LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Novel Word Learning With Verbal Definitions and Images: Tracking Consolidation With Behavioral and Event-Related Potential Measures

Daisy Lei, Yushuang Liu, and Janet G. van Hell

The Pennsylvania State University

Abstract: We examined the impact of images on novel word learning and consolidation, in a conceptual replication of Liu and Van Hell (2020). After participants had learned one set of novel words with definitions and images on Day 1 (remote words) and a different set on Day 2 (recent words), they judged the semantic relatedness of word pairs on Days 2 and 8 while event-related potentials (ERPs) were recorded. Day 2 ERPs showed that remote, but not recent, novel words elicited a late positive component. By Day 8, both remote and recent novel words elicited a late positive component. We observed no N400 on either day. Comparing these learners (definition-image group) with learners trained with definitions only (using data from Liu & Van Hell, 2020) revealed that the groups' ERP patterns did not differ, but definition recall and relatedness judgment performances were higher for the definition-image group than for the definition-only group. Learning novel word meanings through definitions and images strengthened behavioral outcomes but did not affect ERP signatures of learning and consolidation.

Keywords word learning; offline consolidation; semantic integration; event-related potentials (ERPs)

This research was supported by NSF GRFP Grant No. DGE1255832 to Daisy Lei and NSF BCS-1349110 and REU supplement and NSF OISE 1545900 to Janet G. van Hell. The authors would like to thank Amber Liu for creating the novel word images and Dylan Solano, Mayra Pollo, Jaclyn Yuro, and Sarah Paterno for helping with data collection.

Correspondence concerning this article should be addressed to Daisy Lei, Department of Psychology and Center for Language Science, The Pennsylvania State University, Moore Building, University Park, PA 16802. E-mail: daisylei@psu.edu

The handling editor for this article was Kara Morgan-Short.

DOI: 10.1111/lang.12502

Introduction

The complementary learning systems (CLS) account of word learning (Davis & Gaskell, 2009), adapted from the CLS theory of learning and memory (McClelland, McNaughton, & O'Reilly, 1995; for a recent update, see Kumaran, Hassabis, & McClelland, 2016), hypothesizes that two neural systems are involved in word learning. The first system involves the hippocampus, with novel words encoded as episodic memory traces immediately after learning. The second system involves the neocortex such that newly learned novel words are gradually integrated into learners' existing lexicon and become part of their long-term memory. Novel words are thus thought to undergo a postlearning consolidation process during which hippocampal connections decay and neocortical connections are strengthened.

According to the CLS model of word learning, during the initial stages of learning learners can immediately recall information that they have just learned. However, because these newly learned words are encoded as episodic memory traces in the hippocampus and are not yet encoded in the neocortex and integrated into the lexicon, they do not interact with existing words in learners' mental lexical network. Only after a period of offline consolidation do these episodic memory traces become consolidated into memories encoded in the neocortex and integrated into the lexicon or are lexicalized (e.g., Bakker, Takashima, Van Hell, Janzen, & McQueen, 2014; 2015a, 2015b; Bakker-Marshall et al., 2018; Born & Wilhelm, 2012; Dumay & Gaskell, 2007; Gais, Lucas, & Born, 2006; Stickgold & Walker, 2007; for a review of behavioral studies, see Palma & Titone, 2020). To investigate the role of images on the learning and consolidation of novel word meanings, we trained our participants on a set of 40 novel words paired with novel meanings and corresponding images (20 on Day 1, 20 on Day 2) and measured behavioral and neural correlates of consolidation on Day 2 and Day 8.

Background Literature

Learning Novel Word Meanings: Verbal Definition

Pairing novel words with meaning during training has been shown to support novel word learning in studies using electroencephalography (EEG)/event-related potentials (ERPs; e.g., Angwin, Phua, & Copland, 2014; Balass, Nelson, & Perfetti, 2010) and functional magnetic resonance imaging (fMRI) technologies (e.g., Takashima, Bakker, Van Hell, Janzen, & McQueen, 2017; for reviews of behavioral studies, see Palma & Titone, 2020; Rice & Tokowicz, 2020). Some word learning studies (e.g., Angwin et al., 2014; Dobel et al., 2010) paired novel words with meaningful but existing concepts. Training novel words with an existing concept is analogous to learning a new label for an existing concept in a second language (L2). In this study, we focused on learning novel words with novel meanings, which more closely reflects word learning in the native language.

In one of the first EEG/ERP studies to examine the learning and overnight consolidation of novel word meanings, Bakker et al. (2015b) first trained Dutch participants auditorily on two lists of pseudowords (created derivationally from real Dutch words) with definitions, one list on each day such that, by the end of training on the second day, participants were trained on two lists of words with different offline consolidation periods (24 hours and none).

ERPs germane to word learning and lexical sensitivity include the N400 component and the late positive component (LPC). The N400 component is a negative-going waveform that peaks around 400 ms after stimulus onset, and N400 effects have been associated with more automatic processes of semantic activation in the sense that word meanings are invariably activated and do not require the level of awareness needed for controlled processing (for a discussion, see Federmeier, 2021; Kutas & Federmeier, 2011). In particular, in a semantic priming task, a target word following an unrelated prime word typically elicits a more negative N400 response than when a target word follows a semantically related prime word. Also, nonwords have been shown to elicit a more negative-going N400 than do real words (Holcomb, Grainger, & O'Rourke, 2002).

