Effects of Chinese word structure on object categorization in Chinese–English bilinguals*

XUAN PAN

A N D

DEBRA JARED

Department of Psychology, University of Western Ontario

ABSTRACT

We investigated how verbal labels affect object categorization in bilinguals. In English, most nouns do not provide linguistic clues to their categories (an exception is *sunflower*), whereas in Chinese, some nouns provide category information morphologically (e.g., 鸵鸟- *ostrich* and 知更鸟- *robin* have the morpheme鸟- *bird* in their Chinese names), while some nouns do not (e.g., 企鹅- *penguin* and 鸽子- *pigeon*). We examined the effect of Chinese word structure on bilinguals' categorization processes in two ERP experiments. Chinese–English bilinguals and English monolinguals judged the membership of atypical (e.g., *ostrich*, *penguin*) vs. typical (e.g., *robin*, *pigeon*) pictorial (Experiment 1) and English word (Experiment 2) exemplars of categories (e.g., *bird*). English monolinguals showed typicality effects in RT data, and in the N300 and N400 of ERP data, regardless of whether the object name had a category cue in Chinese. In contrast, Chinese–English bilinguals showed a larger typicality effect for objects without category cues in their name than objects with cues, even

when they were tested in English. These results demonstrate that linguistic information in bilinguals' L1 has an effect on their L2 categorization processes. The findings are explained using the label-feedback hypothesis.

1. Introduction

There has been renewed research interest in Whorf's (1956) idea that language influences cognitive processes (Pavlenko, 2014). Here our interest is in how verbal labels (or words) affect object categorization processes in bilinguals. Bilinguals have two potential sources of influence instead of just one. Labels can be constructed differently across languages, and translation-equivalent words often do not refer to precisely the same conceptual and perceptual items. Therefore, the labels from each language could potentially have a somewhat different influence on the conceptual representations of bilinguals. Indeed, studies investigating naming patterns of household objects have suggested that bilinguals who were raised speaking two languages differ from monolinguals in both their first and second languages in the set of objects that is given a particular category label (e.g., Ameel, Storms, Malt, & Sloman, 2005). The current study examined the effects of Chinese word structure on object categorization in Chinese–English bilinguals. Before presenting our experiments, we first briefly discuss research from monolinguals on the impact of labels on categorization; next we present our guiding theoretical framework; and then we describe a study that inspired our work.

1.1. THE EFFECTS OF LABELLING ON CATEGORIZATION

Studies of category learning have shown that categories are learned more effectively when they are accompanied by labels (Fulkerson & Waxman, 2007; Lupyan & Casasanto, 2015; Lupyan, Rakison, & McClelland, 2007; Robinson, Best, Deng, & Sloutsky, 2012). For example, in Lupyan et al. (2007), participants learned to categorize "aliens" as those to be approached or to be avoided with nonsense category labels or other non-linguistic cues present or not. Learning named categories was easier than learning unnamed categories, and this facilitation effect could not be achieved by providing other non-linguistic cues. Labels have also been shown to aid categorization of previously learned familiar items (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012). Once a category is learned, key features of the category are more effectively activated by a verbal label than by other highly associated cues, such as non-linguistic sounds (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015;

Lupyan & Thompson-Schill, 2012), and numbers or symbols (Gervits, Johanson, & Papafragou, 2016). Furthermore, categorization processes are impaired by verbal interference tasks (Lupyan, 2009) and anomic aphasia (Davidoff & Roberson, 2004; Lupyan & Mirman, 2013).

1.2. THE LABEL-FEEDBACK HYPOTHESIS

Lupyan (2012) proposed the label-feedback hypothesis as an account of the influence of word labels on object categorization and perception. It proposes that language is highly interconnected with other cognitive processes, and influences other functional networks in a top-down fashion. Labels can provide feedback in the form of activation to the level of conceptual representations; thus, named concepts should be activated differently under the on-line influence of the label compared to the same concepts activated by non-verbal means, or when the labels are prevented from affecting the concept (see Figure 1). According to this hypothesis, the activation of an object's verbal label results in the activation of the most typical features of the category (e.g., the label "car" activates the feature "has wheels" more strongly than the feature "is black"). This top-down activation from verbal labels to features produces a transient "perceptual warping" in which category members that share those features are drawn closer together and non-members are pushed away.

1.3. THE EFFECTS OF WORD STRUCTURE ON CATEGORIZATION

Most studies investigating the effects of labels on categorization have focused on the advantage of verbal labels over other non-verbal cues. In contrast, Liu et al. (2010) investigated the effects of word structure on categorization by capitalizing on differences in the way that words are constructed in Chinese and English. In English, most nouns do not provide linguistic clues to their categories (exceptions are *sunflower* and *bluebird*), whereas in Chinese, many nouns provide explicit category information either morphologically (e.g., the morpheme 鸟 *bird* in the noun 鸵鸟 *ostrich*, which is pronounced), or orthographically (e.g., the radical 虫 *bug* in the noun 蚊子 *mosquito*, which is not pronounced).[1] In their first experiment, Liu et al. used images of objects that have such category cues in their Chinese names as critical stimuli. Native speakers of Chinese and English were presented with category labels followed by images of typical and atypical exemplars and non-category exemplars of a category. Participants were asked to judge the membership of the pictures while their EEG (electroencephalogram) brainwaves were recorded. Generally, atypical items are categorized with more difficulty than typical items, a finding known as the typicality effect (Rosch, 1973, 1975). Previous studies investigating the typicality effect in pictorial stimuli with ERPs (event-related
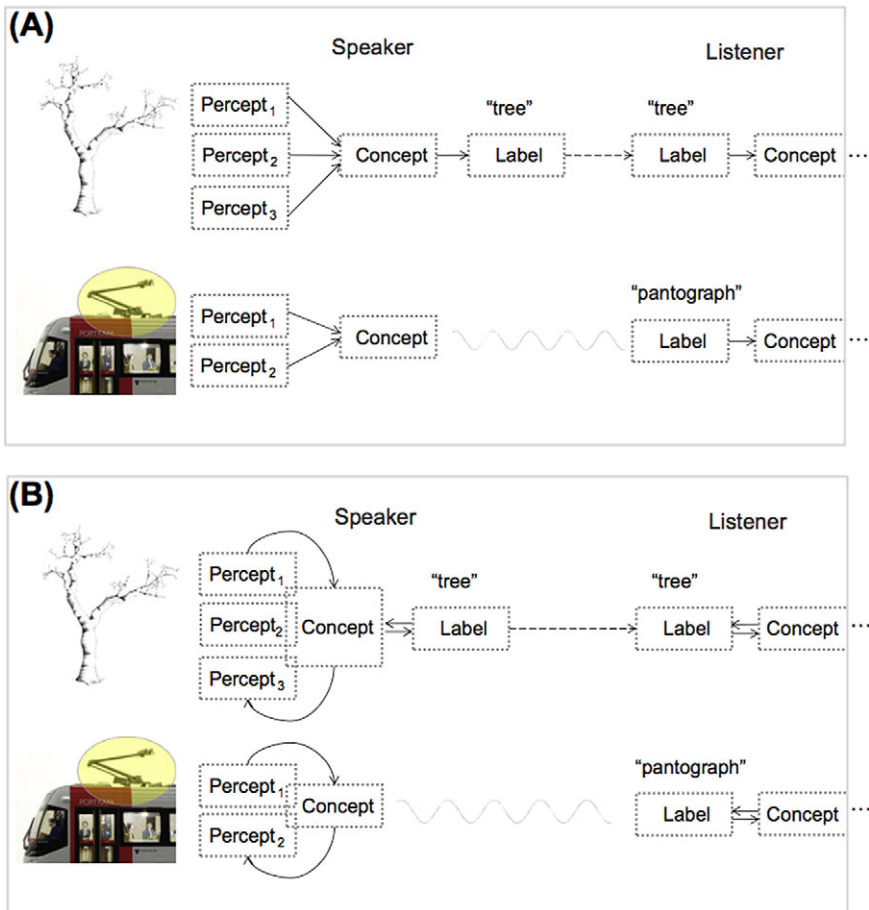
Fig. 1: The Label-feedback Hypothesis (Lupyan, 2012). (A): A schematic view of the standard account in which a word label is simply a means of accessing a concept. Multiple perceptual exemplars of a concept map onto a common conceptual representation. The concept is further mapped onto a word label, which enables a speaker to activate the same concept in a listener using the label. The one-way connections between representational layers prevent the word label to have an influence on the conceptual representations. (B): A schematic view of the label-feedback hypothesis. All representational layers are recurrently connected, which allows the word label to affect the conceptual representations through feedback. Reprinted from: Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. In B. H. Ross (ed.), *The psychology of learning and motivation, Vol. 57, pp. 255–297*. Waltham, MA: Academic Press.

potentials) have observed it in the N300 component (Hamm, Johnson, & Kirk, 2002; Hauk et al., 2007; Kiefer, 2001; McPherson & Holcomb, 1999). Other studies have suggested that N400 and a late positive component are also

471

involved in the typicality effect (Federmeier & Kutas, 2001; Ganis, Kutas, & Sereno, 1996; Hamm et al., 2002; West & Holcomb, 2002).

