Pictorial Referents Facilitate Recognition and Retrieval Speeds of Associations Between Novel Words in a Second Language (L2) and Referents

Xinyi Liu[1], Hiroki Horinouchi[2], Yutao Yang[1], Yan Yan[1], Misa Ando[2], Ukwueze Jonah Obinna[1,3], Shushi Namba[4] and Toshimune Kambara[1]*

[1]Department of Psychology, Graduate School of Education, Hiroshima University, Hiroshima, Japan, [2]Program in Psychology, School of Education, Hiroshima University, Hiroshima, Japan, [3]Citadel International Nursery and Primary School, Enugu, Nigeria, [4]Psychological Process Team, BZP, Robotics Project, RIKEN, Kyoto, Japan

This study reveals associative processes for novel words in a second language (L2) and their referents. Thirty Japanese participants learned associative pairs for novel words in Chinese and pictorial referents (CP), as well as novel words in Chinese and words in Japanese (CJ), against a condition in which they learned only novel words in Chinese (C). After the learning phase, participants conducted two learning condition retrieval tasks for word recognition and three recognition tasks for the source-monitoring of the referents. The correct answers for each recognition task were provided to participants after each trial. Although the correct answers in all conditions increased in both the recognition and learning condition retrieval tasks, there was no significant difference among conditions. In contrast, the response times of the correct trials in all recognition tasks and the first learning condition retrieval tasks were faster for the CP condition than the CJ condition. Additionally, in the second learning condition retrieval task, missed items in associative conditions (CP and CJ conditions) were judged to be learned items more often than unlearned items, whereas missed items in the non-associative condition (C condition) were judged to be unlearned items more than learned items. These findings suggest that pictures contribute to the recognition and retrieval speeds of associations between novel words in L2 and referents, and that associative learning of L2 words and referents could enhance more familiarity effects than the learning of L2 words only.

Keywords: associative learning, word learning, second language, dual coding theory, novel words

INTRODUCTION

Words consist of associations between word forms and referents, including sensorimotor, emotional, and abstract features, gained through the learners' experiences (Paivio, 1986; Pulvermüller, 2003; Kambara et al., 2020). In the dual coding theory of associations between verbal (word forms) and nonverbal features (referents), the Canadian psycholinguist Allan Paivio proposed that the associations between word forms and referents are learned in three differential levels of processing, including representational, referential, and associative processing (Paivio, 1971; Paivio, 1986; Paivio, 2007). First, the representational processing means that the verbal stimuli activate verbal representations, whereas nonverbal stimuli activate nonverbal representations. Second, the referential processing means that

verbal stimuli activate nonverbal representations, whereas nonverbal stimuli activate verbal representations. Third, the associative processing means that verbal representations activate other related verbal representations in the verbal system, whereas nonverbal representations activate other related nonverbal representations in the nonverbal system. According to these three levels of processing, the associative learning of the verbal (word forms) and nonverbal stimuli (referents) would be essential to learn real words. Previous studies have reported that word forms can be associated with referents or with words already associated with referents through associative learning, in which participants learned associative pairs of a word form and referent(s), including pictures or sounds (Paivio and Csapo, 1973; Cornelissen et al., 2004; Breitenstein et al., 2005; Hultén et al., 2009; Tsukiura et al., 2010; Tsukiura et al., 2011; Kambara et al., 2013; Takashima et al., 2014; Ferreira et al., 2015; Grönholm et al., 2015; Hawkins et al., 2015; Hawkins and Rastle, 2016; Takashima et al., 2017; Havas et al., 2018; Li et al., 2020; Horinouchi et al., under review; Yan et al., under review; Yang et al., under review) or lexical conditioning (researchers also call the conditioning classical, semantic, or evaluative), in which participants generalize the referents (evaluative responses; e.g., positive and negative meanings) of real words to referents of pseudowords, real words, or symbols (Razran, 1939; Staats and Staats, 1957; Staats and Staats, 1958; Staats et al., 1959a; Staats et al., 1959b; Staats et al., 1961; Paivio, 1964; Cicero and Tryon, 1989; Tryon and Cicero, 1989; Till and Priluck, 2001; Hughes et al., 2018). In the associative learning of word forms in a first language (L1) or a second language (L2) and referents, differences between modalities of referents could affect task performance in the test phase (Lee et al., 2003; Jeong et al., 2010; Carpenter and Olson, 2012; Emirmustafaoğlu and Gökmen, 2015; Kambara et al., 2013; Takashima et al., 2017). A longitudinal study conducted over 15 days reported that unfamiliar word forms in L1 were easily associated with meaningless pictures, compared to meaningless sounds (Kambara et al., 2013). Further, Paivio and colleagues have shown that the free recall of items is increased by associative learning of verbal and non-verbal inputs, compared to the learning of verbal inputs only or the associative learning of verbal and verbal inputs (Paivio and Csapo, 1973; Paivio and Lambert, 1981; Paivio, 1986). For example, Paivio and Csapo (1973) reported that participants recalled pictures better than words after associative learning of words and pictures by writing or naming the words and pictures. Paivio and Lambert (1981) asked French and English bilinguals to write the English names of pictures (picture condition), translate French words (translated condition), or copy the presented English words (copied condition). The recall of English words was better for the picture condition than for the translated and copied conditions (Paivio and Lambert, 1981; Paivio, 1986). These findings suggest that when words and non-verbal inputs (e.g., visual features, auditory features, and so on) are associatively learned (dual coding theory), the associative learning of words and pictures is superior to the associative learning of words and words (picture superiority effect; Paivio, 1986; Carpenter and Olson, 2012; Emirmustafaoğlu and Gökmen, 2015). In addition, Kroll and colleagues proposed that the translation from L2 to L1 ensures a faster access to the referent than the translation from L1 to L2 in their model (the revised hierarchical model) based on their experiments

(Kroll et al., 2010). Their asymmetrical model means that the translation from L2 to L1 would be lexically associated and the translation from L1 to L2 would be semantically mediated until bilinguals acquire the skill for directly associating L2 with the referent such as strong associations between L1 and referents (Kroll et al., 2010). Thus, when participants directly learn associative pairs of word forms in L2 and referents, pictorial referents could facilitate the learning of these associations.

Previous studies have investigated how participants associated novel word forms in L2 with pictures (Lee et al., 2003; Havas et al., 2018). For example, Havas et al. (2018) reported that native Spanish speakers easily learned Hungarian words associated with familiar pictures, compared to Hungarian words associated with unfamiliar pictures and Hungarian words only. A neuroimaging study identified differences between situation- and text-based learning in this regard: Jeong and colleagues (2010) asked Japanese native speakers to learn 1) spoken Korean words (e.g., "help") with videos that represented events (e.g., a person asking others for help) as situation-based learning, and 2) spoken Korean words with a whiteboard on which Japanese meanings were written (e.g., "help") as text-based learning. The behavioral results of situation tests suggest that task performance on situation-based learning was better than on text-based learning. Other previous findings have also supported the picture superiority effect in new L2 word learning (Kost et al., 1999; Carpenter and Olson, 2012; Emirmustafaoğlu and Gökmen, 2015; Rokni and Karimi, 2013; Morett, 2019; Wang and Lee, 2021). Carpenter and Olson (2012) reported that new L2 words (Swahili words) were acquired better from pictures than from translated L1 words (English words), when task-related biases (overconfidence biases) were removed from the retrieval practice and instructions. Taking these findings together, it seems that associative learning of verbal and non-verbal information in L2 (e.g., pictures, videos, and so on) could be better than learning of only verbal information in L2.

The purpose of the present study was to examine learning differences among the following conditions: 1) associations between word forms in L2 (Chinese) and non-verbal information (pictures), 2) associations between word forms in L2 (Chinese) and word forms in L1, and 3) word forms in L2 (Chinese). In this study, participants were Japanese native speakers, since few studies have directly examined associative learning of word forms in L2 and non-verbal information for native speakers of Japanese (Lee et al., 2003; Jeong et al., 2010). As for the similarities and differences between Japanese and Chinese characters, the Japanese language includes syllabary (hiragana and katakana) and logography (kanji), whereas the Chinese language includes only logography (hanzi; e.g., Akamatsu, 2002). The logographical characters are used as morphemes or words, whereas the syllabic characters are used as syllables or words (e.g., Muljani et al., 1998; Akamatsu, 2002). A previous study reported that morphemes, vocabulary, and grammatical knowledge contribute to the word and meaning inference in L2 Chinese (Chen et al., 2020). L2 vocabulary knowledge also predicted L2 reading comprehension (Prior et al., 2014). Thus, we considered the sample selection and stimuli in this study. First, all participants were native Japanese speakers that had never learned Chinese words. Second, in this study, we did not use words, which are

commonly or similarly used in both Chinese and Japanese languages, since the commonality or similarity effects between words might influence the associative learning of words in L2 and referents. In fact, previous psycholinguistic research did not use similar words in L2 as experimental stimuli (Paivio and Lambert, 1981). We made two predictions, based on previous findings. First, when native speakers of Japanese learn novel word forms in Chinese as an L2, the learning condition retrieval and recognition performance (accuracy rates and response times) would be better for associative pairs of novel word forms in Chinese and pictorial referents (CP) than for those of novel word forms in Chinese and words in Japanese (CJ). This hypothesis is consistent with previous studies that reported picture superiority effects during learning of associative pairs of words and pictures (e.g., Kambara et al., 2013) and concept mediation effects between words in L1 and L2 (Potter et al., 1984). Second, the learning condition retrieval performance of novel word forms only in Chinese (C) would be significantly higher than that of CP and CJ, since attentional loads to referents would decrease the learning condition retrieval performance of items (L2 words) and sources (referents) more than the learning condition retrieval performance of items only (L2 words only; e.g., Troyer et al., 1999).