The LPC is a positive-going waveform that peaks in the 500–700 ms window after stimulus onset in a semantic priming task (e.g., Bakker et al., 2015b, Kandhadai & Federmeier, 2010). The LPC effect has been observed to follow the N400 effect in monolingual speakers and Spanish–English bilingual speakers in a semantic priming task with known existing English words (Hoshino & Thierry, 2012). In particular, the LPC has been associated with explicit memory and a more controlled, conscious, and strategic process of semantic retrieval (e.g., Batterink & Neville, 2011; Hoshino & Thierry, 2012; Liu & Van Hell, 2020; Rohaut et al., 2015); in the recognition memory literature the more frontal LPC has been associated with fast-acting, relatively automatic recognition of information and the parietal LPC with slower, more effortful and conscious recollection of previously presented information (e.g., Rugg & Curran, 2007).

Bakker et al. (2015b) found a reliable centro-parietal LPC priming effect for novel words learned on Day 1 (with a 24-hour consolidation period) and a frontal LPC priming effect for words learned on Day 2. They observed no N400 semantic priming effects. Bakker et al. concluded that novel word meanings can contribute to semantic processing shortly after learning but that word meaning retrieval remains more controlled and strategic (indexed by the presence of an LPC but absence of an N400 semantic priming effect). They further interpreted the shift from frontal to parietal LPC effects to indicate that a qualitative change in lexical-semantic representation takes place during consolidation and that the neural processes supporting novel word meaning retrieval may shift from strategic to more automatic.

Bakker et al.'s (2015b) participants were Dutch undergraduate students who were all bilingual or even multilingual by virtue of the Dutch educational system. To examine the role of prior language learning experience, Liu and Van Hell (2020) tested novel word learning in monolingual native-English participants. They also investigated a longer offline consolidation period (seven days after Day 1) and tested participants on both Day 2 and Day 8. As in Bakker et al. (2015b), Liu and Van Hell (2020) trained their participants on two sets of novel words, one set on Day 1 (remote condition) and one set on Day 2 (recent condition), such that on Day 2, remote novel words had undergone 24 hours of offline consolidation. Each word was paired with a definition. Testing on Day 2 yielded an LPC semantic priming effect for remote but not for recent novel words. ERPs on Day 8 revealed LPC effects for both remote and recent novel words. Similar to Bakker et al. (2015b), Liu and Van Hell (2020) observed no N400 effects on either day, indicating that both after 24-hours and after one-week, learning word meanings was not invariably activated and that word meaning retrieval remained more controlled and strategic. Although Bakker et al.'s (2015b) experienced learners showed a semantic priming LPC effect for novel words immediately after learning, Liu and Van Hell's (2020) inexperienced learners demonstrated an LPC semantic priming effect only after a 24-hour (and one-week) offline consolidation period. The combined findings suggested that lexicalization of novel word meaning is a gradual and protracted process, and this gradual change occurs at a slower pace in inexperienced than in experienced learners.

The studies discussed above presented novel word meanings in a definition or verbal context. However, word meaning may also be represented nonverbally through images. The dual coding theory postulates two independent but interconnected cognitive subsystems: verbal and nonverbal (Paivio, 2014). The verbal system involves language; the nonverbal system involves nonlinguistic objects and events. Specifically, the verbal system represents the world indirectly with language symbols such as written, spoken, or signed words, but the nonverbal system represents the world directly with modality-specific instances (e.g., mental image, environmental sound, actions, emotions). Studies have shown that memory for blacklined drawings and colorful pictures is better than for words alone (e.g., Grady, McIntosh, Rajah, & Craik, 1998; Leach & Samuel, 2007). Extending Liu and Van Hell (2020), the present study examined whether pairing novel words with novel definitions and images strengthens encoding and consolidation over time.

Learning Novel Word Meanings: Visualization

Behavioral studies have shown that the addition of images increases students' retention of words (Smith, Stahl, & Neil, 1987; Smith, Miller, Grossman, & Valeri-Gold, 1994). For example, Smith et al. (1987) presented students with new words in either a definition-only condition, a definition-sentence condition, or a definition-sentence-image condition. The sentences used the words in a meaningful manner, and the images depicted what the sentences conveyed. Students were instructed to copy the material that they were shown. The results showed that the definition-sentence-image group outperformed the definition-only group in a two-week delayed posttest.

Neurocognitive studies have also examined word learning in the context of providing more information with visualizations (e.g., Bermúdez-Margaretto, Beltrán, Cuetos, & Domínguez, 2018; Kaczer et al., 2018; Takashima, Bakker, Van Hell, Janzen, & McQueen, 2014; Takashima et al., 2017). For example, in an fMRI study, Takashima et al. (2017) investigated the neural representations and consolidation trajectories of novel words encoded with meaning (audio with picture or audio with definition) and without meaning (audio only). Their stimuli consisted of Japanese objects not familiar to the Dutch participants as novel words and pictures. Participants completed a recognition task in the MRI scanner on Day 1 (immediately after training) and on Day 8. Behaviorally, free recall was better with meaningful words than with words without meaning. However, there was no systematic behavioral recall performance difference between novel words learned in the two meaning conditions (picture or definition). fMRI results showed hippocampal activation on Day 1 and on Day 8, although less prominently so on Day 8. On Day 8, neocortical activation was significantly greater for the meaningful words (picture or definition) than for audio-only words.