Liu et al. (2010) observed attenuated typicality effects for Chinese speakers compared to English speakers (who were unaware of the category labels in the Chinese names) in their behavioral data and in the N300 and N400 components of their ERP data. The category cue provided in Chinese nouns facilitated the categorization process for Chinese speakers and reduced the influence of typicality, even though the stimuli were pictures, not words. In their second experiment, with Chinese participants only, they specifically compared typicality effects for images with morphological category cues in their names (Ostrich: 鸵鸟, where 鸟 means *bird* and is pronounced as such), and images with orthographic category cues (Penguin: 企鹅, where 鸟 is embedded in the second character and is not pronounced like *bird*), and found somewhat smaller typicality effects for the former (although not in decision latency data). Morphological cues were more helpful in categorization than orthographic cues, perhaps because they more obviously represent the category name. To address the possibility that the results of the first two experiments simply reflect the detection of a match between the labels of a category (e.g., 鸟 "niǎo" *bird*) and an image (e.g., 鸵鸟 "tuóniǎo" *ostrich*), in Experiment 3 they tested English speakers using pictures whose name either did (e.g., *basketball*) or did not (e.g., *vehicle*) contain a category cue. Both produced typicality effects, suggesting that the results for Chinese do not reflect a simple lexical matching process. The authors explained their results by proposing that languages change the way people access semantic information. When categorizing atypical exemplars, English speakers needed additional semantic processing to make their decision, whereas Chinese speakers were able to avoid this additional processing because of the morphological category cue in the objects' Chinese names. However, Liu et al. neither specify what is involved in the additional semantic processing, nor did they propose a mechanism for their findings. The authors did mention that morphological transparency is prevalent in Chinese and is available from the time children first begin to acquire the language, and suggested that it could be an organizing feature of categories.

The label-feedback hypothesis (Lupyan, 2012) can explain Liu et al.'s (2010) findings, with the feedback influencing either on-line processing, as Lupyan suggested, or the structure of conceptual representations. Liu et al. presented a category label and then a picture in their experiments. The processing steps in this task would proceed as follows according to the first view. When participants see the category label, it activates typical features of the category (e.g., *bird* would activate most highly the features that are found in most members of the category, like has wings, has feathers, etc.). Then when the target picture appears, it quickly activates its corresponding label, and the label then activates features of the specific object. For an English monolingual,

a picture of a robin activates the label *robin*, then the label *robin* activates features such as has wings, has feathers, red belly, etc.; an ostrich picture would activate the label *ostrich*, which then activates features like has wings, cannot fly, runs fast, etc. It is easier to categorize a typical object than an atypical one, because the features activated from the category label would overlap more with the features activated for a typical exemplar than an atypical one. For a Chinese monolingual, if an object's Chinese name has the category cue embedded, then this cue would facilitate the activation of the most typical features of the category, even when the object is an atypical exemplar of the category. For example, the morpheme 鸟 *(bird)* embedded in the Chinese label 鸵鸟 *(ostrich)* would make the typical features of the category *bird* more available. Therefore, the perceptual features activated from a category label would have more overlap with the features activated from the feedback from a label with category cue than a label without cue, thus producing a faster response and less negative N300 and N400.

This explanation assumes that the verbal label of the target object is activated quickly when the picture is presented and then the activation of the object label influences the categorization decision. However, it is not clear whether pictures would activate their corresponding names quickly enough in a categorization task. Previous studies observed that pictures are categorized faster than they could be named (e.g., Irwin & Lupker, 1983; Potter & Faulconer, 1975).

An alternative way to understand Liu et al.'s (2010) findings is that the organization of category representations in Chinese speakers changes under the long-term influence of feedback from everyday usage of objects' Chinese labels. More specifically, through daily feedback from Chinese labels, category members that have a category cue in their Chinese names become more strongly associated with the most typical features of the category, resulting in them being stored closer together in the center of the category space, while members that do not have a category cue in their Chinese names are stored in the periphery. For example, every time bilinguals use the Chinese label 鸵鸟 *(ostrich)* to refer to an ostrich, the category cue 鸟 *(bird)* embedded in the label sends feedback to the conceptual representations. This feedback activates the most typical features of the category bird, making *ostrich* share more features with a typical bird (e.g., *robin*). Through years of influence from the Chinese label 鸵鸟 *(ostrich),* the conceptual representation of an *ostrich* in Chinese speakers would be stored closer to typical birds (e.g., *robin*) in the center of the bird category, thus making it easier to categorize an *ostrich* as a bird. On the contrary, because the atypical bird *penguin* does not have a category cue in its Chinese name, the most typical features of the category bird would not get a boost in activation every time the label was used. As a result, the conceptual representation of a *penguin* would be stored at the periphery of the category

473

bird, further away from typical birds, making it difficult to categorize a *penguin* as a bird and producing a slower response and more negative N300 and N400.

In summary, Liu et al.'s (2010) study has suggested that categorization is not only affected by whether or not an object has a verbal label, but also by structure of the verbal labels. The label-feedback hypothesis provides a useful framework to understand the how verbal labels influence categorization.

1.4. THE PRESENT STUDY

We extended the work of Liu et al. (2010) to investigate the impact of verbal labels on categorization in bilinguals. One implication of the label-feedback hypothesis for bilinguals is that there are two potential sources of feedback (one from the label in each language) instead of one. Of interest here was whether bilinguals' performance in a categorization task that is conducted in one language is influenced by knowledge of their other language. Specifically, we were interested in whether Chinese–English bilinguals' performance in an English category decision task is influenced by their knowledge of Chinese word structure. In Experiment 1, picture targets were used that either did or did not have a morphological category cue in their name. Bilinguals were tested in separate English and Chinese sessions. To encourage participants to use the target language in performing the task, each session was designed to appear as monolingual as possible. For the English session, participants were greeted in English by a monolingual Caucasian research assistant, and all conversation and consent forms were in English. Similarly, in the Chinese session, participants were greeted in Chinese by an Asian native Chinese speaker and all conversation and forms were in Chinese. In Experiment 2, to increase the focus on English even further, there was a single English session and English word targets were used. In both experiments, a comparison group of monolingual English speakers was tested. These participants were unaware of the category labels in the Chinese names. To be comparable to the Liu et al. study, we collected both behavioral and ERP data. In the ERP data of Experiment 1, we examined the N300 and N400 components, and in Experiment 2, where targets were words, we examined just the N400 component. As noted previously, the N300 has been observed in studies with picture targets specifically (Hamm et al., 2002; Kiefer, 2001; McPherson & Holcomb, 1999).

Several limitations in the methodology of Liu et al.'s (2010) study were addressed. First, Liu et al. did not include any exemplars without a category cue in their Chinese names. Such stimuli are needed to show that Chinese speakers are indeed sensitive to typicality when no category cue is available. Here, exemplars that did and did not contain a category cue were included. A second limitation is that they used a small number of stimuli and presented them repeatedly to get enough datapoints for the ERP analysis. In their

Experiment 2, items with a morphological cue came from only five categories, with one typical and one atypical item in each category, and each stimulus was presented ten times. Here, more exemplars with a category cue in their Chinese name were added and they came from 11 different categories.

In the Chinese session of Experiment 1, bilinguals were expected to show a smaller typicality effect when the objects had a category cue in their Chinese name than when they did not. Of particular interest was whether Chinese–English bilinguals would show the effects of Chinese category cues even when they were tested in English in Experiment 1 and when targets were English words in Experiment 2. We did not attempt to distinguish between Lupyan's (2012) hypothesis that feedback from labels would produce temporary perceptual warping and the view that feedback from labels results in long-term category restructuring.

2. Experiment 1

2.1. METHOD

2.1.1. *Participants*

Thirty-four Chinese–English bilinguals (mean age 19 years, range 18–29, 25 female) and 28 English monolinguals (mean age 19 years, range 18–22, 20 female) were tested. Participants received course credit or money for their participation. Data from six bilinguals were excluded (three were native speakers of Cantonese, two did not complete the whole session, one had poor ERP recording), leaving 28 Chinese–English bilinguals and 28 English monolinguals in the final sample. The first language of the bilinguals was Mandarin. All bilinguals were born in China or Taiwan, had lived in there for a mean of 16.0 years (range 9–25), and had lived in Canada for a mean of 4.9 years (range 2–9).

2.1.2. *Materials*

Pictures of typical and atypical exemplars of various categories were used. Two pilot studies were done to acquire typicality rating and name agreement data for the stimuli. In Pilot Study 1, 249 category label and item word pairs (e.g., BIRD-robin) were presented to 34 English native speakers without any knowledge of Chinese (mean age 18 years, range 18–26, 23 female) and 18 Chinese native speakers (mean age 18 years, range 18–21, 6 female) one at a time. Participants were asked to rate the typicality of the item using a 0 to 100 slide scale (0-Atypical, 100-Typical). The mean typicality rating for each item was computed. Based on the data from English speakers, who were unaware of the Chinese category cue, 13 categories and 108 items were selected

475

T A B L E 1. *Mean typicality ratings by English and Chinese L1 participants of the 108 words selected from pilot study 1. All of these words were used in Experiment 2.*

Participants and Word Type	Typical	Atypical
English		
Cue	90.10	66.87
No-Cue	90.29	64.37
Chinese		
Cue	90.81	82.87
No-Cue	90.41	75.55

(see Table 1). Half of the items were typical, half were atypical. Items with the highest ratings for a category were selected as typical, items with the lowest ratings were selected as atypical. Half of the items had a morphological category cue in their Chinese names, half did not. Analyses of the typicality ratings showed that there was a main effect of typicality ($\chi^2(1) = 88.34$, $p < .001$). Typical items had higher ratings than atypical items. Importantly, the interaction between typicality (Typical vs. Atypical), language group (English vs. Chinese), and word type (Cue vs. No-Cue) was not significant ($\chi^2(1) = 2.04$, $p > 0.15$), that is, there was no difference in the ratings across language groups regarding the relationship between typicality and word type.

