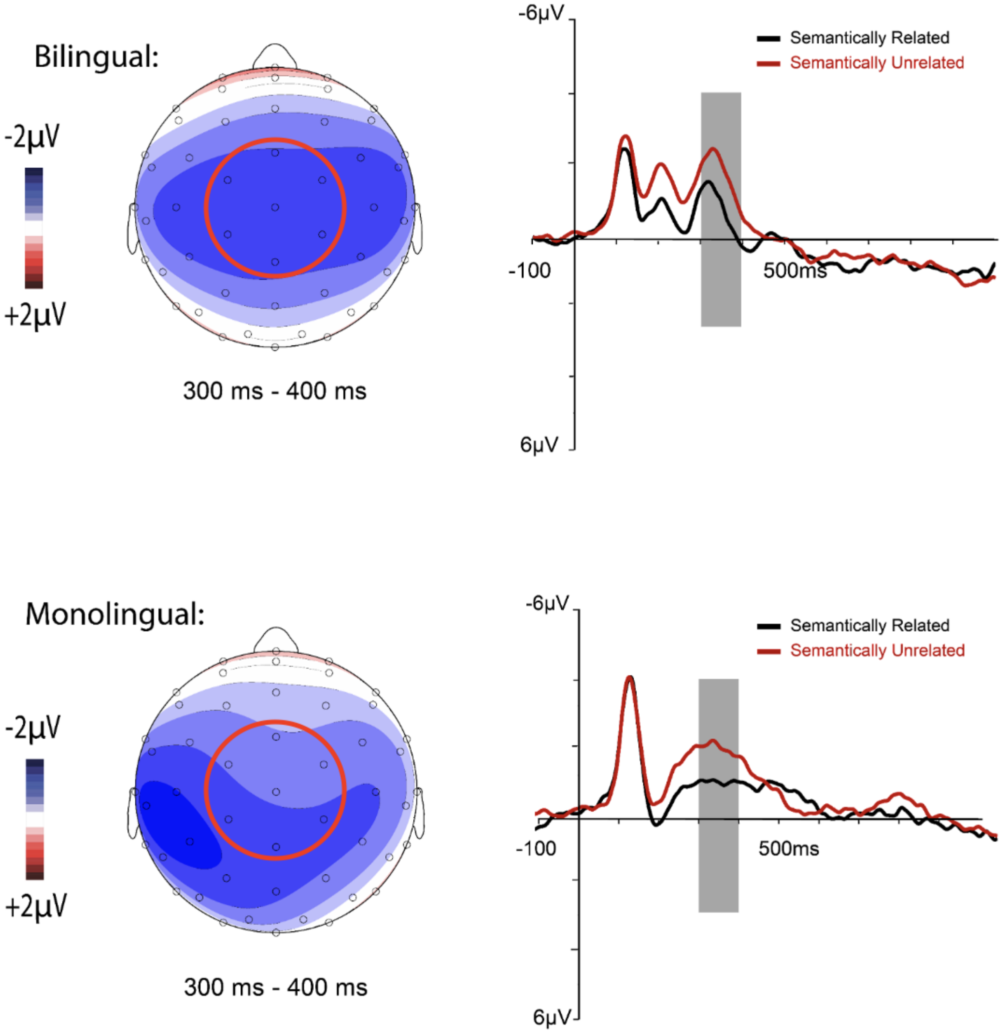


Figure 4. Scalp topography during 300-400ms and ERPs elicited in the semantically related and semantically unrelated conditions, pooled from central channels (in red circle).

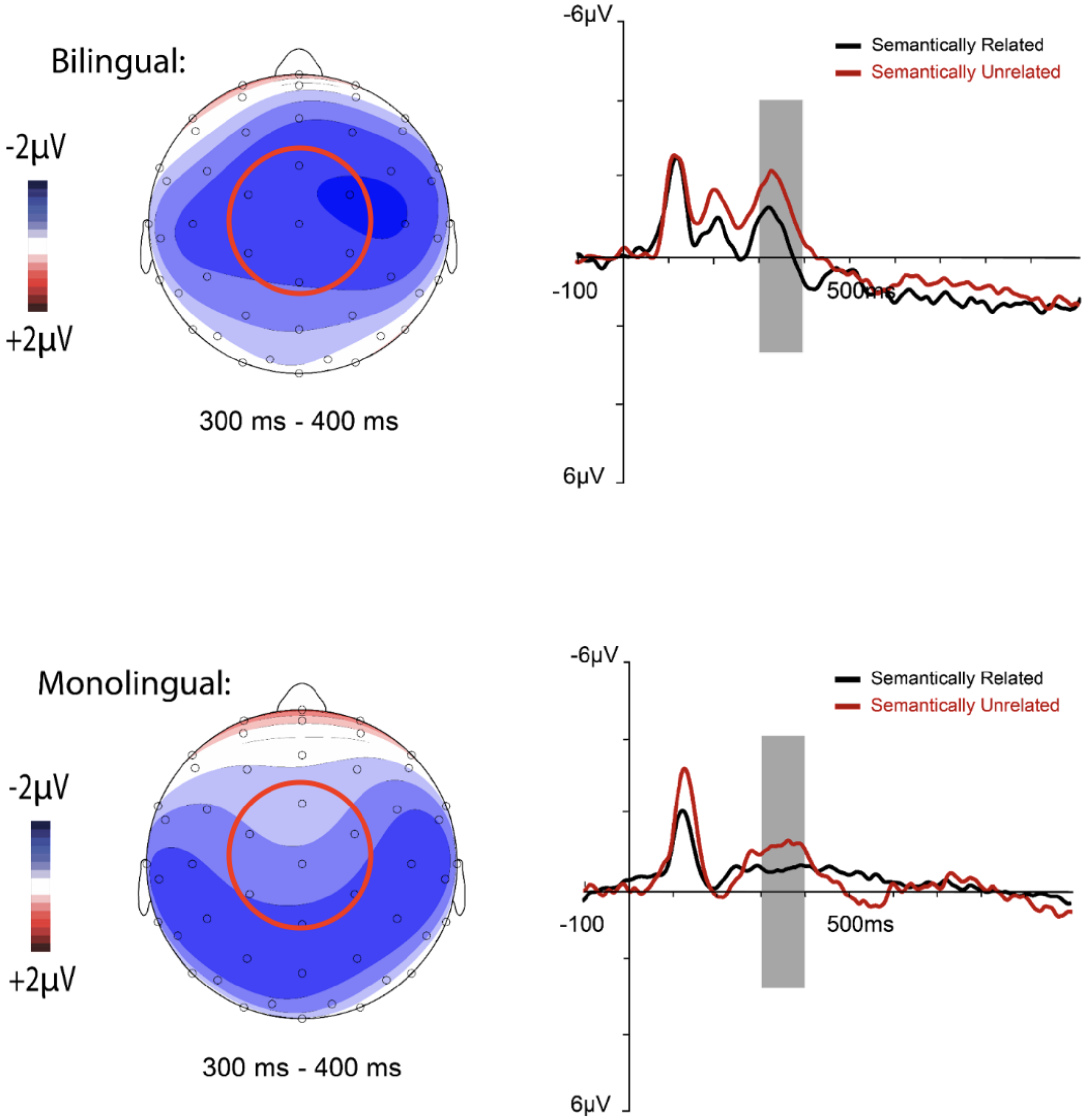
Table 2. Mean amplitude of central electrodes from 300-400ms of semantically related and unrelated trials, for both bilingual and monolingual participants.