The Present Study

To our knowledge, Kaczer et al.'s (2018) study is the only other EEG/ERP study that has examined learning novel words and novel meanings through verbal (text) and visual information. However, Kazer et al.'s verbalizations were compound nouns or adjective-noun pairs referring to existing concepts

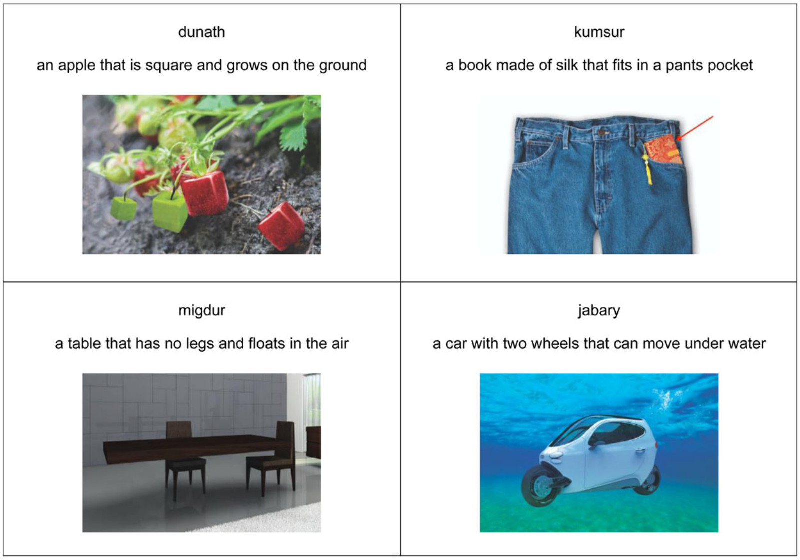

Figure 1 Examples of novel words and their corresponding novel definition and image.

(e.g., *flower pot*, *modern sofa*, *feet cushion*) and the images were single-object black-and-white drawings, whereas we used nonexisting concepts and colorful images. Our study paired novel words with novel definitions (object paired with two unconventional features) and colorful images visualizing these definitions during learning (see Figure 1), thereby extending Liu and Van Hell (2020), who studied learning novel meaning through definitions only. Focusing on the neural correlates of consolidation, we asked two research questions:

1. After novel words are learned with not only definitions but also images, to what extent will novel word meaning consolidation be strengthened such that the activation of novel word meaning achieves more automaticity as indexed by a N400 effect one day and seven days after learning?
2. To what extent does learning novel words with not only definitions but also images significantly improve learning (behaviorally measured) and strengthen the consolidation (neurocognitively measured) of novel words compared to novel words paired with only definitions?

To address the second, overarching question, we compared our behavioral and ERP data directly with Liu and Van Hell's (2020) data to examine the impact of novel word learning with the addition of visualizations (i.e., definition-image) compared to definition alone (i.e., definition-only).

Predictions

Dual coding theory (Paivio, 2014) posits that there are two cognitive subsystems (verbal and nonverbal) involved in representing information. Hence, presenting learners with images and definitions provides two avenues for encoding and representing novel word meaning, and these additional visualizations are predicted to enhance encoding and consolidation. With respect to our first research question, for ERP results for Day 2, we predicted that novel words learned on Day 1 would elicit LPC and N400 semantic priming effects in the definition-image group. Possibly, even words learned on Day 2 might already demonstrate a semantic priming effect on Day 2, and if this were to occur, this most likely would emerge as an LPC effect (reflecting more controlled, conscious semantic retrieval). For Day 8, if combined visual and verbal information indeed enhances encoding and consolidation of novel word meaning as predicted by the dual coding theory, we expected words learned on Day 1 and Day 2 to elicit N400 (and LPC) effects.

In Liu and Van Hell's (2020) study, novel words associated with verbal information elicited LPC effects on Day 2 only for remote novel words and on Day 8 for both remote and recent novel words. Furthermore, with respect to the overarching question comparing the definition-image and the definition-only groups, the dual coding theory predicts better performance on the behavioral definition recall, recognition, and semantic relatedness judgment tasks for a definition-image group than for a definition-only group and predicts that the neural consolidation effects (as indexed by the semantic priming effects) would differ between the two groups.

Alternatively, if visualizations do not enhance encoding and consolidation of novel word meaning, above and beyond verbal definitions, we predicted no differences for the behavioral performances and neural consolidation effects between the definition-image and definition-only groups. This predicted outcome would align with Takashima et al. (2017), who found no difference in neocortical activations between novel words trained with definitions and novel words trained with pictures.