In Pilot Study 2, name agreement data were collected for images corresponding to the 108 items chosen in Pilot Study 1. Images of 108 items were selected from the Internet, all in colour with a white background. Images were presented to 55 English native speakers without any knowledge of Chinese (mean age 22 years, range 18–25, 32 female) and 46 Chinese native speakers (mean age 19 years, range 17–24, 39 female) one at a time. Participants were asked to type in a name for each image. Mean name agreement (percentage of expected name) was computed for each item. Items for which fewer than 30% of participants gave the expected name were excluded (with the exception of 4 items, due to the difficulty in getting the same number of items for each condition). The final stimulus list consisted of 11 categories and 84 images, with 21 images in each of the four experimental conditions (see Table 2). Examples of the stimuli are in Appendix A. In this set, there was a main effect of typicality ($\chi^2(1) = 78.79$, $p < .001$), but importantly, no significant interaction of typicality, language group, and word type was found ($\chi^2(1) = 1.12$, $p > .25$).

2.1.3. *Procedure*

A category label–image matching task was used (see Figure 2). Participants first saw a 500 ms fixation cross, followed by a category label (e.g., *BIRD* in

476

TABLE 2. *Mean percentage naming agreement in Pilot Study 2 and corresponding typicality ratings from Pilot Study 1 for the pictures used in Experiment 1*

Participants and Word Type	Name Agreement		Typicality	
	Typical	Atypical	Typical	Atypical
English				
Cue	76.23	62.36	89.69	66.63
No-Cue				
Chinese	82.76	74.25	91.81	65.89
Cue	83.53	68.58	90.76	82.33
No-Cue	77.23	73.64	91.28	76.25

English; 鸟 in Mandarin) for 500 ms, then followed by an image of an object (e.g., *robin*). Participants were instructed to judge whether the image was an example of the concept represented by the first word. All category label–picture pairs were presented twice to each participant (in random order) to get a clear ERP signal after averaging. There were 348 trials, including 168 critical trials that required a *yes* response (42 trials per condition), 168 filler trials that required a *no* response, and 12 practice trials. Filler trials were created by re-pairing the category label–image pairs from critical trials. That is, the same set of category labels and the same set of pictures was used for the filler (*no*) trials as the critical (*yes*) trials, but in the former case they were shuffled to create unrelated pairs. This means that each target picture was presented four times: two requiring a *yes* response, and two requiring a *no* response. English monolinguals were tested only in English. Chinese–English bilinguals were tested in both Chinese and English in separate sessions. The second session was conducted at least 7 days after the first; half of the participants did the Chinese session first, and half did English session first. The testing environment matched the language of the session, with the Chinese sessions conducted exclusively in Chinese and English sessions conducted exclusively in English. At the end of the second session, participants were asked to fill in a questionnaire about their language background.

2.1.3.1. EEG recording and preprocessing Continuous EEG activity was recorded at 32 scalp sites using ActiveTwo BioSemi active Ag/AgCl electrodes embedded in a custom elastic cap (BioSemi, Amsterdam, The Netherlands). The electro-oculogram (EOG) was recorded with electrodes placed above and below the right eye (vertical), and on the outer canthus of each eye (horizontal). Data were recorded using ActiView software (BioSemi) in the frequency range of 0.1–100 Hz at a sampling rate of 512 Hz. All EEG electrode impedances were maintained below 5 kΩ.
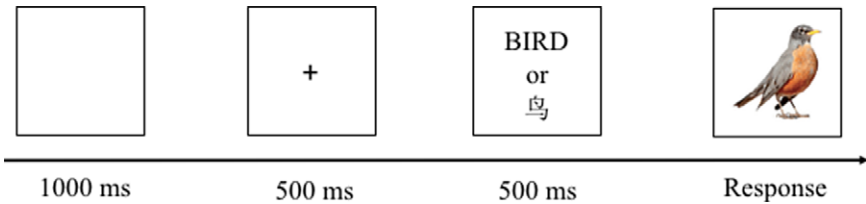
Fig. 2: Experimental procedure in Experiment 1.

Off-line analysis was performed using ERPlab toolbox (Lopez-Calderon & Luck, 2014). All data were re-referenced to the mean electrical activity of the mastoids and bandpass filtered with cut-offs of 0.1 and 30 Hz. The epochs of interest for target images were established to be from −200 to 800 ms post-stimulus onset. Data were baseline corrected to the prestimulus baseline. The data were filtered of eye-movement artifacts that were identified by running an independent component analysis (ICA). Trials contaminated with activity greater than ±75 microvolts (µΩ) were excluded from the analysis (8.9% of the trials for bilinguals in the English session, 9.2% of the trials for bilinguals in the Chinese session, and 10.7% of the trials for English monolinguals).

2.2. RESULTS

Data were analyzed with linear mixed effects (LME) models in R (version 3.4.1; R Development Core Team, 2017) using the lme4 package (version 1.1-18-1; Bates, Mächler, Bolker, & Walker, 2015). The significance of the fixed effects was determined with effect coding and type-II Wald tests using the Anova function provided by the car package (version 2.1-5; Fox & Weisberg, 2011). The latter are reported in the text. Full output from the models appears in Appendix B (see supplementary materials, available at <http://doi.org/10.1017/langcog.2020.8>).

Three sets of analyses were conducted for each dependent variable. To be able to directly compare our findings to those of Liu et al. (2010), the first set included only data from the bilinguals in the Chinese session. Models were fitted with Typicality (Typical vs. Atypical, sum coded) and Word Type (Cue vs. No-Cue, sum coded), as fixed effects, participants and items as random intercepts, and by-participant random slopes for the effects of Typicality and Word Type (without interactions). The second set included data from bilinguals in both sessions and included Test Language as a variable. Specifically, models were fitted with Typicality (Typical vs. Atypical, sum coded), Word Type (Cue vs. No-Cue, sum coded), Test Language (Chinese vs. English, sum coded) as fixed effects, participants and items as random intercepts, and by-participant random slopes for the effects of Typicality, Word Type, and

478

Test Language (without interactions). The third set included only data from the English sessions and included Language Group as a variable. Specifically, models were fitted with Typicality (Typical vs. Atypical, sum coded), Word Type (Cue vs. No-Cue, sum coded), Language Group (Bilingual vs. English Monolingual, sum coded), participants and items as random intercepts, and by-participant random slopes for the effects of Typicality and Word Type (without interaction). The RT data also included Exposure Order (First Exposure vs. Second Exposure, sum coded) as a fixed effect. Of interest were whether there was an overall main effect of Typicality, whether the size of the typicality effect depended on having a category cue in their name (a Typicality × Word Type interaction), and whether this interaction was impacted either by the language of the task for bilinguals (an interaction of Typicality × Word Type × Test Language) or the language group for the English sessions (an interaction of Typicality × Word Type × Language Group).

2.2.1. *Behavioral data*

Incorrect responses (5.4% for the bilingual's English session, 3.6% for the bilingual's Chinese session, and 5.0% for English monolinguals), as well as RTs that were shorter than 200 ms or longer than 1500 ms (2.5% for the bilingual's English session, 1.3% for bilingual's Chinese session, 0.9% for English monolinguals), were excluded from the analyses of the latency data for critical trials. Table 3 shows the mean RTs and error rates for critical trials, overall, and then separately for the first and second exposure to the item.

2.2.1.1. Category decision latency The full output from models of the behavioral data appear in Appendix B (see supplementary materials), Tables 1–5.

2.2.1.1.1. Chinese session. There was a significant main effect of Typicality ($\chi^2(1) = 13.49$, $p < .001$). Typical items were responded to faster than atypical items. The Word Type × Typicality interaction only approached significance ($\chi^2(1) = 2.71$, $p < .09$). However, the 20 ms typicality effect for words with a category cue in their Chinese name was not significant ($\chi^2(1) = 2.47$, $p > .10$), whereas the 49 ms effect for items without cues was highly significant ($\chi^2(1) = 13.45$, $p < .001$). As a point of interest, the correlation between the size of the typicality effect in the Cue condition and length of residence in Canada was $r = .30$, indicating that the typicality effect for these items was smaller for participants who had been in Canada for a shorter period of time.

2.2.1.1.2. Bilinguals: Chinese vs. English sessions. There was a significant main effect of Typicality ($\chi^2(1) = 12.17$, $p < .001$). Typical items were responded to faster than atypical items. The interaction between Word Type

479

T A B L E 3. *Mean response times (in ms) and percentage error rates (between brackets) in Experiment 1*

Participants and Session Language	All trials					
	Cue			No-Cue		
	Typical	Atypical	Typicality effect	Typical	Atypical	Typicality effect
Bilingual						
Chinese session	539 (2.2)	559 (3.8)	20 (1.6)	542 (2.8)	591 (5.5)	49 (2.7)
English session	547 (3.7)	573 (6.2)	26 (2.5)	549 (4.0)	584 (7.8)	35 (3.8)
English monolingual	499 (2.8)	527 (5.1)	28 (2.3)	494 (3.1)	526 (8.9)	32 (5.8)

	First exposure					
	Cue			No-Cue		
	Typical	Atypical	Typicality effect	Typical	Atypical	Typicality effect
Bilingual						
Chinese session	542 (2.6)	559 (4.4)	17 (1.8)	544 (2.9)	599 (4.6)	55 (1.7)
English session	553 (3.6)	574 (6.8)	21 (3.2)	547 (4.4)	600 (7.8)	53 (3.4)
English monolingual	499 (3.7)	541 (5.3)	42 (1.6)	503 (3.2)	538 (10.0)	35 (6.8)

	Second exposure					
	Cue			No-Cue		
	Typical	Atypical	Typicality effect	Typical	Atypical	Typicality effect
Bilingual						
Chinese session	536 (1.9)	560 (3.2)	24 (1.3)	540 (2.7)	582 (6.5)	42 (4.8)
English session	540 (3.7)	571 (5.6)	31 (1.9)	550 (3.6)	568 (7.8)	18 (4.2)
English monolingual	499 (1.9)	514 (4.9)	15 (3.0)	485 (3.1)	514 (7.8)	29 (4.7)

and Typicality was not significant ($\chi^2(1) = 1.16$, $p > .20$). The typicality effect was significant for both items with a cue (23 ms) ($\chi^2(1) = 3.85$, $p = .05$) and items without a cue (42 ms) ($\chi^2(1) = 9.05$, $p < .002$). The Word Type × Typicality × Test Language interaction was not significant ($\chi^2(1) = 1.40$, $p > .20$). While this finding indicates that Chinese–English bilinguals showed the same response pattern regardless of the testing language, we should be cautious about this interpretation because the difference in the typicality effect for the two word types was 29 ms when the task was done in Chinese but only 9 ms in English. There was a significant three-way interaction between Typicality, Word Type, and Exposure Order ($\chi^2(1) = 4.36$, $p = .03$), suggesting that the relationship between Word Type and Typicality was different for the first and second exposures. Therefore, we further analyzed just data from the first exposure to each picture as a YES trial. RTs were fitted with the same variables excluding the fixed effect of Exposure Order.