MATERIALS AND METHODS

Participants

Thirty healthy university students (21 females; M_{age} = 20.70; SD_{age} = 1.60) participated in this experiment. All participants were right-handed native Japanese speakers, who had never learned Chinese words. Thus, the Chinese language was a second language for all participants. Written informed consent was obtained from each participant before starting the experiment. After the experiment, each participant received a gift card (a QUO card) of 1000 Japanese YEN (JPY) from an experimenter. This experiment was approved by the ethics committee of the Graduate School of Education at Hiroshima University (code number: 2019089).

Materials

One hundred and fifty pictures in black and white were collected from a psycholinguistics database (Duñabeitia et al., 2018). All pictures were grayscale drawings of objects. We avoided selecting ambiguous pictures, that is, those in which the experimenters (authors) could not recognize what the objects were. Japanese words associated with these pictures were written down to identify the frequency of each Japanese word. The word frequencies were identified in the Balanced Corpus of Contemporary Written Japanese (BCCWJ: https://pj.ninjal.ac.jp/corpus_center/bccwj/en/freq-list.html), which includes Japanese words in hiragana, katakana, and kanji characters.

Five lists were used in this study. Each list included 30 Chinese words and 30 Japanese words associated with the 30 pictures selected from the psycholinguistic database (Duñabeitia et al., 2018). First, we used BCCWJ to identify each Japanese word's frequency, as noted. Second, we controlled

the stimuli across the five lists. No significant differences among the lists were found for Japanese word lengths [$F (4, 145) = 0.07$, n. s., $f = 0.04$; f here means effect size; see Cohen, 1988; Erdfelder et al., 1996], Japanese word frequencies [$F (4,145) = 0.28$, n. s., $f = 0.09$], and Chinese word lengths [$F (4, 145) = 0.64$, n. s., $f = 0.13$] across the five lists. A Chinese-Japanese bilingual (the first author) and a Japanese native speaker (the last author) selected word stimuli that exhibited differences between Chinese words and translated Japanese words to avoid similarity effects between Chinese and Japanese languages (Paivio and Lambert, 1981).

Finally, we designed four conditions. The first condition, CP, consisted of 30 associations each between a Chinese word and a pictorial referent. The second condition, CJ, comprised 30 associations between a Chinese word and the associated Japanese word. The third condition, C, only involved the 30 Chinese words. The fourth condition, Chinese novel words (CN), included 30 Chinese words that participants did not learn in the learning task and were not presented in recognition tasks.

Procedure

This study employed a within-subjects design. Participants conducted a learning task, recognition tasks, and learning condition retrieval tasks (**Figure 1**). The order of these tasks was learning task, first learning condition retrieval task, first recognition task, second recognition task, third recognition task, and second learning condition retrieval task (**Figure 2**). In each learning task, Chinese words in the CP, CJ, and C conditions were presented once. In each recognition task, Chinese words in the CP, CJ, and C conditions were presented once, whereas each feedback for the correct answers was also presented once. In each learning condition retrieval task, Chinese words in the CP, CJ, C, and CN conditions were presented once. In the first learning task, participants learned each stimulus, specifically, pairs of Chinese words and pictures (CP), pairs of Chinese words and Japanese words (CJ), and Chinese words only (C) (**Figure 1**). Each stimulus in the CP and CJ conditions was presented for 2,000 ms, whereas the presentation time in the C condition was 1,500 ms or 2,000 ms. After each stimulus presentation in the CP, CJ, and C conditions, a fixation cross was presented for 2,000 ms. The presentation order of stimuli and conditions was counterbalanced among participants. Second, in learning condition retrieval tasks, participants judged which condition (CP, CJ, C, and CN) the presented Chinese word was associated with by pushing one of four keys (**Figure 1**). The keys of CP, CJ, C, and CN were associated with the right index, middle, ring, and little fingers, respectively. The duration of each stimulus depended on each participant's judgment. After the judgment, a fixation cross was presented for 2,000 ms. The presentation order of each stimulus was counterbalanced among participants. Third, in the recognition tasks, participants judged to which picture (in the CP condition) or Japanese word (in the CJ condition) each presented Chinese word was associated with (**Figure 1**). In the CP condition, three pictures, including a correct picture, and a Chinese word were simultaneously presented. In the CJ condition, three Japanese
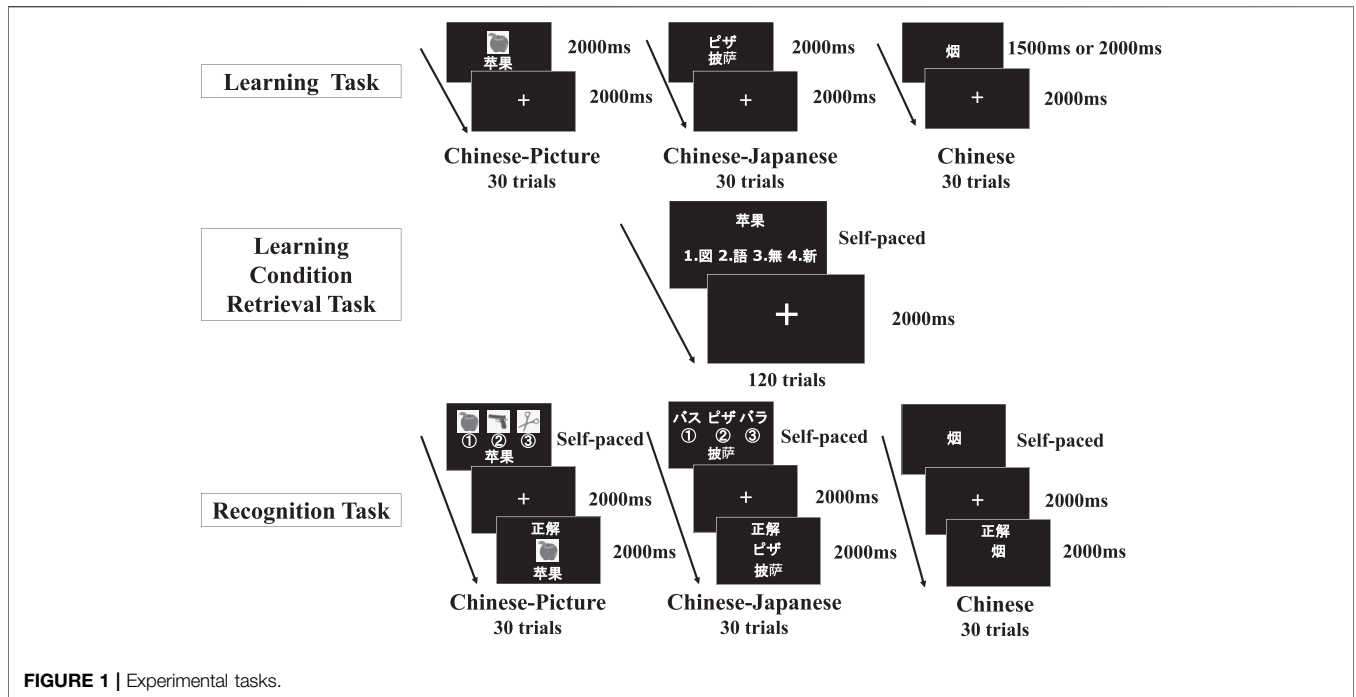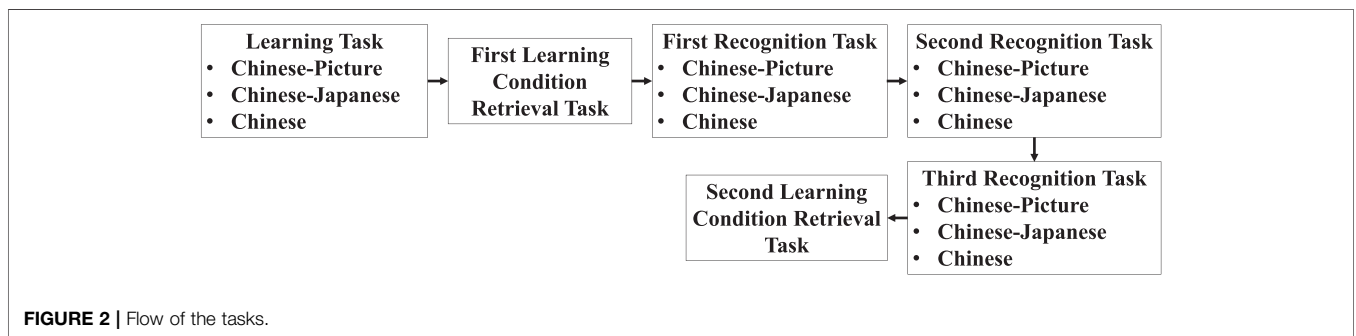
FIGURE 1 | Experimental tasks.