Method

Participants

We recruited 33 monolingual native-English speaking undergraduate students (11 males, 22 females; $M_{\text{age}} = 19.03$ years, $SD = 1.03$, range: 18–21) from the same population that had participated in Liu and Van Hell's (2020) study. The participating students received course credits for their participation. The participants were self-reported right-handed individuals as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), had normal or corrected-to-normal vision (including normal color vision), and had no history of neurological or language-related disorders. We excluded one participant's data from data analysis because of technical issues with the EEG signal on Day 2. We excluded one participant's Day 2 data and seven participants' Day 8 data due to high artifact rejection rates (exclusion criteria: fewer than 20 trials [i.e., < 50%] in any condition). Two additional participants did not return on Day 8 for reasons unrelated to the experiment. Total remaining participant datasets were 31 for Day 2 and 23 for Day 8.

The participants completed a language experience and proficiency questionnaire developed in house (Lei, Liu, & Van Hell, 2022b). We measured proficiency on a 10-point scale (1 = *very low*, 10 = *perfect*) separately for speaking, understanding, reading, and writing. We averaged L2 or third language proficiency across these four domains. The participants had limited knowledge of other languages (self-reported L2 proficiency: $M = 4.08$, $SD = 2.26$; for five participants, third language proficiency: $M = 3.15$, $SD = 2.36$), and had minimal daily exposure to languages other than English ($M = 4.47\%$, $SD = 6.80$; scale = 0–100%). Seventeen participants reported little foreign language classroom experience—from one to three to four semesters.

We assessed the participants' sleep experience the night before Day 2 with post-sleep questionnaires administered on Day 2. The participants reported having regular sleep duration ($M = 7.23$ hr, $SD = 0.83$), and 74% reported having experienced normal or higher than usual sleep quality.

For the definition-only group (from Liu and Van Hell, 2020), the participants included 30 English native speakers ($M_{age} = 19.22$ years, $SD = 1.41$; six males, 23 females), all functionally monolingual right-handed undergraduate students recruited from the same population as the participants in the definition-image group. Analyses included 26 participants on Day 2 and 23 participants on Day 8 for the definition recall and semantic relatedness judgment tasks, 24 participants on Day 2 and 22 participants on Day 8 for the four-alternative-forced-choice (4AFC) word recognition task, and 23 participants on Day 2 and 22 participants on Day 8 for the ERP data. Additional details are reported by Liu and Van Hell (2020).

Stimulus Materials

We took 40 novel words and 40 existing high-frequency English words, with their respective definitions, from Liu and Van Hell (2020; Lei, Liu, & Van Hell, 2022a; Lei, Liu, & Van Hell, 2022c; see also Appendix S1 in the online Supporting Information). We selected novel words from Deacon, Dynowska, Ritter, and Grose-Fifer's (2004) list of nonderivational nonwords that are pronounceable and orthographically legal English nonwords, with mean word length of 5.78 letters (range: 4–7 letters), and no orthographic neighbors according to the CLEARPOND database (Marian, Bartolotti, Chabal, & Shook, 2012). Existing English words consisted of common nouns (e.g., *wheel*). We used the existing English words as control stimuli. As in Liu and Van Hell (2020), we paired novel words with novel definitions: an existing object (e.g., *a car*) as the base for the novel word (e.g., *jabary*) paired with two unconventional features (e.g., two wheels and moves under water). We randomized the 40 novel and 40 existing words and divided them into two lists to be learned on either Day 1 or Day 2. The four lists did not differ in word length, $F(3, 76) = 0.61$, $p = .61$, and the two lists of existing words did not differ in lexical frequency, $t(38) = 0.20$, $p = .84$ (see Liu & Van Hell, 2020). An image accompanied each word and definition (see Figure 1 and Appendix S1 in the online Supporting Information). An undergraduate research assistant majoring in graphic design created the visualizations by editing multiple images to visually represent the novel definition. We found images of existing words online.

Prior to the experiment, we conducted an image-definition rating study. A separate group of 25 undergraduate students (from the same population as the students in the experiment) rated how well each image matched the corresponding definition on a 5-point Likert scale (1 = *does not match*, 5 = *complete match*). Stimuli included 80 pairs of matching definitions and images, corresponding to the 80 critical novel and existing words (congruent items). To ensure that raters used the full scale, we added 40 randomized visualization and definition pairs (38 incongruent items, 2 randomly congruent items). The mean image-definition match rating was significantly higher, $t(83) = 48.65$, $p < .001$, for the aggregated matching pairs (all congruent items: $M = 4.64$, $SD = 0.67$, 95% CI [4.50, 4.79]; existing words: $M = 4.75$, $SD = 0.43$, 95% CI [4.61, 4.88]; novel words: $M = 4.55$, $SD = 0.84$, 95% CI [4.29, 4.81]) than for the aggregated nonmatching pairs (all incongruent items: $M = 1.02$, $SD = 0.05$, 95% CI [1.00, 1.03]; existing words: $M = 1.01$, $SD = 0.04$, 95% CI [.99, 1.03]; novel words: $M = 1.03$, $SD = 0.05$, 95% CI [1.00, 1.05]).