In the first exposure data there was a significant main effect of Typicality ($\chi^2(1) = 10.61$, $p < .001$), and a trend of an interaction between Typicality and Word Type ($\chi^2(1) = 2.66$, $p = .10$). The typicality effect was not significant for items with cues (19 ms) ($\chi^2(1) = 1.52$, $p > .20$), but was highly significant for items without cues (54 ms) ($\chi^2(1) = 11.10$, $p < .001$). The three-way interaction between Typicality, Word Type, and Test Language was again not significant ($\chi^2(1) = 0.04$, $p > .80$), but here the difference in the typicality effect for the two word types was 38 ms when the task was done in Chinese and 32 ms in English, and therefore we can conclude with more confidence that Chinese–English bilinguals showed the same response pattern regardless of the testing language.

2.2.1.1.3.   English sessions: bilinguals vs. monolinguals.   There was a significant effect of Typicality ($\chi^2(1) = 10.53$, $p = .001$). Typical items were responded to faster than atypical items. The key three-way interaction between Typicality, Word Type, and Language Group was not significant ($\chi^2(1) = 0.13$, $p > .70$). However, there was a significant four-way interaction between Typicality, Word Type, Language Group, and Exposure Order ($\chi^2(1) = 5.40$, $p = .02$), suggesting that the relationship between Word Type, Typicality, and Language Group was different for the first and second exposures. Therefore, we further analyzed data from the first exposure separately.

In the first exposure data, there was a significant main effect of Typicality ($\chi^2(1) = 10.71$, $p = .001$), and no interaction of Word Type × Typicality ($\chi^2(1) = 0.40$, $p > .20$). Importantly, there was a significant three-way interaction between Word Type, Typicality, and Language Group ($\chi^2(1) = 3.72$, $p = .05$). For English monolinguals (who were unaware of the category labels in the Chinese names), the typicality effect was similar for items with cues (42 ms) and items without cues (35 ms), but for Chinese–English bilinguals,

481

the typicality effect was 32 ms smaller for items with cues (21 ms) than for items without cues (53 ms).

2.2.1.2. Error data   None of the effects of interest were significant in the error data.

2.2.2. *ERP data*

The data from 22 electrodes (F3, Fz, F4, FC5, FC1, FC2, FC6, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, PO3, PO4, O1, Oz, O2) were included in the analyses (see Figure 3). For each participant, the data from these electrodes were averaged for each condition. Data were included only for trials with a correct response. The negative going N300 component peaked at about 325 ms and was measured in the 250–350 ms time-window. In addition to the N300, we conducted analyses using a 400–500 ms time-window. We analyzed the
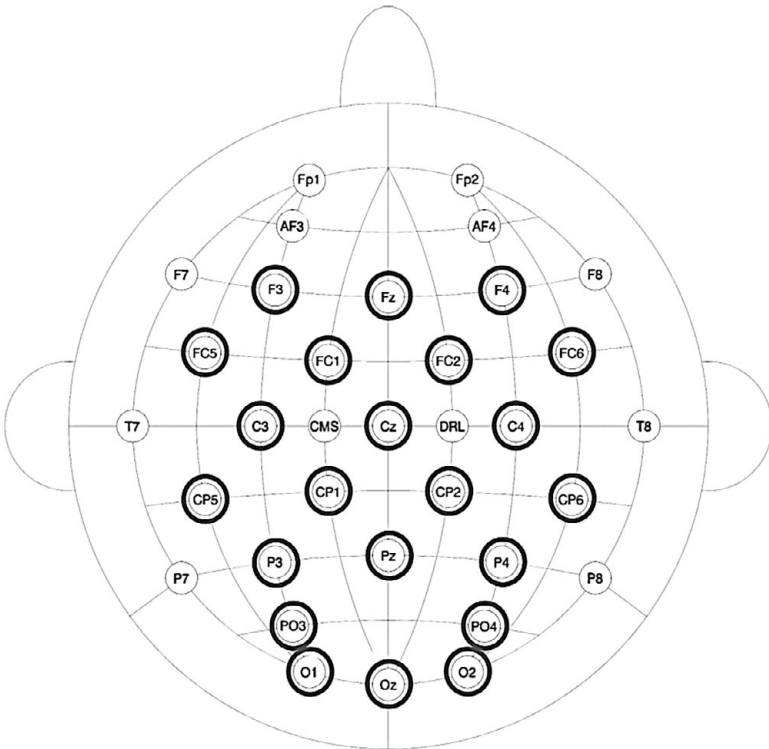
Fig. 3: Electrode montage for Experiment 1 and Experiment 2. Circles indicate electrodes included in the analysis.
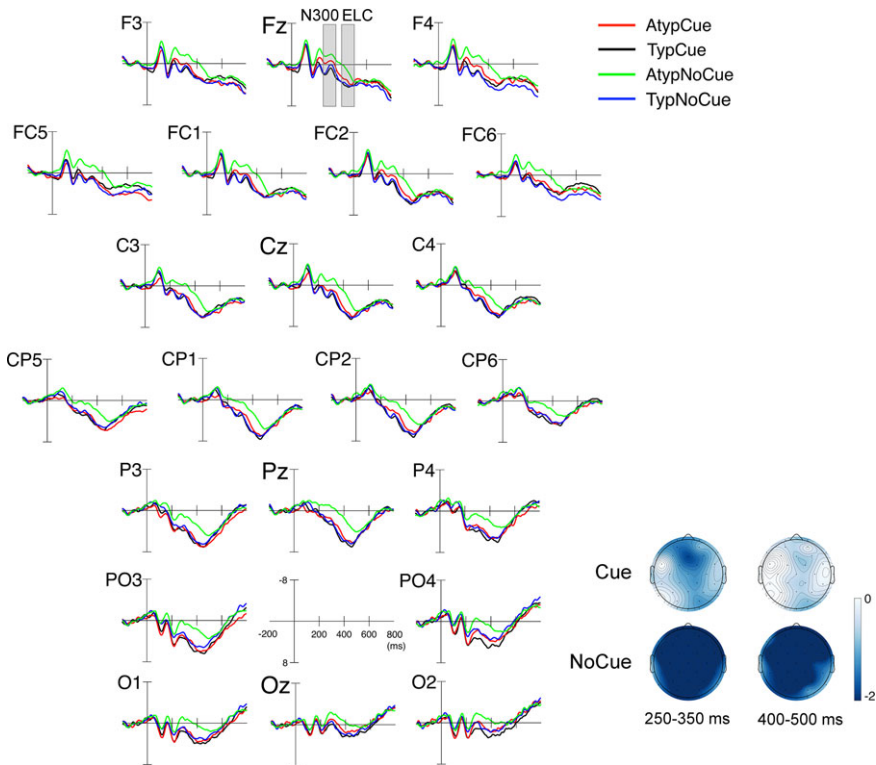
482

Fig. 4: Grand average waveforms in microvolts (μV) and voltage maps of the typicality effect (Atypical - Typical) in N300 and ELC components for bilinguals in the Chinese session in Experiment 1.

waveforms in this time-window separately from the previous time-window, although in our data it may not be a distinct component from the N300 but rather a continuation of that component. This was also the case for Chinese speakers in Liu et al.'s (2010) Experiment 2. We refer to it here as an extended late component (ELC). Figures 4, 5, and 6 show the grand average waveforms in microvolts (μV) evoked in response to the four conditions, and voltage maps showing the typicality effect on N300 and ELC components, for the bilingual Chinese session, the bilingual English session, and English monolinguals, respectively. Analyses were done only on data from both presentations of the pictures. The coding of the ERP component of the experiment did not permit the separation of data from the two presentations.
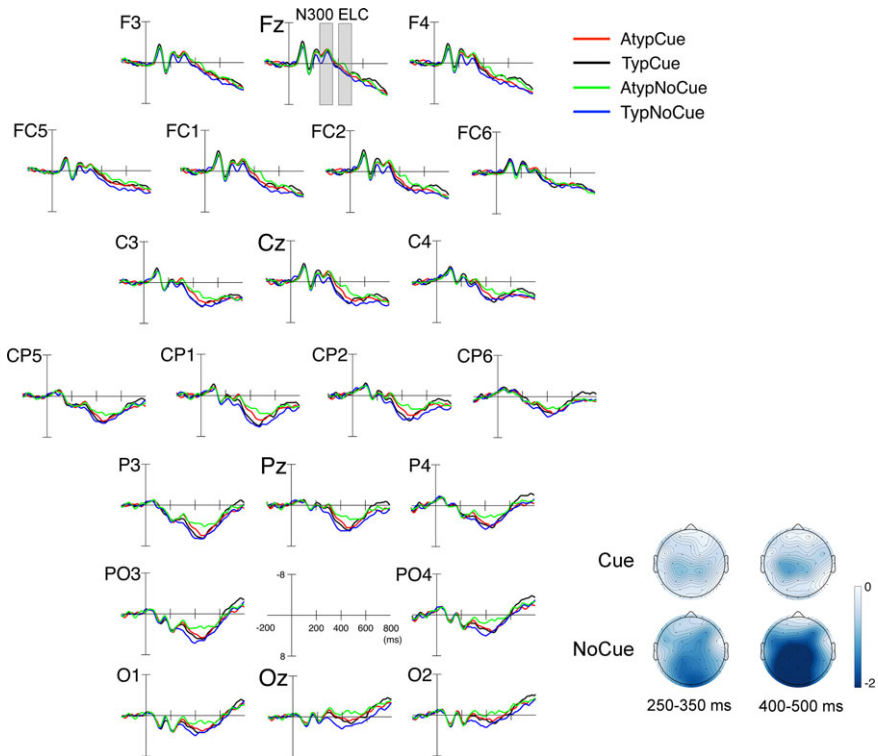
483

Fig. 5: Grand average waveforms in microvolts (μV) and voltage maps of the typicality effect (Atypical - Typical) in N300 and ELC components for bilinguals in the English session in Experiment 1.