FIGURE 2 | Flow of the tasks.

words, including a correct Japanese word and a Chinese word, were simultaneously presented. The duration of each stimulus depended on each participant's judgment. After each judgment, pushing a key associated with the right index, middle, or ring finger, a fixation was followed for 2,000 ms. The presentation order of stimuli and conditions was counterbalanced among participants. The positions of the correct referents in the recognition tasks were randomized and counterbalanced among participants. In addition, participants conducted an additional recognition task for the C condition to control stimulus presentation times in recognition tasks. In this task, when participants recognized each Chinese word in the C condition, they pushed a key associated with the right index finger. The duration of each stimulus depended on each participant's response. After each judgment, a fixation was followed for 2,000 ms. When participants needed to take a break between tasks, they had a 5 min break between the first learning condition retrieval task and the first recognition task. Except for this break, there was no other interruption between tasks.

All the participants conducted these tasks on a Windows based laptop (ProBook 650 G4) with SuperLab 4.5 software for task presentation and recording of behaviors.

Statistical Analysis

We conducted generalized linear mixed effects models for performances and linear mixed effects models for response times, including two random effects of participant and word (item), two fixed effects of conditions (CP and CJ in the recognition tasks, and CP, CJ, C, and CN in the learning condition retrieval tasks), and times (the first, second, and third recognition tasks, and the first and second learning condition retrieval tasks) to represent performances on all trials (0: incorrect; 1: correct) and the response times of the correct trials in both recognition and learning condition retrieval tasks (see Baayen et al., 2008 for linear mixed effects modeling in psycholinguistic research). In the generalized linear mixed effects models for performances, the independent variables were conditions and times in recognition and learning condition retrieval tasks,

TABLE 1 | Mean proportions and standard deviations (*SD*) in recognition tasks.

	CP		CJ	
	M	*SD*	*M*	*SD*
First time	0.91	0.07	0.92	0.05
Second time	0.96	0.06	0.98	0.03
Third time	0.98	0.04	0.98	0.04

M, mean proportion; SD, standard deviation; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs. These mean proportions and standard deviations associate with scores in which participants correctly identified referents. The mean proportions and standard deviations were calculated on SPSS.

whereas the dependent variables were performances on all trials (0: incorrect; 1: correct). On the other hand, in the linear mixed effects models for response times, the independent variables were conditions and times in recognition and learning condition retrieval tasks, whereas the dependent variables were the response times of the correct answers. Random slopes were used in the models to decrease type I error (Barr, 2013). The syntax for the generalized linear mixed effects modeling of the performances in each task was glmer {performance ~ time*condition + (1 + time + condition | participant) + (1 | word), data = data, family = "binomial", control = glmerControl [optimizer = "bobyqa", optCtrl = list (maxfun = 100000)]}. The syntax for the linear mixed effects modeling of the response times in each task was lmer {response time ~ time*condition + (1 + time + condition | participant) + (1 | word), data = data, control = lmerControl [optimizer = "bobyqa", optCtrl = list (maxfun = 100000)]}. The analyses of the linear mixed-models were conducted by using a statistical program R (R Core Team, 2019), "lme4" (Bates, 2005; Bates et al., 2015), "lmerTest" (Kuznetsova et al., 2017) and "lsmeans" packages (Lenth, 2016). In addition, we conducted analyses of variance (ANOVAs) including two factors of condition and time to represent correct response scores (CRS; see Tulving and Thomson, 1971) for each condition in learning condition retrieval tasks. The independent variables were conditions (CP, CJ, and C in learning condition retrieval tasks) and times (the first and second learning condition retrieval tasks) in learning condition retrieval tasks, whereas the dependent variables were the CRS. We calculated the CRS as adjusted hit rate (i.e., the participant correctly judged the learned stimuli as learned stimuli) excluding false alarm rate for each participant (i.e., the participant incorrectly judged unlearned stimuli as learned stimuli; see Tulving and Thomson, 1971). Results corrected with Greenhouse-Geisser correction were reported only when Mauchly's test of sphericity was significant. Finally, to clarify whether each missed item should be judged as a learned item (e.g., a missed item in CP would be judged as an item in CJ or C, not CN) or an unlearned item (e.g., a missed item in CP would be judged as an item in CN, not CJ or C), we calculated the learned and unlearned rates for each condition (CP, CJ, and C). After the calculations of learned and unlearned rates for each condition, we performed paired *t*-tests to establish whether there was any significant difference between the learned and unlearned rates in each condition. Since the paired *t*-tests were conducted 6 times, we adjusted the *p*-value to $p < 0.05/6$ to avoid type I error. We conducted these analyses using SPSS

TABLE 2 | Results of the generalized linear mixed effects model for the performances in Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ).

Random effects	Variance	SD	
Words (intercept)	0.60	0.77	
Participants (intercept)	0.39	0.62	
Time 2	0.13	0.37	
Time 3	0.40	0.63	
Condition 2	0.03	0.16	

Fixed effects	Estimate	SE	z value
Intercept***	2.68	0.20	13.68
Time 2***	1.16	0.25	4.67
Time 3***	1.77	0.31	5.69
Condition 2	0.09	0.18	0.47
Time 2: Condition 2	0.34	0.32	1.05
Time 3: Condition 2	−0.06	0.36	−0.16

***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; SD, standard deviation; SE, standard error. These results were generated from the "lme4" (Bates, 2005; Bates et al., 2015) and "lmerTest" packages (Kuznetsova et al., 2017) on R (R Core Team, 2019). In the generalized linear mixed effect model, times include the first, second, and third recognition tasks as 1, 2, and 3, respectively, whereas conditions include Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ) in the recognition tasks as 1 and 2, respectively.

software. To control the stimuli in the Material sections, we used js-STAR (http://www.kisnet.or.jp/nappa/software/star/).

RESULTS

Recognition Tasks

First, to examine the differences between the performances in CP and CJ, we applied a generalized linear mixed-effects model to analyze conditions (CP and CJ) and times (the first, second, and third time) of recognition tasks as fixed effects, words and participants as random effects, and performance (0: incorrect; 1: correct) as binomially dependent variables (**Tables 1–3**). In the generalized linear mixed effects model, times included the first, second, and third recognition tasks as 1, 2, and 3, respectively, whereas conditions included CP and CJ as 1 and 2, respectively. The results showed that performances of the second and third recognition tasks were significantly higher than those on the first recognition task in the CP and CJ conditions, whereas there was no significant difference between the performances of the second and third recognition tasks in the CP and CJ conditions. In addition, there was no significant difference between the performances in CP and CJ in the first, second, and third recognition tasks.

Second, to investigate the differences among the response times of the correct answers in the CP and CJ conditions in the retrieval tasks, we also conducted a linear mixed effects model to analyze conditions (CP and CJ) and times (the first, second, and third time) of recognition tasks as fixed effects, words and participants as random effects, and the response times (ms) of correct answers as dependent variables (**Tables 4–6**). In the linear mixed effects model, times included the first, second, third recognition tasks as 1, 2, and 3, respectively, whereas

TABLE 3 | All the contrasts of the performances of Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ) in recognition tasks.

Contrast	Estimate	SE	z ratio	p
CP1 vs. CP2****	−1.16	0.25	−4.67	<0.0001
CP1 vs. CP3****	−1.77	0.31	−5.69	<0.0001
CP1 vs. CJ1	−0.09	0.18	−0.47	0.9972
CP1 vs. CJ2****	−1.58	0.28	−5.58	<0.0001
CP1 vs. CJ3****	−1.79	0.33	−5.47	<0.0001
CP2 vs. CP3	−0.61	0.34	−1.80	0.4677
CP2 vs. CJ1***	1.08	0.27	4.03	0.0008
CP2 vs. CJ2	−0.42	0.29	−1.46	0.6872
CP2 vs. CJ3	−0.63	0.36	−1.75	0.502
CP3 vs. CJ1****	1.68	0.32	5.24	<0.0001
CP3 vs. CJ2	0.18	0.37	0.50	0.9963
CP3 vs. CJ3	−0.03	0.33	−0.08	1
CJ1 vs. CJ2****	−1.50	0.28	−5.42	<0.0001
CJ1 vs. CJ3****	−1.71	0.32	−5.33	<0.0001
CJ2 vs. CJ3	−0.21	0.36	−0.58	0.9924

*****: p < 0.0001; ***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; SE, standard error. Numbers in contrasts mean the times of the recognition tasks. These results were generated from the "lsmeans" package (Lenth, 2016) on R (R Core Team, 2019). These contrasts were adjusted by Tukey's method.*

TABLE 4 | Mean response times (ms) and standard deviations (SD) in recognition tasks.