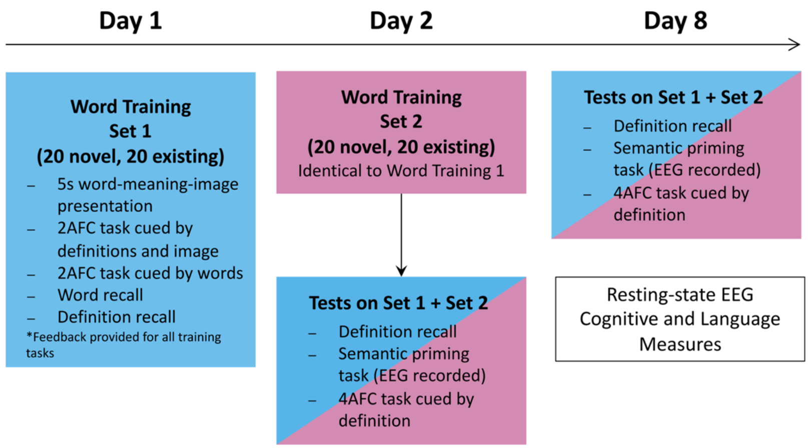

Figure 2 Study design. Tasks for each session are listed. The color(s) of each box indicate(s) the word training condition. Blue indicates Wordlist 1 (remote words); purple indicates Wordlist 2 (recent words). The white box lists tasks unrelated to the trained words. 2AFC = two-alternative-forced-choice; 4AFC = four-alternative-forced-choice; EEG = electroencephalography.

The participants also rated the visual complexity (1 = *visually simple*; 5 = *visually complex*) and familiarity (1 = *not familiar*; 5 = *very familiar*) of all 80 images on a 5-point Likert scale. Visual complexity ratings for novel words ($M = 3.50$, $SD = 0.46$, 95% CI [3.36, 3.65]) and existing words ($M = 3.30$, $SD = 0.67$, 95% CI [3.11, 3.54]) did not differ significantly, $t(39) = 1.42$, $p = .16$. As we had expected, familiarity ratings were significantly lower, $t(39) = 22.37$, $p < .001$, for novel words ($M = 2.47$, $SD = 0.66$, 95% CI [2.26, 2.68]) than for existing words ($M = 4.82$, $SD = 0.16$, 95% CI [4.77, 4.87]).

Procedure

The participants provided informed consent prior to participating in the study. They came into the lab for three sessions on Day 1, Day 2, and Day 8 (see Figure 2) for a total of 7–7.5 hours. They partook in a training session on one list of words (20 novel words and 20 existing words) on Day 1 (remote condition) and another list (new set of 20 novel and 20 existing words) on the following day after a night of sleep on Day 2 (recent condition). For Day 8 data, we continued to refer to words trained on Day 1 as remote and to the words trained on Day 2 as recent. After the Day 2 training session, the participants completed the sleep-related surveys while the EEG-system was set up. They then completed the testing session (on all trained novel and existing words). On Day 8, the participants completed the language history questionnaire while the EEG system was set up.

Training Session

Training sessions began with a 5-s visual presentation of each word (20 novel, 20 existing) with its definition and image in pseudorandom order (during which the participants were not asked to do anything other than watch the presentations). This was followed by four untimed training tasks, administered in the following order:

1. a two-alternative-forced-choice word recognition task in which the participants read the definition and image of a word and had to select the matching word from two options (three trials per word, 120 total trials);
2. a two-alternative-forced-choice definition-image recognition task in which the participants read a word and had to select the matching definition and image from two choices (three trials per word);
3. a word recall task in which the participants typed the word matching the presented definition and corresponding image (one trial per word); and
4. a definition recall task in which the participants typed the definition corresponding to the presented word (one trial per word).

For all training tasks, we presented items pseudorandomly, and the participants provided their responses using a keyboard. We presented feedback (the cue and correct answer) for 3 s for each item, regardless of whether the response was correct or not. For all tasks, we recorded responses for each trial. We recorded and presented response accuracy on the feedback screen for the recognition tasks. We scored accuracy for the recall tasks offline.

Testing Session

Our testing tasks were identical to those used in Liu and Van Hell's (2020) study. The participants performed an untimed definition recall task on all 40 novel and 40 existing words trained on Day 1 and Day 2 and trials were presented to them pseudorandomly (80 total trials). The definition recall allowed for reactivation of previously learned words to eliminate any recency effects (see also Bakker et al., 2015b; Liu & Van Hell, 2020). Then, the participants performed an EEG-recorded semantic priming task on all trained novel and existing words. They received no feedback for any testing task.

For the semantic priming task, the target words were the words trained on Day 1 and Day 2. We paired each target word with two related and two unrelated prime words from Liu and Van Hell (2020). The prime words were existing English words selected from the Florida Free Association Norms Database (Nelson, McEvoy, & Schreiber, 2004) based on the prime-target forward association strength value. For novel words (e.g., *jabary*), we used the base word (in this instance, *car*) to search for semantically related words as primes. The selected related prime words (listed in Liu and Van Hell, 2020's Appendix S1) did not appear in the novel word definitions. We pseudorandomly selected the two semantically unrelated primes for each target word from the set of prime words. Each trial began with a 600 ms fixation cross followed by a related or unrelated prime word displayed for 250 ms and a blank screen for 250 ms. Next, the target word was shown for 1 s followed by the question "Semantically related?" presented for 2,000 ms or until the participant responded "yes" or "no" by pressing a button. The participants' response accuracy was recorded.