2.2.2.1. N300 (250–350 ms) The full output from models of the N300 data appear in Appendix B (see supplementary materials), Tables 6–8.

2.2.2.1.1. Chinese session. There was a significant main effect of Typicality ($\chi^2(1) = 10.06$, $p < .001$). Typical items had a smaller N300 than atypical items. The Typicality × Word Type interaction was significant ($\chi^2(1) = 4.09$, $p = .04$). The typicality effect was smaller for items with a cue than items without a cue. The correlation between the size of the typicality effect in the Cue condition and length of residence in Canada was $r = .11$.

2.2.2.1.2. Bilinguals: Chinese and English sessions. There was a significant main effect of Typicality ($\chi^2(1) = 16.56$, $p = .001$). Atypical items elicited a more negative N300 than typical items. Importantly, there was a significant interaction between Typicality and Word Type ($\chi^2(1) = 5.96$, $p = .01$). The typicality effect was smaller for items with a cue than items without a cue. The
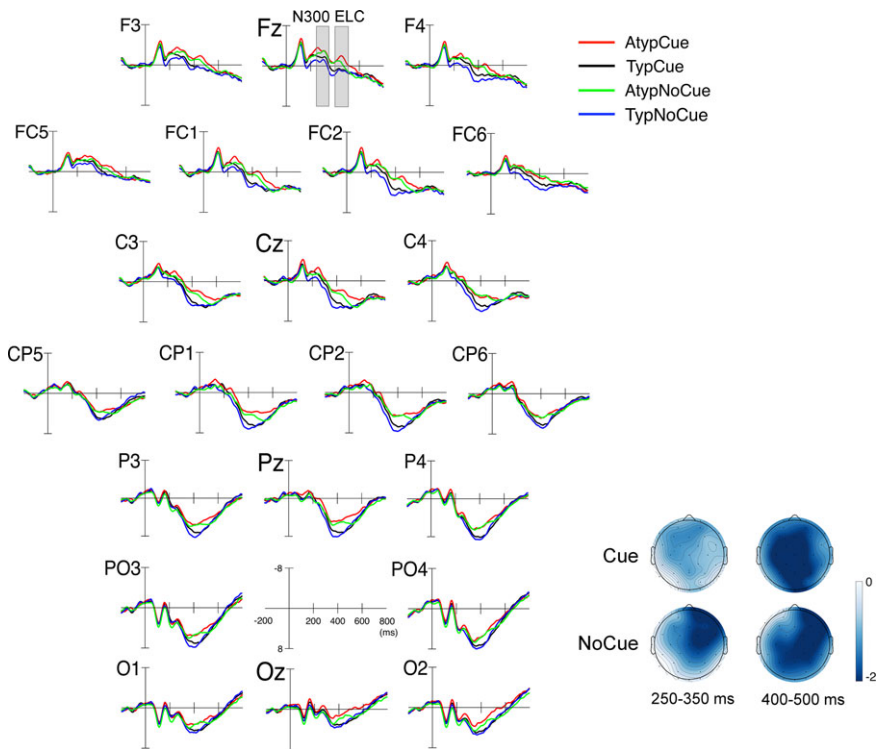
484

Fig. 6: Grand average waveforms in microvolts (μV) and voltage maps of the typicality effect (Atypical - Typical) in N300 and ELC components for the English monolinguals in Experiment 1.

three-way interaction between Typicality, Word Type, and Test Language was not significant ($\chi^2(1) = 0.74$, $p > .30$). Chinese–English bilinguals showed the same response pattern regardless of the language used in testing. Indeed, a separate model for the English session confirmed that the Typicality × Word Type interaction was significant ($\chi^2(1) = 5.31$, $p = .02$), as it was for the Chinese session (reported above).

2.2.2.1.3. English sessions: bilinguals and monolinguals. There was a significant main effect of Typicality ($\chi^2(1) = 23.33$, $p < .001$). Atypical items elicited a more negative N300 than typical items. The Typicality × Word Type interaction neared significance ($\chi^2(1) = 2.77$, $p < .09$). The three-way interaction between Typicality, Word Type, and Language Group was not significant ($\chi^2(1) = 0.47$, $p > .40$). However, separate models for each language group revealed that the Typicality × Word Type interaction was significant for bilinguals ($\chi^2(1) = 5.31$, $p = .02$), as previously noted, but not for monolinguals

485

($\chi^2(1) = 0.34$). At this early time-point, although different patterns are beginning to arise for bilinguals and monolinguals, there appears to have been too much variability across participants and electrodes to produce a significant triple interaction.

2.2.2.2.   ELC (400–500 ms)   The full output from models of the ELC data appear in Appendix B (see supplementary materials), Tables 9–11.

2.2.2.2.1.   Chinese session.   There was a significant main effect of Typicality ($\chi^2(1) = 6.90$, $p < .01$). Typical items showed a smaller ELC than atypical items. The Typicality × Word Type interaction was significant ($\chi^2(1) = 9.26$, $p = .002$). The typicality effect was smaller for items with a cue than items without a cue. The correlation between the size of the typicality effect in the Cue condition and length of residence in Canada was $r = .38$, indicating that the typicality effect for these items was smaller for participants who had been in Canada for a shorter period of time.

2.2.2.2.2.   Bilinguals: Chinese and English sessions.   There was a significant main effect of Typicality ($\chi^2(1) = 23.43$, $p < .001$). Atypical items elicited a more negative ELC than typical items. Importantly, there was a significant interaction between Typicality and Word Type ($\chi^2(1) = 12.50$, $p < .001$). The three-way interaction between Typicality, Word Type, and Test Language was not significant ($\chi^2(1) = 0.35$, $p > .40$). Chinese–English bilinguals showed the same response pattern regardless of the language used in testing. Indeed, a separate model for the English session confirmed that the Typicality × Word Type interaction was significant ($\chi^2(1) = 5.59$, $p = .01$), as it was for the Chinese session (reported above).

2.2.2.2.3.   English sessions: bilinguals and monolinguals.   There was a significant main effect of Typicality ($\chi^2(1) = 48.11$, $p < .001$). Importantly, the three-way interaction between Typicality, Word Type, and Language Group was nearly significant ($\chi^2(1) = 3.08$, $p = .07$). A separate analysis on the English monolinguals revealed that the Typicality × Word Type interaction was nowhere near significant ($\chi^2(1) = 0.001$), unlike for the bilinguals (reported above). The typicality effects were smaller for items with cues than items without cues in bilinguals but were similar for the two types of items in English monolinguals.

2.3. DISCUSSION

Experiment 1 explored the effect of Chinese word structure on bilinguals' categorization processes with pictorial stimuli. The first question we addressed was whether our data support Liu et al.'s (2010) claim that typicality effects in

486

Chinese speakers are attenuated for pictures whose names contain a morpho-logical category cue (i.e., the cue facilitates categorization). Here we improved on their design by including pictures with names that do not have a morpho-logical category cue for comparison purposes, and we also increased the number of stimuli used. In the Chinese session, Chinese–English bilinguals indeed showed a smaller typicality effect in a semantic categorization task for pictures with a morphological category cue in their name than for pictures without that cue in their name. The interaction between typicality and word type was significant in the N300 and the ELC, and although not significant in the decision latency data, the typicality effect in those data were only signif-icant for items without a category cue in their name. Furthermore, for pictures with a category cue in their name, there was a correlation of the size of the typicality effect with years of residence in Canada in the RT and ELC data, such that the typicality effect was smaller for those who had been in Canada for less time, and who had, presumably, less exposure to English.

Having established that our stimuli show an effect of Chinese word structure on picture categorization when the experimental session was conducted in Chinese, we then addressed our question of interest, which was whether Chinese–English bilinguals show the effects of Chinese category cues even when they were tested in English. First we investigated whether the effect of category cues differed for bilinguals when they did the task in English com-pared to Chinese. They did not. The triple interactions between typicality, word type, and test language were not significant in the RT, N300, or ELC data. In the ERP data, significant interactions of typicality and word type were observed both when bilinguals completed the task in Chinese and when they completed it in English. In the RT data, this pattern was more evident when just the first presentation of each stimulus was considered. We then examined whether the effect of Chinese category cues on the English version of the task differed for bilinguals and for English monolinguals, the latter of whom are unaware of the category cues. We obtained some evidence that it did. The triple interaction between typicality, word type, and language group was significant in the RT (first exposure) data and was very close to significant in the ELC data. In both the N300 and the ELC data, the interaction of typicality and word type was significant for bilinguals but nowhere close to significant for English monolinguals.