Language	Semantically	Mean	SD	N
Bilingual	Related	-0.916	1.352	18
	Unrelated	-1.819	1.546	18
Monolingual	Related	-1.027	1.798	18
	Unrelated	-1.651	2.076	18

Clear differences between matching and mismatching conditions are visually evident in both groups. In order to quantify and analyze this effect, individual mean amplitudes were measured in a time window from 300-400ms (see the shaded area in Figure. 3) and were submitted to a 2x2 Mixed design ANOVA with condition (match/mismatch) as the within-subjects variable and Group (monolingual/bilingual) as the between-subjects variable. This analysis revealed only a main effect of condition ($F(1, 34) = 21.214$, $p < .001$, $\eta_p^2 = 0.384$), but no effect or group (STATS) nor a significant interaction between condition and group ($F(1, 34) = 0.710$, $p = 0.450$, $\eta_p^2 = 0.020$). As shown in Table 2, the mean amplitude of semantically unrelated trials was significantly more negative than the semantically related trials, for both groups.

It is important to note that both, matching and mismatching semantic category waveforms shown in Figure. 4 are formed by collapsing across matching and mismatching classifier trials. Therefore, any classifier effect should be equally represented in both waveforms. However, in order to investigate the N400 effect without any influence from the same/different classifier, we repeated the exact same analysis including *only* trials where the object depicted in the third picture used the same classifier as the first two). The corresponding waveforms and maps can be seen in Figure 4. This analysis again revealed a main effect of semantic relatedness ($F(1, 34) = 14.494$, $p < .001$, $\eta_p^2 = 0.299$), but no main effect or group (STATS), nor a significant interaction between group and condition ($F(1, 34) = 0.438$, $p = 0.513$, $\eta_p^2 = 0.013$).

Figure 5. Scalp topography during 300-400ms and ERPs elicited by semantically matching and mismatching conditions restricted to trials when the third picture is grammatically congruent.



ERPs elicited by pictures matching/mismatching classifier:

Again, based on visual inspection of our waveforms, we pooled seven frontal-central channels (1, 2, 3, 7, 8, 9, 19) of the scalp to create individual averages for each condition (classifier match and mismatch). Figure. 6 shows grand averages across all individuals for each condition and for each group (see Figure. 6, right side).