The semantic relatedness judgment task referred to the behavioral aspect (accuracy) of the semantic priming task. Because we instructed the participants not to respond until prompted (the prompt appeared 1 s after the target word onset), we did not analyze latency. The next trial automatically began if there was no response for 2,000 ms. After each trial, a smiley face was displayed for 1,000 ms, during which the participants could blink and rest their eyes before the next trial. We counterbalanced the "yes" and "no" left and right finger-response mapping among the participants. The semantic priming task was a 2 × 2 × 2 design—Relatedness (related or unrelated prime) × Condition (remote or recent) × Word-type (existing or novel word)—with 40 trials for each Relatedness × Condition × Word-type condition, presented in four blocks of 80 trials and pseudorandomized across trials. Prior to the experimental trials, to familiarize the participants with the task, we asked the participants to complete 20 practice trials on untrained existing word pairs that were not related to any trained items (e.g., *son–daughter*).

After the semantic priming task, the participants completed a 4AFC word recognition task in which they selected the corresponding word from four options based on the presented definition.

Sleep Measures, Language History Questionnaire, Executive Function Tasks, and Vocabulary Tasks

The participants completed a sleep-related survey on Day 2 with 44 questions addressing different aspects of sleep quality and quantity (see Appendix S15 in the online Supporting Information) adapted from the Pittsburgh Sleep Quality Index (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989), Epworth Sleepiness Scale (Johns, 1991), Morningness-Eveningness Questionnaire (Horne &

Östberg, 1976), and Stanford Sleepiness Scale (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). The participants also completed a language history questionnaire (see Appendix S16 in the online Supporting Information) that asked about their language learning history and use, percentage of time that they had been exposed to the language(s) that they knew, and their foreign language classroom learning experience.

On Day 8, we recorded 5 min eyes-open and eyes-closed resting-state EEG prior to the testing tasks. After the participants had completed the tests, we administered two general cognitive ability and two vocabulary tasks. We have not reported these data in this paper.

Electroencephalography Data Acquisition and Preprocessing

EEG data acquisition and preprocessing procedures were identical to those used in Liu and Van Hell's (2020) study. The participants were seated comfortably in a chair in a sound-attenuated, dimly lit room, approximately three feet away from a computer monitor displaying the stimuli. A actiCAP system (Brain Products GmbH, n.d.), an elastic cap with 31 active electrodes, was placed on their heads: five electrodes located along the midline (Fz, FCz, Cz, Pz, Oz) and 26 electrodes on the lateral sites (FP1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP5/6, CP1/2, P7/8, P3/4, O1/2, PO9/10). Two electrodes were placed on the mastoids (TP9/10) for offline re-reference. An additional four electrodes were placed above and below the left eye and on the canthus of each eye to monitor eye movements and blinking. Electrode impedances were kept below 10 kΩ. Electrodes were referenced to a vertex reference (electrode FCz). EEG signals were amplified by a SynAmps RT amplifier (Compumedics NeuroScan, n.d.) using a .05–100 Hz bandpass filter (first-order Butterworth with a 6 dB/octave roll-off), digitized and continuously sampled at a rate of 500 Hz.

We preprocessed the EEG data using the EEGLAB (Version 14.1.2b) toolbox (Delorme & Makeig, 2004) and the ERPLab (Version 7.0.0) software (Lopez-Calderon & Luck, 2014). We re-referenced the recorded EEG signals offline to an average of the left and right mastoids. We removed EEG recordings of breaks. We used ERPLab for data cleaning and processing by applying an offline 30 Hz low-pass filter. We visually inspected the EEG data for any flat or extremely noisy channel signal. We interpolated channels with flat signals or noisy data, which resulted in the interpolation of 17 channels across all the participants (14 channels for Day 2 and 3 channels for Day 8 data). For each participant, we averaged ERPs from 100 ms prior to stimulus onset (prestimulus baseline) to 800 ms after stimulus onset offline in different experimental conditions at each electrode site by participant and then across all participants for each day separately. We calibrated artifact rejection thresholds individually to each participant. We applied ERPLAB's moving window peak-to-peak artifact detection function (voltage threshold range: 70–100 μV) to the vertical eye electrode to detect blinks, ERPLAB's step-like artifact detection function (voltage threshold range: 15–50 μV) to the horizontal eye electrode to detect horizontal eye movements, and ERPLAB's moving window peak-to-peak artifact detection function (voltage threshold range: 70–100 μV) to all other electrodes to detect channel drift or large amplitude activity. We excluded trials containing artifacts from further data analysis (17.91% of trials rejected for Day 2, 25.22% of trials for Day 8). We excluded the participants with fewer than 20 trials (< 50%) in one of the conditions from further analyses (one participant from Day 2, seven from Day 8). We have reported the mean number of trials for each condition in Appendix S2 in the online Supporting Information.

Analysis

We conducted all statistical analyses using the R software (R Core Team, 2020). Due to word limitations, we have reported the complete statistical results, including the confidence intervals for the means and effects sizes, in the appendices in the online Supporting Information. To determine statistical significance, we set alpha at .05. We reported partial eta squared as a measure of effect size.

Research Question on Word Meaning Activation After Definition-Image Training

After novel words are learned with definitions and images, to what extent is novel word meaning invariably activated (achieve more automaticity) as indexed by a N400 effect one day and seven days after learning? Or does word meaning retrieval remain more controlled and strategic as indexed by an LPC effect?