In previous studies, the N300 component has been found to be related to how integral the meaning of a non-verbal stimulus (e.g., picture, video) is to the whole context, which highly resembles the categorization process (Sitnikova, Kuperberg, & Holcomb, 2003; West & Holcomb, 2002). The categorization process can be described as making judgments on how integral the meaning of a category member is to the category as a whole. In addition, the ELC has also been found to be involved in the typicality effect with non-verbal

stimuli (Liu et al., 2010; West & Holcomb, 2002). Researchers have suggested that the ELC might indicate different levels of decision-making and evaluative processes (Heinze, Muente, & Kutas, 1998; Stuss, Picton, & Cerri, 1988) or violations of rules or goal-related requirements (Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Sitnikova et al., 2003). Therefore, findings in the current experiment suggest that bilinguals find it easier to integrate the semantic information of an object with a category cue in its name into the category to which it belongs.

To summarize, both behavioral and ERP results in Experiment 1 demonstrated that Chinese word structure influences picture categorization in Chinese–English bilinguals. A category cue embedded in an object's Chinese name facilitates categorization of the object. Furthermore, the facilitation from objects' Chinese names in bilinguals occurred even when English was the language used for testing, suggesting that labels in the inactive language have an influence on a bilingual's categorization processes.

There are, however, several limitations to Experiment 1. Because of the relatively small number of items (some items we had hoped to use were excluded after Pilot Study 1 due to low name agreement for pictures), each stimulus was presented twice in the categorization task to get a clear ERP signal after averaging. The behavioral results showed that this repeated presentation of stimuli influenced participants' responses. The faciliatory effect of a category cue embedded in objects' Chinese names only appeared in the first exposure data. This could be due to a familiarity effect: as participants became more familiar with the stimuli, they made faster responses, especially for atypical items. However, in the ERP data analyses, data were collapsed across both the first and the second exposure because the coding method we used for the ERP data did not allow us to separate data from the first and second exposures. This might have had an influence on the ERP results, and could be the reason that, in the analyses of English sessions, no significant triple interaction in the N300 component was observed. In Experiment 2, word stimuli instead of pictures were used as targets. Because name agreement was not a problem, more items could be included, and there was no need to repeat them in the experimental task.

The filler pairs in Experiment 1 were created by re-pairing the critical category label–image pairs. This was done to prevent participants from developing a link between an image and a response. For example, participants might link a picture of a robin with a *yes* response in the first presentation, and then might quickly make a *yes* response when they saw a robin picture for the second time without categorization. This re-pairing method resulted in each target picture being presented four times in the categorization task, which could have influenced participants' responses on critical trials and weakened the results. In Experiment 2, because there was no need to repeat stimuli, new items that

488

were different from critical stimuli were used as fillers, so that each critical target was presented only once in the categorization task.

Another limitation in Experiment 1 is that, although bilinguals did the English session in a pure English environment, half of the bilinguals did the Chinese session first. This could have given them some clues that bilingualism and Chinese were of interest in the study and possibly had some influence on their results in the English session. When the bilingual participants were tested in English, the knowledge that Chinese was relevant to the study could have encouraged them to keep their Chinese active and made it more likely that we would observe the effect of a category level cue in objects' Chinese names. Therefore, in Experiment 2, bilingual participants were put into an English monolingual mode to the fullest possible extent. The use of word targets instead of pictures made it possible to make it clearer to participants that only their knowledge of English was required.

3. Experiment 2

In Experiment 2, participants were tested only in English, they were greeted in English, and all conversation and consent forms were in English. In addition, bilingual participants were recruited via a filter system in the University research participation pool, and advertisements posted on social media groups for Chinese students. Therefore, the study was directed only to native Chinese speakers without the requirements of bilingualism being listed in study information. A monolingual English comparison group was also tested. Again, this group was unaware of the category labels in Chinese.

The same experimental paradigm was used as in Experiment 1, but English word labels of target items were used instead of images. Previous studies investigating the typicality effect with ERPs have found that typicality effects in linguistic stimuli are marked by the N400 component, such that atypical items of a category elicit a larger N400 than typical items (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980). The N300 component was not of interest in Experiment 2 because studies have shown that it is specifically elicited for pictorial stimuli (Hauk et al., 2007; Kiefer, 2001).

3.1. METHOD

3.1.1. *Participants*

Thirty-nine Chinese–English bilinguals (mean age 22, range 18–46, 23 female) and 29 English monolinguals (mean age 18, range 18–21, 10 female) were recruited. Participants received course credit or money for their participation. None of the participants had participated in Experiment 1. Data from eleven

489

bilinguals (ten with low accuracy on the categorization task (< 63%), one with poor ERP recording) and one English monolingual (poor ERP recording) were excluded, leaving 28 Chinese–English bilinguals and 28 English monolinguals in the final sample. The first language of bilinguals in the final sample was Mandarin. All bilinguals were born in China or Taiwan, had lived in China or Taiwan for a mean of 15.8 years (range 2–25), and had lived in Canada for a mean of 7.4 years (range 2–21).

3.1.2. *Materials*

Critical stimuli for Experiment 2 were the stimuli selected from Pilot Study 1 for norming in Pilot Study 2, which consisted of 13 categories and 108 items (in contrast to the subset of 11 categories and 84 items used in Experiment 1). Half of the objects were typical, half were atypical (see Table 1). Half of the objects had a category label in their Chinese names, half did not. All of the critical stimuli required *yes* decisions. Another 108 category label–object name pairs were created as filler stimuli to require *no* decisions. The same set of category labels were used in filler pairs as in critical pairs. The breakdown of the target words used in filler pairs was as follows: 56 items from the 13 categories, and 52 items from other categories (this was done because not enough filler stimuli could be found within the 13 categories). Half of the filler items were typical, half were atypical. Half of the filler items had a category label in their Chinese names, half did not.

3.1.3. *Procedure*

A category label–object name matching task was used (see Figure 7). Participants first saw a 500 ms fixation cross, followed by a category label (e.g., *BIRD*) for 500 ms, then followed by a word (e.g., *robin*). Participants were instructed to judge whether the concept represented by the second word is an example of the category represented by the first word. All category label–word pairs were presented once to each participant in a random order. There were 216 trials, including 108 critical trials that required a *yes* response (27 trials
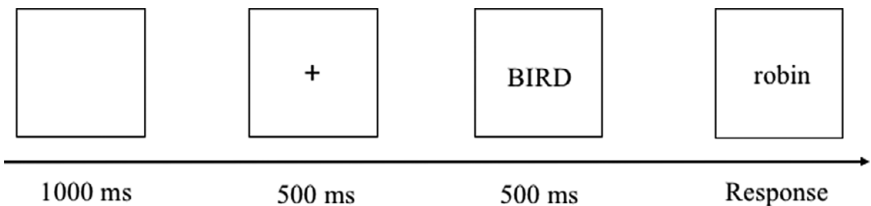

Fig. 7: Experimental procedure in Experiment 2.

490

per condition), 108 filler trials that required a *no* response, and 12 practice trials. All conversation and experimental materials were in English. At the end of the experiment, participants were asked to fill in a questionnaire about their language background.

3.1.3.1. EEG recording and preprocessing Recording, digitization of the EEG activity, and off-line analysis were done as in Experiment 1. The epochs of interest for target words were established to be from −200 to 1000 ms post-stimulus onset. Trials contaminated with activity greater than ±75 microvolts (μΩ) were excluded from the analysis (10.5% for Chinese–English bilinguals; 9.8% for English monolinguals).

3.2. RESULTS

3.2.1. *Behavioural analyses*

Incorrect responses (21.0% for bilinguals, 9.2% for English monolinguals), as well as RTs that were shorter than 200 ms or longer than 2500 ms for bilinguals (3.8%), and RTs that were shorter than 200 ms or longer than 1500 ms for English monolinguals (1.9%), were excluded from the analyses of the latency data. The mean response latencies and error rates are presented in Table 4. The higher error rate for bilinguals compared to English monolinguals is likely due to the fact that there were some very low-frequency targets (e.g., tuxedo, quartz) that the bilinguals may not have known. Correct responses most likely reflect known words, and therefore only RT data for correct responses, and not error data, were analyzed.

RTs from bilinguals and English monolinguals were analyzed with LME models. RTs were fitted with Typicality (Typical vs. Atypical, sum coded), Word Type (Cue vs. No-Cue, sum coded), Language Group (Bilingual vs. English Monolingual, sum coded), and Word Frequency (CELEX_W, without interactions with other fixed factors) as fixed effects, participants and

TABLE 4. *Mean response times (in ms) and percentage error rates (between brackets) in Experiment 2*

	Cue			No-Cue		
	Typical	Atypical	Typicality effect	Typical	Atypical	Typicality effect
Bilingual	932 (11.4)	1025 (31.5)	93 (20.1)	915 (14.3)	1062 (26.6)	147 (12.3)
English monolingual	595 (4.1)	668 (13.5)	73 (9.4)	596 (5.7)	663 (13.4)	67 (7.7)

items as random intercepts, and by-participant random slopes for the effects of Typicality and Word Type (without interaction). The full output from model of the RT data appears in Appendix B (see supplementary materials), Table 12.

There was a significant main effect of Typicality ($\chi^2(1) = 24.89$, $p < .001$). Typical items were responded to 83 ms faster than atypical items. Importantly, there was a significant three-way interaction between Typicality, Word Type, and Language Group ($\chi^2(1) = 4.16$, $p < .05$). The typicality effects were smaller for items with cues (93 ms) than items without cues in bilinguals (147 ms), but were similar for items with cues (73 ms) and without cues in English monolinguals (67 ms).

3.2.2. *ERP analyses*

The data from the same set of electrodes as in Experiment 1 were included in the analysis. Data were included only for trials with a correct response. The N400 component peaked at around 400 ms and was measured in the 375–500 ms time-window. Figures 8 and 9 show the grand average waveforms in microvolts (μV) evoked in response to the four conditions, and voltage maps showing the typicality effect on N400 components, for Chinese–English bilinguals and English monolinguals, respectively.