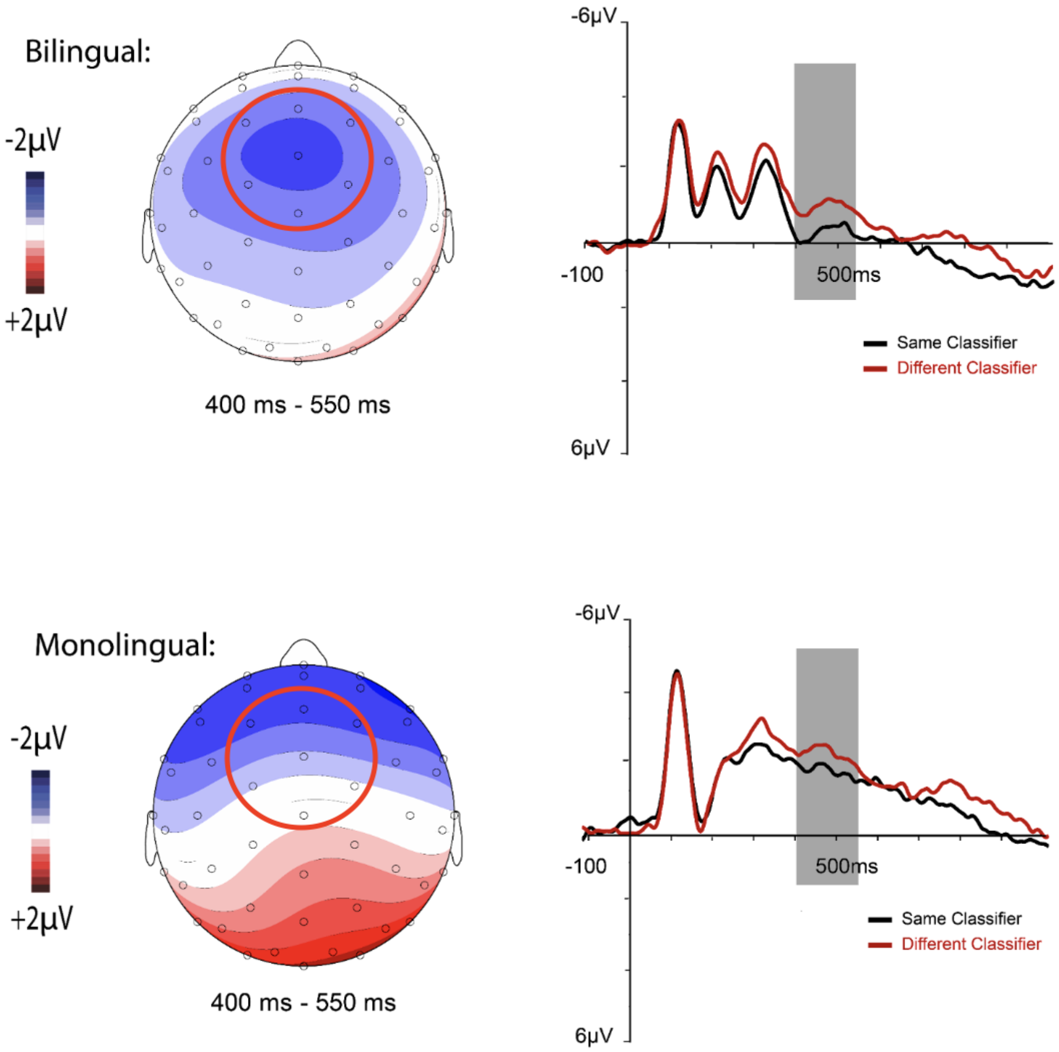


Figure 6. Scalp topography during 400-550ms and ERPs elicited in the classifier matching and mismatching conditions, pooled from seven frontal-central electrode sites (in the red circle).

A clear difference between matching and mismatching conditions is visually evident in the bilingual group and a smaller difference in the monolingual group. In order to quantify and analyze this effect, individual mean amplitudes were measured in a time window from 400 ms. to 550 ms. (see the shaded area in Figure 6.) and were submitted to a 2x2 Mixed-design ANOVA with condition (match/mismatch) as the within-subjects variable and Group (monolingual/bilingual) as the between-subjects variable. This analysis revealed a main effect of condition (F (1, 34) = 7.961, p = .008, η_p^2 = 0.190) but no main effect of group. Importantly, however, the interaction between group and condition was not significant (F (1, 34) = 0.418, p = 0.522, η_p^2 = 0.012). As shown in Table 4., the mean amplitude of grammatically incongruent trials is significantly more negative than the grammatically congruent trials for both the bilinguals and the monolinguals.

Table 3. Mean amplitude of pooled frontal-central electrodes from 400-550ms of matching and mismatching classifier trials, for each group.

Language	Grammatically	Mean	SD	N
Bilingual	Congruent	-0.432	1.395	18
	Incongruent	-1.141	1.865	18
Monolingual	Congruent	-1.618	2.107	18
	Incongruent	-2.063	2.388	18

Again, matching and mismatching classifier waveforms shown in Figure 3 are formed by collapsing across matching and mismatching semantic category trials. Therefore, any category effect should be equally represented in both waveforms. Therefore, in order to investigate the classifier effect without any influence from the same/different category, again here we repeated the exact same analysis but included *only* trials where the object depicted in the third picture belonged in the same category as the first two). The corresponding waveforms and maps can be seen in Figure 6.

In contrast with the previous analysis, this one revealed not only a main effect of condition (F (1, 34) = 6.003, p = .020, η_p^2 = 0.150) but also a significant interaction between Group and Condition ($F(1, 34) = 7.440$, $p = 0.004$, η_p^2 = 0.225). Post hoc analysis with a Bonferroni correction revealed that the mean amplitude of mismatching classifier trials was significantly more negative than the matching classifier trials *but only* for Chinese-English bilinguals (Mean Difference = 1.056, SE = 0.334, t = 2.160, $P_{bonf} = 0.02$), and not for English monolinguals (Mean Difference = 0.743, SE = 0.334, t = 2.224, $P_{bonf} = 0.19$).

Table 4. Mean amplitude of frontal-central electrodes from 400-550ms of matching and mismatching classifier trials <u>when the third picture is semantically related to the first two</u>, for both bilingual and monolingual participants.

Language	Grammatically	Mean	SD	N
Bilingual	Congruent	-0.221	1.620	18
	Incongruent	-1.366	1.552	18
Monolingual	Congruent	-1.887	2.384	18
	Incongruent	-2.028	2.211	18