Event-Related Potential Analysis: Word Meaning Activation After Definition-Image Training

Our statistical analyses followed those used in Liu and Van Hell's (2020) study. To address our first research question, we conducted semantic priming ERP analyses on two time-windows: 300–500 ms for the N400 component and 500–700 ms for the LPC component, both post target word onset. Due to different numbers of participants for each day, we conducted separate analyses for the

Day 2 and the Day 8 data. Overall analyses including the variable time (Day 2, Day 8) on the same 23 participants yielded similar results (see Appendix S13 in the online Supporting Information). For each time-window of interest, we performed two omnibus repeated measures ANOVAs with the variables word-type (existing, novel), condition (remote, recent), semantic relatedness (related prime, unrelated prime). One ANOVA included a variable of midline electrodes (Fz, Cz, Pz) and the other ANOVA included two variables of anteriority (anterior region, posterior region) and laterality (right hemisphere, left hemisphere). For the variables anteriority and laterality, we grouped lateral electrodes into regions of interest: right frontal (F4, F8, FC2, FC6), left frontal (F3, F7, FC1, FC5), right posterior (CP2, CP6, P4, P8), and left posterior (CP1, CP5, P3, P7). We applied the Greenhouse-Geisser correction for violation of sphericity to all F tests with more than one degree of freedom in the numerator. We have reported only significant main effects of word-type (reflecting lexicality effects) and relatedness (reflecting semantic priming effects) and interaction effects involving the variable word-type and/or the variable relatedness ($p < .05$) because they reflect theoretically relevant semantic priming effects.

To test a priori and critical hypotheses on the lexical consolidation of novel words, we conducted planned ANOVAs on Day 2 novel word EEG data to examine the simple main effect of relatedness in the remote and recent conditions separately. We did not conduct these planned tests for the Day 8 EEG data because the distinction between remote and recent conditions was no longer critical on Day 8. For significant interactions with word-type or relatedness, we conducted theoretically driven post-hoc tests, applying the Bonferroni correction for multiple comparisons. When a theoretically relevant higher order interaction effect achieved significance, we have not reported significant main effects.

Research Question on Training-Type

To what extent can images paired with definitions significantly improve learning and strengthen the consolidation of novel words compared to novel words paired with only definitions? To address our second research question, we examined the impact of training-type by comparing our behavioral and ERP semantic priming data with the parallel data from Liu and Van Hell's (2020) study. The participants in our study and those in Liu and Van Hell's (2020) study were undergraduate students at the same university with similar language backgrounds, mean ages, and gender ratios (see Participants section). In the online Supporting Information, we have reported for the definition-image group a summary of behavioral results in Appendix S4, the behavioral data descriptive statistics in Appendix S5, and the behavioral statistical analyses in Appendix S6.

Behavioral Analysis: Training-Type

To examine whether images significantly improved the learning and consolidation of novel words compared to definitions only, we added training-type (definition-only, definition-image) as a between-subjects variable to the word-type (novel, existing) by condition (remote, recent) repeated-measures ANOVAs, separately for the definition recall, 4AFC word recognition, and semantic relatedness judgment tasks and separately for Day 2 data and Day 8 data. For the semantic relatedness judgment task, we included a within-subject variable of semantic relatedness (related, unrelated) to the repeated-measures ANOVAs. When significant interactions with training-type emerged, we conducted post-hoc ANOVAs on the definition-only and the definition-image data. We applied the Bonferroni correction to the p values for the post-hoc tests and have reported these corrected p values.

Event-Related Potential Analysis: Training-Type

To examine the impact of training-type on the semantic priming effect in the N400 and LPC time windows, we added training-type as a between-subjects variable to the word-type (novel, existing) by condition (remote, recent) repeated-measures ANOVAs, for each day separately. We also performed planned analyses with the between-subjects variable for the ERP data for Day 2. We conducted post-hoc analyses when there was a theoretically relevant interaction, namely any significant interaction(s) involving the training-type and relatedness variables. We applied the Bonferroni correction for multiple comparisons to any posthoc tests. We applied the Greenhouse-Geisser correction for violation of sphericity to all F tests with more than one degree of freedom in the numerator. We have reported only the Training-type × Relatedness interaction and any significant interaction effects involving at least both the training-type variable and the relatedness variable ($p < .05$) because they reflected theoretically relevant training-type effects on the semantic priming effect.

Results

Word Meaning Activation After Learning Novel Words With Definitions and Images: Invariable or More Controlled? (Question 1). Event-Related Potential Measurements

We included 31 participants in the Day 2 and 23 participants in the Day 8 data. In the online Supporting Information, Appendix S7 presents complete ERP results tables and Appendix S8 complete descriptive statistics for ERP amplitudes. Due to word limitations, some analyses that we have reported below do not include the descriptive data because we have reported them in Appendix S8 in the online Supporting Information. To examine semantic priming effects, we discuss analyses involving the variable relatedness. We have reported lexicality effects (ERP differences between novel and existing words) that involved the variable word-type in Appendix S9 in the online Supporting Information.