Mean amplitudes in the N400 were analyzed with LME models. The model was fitted with Typicality (Typical vs. Atypical, sum coded), Word Type (Cue vs. No-Cue, sum coded), and Language Group (Bilingual vs. English Monolingual, sum coded) as fixed effects, participants as random intercept, and by-participant random slopes for the effects of Typicality and Word Type (without interaction). The full output from the model of the N400 data appears in Appendix B (see supplementary materials), Table 13.

There was a significant main effect of Typicality ($\chi^2(1) = 11.37$, $p < .001$). Atypical items elicited a more negative N400 than typical items. Importantly, there was a significant three-way interaction between Word Type, Typicality, and Language Group ($\chi^2(1) = 4.50$, $p < .05$). The typicality effect was smaller for items with cues than without cues in bilinguals ($\chi^2(1) = 7.30$, $p = .006$), but were similar for items with cues and without cues in English monolinguals ($\chi^2(1) = 0.12$).

3.3. DISCUSSION

In Experiment 2, we further explored the effects of Chinese word structure on bilinguals' categorization processes with English word stimuli. The use of word targets instead of pictures allowed more items to be included in the stimulus list, so there was no need to repeat critical targets. Different words were used to create filler trials with *no* response, so that each critical target word
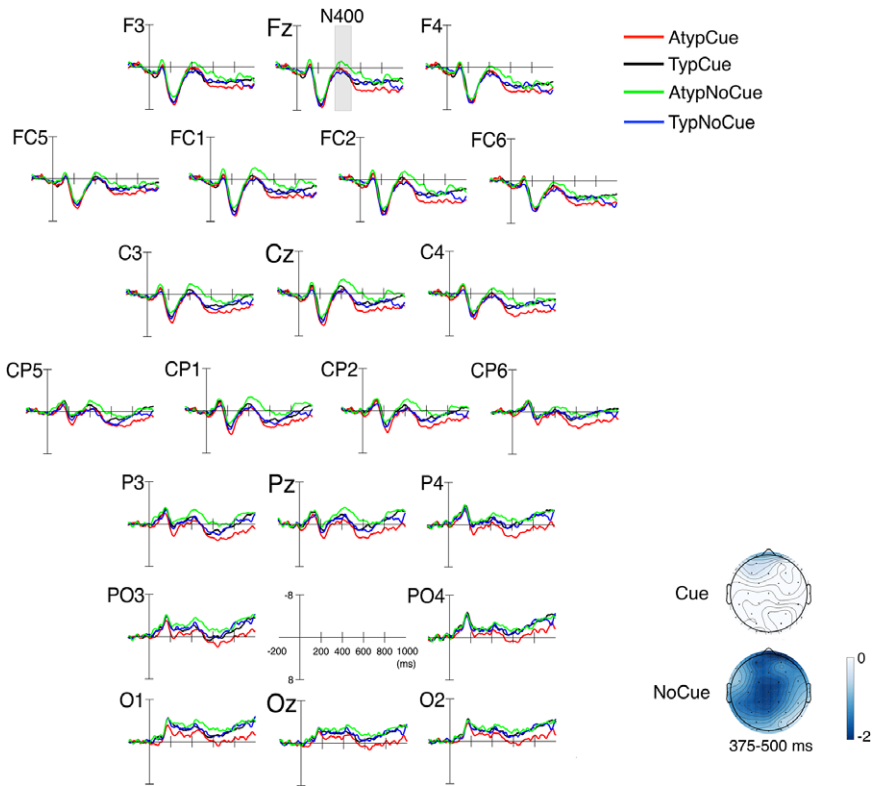
492

Fig. 8: Grand average waveforms in microvolts (μV) and voltage maps of the typicality effect (Atypical - Typical) in the N400 component for Chinese-English bilinguals in Experiment 2.

was presented only once. In addition, the use of word targets further reinforced the English nature of the experiment. Bilingual participants were put into an English monolingual mode to the fullest extent possible. RTs and ERP responses were measured as Chinese–English bilinguals and English monolinguals categorized English word labels of typical and atypical objects. In the RT and N400 data, typicality effects were smaller for items with cues than items without cues in bilinguals, whereas English monolinguals showed no such difference (because they were unaware of the category labels in the Chinese names).

The N400 component has been broadly used as an index of the semantic congruency of a word to the whole context (see Kutas & Federmeier, 2011, for a review). A more negative N400 is thought to be associated with more semantic violation. Categorizing an atypical item produces more sense of semantic violation than categorizing a typical item, because atypical items usually
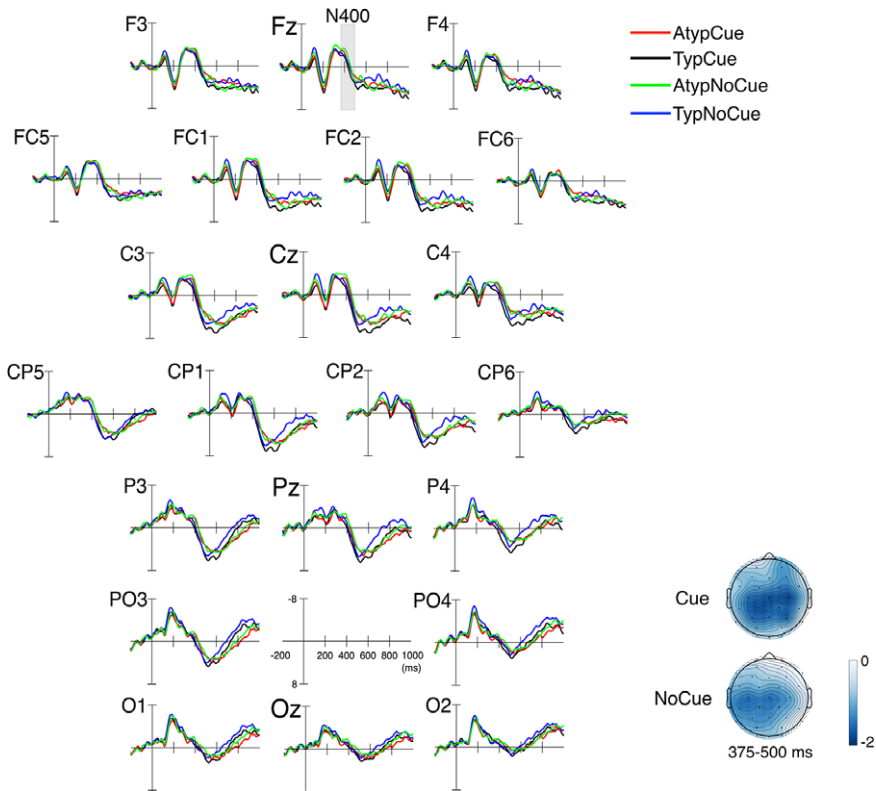
493

Fig. 9: Grand average waveforms in microvolts (μV) and voltage maps of the typicality effect (Atypical - Typical) in the N400 component for English monolinguals in Experiment 2.

contain more semantic features that are not commonly seen in the category members. Thus, the N400 component has also been used as an ERP marker for the typicality effects in linguistic stimuli (Kutas & Federmeier, 2000; Kutas & Hillyard, 1980).

In summary, these results provide further evidence that the category information imbedded in objects' Chinese names facilitates categorization, and demonstrates that this facilitation occurs even when bilinguals categorize English words.

4. General discussion

The current study examined the effects of word structure on bilinguals' categorization processes. More specifically, since bilinguals know two different languages, we investigated the question of whether word structure in one

language could have an impact when bilinguals perform a categorization task in the other language. Chinese–English bilinguals and English monolinguals completed two categorization tasks, one with picture targets and one with word targets. Results from Experiments 1 and 2 showed that the typicality effect was smaller for items with cues in their Chinese names than items without cues in bilinguals, while English monolinguals showed no such difference (because they were unaware of the category labels in the Chinese names), suggesting that category information in an object's Chinese name facilitated categorization of the object in Chinese–English bilinguals. Most interestingly, the facilitation from objects' Chinese names in bilinguals existed even when they were tested in a pure English-speaking environment where no clue showed that Chinese was involved.

4.1. THE LABEL-FEEDBACK ACCOUNT

The label-feedback hypothesis (Lupyan, 2012) provides a mechanism to account for our findings. Previous research investigating this hypothesis considered the influence of having a verbal label for an object on object categorization. However, the hypothesis can also provide an account of the influence of different characteristics of labels, like the structure of a label. Furthermore, previous research was conducted with monolinguals. Bilinguals have two labels for each object, which potentially makes the influence of labels on perception in bilinguals more complex than in monolinguals. As was discussed previously for Liu et al.'s (2010) study, two explanations of our findings can be derived from the label-feedback hypothesis.

In the first explanation, the category label and the category cue in an object's Chinese name temporarily warp the semantic space by activating the most typical features of the category to a higher degree than the less typical features. In both of our tasks, participants first saw a category label and then saw a target that was either a picture or an English word. The processing steps in these tasks would proceed as follows according to the first view. When Chinese–English bilinguals saw the category label, it would have activated typical features of the category. Then, when they saw a target, both the English and Chinese names of the target would have been activated, and these target labels then would have sent feedback to the conceptual level. If the category cue in a target object's Chinese name highly activates the most typical features of the category, then for atypical exemplars with a category cue, there would be more overlap between the features activated from the category label and the target than when there was no category cue in the Chinese name, thus facilitating categorization. Based on this explanation, the current findings also provide supporting evidence for the language non-selective activation view in bilinguals (Dijkstra & Van Heuven, 2002), which claims that bilinguals activate

495

information from both of their languages simultaneously even when they are using only one of their languages. However, this explanation assumes that in Experiment 1, the Chinese label for the picture stimuli was available by the time of the N300 response in the English task. As aforementioned, various studies have suggested that pictures can be categorized faster than they are named (e.g., Irwin & Lupker, 1983; Potter & Faulconer, 1975). Other studies have found that automatic translation from L2 to L1 in bilinguals takes place at a late, post-lexical processing stage (around 400 ms), after word meaning retrieval (Thierry & Wu, 2007). Therefore, in Experiment 1, it is possible that participants categorized a target picture before its label was highly activated.