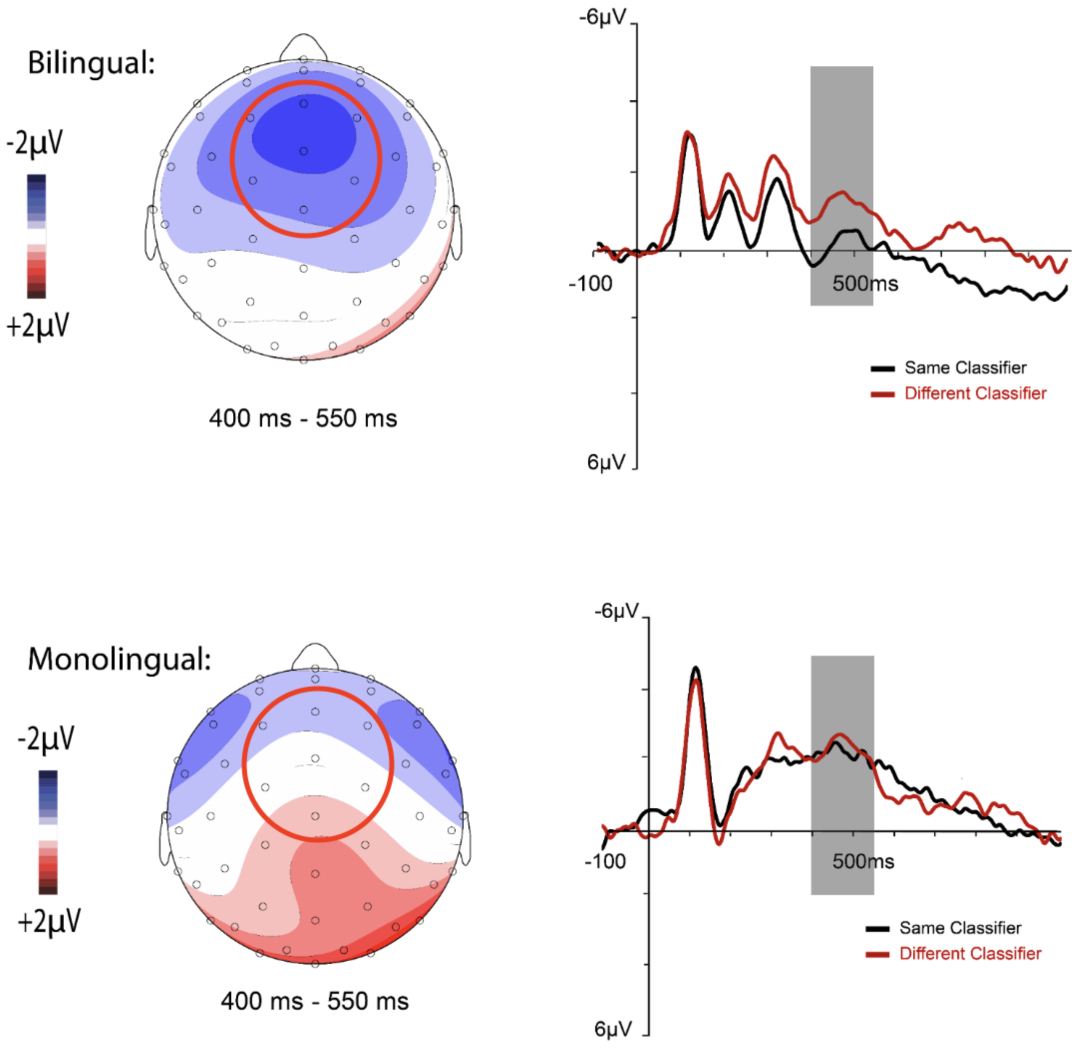


Figure 7. Scalp topography during 400-550ms and ERPs elicited in the matching and mismatching classifier conditions when the third picture is also semantically related to the first two, pooled from frontal-central channels.

3.3 Discussion:

The purpose of study 2 was to investigate whether grammatical features of a language not currently in use could affect how speakers of that language perceive and process images of objects. Previous research like Boutonnet et al. (2012) has suggested that arbitrary grammatical features like grammatical gender were unconsciously activated during an object categorization task. However, whether this could be generalized to all grammatical features remained unclear. In study 2, we investigated whether Boutonnet et al. (2012)'s results could be replicated with a less arbitrary grammatical feature like the classifier system of Chinese.

In terms of behavioral data, we replicated Boutonnet et al. (2012). Like them, we found no effect of semantic category, nor grammatical classifier on reaction time. The only behavioral finding of Boutonnet et al. that we failed to replicate was the main effect of Group, as our monolingual and bilingual participants did not differ significantly in their reaction time. Additionally, our reaction times were, on average, longer than in Boutonnet et al. (2012). It could be that our pictures were visually more complex than theirs, or our categories more difficult to discern. However, participants in the present study, with an accuracy of 96.5%, did significantly better in the categorization task than those from Boutonnet et al. (2012), who reported an accuracy of 79%.

Although we did not observe a category priming effect in our behavioral data, the electrophysiological data do suggest a semantic priming effect: That is, the N400 elicited by category (semantically) mismatching pictures was more negative than that elicited by in the category matching condition. This supports our behavioral findings in terms of accuracy, showing that all participants were correctly engaged in the semantic categorization task as requested.

The most important finding for our experimental question is whether we would find any evidence of an effect of the Chinese classifiers restricted to our bilingual participants. However, when comparing all matching classifier trials to mismatching classifier trials, we failed to observe a significant difference between groups in the mean amplitude of the frontal central channels from 400 to 550ms. As predicted, we expected a difference between groups because the manipulation of a grammatical Chinese feature should only be accessible to the Chinese-English bilinguals, and not the English monolinguals. However, as mentioned in the results section, when including all trials in the grammatical (classifier) comparison, half of the trials contributing to the ERPs were semantically (category) mismatching. Considering that the classifier feature is less arbitrary than gender, and it may reflect the visual similarity or proximity in function, monolingual English speakers might be picking up on those similarities even if they don't speak a language that contains and uses that grammatical information. This possibility is supported by the results from study 1, where all participants, including monolingual English speakers, rated the pairs sharing the same classifier as more similar than those that use a different classifier. In this analysis, we included both the trails that were semantically related and trials that were semantically unrelated. Considering that the semantic effect could potentially confound the grammatical effect, we decided to control for semantic relatedness when comparing grammatically (classifier) matching and mismatching trials. Although leaving us with only half of the trials and therefore, less power, this helps remove the possible confound of category match/mismatch.

When including only trials that were semantically related, we found a significant main effect of classifier in the frontal central region around 400 to 550 ms. that is only present in bilingual participants: the ERP amplitude in this time range was significantly more negative in grammatically mismatching trials than in the matching condition. This suggests that bilingual participants were activating classifier information while performing an object categorization task in English, with no mention of their native language (Chinese). Since all participants were not explicitly told about the recruitment criteria after filling out the screening survey weeks prior to the experiment, and the entire experiment was carried out in an English context, the participants could not have been aware of the classifier manipulation during the experiment. This was confirmed by a verbal post-experiment manipulation check. Most participants indicated that they thought the experiment was about how people speaking different languages categorize things differently, with no mention of the classifier manipulation, confirming that the design was successful in masking the actual goal of the investigation and in engaging the participants to perform the category task we asked them to do.

Similar to Boutonett et al. (2012), we interpret these findings as evidence that Chinese classifier information is unconsciously activated during the semantic (category) processing of the images of objects, even with no explicit mention of the grammatical feature being manipulated. This suggests that while looking at an object, the Chinese speakers are activating classifier information automatically in real life without intentionally doing so. This shows support for the label-feedback hypothesis (Lupyan, 2012), as it suggests by speaking a language that contains categorizing information like the classifier, the linguistic “label” (even grammatical ones) assigned to the object is activated when during perceptual processing of the object (during the object categorization task), even when the grammatical feature is not mentioned, and participants were not tested in the language containing the grammatical information.