	CP		CJ	
	M	SD	M	SD
First time	3976.14	272.47	4217.84	573.62
Second time	3631.52	273.66	3860.49	447.19
Third time	3491.26	401.15	3661.73	371.00

M, mean response time of correct answers; SD, standard deviation; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs. These mean response times and standard deviations associate with response times for correct answers in which participants correctly identified referents. The mean response times and standard deviations were calculated on SPSS.

conditions included CP and CJ as 1 and 2, respectively. The results showed that the response times of the second and third recognition tasks were significantly faster than those of the first recognition task in the CP and CJ conditions, whereas the response times in the third recognition task were also significantly faster than in the second recognition task in the CP and CJ conditions. Additionally, the response times in CP were significantly faster than CJ in the first, second, and third recognition tasks.

Learning Condition Retrieval Tasks

First, to examine the differences between the performances in CP, CJ, C, and CN in learning condition retrieval tasks, we conducted a generalized linear mixed effects model to analyze conditions (CP, CJ, C, and CN) and times (the first and second times) of learning condition retrieval tasks as fixed effects, words and participants as random effects, and performance (0: incorrect; 1: correct) as binomially dependent variables (**Tables 7–10**). In the generalized linear mixed effects model, times included the first and second learning condition retrieval tasks as 1 and 2, respectively, whereas conditions included CP,

TABLE 5 | Results of the linear mixed effects model for response times of correct answers in Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ).

Random effects	Variance	SD
Words (intercept)	92140	303.5
Participants (intercept)	103236	321.3
Time 2	21068	145.1
Time 3	19598	140.0
Condition 2	29886	172.9
Residual	665869	816.0

Fixed effects	Estimate	SE	t value
Intercept***	3974.18	72.77	54.61
Time 2***	−354.89	47.88	−7.41
Time 3***	−482.98	47.24	−10.23
Condition 2***	258.14	51.18	5.04
Time 2: Condition 2	−11.99	56.16	−0.21
Time 3: Condition 2+	−106.12	56.03	−1.89

*****: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; SD, standard deviation; SE, standard error. These results were generated from the "lme4" (Bates, 2005; Bates et al., 2015) and "lmerTest" packages (Kuznetsova et al., 2017) on R (R Core Team, 2019). In the linear mixed effect model, times include the first, second, and third recognition tasks as 1, 2, and 3, respectively, whereas conditions include Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ) in the recognition tasks as 1 and 2, respectively.*

TABLE 6 | All the contrasts of the response times of correct answers of Chinese word and picture pairs (CP) and Chinese word and Japanese word pairs (CJ) in recognition tasks.

Contrast	Estimate	SE	t ratio	p
CP1 vs. CP2****	354.9	47.9	7.41	<0.0001
CP1 vs. CP3****	483.0	47.2	10.22	<0.0001
CP1 vs. CJ1****	−258.1	51.2	−5.04	<0.0001
CP1 vs. CJ2	108.7	57.2	1.90	0.4188
CP1 vs. CJ3****	331.0	50.3	6.58	<0.0001
CP2 vs. CP3*	128.1	43.7	2.93	0.0485
CP2 vs. CJ1****	−613.0	57.3	−10.70	<0.0001
CP2 vs. CJ2***	−246.1	50.3	−4.90	0.0001
CP2 vs. CJ3	−23.9	47.1	−0.51	0.9955
CP3 vs. CJ1****	−741.1	62.6	−11.85	<0.0001
CP3 vs. CJ2****	−374.2	59.9	−6.25	<0.0001
CP3 vs. CJ3*	−152.0	50.1	−3.03	0.0375
CJ1 vs. CJ2****	366.9	47.6	7.70	<0.0001
CJ1 vs. CJ3****	589.1	47.1	12.51	<0.0001
CJ2 vs. CJ3****	222.2	43.5	5.10	<0.0001

*****: p < 0.0001; ***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; SE, standard error. Numbers in contrasts mean the times of the recognition tasks. These results were generated from the "lsmeans" package (Lenth, 2016) on R (R Core Team, 2019). These contrasts were adjusted by Tukey's method.*

CJ, C, and CN as 1, 2, 3, and 4, respectively. The results showed that in all conditions (CP, CJ, C, and CN), the performances of the second learning condition retrieval task were significantly higher than those of the first learning condition retrieval task. In addition, in the first learning condition retrieval task, the accuracy rate of CN was significantly higher than those of the other conditions (CP, CJ, and C), whereas there was no significant difference of other comparisons. In the second learning condition retrieval task, there was a marginally

TABLE 7 | Mean proportions and standard deviations (SDs) in learning condition retrieval tasks.

	CP		CJ		C		CN	
	M	SD	M	SD	M	SD	M	SD
First time	0.39	0.16	0.45	0.15	0.47	0.19	0.68	0.19
Second time	0.81	0.12	0.80	0.14	0.84	0.15	0.87	0.18

M, mean proportion; SD, standard deviation; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; C, Chinese words only; CN, novel Chinese words. These mean proportions and standard deviations associate with scores in which participants correctly identified learning conditions. The mean proportions and standard deviations were calculated on SPSS.

TABLE 8 | Results of the generalized linear mixed effects model for the performances of Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in learning condition retrieval tasks.

Random effects	Variance	SD
Words (intercept)	0.80	0.89
Participants (intercept)	0.39	0.62
Time 2	0.34	0.59
Condition 2	0.41	0.64
Condition 3	0.20	0.45
Condition 4	1.33	1.15

Fixed effects	Estimate	SE	z value
Intercept**	−0.54	0.17	−3.22
Time 2***	2.51	0.17	14.60
Condition 2	0.30	0.21	1.44
Condition 3+	0.36	0.19	1.89
Condition 4***	1.64	0.28	5.77
Time 2: Condition 2**	−0.55	0.18	−3.09
Time 2: Condition 3	−0.22	0.19	−1.21
Time 2: Condition 4***	−1.03	0.20	−5.25

***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 < p < 0.05; SD, standard deviation; SE, standard error. These results were generated from the "lme4" (Bates, 2005; Bates et al., 2015) and "lmerTest" packages (Kuznetsova et al., 2017) on R (R Core Team, 2019). In the generalized linear mixed effect model, times include the first and second learning condition retrieval tasks as 1 and 2, respectively, whereas conditions include Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in the learning condition retrieval tasks as 1, 2, 3, and 4, respectively.

TABLE 9 | All the contrasts of the performances of Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in learning condition retrieval tasks.

Contrast	Estimate	SE	z ratio	p
CP1 vs. CP2****	−2.51	0.17	−14.60	<0.0001
CP1 vs. CJ1	−0.30	0.21	−1.44	0.8368
CP1 vs. CJ2****	−2.26	0.23	−9.97	<0.0001
CP1 vs. C1	−0.36	0.19	−1.89	0.5599
CP1 vs. C2****	−2.65	0.21	−12.83	<0.0001
CP1 vs. CN1****	−1.64	0.28	−5.77	<0.0001
CP1 vs. CN2****	−3.12	0.31	−9.95	<0.0001
CP2 vs. CJ1****	2.21	0.26	8.55	<0.0001
CP2 vs. CJ2	0.25	0.23	1.11	0.9542
CP2 vs. C1****	2.15	0.25	8.47	<0.0001
CP2 vs. C2	−0.13	0.22	−0.63	0.9985
CP2 vs. CN1+	0.88	0.32	2.78	0.0994
CP2 vs. CN2	−0.61	0.31	−2.00	0.4849
CJ1 vs. CJ2****	−1.96	0.16	−12.05	<0.0001
CJ1 vs. C1	−0.06	0.23	−0.27	1
CJ1 vs. C2****	−2.35	0.25	−9.28	<0.0001
CJ1 vs. CN1***	−1.34	0.29	−4.68	0.0001
CJ1 vs. CN2****	−2.82	0.33	−8.64	<0.0001
CJ2 vs. C1****	1.90	0.27	7.16	<0.0001
CJ2 vs. C2	−0.39	0.24	−1.58	0.7636
CJ2 vs. CN1	0.62	0.30	2.06	0.442
CJ2 vs. CN2+	−0.86	0.31	−2.82	0.0912
C1 vs. C2****	−2.29	0.17	−13.32	<0.0001
C1 vs. CN1***	−1.28	0.30	−4.22	0.0006
C1 vs. CN2****	−2.76	0.35	−7.94	<0.0001
C2 vs. CN1*	1.01	0.32	3.20	0.0301
C2 vs. CN2	−0.47	0.32	−1.46	0.829
CN1 vs. CN2****	−1.48	0.18	−8.17	<0.0001

****: p < 0.0001; ***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; C, Chinese words only; CN, novel Chinese words; SE, standard error. Numbers in contrasts mean the times of the learning condition retrieval tasks. These results were generated from the "lsmeans" package (Lenth, 2016) on R (R Core Team, 2019). These contrasts were adjusted by Tukey's method.