The second explanation for the current findings is that bilingual participants' organization of category representations is permanently changed under the long-term effects of everyday usage of objects' Chinese names. More specifically, objects that have a category cue in their Chinese names are more strongly associated with the most typical features of the category through the feedback from frequent usage of Chinese labels. As a consequence, they are stored in the centre of the category, even an atypical exemplar, making them easier to categorize. In contrast, atypical objects that do not have a category cue in their Chinese names are stored in the periphery of the category space, thus making them difficult to categorize. This hypothesis that category representations are permanently changed under the long-term effects of labels is not assumed by Lupyan (2012), but it is consistent with findings from object naming studies in bilinguals (e.g., Ameel et al., 2005). It was beyond the scope of the present study to distinguish between the short-term perceptual warping and the long-term category restructuring explanations. However, in future research it may be possible to do so using transcranial direct current stimulation (e.g., Lupyan, Mirman, Hamilton, & Thompson-Schill, 2012; Perry & Lupyan, 2014).

4.2. OTHER INTERPRETATIONS OF THE CURRENT FINDINGS

Participants did categorization tasks in Experiment 1 and 2, which we assumed involved extensive semantic processing. While it is most likely that participants made decisions based on the semantic congruency between the category labels and the target objects, it could be argued that participants did the categorization task based only on phonological or lexical overlap between the category labels and the objects' Chinese names. For example, in Experiment 1, when bilingual participants saw the picture of an ostrich, the Chinese name 鸵鸟 was activated, which had the category label 鸟 embedded in it. Bilingual participants could have made the decision that an *ostrich* is a *bird* solely based on the overlap of the character 鸟 in the category label and the object's name. Similarly, in Experiment 2, bilinguals could have automatically translated the

496

category labels and the object words into Chinese, then performed the categorization task based on phonological or lexical overlap between the category's and the object's Chinese labels. Alternatively, the faciliatory effects observed here could have been due to priming from the category label to items with the corresponding label in their names, thus causing them to be responded to faster than items without the label in their names. Although this argument is less likely when bilinguals were tested in English, especially in Experiment 2 where both category labels and targets were presented in English, bilingual participants could have automatically translated the category labels into Chinese, then activated items with the corresponding category label in their names. If indeed this is the case that the faciliatory effects observed in these experiments were lexical, the current findings would still be interesting in that they provide evidence for the language non-selective activation view of bilinguals.

However, Liu et al. (2010) provided evidence against a lexical interpretation of their Chinese findings by conducting an experiment with English monolinguals. Critical pictures had a name that either did (e.g., BALL-*basketball*) or did not (e.g., VEHICLE-*car*) contain a category cue. Both of the lexical accounts would predict that English monolinguals would show reduced typicality effects in the former compared to the latter, whereas both produced significant typicality effects in behavioural and ERP data. Furthermore, in our data the faciliatory effects from an object's Chinese name were observed in ERP components related to semantic processing. The N300 component has been found to be related to how integral the meaning of a non-verbal stimulus (e.g., picture, video) is to the whole context, which highly resembles the categorization process (Sitnikova et al., 2003; West & Holcomb, 2002), the N400 component has been broadly used as an index of the semantic congruency of a word to the whole context (Kutas & Federmeier, 2011), and the ELC appears to be associated with different levels of decision-making and evaluative processes (Heinze et al., 1998; Stuss et al., 1988) or violations of rules or goal-related requirements (Sitnikova et al., 2008; Sitnikova et al., 2003). Therefore, it is unlikely that the effects observed in the current study are only due to the overlap at the phonological or lexical level.

4.3. CONCLUSION

The current findings add to our understanding of language–cognition interaction in bilinguals, providing evidence that word labels from both languages can have an important effect on object categorization. An interesting characteristic of Chinese labels was used, and the results demonstrated that object categorization is influenced by how a verbal label is constructed. Furthermore, verbal labels in Chinese influenced categorization processes even when the task was conducted entirely in English, providing evidence that bilinguals'

497

cognitive processes are constantly under the influence of two sets of labels, even when only one language is being used. The label-feedback hypothesis provides a useful framework in which to understand the mechanisms of language–cognition interactions in bilinguals.

Supplementary materials

For supplementary materials for this paper, please visit <http://doi.org/10.1017/langcog.2020.8>.

REFERENCES

Ameel, E., Storms, G., Malt, B. C. & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language* **53**(1), 60–80.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48.

Boutonnet, B. & Lupyan, G. (2015). Words jump-start vision: a label advantage in object recognition. *Journal of Neuroscience* **35**(25), 9329–9335.

Davidoff, J. & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: a case study. *Language and Cognitive Processes* **19**(1), 137–174.

Dijkstra, T. & Van Heuven, W. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism: Language and Cognition* **5**(3), 175–197.

Edmiston, P. & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition* **143**, 93–100.

Federmeier, K. D. & Kutas, M. (2001). Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **27**(1), 202–224.

Fox, J. & Weisberg, S. (2011). *An {R} companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Fulkerson, A. L. & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: evidence from 6- and 12-month-olds. *Cognition* **105**(1), 218–228.

Ganis, G., Kutas, M. & Sereno, M. I. (1996). The search for 'common sense': an electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience* **8**(2), 89–106.

Gervits, F., Johanson, M. & Papafragou, A. (2016). *Intentionality and the role of labels in categorization*. Retrieved from <https://mindmodeling.org/cogsci2016/papers/0206/paper0206.pdf>.

Hamm, J. P., Johnson, B. W. & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology* **113**(8), 1339–1350.

Hauk, O., Patterson, K., Woollams, A., Cooper-Pye, E., Pulvermüller, F. & Rogers, T. T. (2007). How the camel lost its hump: the impact of object typicality on event-related potential signals in object decision. *Journal of Cognitive Neuroscience* **19**(8), 1338–1353.

Heinze, H. J., Muente, T. F. & Kutas, M. (1998). Context effects in a category verification task as assessed by event-related brain potential (ERP) measures. *Biological Psychology* **47**(2), 121–135.

Irwin, D. J. & Lupker, S. J. (1983). Semantic priming at pictures and words: a levels of processing approach. *Journal of Verbal Learning and Verbal Behavior* **22**(1), 45–60.

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory & Cognition* **29**(1), 100–116.

Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* **4**(12), 463–470.

Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology* **62**, 621–647.

Kutas, M. & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology* **11**(2), 99–116.

Liu, C., Tardif, T., Mai, X., Gehring, W. J., Simms, N. & Luo, Y. J. (2010). What's in a name? Brain activity reveals categorization processes differ across languages. *Human Brain Mapping* **31**(11), 1786–1801.

Lopez-Calderon, J. & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience* **8**, e00213.

Lupyan, G. (2009). Extracommunicative functions of language: verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review* **16**(4), 711–718.

Lupyan, G. (2012). What do words do? Toward a theory of language-augmented thought. In B. H. Ross (ed.), *The psychology of learning and motivation* (Vol. **57**, pp. 255–297). Waltham, MA: Academic Press.

Lupyan, G. & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition* **7**(2), 167–193.

Lupyan, G. & Mirman, D. (2013). Linking language and categorization: evidence from aphasia. *Cortex* **49**(5), 1187–1194.

Lupyan, G., Mirman, D., Hamilton, R. & Thompson-Schill S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition* **124**(1), 36–49.

Lupyan, G., Rakison, D. H. & McClelland, J. L. (2007). Language is not just for talking: redundant labels facilitate learning of novel categories. *Psychological Science* **18**(12), 1077–1083.

Lupyan, G. & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General* **141**(1), 170–186.

McPherson, W. B. & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* **36**(1), 53–65.

Pavlenko, A. (2014). *The bilingual mind: and what it tells us about language and thought*. Cambridge/New York: Cambridge University Press.

Perry, L. K. & Lupyan, G. (2014). The role of language in multi-dimensional categorization: evidence from transcranial direct current stimulation and exposure. *Brain & Language* **135**, 66–72.

Potter, M. C. & Faulconer, B. A. (1975). Time to understand pictures and words. *Nature* **253**(5491), 437.

R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online <https://www.R-project.org/>.

Robinson, C. W., Best, C. A., Deng, W. S. & Sloutsky, V. (2012). The role of words in cognitive tasks: What, when, and how? *Frontiers in Psychology*, **3**, e00095.

Rosch, E. (1973). Natural categories. *Cognitive Psychology* **4**(3), 328–350.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* **104**(3), 192–233.

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A. & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience* **20**(11), 2037–2057.

Sitnikova, T., Kuperberg, G. & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: an electrophysiological investigation. *Psychophysiology* **40**(1), 160–164.

Stuss, D. T., Picton, T. W. & Cerri, A. M. (1988). Electrophysiological manifestations of typicality judgment. *Brain and Language* **33**(2), 260–272.

Thierry, G. & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences* **104**(30), 12530–12535.

West, W. C. & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research* **13**(3), 363–375.

Whorf, B. L. (1956). *Language, thought, and reality: selected writings of Benjamin Lee Whorf*. Edited by J. Carroll, S. Levinson & P. Lee. Cambridge, MA: MIT Press.

Appendix A

Example stimuli in Experiment 1.

Category		Item	Typicality	Word Type
English	Chinese			
BIRD	鸟		Typical	Cue
BIRD	鸟		Atypical	Cue
BIRD	鸟		Typical	No-Cue
BIRD	鸟		Atypical	No-Cue

500