One major finding that we failed to replicate from Boutonett et al. (2012) is the LAN they found in the Spanish-English bilingual participants during trials where there was a mismatch in grammatical gender. Although we did observe a general frontal central negativity in the bilingual participants when there was a mismatch in classifier, what we have observed could not be concluded as a LAN. Boutonett et al. (2012) expected to observe LAN because it is a marker for morphosyntactic processing in ERP. They interpreted the finding of LAN being significantly more negative in grammatical mismatch conditions as evidence for the activation of gender information during the semantic object categorization task. As previously mentioned, classifier represents less arbitrary conceptual categories that could be picked up even when people do not have access to it as a grammatical feature. The failure to replicate Boutonett et al. (2012)'s finding of LAN could be explained if grammatical gender and classifier are processed differently. Since classifier represent a system of several dimensions that pre-exist in perceptual processing, the activation of classifier information might not be purely grammatical as it might elicit conceptual categorization, congruent to classifier categories. Since the frontal central negativity in the present study is observed exclusively in the Chinese-English bilingual participants, it still suggests that classifier information, a grammatical feature in Chinese, was unconsciously activated during semantic processing of objects in an English context. We interpret this result as supportive of the label-feedback hypothesis, because it suggests that the classifier assigned to the object is actively modulating perceptual processing during a semantic categorization task.

Appendix A: Table of stimuli

Semantically related, Grammatically congruent

colorL01(1): T-shirt	colorL02(2)-color-cloths005: Shirt	colorL16(2): sleeveless shirt
colorL03(2): trench coat	colorL08(3): coat	colorL18(2): jacket
colorL21(2): dress shoe	colorL21(3)L: sneaker	colorL21(4)-color-cloths002: shoe
colorB08(1): cat	colorB01(2): rabbit	colorB13(1): fox
colorE08(1): crab	colorE09(2): starfish	colorE11(1): octopus
colorC13(1): hawk	colorC14(1): swan	colorC18: pigeon
colorC23-color-animal030: goose	colorC24-color-animal034: rooster	colorC25-color-animal035: duck
colorB17(2): chimpanzee	colorB11(1): leopard	colorB10(1): panda
colorF01: mosquito	colorF03(1): fly	colorF05: bee
colorP11(1): microwave	colorP13(1): refrigerator	colorP03(2): rice cooker
colorB12(1): elephant	colorB03(1)color_animal016: lion	colorB23(1): bear
colorB4(2): cattle	colorB21(1)-color-animal017: sheep	colorB29-color-animal024: camel
colorB04(1)-color-animal015: cow	colorB12(2): elephant	colorB07-color-animal012: deer
colorN02(1): computer desk	colorN18(1): coffee table	colorN04(1): desk
colorV16(2): scissors	colorV20(1)-color-tool025: nutcracker	colorV14(1): pliers
colorV11-color-tool018: wrench	colorV08(2): screwdriver	colorV07(1): tweezers
colorW12: dagger	colorW21(1): single-edged sword	colorW24(2): knife
colorM06(2): rose	colorM15(1): lily	colorM17(2): lotus flower
colorD07(1): goldfish	colorD12(2): cutlassfish	colorD09(2): catfish
colorF02(2): centipede	colorF07(2): caterpillar	colorF10(1): earthworm
colorH17: carrot	colorH19-color-vege002: eggplant	colorH23(2): cucumber
colorH26(2): celtuce	colorH20-color-vege007: celery	colorH21: garlic (long green)
colorQ01: car	colorQ04(4): bus	colorQ18: sports car
colorG05(1): lychee	colorG09(3): Hawthorn	colorG20-color-fruit001: strawberry

Semantically unrelated, Grammatically congruent

colorB09(1): pig	colorB22(1): donkey	colorH18: onion
colorF04: butterfly	colorF09(1): spider	colorL11(1): boot
colorC08(2): sparrow	colorC10(2): parrot	colorX14(2): bowl
colorB5(1): monkey	colorB17(1): gorilla	colorG06-color-fruit011: pear
colorL19-color-cloths015: socks	colorL20(1): hat	colorE16(1): shrimp
colorB14(1): tiger	colorB15(1): kangaroo	colorJ19(04): clip
colorK27: comb	colorK44: makeup brush	colorH03(2): spring onion
colorP08(1): TV	colorJ32: computer	colorQ05(1): tractor
colorJ01(1): ballpoint pen	colorJ13: pencil	colorW04(1): sniper rifle
colorX02(1): fork	colorX11: spoon	colorK12: flashlight
colorK17(1)-color-tool028: lock	colorK18: key	colorH08: spinach
colorF14: cricket	colorF18(1): mantis	colorJ18: calculator
colorE13(2): cuttlefish	colorD15: seahorse	colorJ20: mouse (computer)
colorF15: moth	colorF17: ant	colorK04(2): watch
colorQ05 (2): tractor	colorQ07 (2): motorbike	colorU11(1): piano
colorV15(2): shovel	colorV12-color-tool017: axe	colorH15(2): coriander
colorJ10: ruler	colorJ30(1): tape measure	colorU23(2): guitar
colorN04(2): desk	colorN18(3): coffee table	colorW07(1): bow and arrow
colorN12(2): sofa	colorN17(1): bed	colorJ31: optical disk
colorV25-color-tool014: brush	colorV16--color-tool024: scissors	colorH01: garlic chives
colorU03(2): violin	colorU23(3): guitar	colorN13(1): chair
colorJ30(2)-color-tool015: tape measure	colorJ36-color-tool023: triangular ruler	colorU03(1): violin
colorM01(1): chrysanthemum	colorM16(2): peony	colorH31-color-vege001: mushroom
colorL25-color-cloths003: tie	colorL27-color-cloths013: scarf	colorD14: shark