TABLE 10 | Means and standard deviations of corrected recognition scores (CRS) in learning condition retrieval tasks.

		CP1	CJ1	C1	CP2	CJ2	C2
HIT	M	0.39	0.45	0.47	0.81	0.80	0.84
	SD	0.16	0.15	0.19	0.12	0.14	0.15
FA	M	0.32	0.32	0.32	0.13	0.13	0.13
	SD	0.19	0.19	0.19	0.18	0.18	0.18
CRS	M	0.08	0.13	0.15	0.68	0.66	0.70
	SD	0.26	0.21	0.27	0.25	0.28	0.26

HIT, hit rate; FA, false alarm rate; CRS, corrected recognition score (Tulving and Thomson, 1971); M, mean accuracy rate; SD, standard deviation; CP1, Chinese word and picture pairs in the first learning condition retrieval task; CJ1, Chinese word and Japanese word pairs in the first learning condition retrieval task; C1, Chinese words only in the first learning condition retrieval task; CP2, Chinese word and picture pairs in the second learning condition retrieval task; CJ2, Chinese word and Japanese word pairs in the second learning condition retrieval task; C2, Chinese words only in the second learning condition retrieval task. In the learning condition retrieval tasks, participants identified learning conditions associated with presented words. We calculated CRS scores for participants by using a syntax (CRS = HIT − FA) based on a previous study (Tulving and Thomson, 1971) The means and standard deviations were calculated on Microsoft Excel.

significant difference between the performance in CJ and CN, whereas there was no significant difference of other comparisons. Further, based on previous findings (e.g., Tulving and Thomson, 1971), we also conducted another ANOVA test to analyze condition and number of learning condition retrieval tasks as independent variables and corrected recognition scores (CRSs) as dependent variables (**Table 4**). In ANOVA results for the CRS of CP, CJ, and C conditions, the main effect of time was significant [$F(1, 29) = 277.47$, $p < 0.001$, $\eta_p^2 = 0.91$], while the main effect of condition [Greenhouse-Geisser corrected: $F(1.645, 47.716) = 2.06$, $p = 0.15$, $\eta_p^2 = 0.07$] and the interaction [$F(2, 58) = 1.41$, $p = 0.25$, $\eta_p^2 = 0.05$] were not significant. Based on a post-hoc analysis (Bonferroni's method), the accuracy rate of the second learning condition retrieval task was significantly higher than those of the first learning condition retrieval task.

Second, to examine the differences between the response times of the correct answers in CP, CJ, C, and CN in learning condition retrieval tasks, we conducted a linear mixed effects model to

TABLE 11 | Mean response times (ms) and standard deviations (SD) in learning condition retrieval tasks.

	CP		CJ		C		CN	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
First time	2823.05	686.59	3272.93	1231.78	2722.01	756.67	2016.01	379.36
Second time	1826.11	360.23	2084.85	556.32	1546.62	313.69	1462.16	284.29

M, mean response time; SD, standard deviation; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; C, Chinese words only. These mean response times and standard deviations associate with response times for correct answers in which participants correctly identified learning conditions. The mean response times and standard deviations were calculated on SPSS.

TABLE 12 | Results of the linear mixed effects model for the response times of correct answers of Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in learning condition retrieval tasks.

Random effects	Variance	*SD*
Words (intercept)	127170	356.6
Participants (intercept)	289973	538.5
Time 2	213335	461.9
Condition 2	209675	457.9
Condition 3	19761	140.6
Condition 4	78936	281.0
Residual	1378223	1174.0

Fixed effects	Estimate	*SE*	*t* value
Intercept***	2845.71	124.18	22.92
Time 2***	−989.64	115.07	−8.60
Condition 2***	491.69	133.82	3.67
Condition 3	−141.21	106.03	−1.33
Condition 4***	−829.49	113.18	−7.33
Time 2: Condition 2*	−237.80	108.40	−2.19
Time 2: Condition 3	−159.74	106.50	−1.50
Time 2: Condition 4***	462.89	100.91	4.59

***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 < p < 0.05; SD, standard deviation; SE, standard error. These results were generated from the "lme4" (Bates, 2005; Bates et al., 2015) and "lmerTest" packages (Kuznetsova et al., 2017) on R (R Core Team, 2019). In the linear mixed effect model, times include the first and second learning condition retrieval tasks as 1 and 2, respectively, whereas conditions include Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in the learning condition retrieval tasks as 1, 2, 3, and 4, respectively.

TABLE 13 | All the contrasts of the response times of correct answers of Chinese word and picture pairs (CP), Chinese word and Japanese word pairs (CJ), Chinese words only (C), and novel Chinese words (CN) in learning condition retrieval tasks.

Contrast	Estimate	*SE*	*t* ratio	*p*
CP1 vs. CP2****	989.6	115.3	8.58	<0.0001
CP1 vs. CJ1**	−491.7	134.2	−3.67	0.0094
CP1 vs. CJ2****	735.8	119.6	6.15	<0.0001
CP1 vs. C1	141.2	106.7	1.32	0.8893
CP1 vs. C2****	1290.6	122.4	10.55	<0.0001
CP1 vs. CN1****	829.5	113.6	7.30	<0.0001
CP1 vs. CN2****	1356.2	158.2	8.57	<0.0001
CP2 vs. CJ1****	−1481.3	178.3	−8.31	<0.0001
CP2 vs. CJ2	−253.9	117.6	−2.16	0.3912
CP2 vs. C1****	−848.4	134.7	−6.30	<0.0001
CP2 vs. C2*	300.9	85.8	3.51	0.0143
CP2 vs. CN1	−160.2	110.2	−1.45	0.8285
CP2 vs. CN2**	366.6	99.9	3.67	0.0091
CJ1 vs. CJ2****	1227.4	113.0	10.86	<0.0001
CJ1 vs. C1****	632.9	123.6	5.12	<0.0001
CJ1 vs. C2****	1782.3	166.8	10.69	<0.0001
CJ1 vs. CN1****	1321.2	153.8	8.59	<0.0001
CJ1 vs. CN2****	1847.9	211.6	8.74	<0.0001
CJ2 vs. C1***	−594.5	116.2	−5.12	0.0001
CJ2 vs. C2***	554.8	108.6	5.11	0.0001
CJ2 vs. CN1	93.7	120.1	0.78	0.9935
CJ2 vs. CN2**	620.5	146.1	4.25	0.0023
C1 vs. C2****	1149.4	111.9	10.27	<0.0001
C1 vs. CN1****	688.3	110.1	6.25	<0.0001
C1 vs. CN2****	1215.0	161.7	7.51	<0.0001
C2 vs. CN1***	−461.1	100.6	−4.58	0.0004
C2 vs. CN2	65.7	99.5	0.66	0.9978
CN1 vs. CN2***	526.8	106.0	4.97	0.0002

****: p < 0.0001; ***: p < 0.001; **: p < 0.01; *: p < 0.05; +: 0.1 > p > 0.05; CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; C, Chinese words only; CN, novel Chinese words; SE, standard error. Numbers in contrasts mean the times of the recognition tasks. These results were generated from the "lsmeans" package (Lenth, 2016) on R (R Core Team, 2019). These contrasts were adjusted by Tukey's method.

analyze conditions (CP, CJ, C, and CN) and times (the first and second times) of learning condition retrieval tasks as fixed effects, words and participants as random effects, and the response times (ms) of correct answers as dependent variables (**Tables 11–13**). In the linear mixed effects model, times included the first and second learning condition retrieval tasks as 1 and 2, respectively, whereas conditions included CP, CJ, C, and CN as 1, 2, 3, and 4, respectively. The results showed that in all conditions (CP, CJ, C, and CN), the response times on the second learning condition retrieval task were significantly faster than those on the first learning condition retrieval task. Regarding the results of the first learning condition retrieval task, the response times of CN were significantly faster than those of the other three conditions (CP, CJ, and C), whereas the response times of C were significantly faster than those of CJ. In addition, the response times in CP were significantly faster than in CJ. In contrast, there was no significant difference between CP and C. On the second learning condition retrieval task, the response times of C and CN were significantly

faster than those of CP and CJ. In contrast, there was no significant difference between CP and CJ or between C and CN.

Finally, we analyzed the missed items in CP, CJ, and C to identify whether each missed item should be judged as a learned item (e.g., a missed item on CP would be judged as an item in CJ or C, not CN) or an unlearned item (e.g., a missed item of CP would be judged as an item in CN, not CJ or C; **Table 14**). First, we calculated the learned and unlearned rates for each condition (CP, CJ, and C). Second, paired *t*-tests were performed to the learned and unlearned rates in each condition. Considering that paired *t*-tests were conducted 6 times, we adjusted the *p*-value to

TABLE 14 | Mean learned and unlearned rates (standard deviations) of missed items in leaning condition retrieval tasks.