Semantically related, Grammatically incongruent

colorB23(2)-color-animal022: bear	colorB21(1)-color-animal017: sheep	colorB02(2): horse
colorL14(1): jeans	colorL12(1)-color-cloths010: dress	colorL10(1): sweater
colorB11(2): leopard	colorB14(2): tiger	colorB06(2): wolfdog
colorB08(2): cat	colorB01(4): rabbit	colorB06(3)color-animal010: dog
colorW13(1): handgun	colorW15(2): machine gun	colorW01(1): gun
colorB10(2)-color-animal023: panda	colorB15(2)-color-animal021: kangaroo	colorB09(2)-color-animal018: pig
colorB28-color-animal018: giraffe	colorB14(3): tiger	colorB12(4): elephant
colorB5(2): monkey	colorB13(2): fox	colorB22(3)-color-animal013: donkey
colorB01(01)color-animal006:rabbit	colorB11(3): leopard	colorB26-color-animal008: zebra
colorB11(4)color-animal009: leopard	colorB13(4)color-animal002: fox	colorB18(3): antelope
colorG17 (3): papaya	colorG21 (2): Melon	colorG24-color-fruit003: grape
colorN09(5): drawer	colorN10(1): wardrobe	colorN03(2): dresser
colorV02: file	colorV03: sickle	colorV17(1): saw
colorF20: dragonfly	colorF22: cicada	colorF07: caterpillar
colorW21(2): single-edged sword	colorW24(1)-color-tool021:knife	colorW08(1): sword
colorG25(1): watermelon	colorG07(2): mango	colorG28(3): durian
colorG10(2): pineapple	colorG16(1): kiwi	colorG13-color-fruit020: banana
colorG18: peach	colorG30-color-fruit010: pomegranate	colorG01(1): lemon
colorG22(2): apricot	colorG05-color-fruit015: lychee	colorG26(2): plum
colorA10(2): turtle	colorA03: gecko	colorA01(1): snake
colorG25-color-fruit018: watermelon	colorG12 (1): apple	colorG04-color-fruit017: cherry
colorH16(2): tomato	colorH07(1): green pepper	colorH04(2): water spinach
colorH23-color-vege013: cucumber	colorH28: zucchini	colorH14(1): lettuce
colorE11: octopus	colorE16(2): shrimp	colorD12(4): cutlassfish

Semantically unrelated, Grammatically incongruent

colorC17(2): gull	colorC21: Gyps	colorH10(2): cauliflower
colorE16(3): shrimp	colorE08(2): crab	colorH19(2): eggplant
colorH15(3): coriander	colorH03-color-vege006: welsh onion	colorK09(1): mirror
colorM06(1): rose	colorM14(2): sunflower	colorK22(1): broom
colorA06(1): crocodile	colorA01(2): snake	colorK42: sunglasses
colorG12-color-fruit005: apple	colorG16-color-fruit019: kiwi	colorW10: tank
colorG07-color-fruit009: mango	colorG08(2): coconut	colorL28: button
colorG17(1): papaya	colorG23: orange	colorW19(2): grenade
colorA08(1): turtle	colorA09(1): Toad	colorG04: cherry
colorA02(1): tortoise	colorA08(2): turtle	colorG24(2): grape
colorL05(2)-color-cloths007: trousers	colorL23: shorts	colorG14(1): persimmon
colorL10(2)-color-cloths008: sweater	colorL24(2): suit	colorF13(1): ladybug
colorV27: nail	colorV26: screw	colorL09(1): waistcoat
colorV04-color-tool016: hammer	colorV14(2): pliers	colorG15-color-fruit008 (pitaya)
colorP17(3): radio	colorP21: speaker	colorL17(1): down coat
colorN15(3): Bookshelf	colorN01(1): TV cabinet	colorH12: potato
colorB18(1): deer	colorB25color-animal007: polar bear	colorP04-color-elect004: telephone
colorB16(1): rat	colorB16(2): mice	colorJ12: notebook
colorB5(3)-color-animal020: monkey	colorB08(3): cat	colorI06: peanut
colorJ05(2): fax	colorJ14(2): printer	colorB24(2)-color-animal001: wolf
colorQ02(1): bike	colorQ13(2): tricycle	colorI01(1): walnut
colorB12(3)color-animal005: elephant	colorB19(1): whale	colorP16-color-elect008: air conditioner
colorC07(2):crow	colorC03(1): swallow	colorH27-color-vege011: pea
colorC12-color-animal029: owl	colorC19(1): ostrich	colorM04(1): cactus

Bibliography

A debate on language and thought. (n.d.). *The Economist*. Retrieved May 9, 2022, from https://www.economist.com/johnson/2010/12/13/a-debate-on-language-and-thought

A little labeling goes a long way: Semi-supervised learning in infancy—LaTourrette—2019—Developmental Science—Wiley Online Library. (n.d.). Retrieved March 21, 2022, from https://onlinelibrary.wiley.com/doi/full/10.1111/desc.12736

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Antunes, M., & Biala, G. (2012). The novel object recognition memory: Neurobiology, test procedure, and its modifications. *Cognitive Processing*, *13*(2), 93–110. https://doi.org/10.1007/s10339-011-0430-z

Aslanov, I. A., Sudorgina, Y. V., & Kotov, A. A. (2022). The Explanatory Effect of a Label: Its Influence on a Category Persists Even If We Forget the Label. *Frontiers in Psychology*, *12*, 745586. https://doi.org/10.3389/fpsyg.2021.745586

Athanasopoulos, P., & Casaponsa, A. (2020a). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Athanasopoulos, P., & Casaponsa, A. (2020b). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, *37*(5–6), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Barber, H., & Carreiras, M. (2005). Grammatical Gender and Number Agreement in Spanish: An ERP Comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153. https://doi.org/10.1162/0898929052880101

Beatty-Martínez, A. L., & Dussias, P. E. (2019). Revisiting Masculine and Feminine Grammatical Gender in Spanish: Linguistic, Psycholinguistic, and Neurolinguistic Evidence. *Frontiers in Psychology*, *10*, 751. https://doi.org/10.3389/fpsyg.2019.00751

Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research*, *1479*, 72–79. https://doi.org/10.1016/j.brainres.2012.08.044

Boutonnet, B., McClain, R., & Thierry, G. (2014). Compound words prompt arbitrary semantic associations in conceptual memory. *Frontiers in Psychology*, *5*, 222. https://doi.org/10.3389/fpsyg.2014.00222

Bowerman, M., & Levinson, S. (2001). *Language Acquisition and Conceptual Development*. https://doi.org/10.1017/CBO9780511620669.018

Cheng, L. L.-S., & Sybesma, R. (2008, October 16). *Classifiers in Four Varieties of Chinese*. The Oxford Handbook of Comparative Syntax. https://doi.org/10.1093/oxfordhb/9780195136517.013.0007

Dai, G. (2016). Classifier constructions in translated Chinese. *New Frontiers in Translation Studies*, 109–132. Scopus. https://doi.org/10.1007/978-981-10-0742-2_8

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, *19*(2), 189–195. https://doi.org/10.1111/j.1467-9280.2008.02066.x

Frank, M. C., Fedorenko, E., Lai, P., Saxe, R., & Gibson, E. (2012). Verbal interference suppresses exact numerical representation. *Cognitive Psychology*, *64*(1), 74–92. https://doi.org/10.1016/j.cogpsych.2011.10.004

Friederici, A. D., & Jacobsen, T. (1999). Processing Grammatical Gender During Language Comprehension. *Journal of Psycholinguistic Research*, *28*(5), 467–484. https://doi.org/10.1023/A:1023264209610

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183–192. https://doi.org/10.1016/0926-6410(93)90026-2

Gao, M. Y., & Malt, B. C. (2009). Mental representation and cognitive consequences of Chinese individual classifiers. *Language and Cognitive Processes*, *24*(7–8), 1124–1179. https://doi.org/10.1080/01690960802018323

Gleitman, L., & Papafragou, A. (2005). Language and Thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge Handbook of Thinking and Reasoning* (pp. 633–661). Cambridge University Press.

Gliga, T., Volein, A., & Csibra, G. (2010). Verbal Labels Modulate Perceptual Object Processing in 1-Year-Old Children. *Journal of Cognitive Neuroscience*, *22*, 2781–2789. https://doi.org/10.1162/jocn.2010.21427

Grüter, T., Lau, E., & Ling, W. (2020). How classifiers facilitate predictive processing in L1 and L2 Chinese: The role of semantic and grammatical cues. *Language, Cognition and Neuroscience*, *35*(2), 221–234. https://doi.org/10.1080/23273798.2019.1648840

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

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660–676. https://doi.org/10.1111/j.1469-8986.1997.tb02142.x

Hsu, C.-C., Tsai, S.-H., Yang, C.-L., & Chen, J.-Y. (2014). Processing classifier–noun agreement in a long distance: An ERP study on Mandarin Chinese. *Brain and Language*, *137*, 14–28. https://doi.org/10.1016/j.bandl.2014.07.002

Huang, S., & Schiller, N. O. (2021). Classifiers in Mandarin Chinese: Behavioral and electrophysiological evidence regarding their representation and processing. *Brain and Language*, *214*, 104889. https://doi.org/10.1016/j.bandl.2020.104889

Imai, M. & Mazuka. (2003). Re-evaluating linguistic relativity: Language-specific categories and the role of universal ontological knowledge in the construal of individuation. In *Language in mind: Advances in the study of language and thought* (pp. 430–464).

Imai, M., Saalbach, H., & Stern, E. (2010). Are Chinese and German Children Taxonomic, Thematic, or Shape Biased? Influence of Classifiers and Cultural Contexts. *Frontiers in Psychology*, *1*, 194. https://doi.org/10.3389/fpsyg.2010.00194

Jiang, S. (2017). The Semantics of Chinese Classifiers and Linguistic Relativity. In *The Semantics of Chinese Classifiers and Linguistic Relativity* (p. 215). https://doi.org/10.4324/9781315265483

Kemmer, L., Coulson, S., De Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology*, *41*(3), 372–384. https://doi.org/10.1111/1469-8986.2004.00180.x

Kotz, S. A. (2009). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain and Language*, *109*(2–3), 68–74. Scopus. https://doi.org/10.1016/j.bandl.2008.06.002

Kuo, J. Y.-C., & Sera, M. D. (2009). Classifier Effects on Human Categorization: The Role of Shape Classifiers in Mandarin Chinese. *Journal of East Asian Linguistics*, *18*(1), 1–19.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463–470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & Cognition*, *11*(5), 539–550. https://doi.org/10.3758/BF03196991

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in Chinese: Evidence from ERPs. *Cognition*, *166*, 433–446. https://doi.org/10.1016/j.cognition.2017.06.010

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920–933. https://doi.org/10.1038/nrn2532

Lee, C., Kim, Y.-W., & Yi, B.-U. (2021). *Numeral classifiers and classifier languages: Chinese, Japanese, and Korean*. Scopus. https://doi.org/10.4324/9781315166308

Li, F., Hong, X., & Wang, Y. (2021). The N400 and Post-N400 positivity effect in Mandarin classifier-noun congruence: An ERP study. *Journal of Neurolinguistics*, *57*, 100958. https://doi.org/10.1016/j.jneuroling.2020.100958

Li, P., Dunham, Y., & Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, *58*(4), 487–524. https://doi.org/10.1016/j.cogpsych.2008.12.001

Lucy, J. A. (2001). Sapir–Whorf Hypothesis. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 13486–13490). Pergamon. https://doi.org/10.1016/B0-08-043076-7/03042-4

Maciuszek, J., Polak, M., & Świątkowska, N. (2019). Grammatical Gender Influences Semantic Categorization and Implicit Cognition in Polish. *Frontiers in Psychology*, *10*, 2208. https://doi.org/10.3389/fpsyg.2019.02208