	CP		CJ		C	
	Learned	**Unlearned**	**Learned**	**Unlearned**	**Learned**	**Unlearned**
First time	0.46 (0.25)	0.54 (0.25)	0.53 (0.23)	0.47 (0.23)	0.32 (0.20)	0.68 (0.20)
Second time	0.68 (0.21)	0.32 (0.21)	0.71 (0.30)	0.29 (0.30)	0.28 (0.34)	0.72 (0.34)

CP, Chinese word and picture pairs; CJ, Chinese word and Japanese word pairs; C, Chinese words only; these rates show miss rates as learned or unlearned rates (standard deviations), with each number of learned or unlearned items divided by the respective sum of learned and unlearned items. These mean rates and standard deviations associate with scores in which participants correctly or incorrectly identified learning conditions. The mean rates and standard deviations were calculated on SPSS.

$p < 0.05/6$ to avoid type I error. Additionally, all the conditions in which all trials in each condition were correct items (i.e., all the conditions did not involve any missed trials) were excluded as missing values. In paired *t*-tests for learned and unlearned rates of CP, CJ, and C in the first learning condition retrieval task, the learned rate was significantly lower than the unlearned rate in C [$t (29) = -4.99$, $p < 0.001$, $d = 0.90$], while no significant difference between the learned and unlearned rates was seen in CP or CJ. In the second learning condition retrieval task, the learned rate was significantly higher than the unlearned rate in CP and CJ [CP: $t (28) = 4.69$, $p < 0.001$, $d = 0.86$; CJ: $t (28) = 3.75$, $p < 0.005$, $d = 0.70$] and significantly lower in C [$t (24) = -3.25$, $p < 0.005$, $d = 0.65$]. These results indicate that associative conditions (CP and CJ conditions) were easier to learn than the non-associative condition (C condition).

DISCUSSION

Training Effects

We found that the accuracy rates on recognition and learning condition retrieval tasks gradually increased in each condition (CP, CJ, C, and CN). These findings would be affected by training effects, including repeated training or feedback on correct answers in recognition tasks. These findings were consistent with previous studies (Breitenstein et al., 2005; Cornelissen et al., 2004; Kambara et al., 2013; Havas et al., 2018; Horinouchi et al., under review). For instance, Kambara et al. (2013) examined the associative learning of unfamiliar words in L1 and meaningless perceptual features by repeating training and feedback on correct answers for 15 days. They found that the accuracy rates of the associative pairs of unfamiliar words in L1 and meaningless referents (meaningless pictures or sounds) increased with presentation time and feedback on correct answers. This study suggests that training that is repeated or includes feedback on correct answers could enhance associative learning of both words in L2 and referents (pictures and words in L1).

Effects of Picture Superiority on Associative Learning of Words in L2 and Referents

The response times on the first learning condition retrieval task and all the recognition tasks supported picture superiority effects, in which pictorial referents promote recognition and learning condition retrieval speed for associative pairs of word forms in L2 and referents. These findings are consistent with previous studies on the effects of picture superiority (Tulving and Thomson, 1971; Paivio and Csapo, 1973; Paivio and Lambert, 1981; Kambara et al., 2013). For example, Paivio and Csapo (1973) asked participants to learn stimuli under picture, abstract word, and concrete word conditions and also to recall words (word condition) or names of pictures (picture naming condition) presented during the learning session. They reported that the performance on the picture naming condition was significantly better than that on the word condition. In addition, Paivio and Lambert (1981) asked participants to write the English names of pictures (picture condition), translate words in L2 (translated condition), and write presented words in L1 (copied condition). Their findings showed that the performance on the picture condition was higher than that on the translated and copied conditions during a free recall phase. Matsumi (1994) also reported an effect of picture superiority, without using a pictorial stimulus, but asking participants to imagine objects or events of given words while translating or copying given words (imagined condition) or just to translate or copy given words (non-imagined condition). The results showed that participants could present words in the imagined condition more than in the non-imagined condition. Kambara and colleagues reported that task performance (accuracy rates and response times in recognition tasks) on associative learning of unfamiliar words and meaningless pictures was better than that on associative learning of unfamiliar words and meaningless sounds (Kambara et al., 2013). Taken together, these findings suggest that pictures as referents could promote recognition and learning condition retrieval speeds for associative pairs of words in L2 and referents.

On the other hand, the accuracy rates on learning condition retrieval and recognition tasks did not support the picture superiority effects. One reason accuracy rates did not involve picture superiority effects could be the high accuracy rates (more than 0.90) on both CP and CJ conditions in recognition tasks (a ceiling effect). In addition, the logographic Chinese characters (hanzi) might have looked similar to the logographic Japanese characters (kanji) to the participants, which might have affected the ceiling effect of the accuracy rates in the CP and CJ conditions, even though we did not use the same or similar words in Chinese and Japanese to avoid similarity effects between words in L1 and L2 (e.g., Paivio and Labmert, 1981). Thus, adjusting the difficulty of the task (the number of stimuli and stimulus selection), the results of the accuracy rates might be consistent with the response times. Future studies should consider adjusting the task in terms of difficulty.

In addition, the response times on the second learning condition retrieval task did not also support the picture superiority effects. In the CJ condition, continuously associative

learning of L2 and L1 words might promote directly associations between L2 and the referents. Kroll and colleagues proposed that the translation from L2 to L1 would be lexically associated until bilinguals acquire the skill for directly associating L2 with the referent (Kroll et al., 2010). Therefore, if participants perform more than three times of the recognition tasks, there might be also no significant difference between response times of the CP and CJ conditions in the recognition tasks.

Our findings supported Paivio's dual coding theory, in which a word and the referent (visual, auditory, haptic, olfactory, and gustatory features) are individually learned and associated (Paivio, 1986). In the dual coding theory, the referents are shared between words in the first and second languages (Paivio and Desrochers, 1980; Paivio, 1986). This referential sharing between the first and second language implies that the associative learning of pairs of Chinese words and Japanese words would indirectly associate with the referents, whereas the associative learning of pairs of Chinese words and pictures would directly associate with the referents. In the revised hierarchical model, Kroll and colleagues have proposed that the translation from L2 to L1 ensures a faster access to the referent than the translation from L1 to L2 (Kroll et al., 2010). The revised hierarchical model means that the backward translation from L2 to L1 would be lexically associated and the forward translation from L1 to L2 would be semantically mediated until bilinguals acquire the skill for directly associating L2 with the referent such as strong associations between L1 and referents (Kroll et al., 2010). Hence, our findings suggest that the associative learning of words in L2 and referents is more effective than the associative learning of words in L2 and words in L1, even if the L2 learners are beginners of L2.

Differences Between Associative Learning of Words in L2 and Referents and Learning of Words in L2 Only

Although there was no significant difference among the accuracy rates of conditions in the learning condition retrieval task, response times on the C condition were significantly faster than those on the CP and CJ conditions. These findings indicate that if participants learn both associative pairs of words in L2 and referents and words in L2 only, participants could more easily detect words in L2 than associative pairs. These results might also mean that divided attention would affect retrieval of items (L2 words in this study) and sources (referents in this study) more than retrieval of items only (e.g., Troyer et al., 1999).

In addition, we found that missed trials in the C condition were judged to be unlearned trials more than learned trials in the first and second learning condition retrieval tasks, whereas missed trials in CP and CJ conditions were judged to be learned trials more than unlearned trials in the second learning condition retrieval task. These findings suggest that missed trials in associative conditions (CP and CJ conditions) could be judged to be learned trials more than unlearned trials, while missed trials in the non-associative condition (C condition) could not. The differences would be affected by the relationship between the familiarity and

recollection of words and referents (Yonelinas, 2002; Diana et al., 2007; Wixted, 2007). Associative conditions, in which words associate with referents, would involve both familiarity and recollection (i.e., referents) of words. Participants would retrieve the familiarity of words even if they could not retrieve the recollection of words. In contrast, non-associative conditions would only involve familiarity with words, and so when participants could not retrieve the familiarity of words, they could not judge them as learned words.