Maier, M., Glage, P., Hohlfeld, A., & Abdel Rahman, R. (2014). Does the semantic content of verbal categories influence categorical perception? An ERP study. *Brain and Cognition*, *91*, 1–10. https://doi.org/10.1016/j.bandc.2014.07.008

Molinaro, N., Barber, H. A., Caffarra, S., & Carreiras, M. (2015). On the left anterior negativity (LAN): The case of morphosyntactic agreement: a reply to Tanner et al. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *66*, 156–159. https://doi.org/10.1016/j.cortex.2014.06.009

Montefinese, M., Ambrosini, E., & Roivainen, E. (2019). No grammatical gender effect on affective ratings: Evidence from Italian and German languages. *Cognition and Emotion*, *33*(4), 848–854. https://doi.org/10.1080/02699931.2018.1483322

Ni, L., Liu, Y., Yu, W., & Fu, X. (2019). The China Image Set (CIS): A New Set of 551 Colored Photos With Chinese Norms for 12 Psycholinguistic Variables. *Frontiers in Psychology*, *10*, 2631. https://doi.org/10.3389/fpsyg.2019.02631

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. *Scientific Reports*, *8*, 14156. https://doi.org/10.1038/s41598-018-32483-2

O'Rourke, E., & Coderre, E. L. (2021). Implicit Semantic Processing of Linguistic and Non-linguistic Stimuli in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *51*(8), 2611–2630. https://doi.org/10.1007/s10803-020-04736-5

Plunkett, K., Hu, J.-F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665–681. https://doi.org/10.1016/j.cognition.2007.04.003

Qian, Z., & Garnsey, S. M. (2016). A sheet of coffee: An event-related brain potential study of the processing of classifier-noun sequences in English and Mandarin. *Language, Cognition and Neuroscience*, *31*(6), 761–784. https://doi.org/10.1080/23273798.2016.1153116

Rösler, F., Pechmann, T., Streb, J., Röder, B., & Hennighausen, E. (1998). Parsing of Sentences in a Language with Varying Word Order: Word-by-Word Variations of

Processing Demands Are Revealed by Event-Related Brain Potentials. *Journal of Memory and Language*, *38*(2), 150–176. https://doi.org/10.1006/jmla.1997.2551

Schmid, M. S., & Yılmaz, G. (2018). Predictors of Language Dominance: An Integrated Analysis of First Language Attrition and Second Language Acquisition in Late Bilinguals. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01306

Speed, L. J., Chen, J., Huettig, F., & Majid, A. (2021). Classifier categories reflect but do not affect conceptual organization. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *47*(4), 625–640. https://doi.org/10.1037/xlm0000967

Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and Language*, *120*(2), 135–162. https://doi.org/10.1016/j.bandl.2011.07.001

Tanner, D. (2015). On the left anterior negativity (LAN) in electrophysiological studies of morphosyntactic agreement: A commentary on “grammatical agreement processing in reading: ERP findings and future directions” by Molinaro et al., 2014. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *66*, 149–155. https://doi.org/10.1016/j.cortex.2014.04.007

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301. https://doi.org/10.1016/j.neuropsychologia.2014.02.002

Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Tivadar, R. I., & Murray, M. M. (2019). A Primer on Electroencephalography and Event-Related Potentials for Organizational Neuroscience. *Organizational Research Methods*, *22*(1), 69–94. https://doi.org/10.1177/1094428118804657

Vainapel, S., Shamir, O. Y., Tenenbaum, Y., & Gilam, G. (2015). The dark side of gendered language: The masculine-generic form as a cause for self-report bias. *Psychological Assessment*, *27*(4), 1513–1519. https://doi.org/10.1037/pas0000156

Vigliocco, G., Vinson, D., Paganelli, F., & Dworzynski, K. (2005). Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use. *Journal of Experimental Psychology. General*, *134*, 501–520. https://doi.org/10.1037/0096-3445.134.4.501

von Grebmer zu Wolfsthurn, S., Pablos, 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

Vos, S. H., Gunter, T. C., Kolk, H. H. J., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*(1), 41–63. https://doi.org/10.1111/1469-8986.3810041

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, R., Zhang, C., 王瑞晶, & 张侮偲. (2014). EFFECT OF CLASSIFIER SYSTEM ON OBJECT SIMILARITY JUDGMENT: A CROSS-LINGUISTIC STUDY / 量词系统对物体相似度判断的影响：一项跨语言研究. *Journal of Chinese Linguistics*, *42*(1), 188–217.

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf* (J. B. Carroll, Ed.). MIT Press.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 253–265. https://doi.org/10.1002/wcs.104

Yang, K. (2015). On the Semantic Functions of Three Different Types of Classifier Reduplication in Mandarin Chinese. In D. Xu & J. Fu (Eds.), *Space and Quantification in Languages of China* (pp. 207–231). Springer International Publishing. https://doi.org/10.1007/978-3-319-10040-1_11

Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonalities and differences in the neural representations of English, Portuguese, and Mandarin sentences: When knowledge of the brain-language mappings for two languages is better than one. *Brain and Language*, *175*, 77–85. https://doi.org/10.1016/j.bandl.2017.09.007

Yi, B.-U. (2021). Numeral classifiers and diversity of classifier systems. In *Numeral Classifiers and Classifier Languages: Chinese, Japanese, and Korean* (pp. 6–39). Scopus.

Younger, B. A. (1985). The Segregation of Items into Categories by Ten-Month-Old Infants. *Child Development*, *56*(6), 1574–1583. https://doi.org/10.2307/1130476

Yum, Y. N., & Law, S.-P. (2019). Impact of characteristics of L1 literacy experience on picture processing: ERP data from trilingual non-native Chinese and English readers. *Cognition*, *183*, 213–225. https://doi.org/10.1016/j.cognition.2018.11.009

Zhu, S. (2021). An empirical study of the classifier effect in advanced L2 users of Chinese. *International Journal of Bilingualism*, *25*(1), 151–167. Scopus. https://doi.org/10.1177/1367006920943301