Future Directions

This study has limitations to its experimental methods. First, all participants were native speakers of Japanese and university students. Future studies would also need to examine native speakers of other languages and people in other age groups. Second, the results of this study might only show associative learning for one day, since participants conducted all the tasks on the same day. To investigate the effects of memory consolidation or decay, researchers would also need to conduct experiments for more than two days (Kambara et al., 2013). Third, the associative learning of words and referents might be affected by the sound-symbolic effects between words and referents. For example, when participants judge whether the name of a presented meaningless round figure is *bouba* (*malma*) or *kiki* (*takete*), they consistently choose *bouba* as the name of the round figure (*bouba-kiki* effects; Köhler, 1947; Ramachandran and Hubbard 2001; Westbury, 2005; Aryani et al., 2020). In addition, each sound-symbolic word, each written vowel, or each mouth shape to pronounce a vowel is associated with specific, subjective evaluations (perceptual imageabilities and emotional features; e.g., Namba and Kambara, 2020; Ando et al., 2021; Kambara and Umemura, 2021). Future studies should consider the sound-symbolic effects on the associative learning of words and referents. Fourth, since the maintenance and retrieval (scanning) of verbal information (verbal working memory; e.g., Kambara et al., 2017; Kambara et al., 2018) would be associated with new word learning (Gathercole and Baddeley, 1989; Gathercole and Baddeley, 1990; Horinouchi et al., under review), future studies should examine the relationship between the verbal working memory and associative learning of L2 words and referents. Finally, we did not collect the information concerning the educational background of the participants in this study. People in Japan have learned English as a second language at least from their junior high school. Future studies should consider the educational background of the participants about their language learning, and also examine the effects of their language learning experiences on the learning of a new language (e.g., Potter et al., 2017).

CONCLUSION

We conducted an experiment to examine differences in task performance (accuracy rates and response times) for CP (associative pairs of L2 words and pictures), CJ (associative pairs of L2 words and L1 words), and C conditions (L2 words only) in recognition and learning condition retrieval tasks. Three main results were obtained. First, the accuracy rate on recognition

and learning condition retrieval tasks increased in each condition, whereas no significant difference was shown among CP, CJ, and C. Second, the response time of CP was faster than that of CJ in all the recognition tasks and the first learning condition retrieval task. Third, in the second learning condition retrieval task, missed items in associative conditions (CP and CJ conditions) were judged as learned items more than unlearned items, whereas missed items in the non-associative condition (C condition) were judged as unlearned items more than learned items. In sum, these findings suggest that picture superiority effects could facilitate recognition and retrieval speeds of associative pairs of word forms in L2 and referents, and that associative learning of L2 words and referents could enhance more familiarity effects than learning of L2 words only.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the dataset includes personal information. Requests to access the datasets should be directed to Toshimune Kambara, tkambara@hiroshima-u.ac.jp.

ETHIC STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Graduate School of Education at Hiroshima University. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

REFERENCES

Akamatsu, N. (2002). A similarity in word-recognition procedures among second language readers with different first language backgrounds. *Appl. Psycholinguist.* 23, 117–133. doi:10.1017/S0142716402000061

Ando, M., Liu, X., Yan, Y., Yutao, Y., Namba, S., Abe, K., et al. (2021). Sound-symbolic semantics of written Japanese vowels in a paper-based survey study. *Front. Commun.* 6, 617532. doi:10.3389/fcomm.2021.617532

Aryani, A., Isbilen, E. S., and Christiansen, M. H. (2020). Affective arousal links sound to meaning. *Psychol. Sci.* 31 (8), 978–986. doi:10.1177/0956797620927967

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. J. Mem. Lang. 59 (4), 390–412.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Front. Psychol.* 4, 328. doi:10.3389/fpsyg.2013.00328

Bates, D. M. (2005). Fitting linear mixed models in. *R. R. News* 5, 27–30. Retrieved from https://cran.opencpu.org/doc/Rnews/Rnews_2005-1.pdf#page=27

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models Usinglme4. *J. Stat. Softw.* 67 (1), 1–48. doi:10.18637/jss.v067.i01

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A. F., Sommer, J., Wolbers, T., et al. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage* 25 (3), 958–968. doi:10.1016/j.neuroimage.2004.12.019

Carpenter, S. K., and Olson, K. M. (2012). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *J. Exp. Psychol. Learn. Mem. Cogn.* 38 (1), 92–101. doi:10.1037/a0024828

Chen, T., Koda, K., and Wiener, S. (2020). Word-meaning inference in L2 Chinese: an interactive effect of learners' linguistic knowledge and words' semantic transparency. *Read. Writ.* 33, 2639–2660. doi:10.1007/s11145-020-10058-w

Cicero, S. D., and Tryon, W. W. (1989). Classical conditioning of meaning--II. A replication and triplet associative extension. *J. Behav. Ther. Exp. Psychiatry* 20 (3), 197–202. doi:10.1016/0005-7916(89)90023-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum.

Cornelissen, K., Laine, M., Renvall, K., Saarinen, T., Martin, N., and Salmelin, R. (2004). Learning new names for new objects: cortical effects as measured by magnetoencephalography. *Brain Lang.* 89 (3), 617–622. doi:10.1016/j.bandl.2003.12.007

Diana, R. A., Yonelinas, A. P., and Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci.* 11 (9), 379–386. doi:10.1016/j.tics.2007.08.001

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., et al. (2018). MultiPic: a standardized set of 750 drawings with norms for six European languages. *Q. J. Exp. Psychol.* 71 (4), 808–816. doi:10.1080/17470218.2017.1310261

Emirmustafaoğlu, A., and Gökmen, D. U. (2015). The Effects of Picture vs. Translation Mediated Instruction on L2 Vocabulary Learning. Proc. Soc. Behav. Sci. 199, 357–362. doi:10.1016/j.sbspro.2015.07.559

Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: a general power analysis program. *Behav. Res. Methods Instr. Comput.* 28 (1), 1–11. doi:10.3758/BF03203630

Ferreira, R. A., Göbel, S. M., Hymers, M., and Ellis, A. W. (2015). The neural correlates of semantic richness: evidence from an fMRI study of word learning. *Brain Lang.* 143, 69–80. doi:10.1016/j.bandl.2015.02.005

Gathercole, S. E., and Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *J. Mem. Lang.* 28 (2), 200–213. doi:10.1016/0749-596X(89)90044-2

Gathercole, S. E., and Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *J. Mem. Lang.* 29 (3), 336–360. doi:10.1016/0749-596X(90)90004-J

Grönholm, P., Rinne, J. O., Vorobyev, V., and Laine, M. (2015). Naming of newly learned objects: a PET activation study. *Brain Res. Cogn. Brain Res.* 25 (1), 359–371. doi:10.1016/j.cogbrainres.2005.06.010

Havas, V., Taylor, J., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., and Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Q. J. Exp. Psychol.* 71 (6), 1469–1481. doi:10.1080/17470218.2017.1329325

Hawkins, E., Astle, D. E., and Rastle, K. (2015). Semantic advantage for learning new phonological form representations. *J. Cogn. Neurosci.* 27, 775–786. doi:10.1162/jocn_a_00730

Hawkins, E. A., and Rastle, K. (2016). How does the provision of semantic information influence the lexicalization of new spoken words? *Q. J. Exp. Psychol.* 69, 1322–1339. doi:10.1080/17470218.2015.1079226

Horinouchi, H., Liu, X., Kabir, R. S., Kobayashi, R., Haramaki, Y., and Kambara, T. Associative pairs of a new word form and single referent are learned better than associative pairs of a new word form and multiple referents in a single-day learning context. *J. Psycholing. Res.* (under review).

Hughes, S., Barnes-Holmes, D., Van Dessel, P., de Almeida, J. H., Stewart, I., and De Houwer, J. (2018). On the symbolic generalization of likes and dislikes. *J. Exp. Soc. Psychol.* 79, 365–377. doi:10.1016/j.jesp.2018.09.002

Hultén, A., Vihla, M., Laine, M., and Salmelin, R. (2009). Accessing newly learned names and meanings in the native language. *Hum. Brain Mapp.* 30, 976–989. doi:10.1002/hbm.20561

Jeong, H., Sugiura, M., Sassa, Y., Wakusawa, K., Horie, K., Sato, S., et al. (2010). Learning second language vocabulary: neural dissociation of situation-based learning and text-based learning. *Neuroimage* 50, 802–809. doi:10.1016/j.neuroimage.2009.12.038

Kambara, T., Brown, E. C., Jeong, J. W., Ofen, N., Nakai, Y., and Asano, E. (2017). Spatio-temporal dynamics of working memory maintenance and scanning of verbal information. *Clin. Neurophysiol.* 128 (6), 882–891. doi:10.1016/j.clinph.2017.03.005

Kambara, T., Brown, E. C., Silverstein, B. H., Nakai, Y., and Asano, E. (2018). Neural dynamics of verbal working memory in auditory description naming. *Sci. Rep.* 8, 15868. doi:10.1038/s41598-018-33776-2

Kambara, T., Tsukiura, T., Shigemune, Y., Kanno, A., Nouchi, R., Yomogida, Y., et al. (2013). Learning-dependent changes of associations between unfamiliar words and perceptual features: a 15-day longitudinal study. *Lang. Sci.* 35, 80–86. doi:10.1016/j.langsci.2012.05.001

Kambara, T., and Umemura, T. (2021). The relationships between initial consonants in Japanese sound symbolic words and familiarity, multi-sensory imageability, emotional valence, and arousal. *J. Psycholinguist Res.* doi:10.1007/s10936-020-09749-w

Kambara, T., Umemura, T., Ackert, M., and Yang, Y. (2020). The relationship between psycholinguistic features of religious words and core dimensions of religiosity: a survey study with Japanese participants. *Religions* 11 (12), 673. doi:10.3390/rel11120673

Köhler, W. (1947). *Gestalt psychology*. New York, NY: Liveright.

Kost, C. R., Foss, P., and Lenzini, J. J., Jr (1999). Textual and pictorial glosses: effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Lang. Ann.* 32 (1), 89–97. doi:10.1111/j.1944-9720.1999.tb02378.x

Kroll, J. F., van Hell, J. G., Tokowicz, N., and Green, D. W. (2010). The revised hierarchical model: a critical review and assessment. *Biling. (Camb Engl)* 13 (3), 373–381. doi:10.1017/S136672891000009X

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82 (13), 1–26. doi:10.18637/jss.v082.i13

Lee, H. S., Fujii, T., Okuda, J., Tsukiura, T., Umetsu, A., Suzuki, M., et al. (2003). Changes in brain activation patterns associated with learning of Korean words by Japanese: an fMRI study. *Neuroimage* 20 (1), 1–11. doi:10.1016/S1053-8119(03)00254-4

Lenth, R. V. (2016). Least-squares means: The R Package lsmeans. *J. Stat. Softw.* 69, 1–33. doi:10.18637/jss.v069.i01

Li, C., Fan, L., and Wang, B. (2020). Post-encoding positive emotion impairs associative memory for English vocabulary. *PLoS One* 15 (4), e0228614. doi:10.1371/journal.pone.0228614

Matsumi, N. (1994). Processes of words memory in second language acquisition: a test of bilingual dual coding theory. *Jpn. J. Psychol.* 64 (6), 460–468. doi:10.4992/jjpsy.64.460

Morett, L. M. (2019). The power of an image: images, not glosses, enhance learning of concrete L2 words in beginning learners. *J. Psycholinguist. Res.* 48 (3), 643–664. doi:10.1007/s10936-018-9623-2

Muljani, D., Koda, K., and Moates, D. R. (1998). The development of word recognition in a second language. *Appl. Psycholinguist.* 19, 99–113. doi:10.1017/S0142716400010602

Namba, S., and Kambara, T. (2020). Semantics based on the physical characteristics of facial expressions used to produce Japanese vowels. *Behav. Sci.* 10 (10), 157. doi:10.3390/bs10100157

Paivio, A. (1964). Generalization of verbally conditioned meaning from symbol to referent. *Can. J. Psychol.* 18 (2), 146–155. doi:10.1037/h0083289

Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: a dual coding approach*. Oxford, United Kingdom: Oxford University Press.

Paivio, A. (2007). *Mind and its evolution: a dual coding theoretical approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Paivio, A., and Csapo, K. (1973). Picture superiority in free recall: imagery or dual coding? *Cogn. Psychol.* 5 (2), 176–206. doi:10.1016/0010-0285(73)90032-7

Paivio, A., and Desrochers, A. (1980). A dual-coding approach to bilingual memory. *Can. J. Psychol./Revue canadienne de Psychol.* 34, 388–399. doi:10.1037/h0081101

Paivio, A., and Lambert, W. (1981). Dual coding and bilingual memory. *J. Verbal Learn. Verbal Behav.* 20, 532–539. doi:10.1016/S0022-5371(81)90156-0

Potter, C. E., Wang, T., and Saffran, J. R. (2017). Second language experience facilitates statistical learning of novel linguistic materials. *Cogn. Sci.* 41, 913–927. doi:10.1111/cogs.12473

Potter, M. C., So, K.-F., Eckardt, B. V., and Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *J. Verbal Learn. Verbal Behav.* 23, 23–38. doi:10.1016/S0022-5371(84)90489-4

Prior, A., Goldina, A., Shany, M., Geva, E., and Katzir, T. (2014). Lexical inference in L2: predictive roles of vocabulary knowledge and reading skill beyond reading comprehension. *Read. Writ.* 27, 1467–1484. doi:10.1007/s11145-014-9501-8

Pulvermüller, F. (2003). *The neuroscience of language: on brain circuits of words and serial order*. New York, NY: Cambridge University Press.

R Core Team (2019). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ramachandran, V. S., and Hubbard, E. M. (2001). Synaesthesia: a window into perception, thought, and language. *J. Consc. Stud.* 8 (12), 3–34. Retrieved from https://www.ingentaconnect.com/content/imp/jcs/2001/00000008/00000012/1244

Razran, G. H. S. (1939). A quantitative study of meaning by a conditioned salivary technique (semantic conditioning). *Science* 90, 89–90. doi:10.1126/science.90.2326.89-a

Rokni, S. J. A., and Karimi, N. (2013). Visual instruction: an advantage or a disadvantage? what about its effect on EFL learners' vocabulary learning. *As. J. Soc. Sci. Hum.* 2 (4), 236–243. Retrieved from http://www.ajssh.leena-luna.co.jp/AJSSHPDFs/Vol.2(4)/AJSSH2013(2.4-23).pdf

Staats, A. W., and Staats, C. K. (1958). Attitudes established by classical conditioning. *J. Abnorm. Psychol.* 57 (1), 37–40. doi:10.1037/h0042782

Staats, A. W., Staats, C. K., and Heard, W. G. (1959a). Language conditioning of meaning to meaning using a semantic generalization paradigm. *J. Exp. Psychol.* 57 (3), 187–192. doi:10.1037/h0042274

Staats, A. W., Staats, C. K., and Heard, W. G. (1961). Denotative meaning established by classical conditioning. *J. Exp. Psychol.* 61 (4), 300–303. doi:10.1037/h0046306

Staats, A. W., Staats, C. K., Heard, W. G., and Nims, L. P. (1959b). Replication report: meaning established by classical conditioning. *J. Exp. Psychol.* 57 (1), 64. doi:10.1037/h0048859

Staats, C. K., and Staats, A. W. (1957). Meaning established by classical conditioning. *J. Exp. Psychol.* 54 (1), 74–80. doi:10.1037/h0047716

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., and McQueen, J. M. (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *Neuroimage* 84, 265–278. doi:10.1016/j.neuroimage.2013.08.023

Takashima, A., Bakker, I., van Hell, J. G., Janzen, G., and McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain Lang.* 167, 44–60. doi:10.1016/j.bandl.2016.05.009

Till, B. D., and Priluck, R. L. (2001). Conditioning of meaning in advertising: brand gender perception effects. *J. Curr. Issues Res. Advert.* 23 (2), 1–8. doi:10.1080/10641734.2001.10505116

Troyer, A. K., Winocur, G., Craik, F. I., and Moscovitch, M. (1999). Source memory and divided attention: reciprocal costs to primary and secondary tasks. *Neuropsychology* 13 (4), 467–474. doi:10.1037/0894-4105.13.4.467

Tryon, W. W., and Cicero, S. D. (1989). Classical conditioning of meaning–I. A replication and higher-order extension. *J. Behav. Ther. Exp. Psychiatry* 20 (2), 137–142. doi:10.1016/0005-7916(89)90046-3

Tsukiura, T., Mano, Y., Sekiguchi, A., Yomogida, Y., Hoshi, K., Kambara, T., et al. (2010). Dissociable roles of the anterior temporal regions in successful encoding of memory for person identity information. *J. Cogn. Neurosci.* 22 (10), 2226–2237. doi:10.1162/jocn.2009.21349

Tsukiura, T., Sekiguchi, A., Yomogida, Y., Nakagawa, S., Shigemune, Y., Kambara, T., et al. (2011). Effects of aging on hippocampal and anterior temporal activations during successful retrieval of memory for face-name associations. *J. Cogn. Neurosci.* 23 (1), 200–213. doi:10.1162/jocn.2010.21476

Tulving, E., and Thomson, D. M. (1971). Retrieval processes in recognition memory: effects of associative context. *J. Exp. Psychol.* 87 (1), 116–124. doi:10.1037/h0030186

Wang, S., and Lee, C. I. (2021). Multimedia gloss presentation: learners' preference and the effects on EFL vocabulary learning and reading comprehension. *Front. Psychol.* 11, 602520. doi:10.3389/fpsyg.2020.602520

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain Lang.* 93, 10–19. doi:10.1016/j.bandl.2004.07.006

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychol. Rev.* 114 (1), 152–176. doi:10.1037/0033-295X.114.1.152

Yan, Y., Yang, Y., Ando, M., Liu, X., and Kambara, T. Associative learning of new word forms in a first language and gustatory stimuli: a behavioral experiment. *Anal. Verbal. Behav.* (under review).

Yang, Y., Yan, Y., Ando, M., Liu, X., and Kambara, T. Associative learning of new word forms in a first language and haptic features in a single-day experiment. *Biosemiotics.* (under review).

Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *J. Mem. Lang.* 46 (3), 441–517. doi:10.1006/jmla.2002.2864