N400 Semantic Priming Effect on Day 2

Midline ANOVA. In the N400 window, the midline ANOVA revealed a Condition × Relatedness interaction, $F(1, 30) = 8.01, p < .01, \eta_p^2 = .21$. Post-hoc relatedness ANOVAs conducted separately for recent and remote learning conditions revealed a significant relatedness effect for both the remote condition, $F(1, 30) = 61.86$, $p < .001$, $\eta_p^2 = .67$ (related: $M = 1.79$ μV, $SD = 4.18$, 95% CI [1.18, 2.39]; unrelated: $M = -151$ μV, $SD = 3.61$, 95% CI [−2.03,−0.99]), and the recent condition, $F(1, 30) = 30.79$, $p < .001$, $\eta_p^2 = .51$ (related: $M = 0.29$ μV, $SD = 4.28$, 95% CI [−0.32, 0.91]; unrelated: $M = -1.81$ μV, $SD = 3.2$, 95% CI [−2.27, −1.35]), such that N400 amplitude was significantly more negative-going for unrelated trials than for related trials (see Figures 3 and 4).

The Word-type × Midline × Relatedness interaction was also significant, $F(1.30, 38.94) = 5.96$, $p = .01$, $\eta_p^2 = .17$. To follow up on this interaction, we conducted separate ANOVAs for existing and for novel words. For existing words, there was a significant Midline × Relatedness interaction, $F(1.28, 38.41) = 6.53$, $p = .02$, $\eta_p^2 = .18$. Follow-up relatedness ANOVAs conducted separately for the three midline channels (Fz, Cz, and Pz) yielded relatedness effects in all channels: Fz, $F(1, 30) = 56.97$, $p < .001$, $\eta_p^2 = .66$; Cz, $F(1, 30) = 117.81$, $p < .001$, $\eta_p^2 = .80$; Pz, $F(1, 30) = 117.48$, $\eta_p^2 = .80$, $p < .001$, such that unrelated trials elicited significantly more negative-going N400 amplitudes than related trials at all three channels (see ERP descriptive data in Appendix S8 in the online Supporting Information). As we had expected, we found a significant N400 semantic priming effect for existing words across the midline

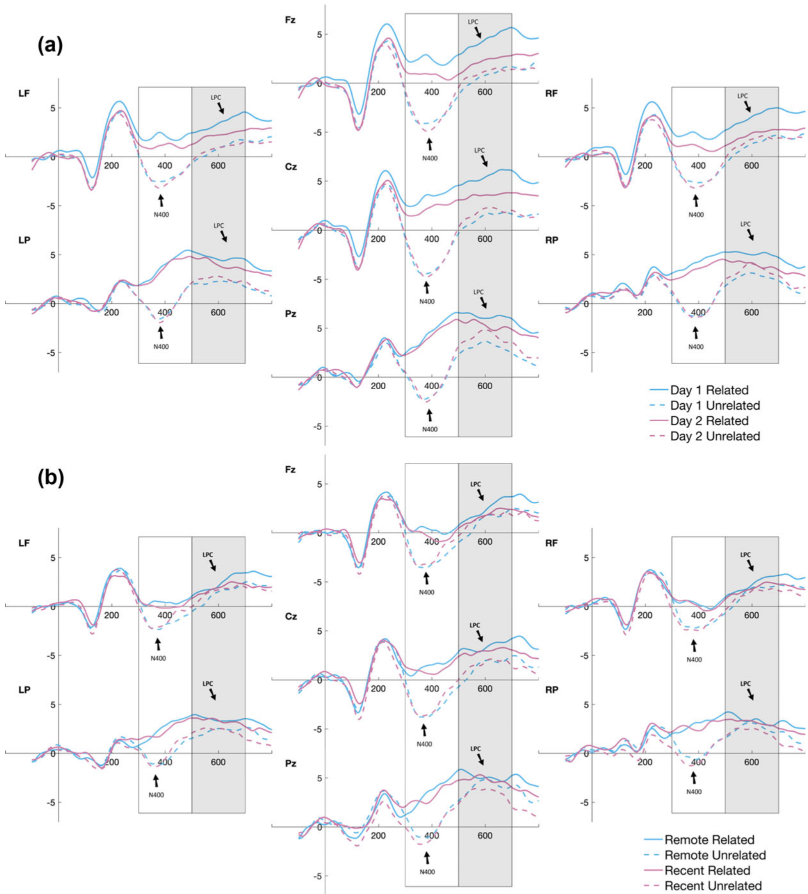

Figure 3 Semantic priming effects for existing words tested on Day 2 (Panel A) and Day 8 (Panel B). Event-related potentials time-locked to existing words preceded by a semantically related or unrelated prime word. Black lettered effect annotations (N400 or late positive component) indicate the effect was significant for both remote and recent learning conditions. A 15 Hz low-pass filter was applied to the waveforms for presentation purposes. Fz, Cz, and Pz are midline electrodes. LF = left frontal; LP = left posterior; RF = right frontal; RP = right posterior; LPC = late positive component

channels. For novel words, there was also a significant Midline × Relatedness interaction, $F(1.36, 40.74) = 6.43$, $p = .02$, $\eta_p{}^2 = .18$. Subsequent relatedness ANOVAs for the midline channels yielded no significant effect in any of the midline channels (Fz and Cz: both p's = 1.00; Pz: $p = .16$).

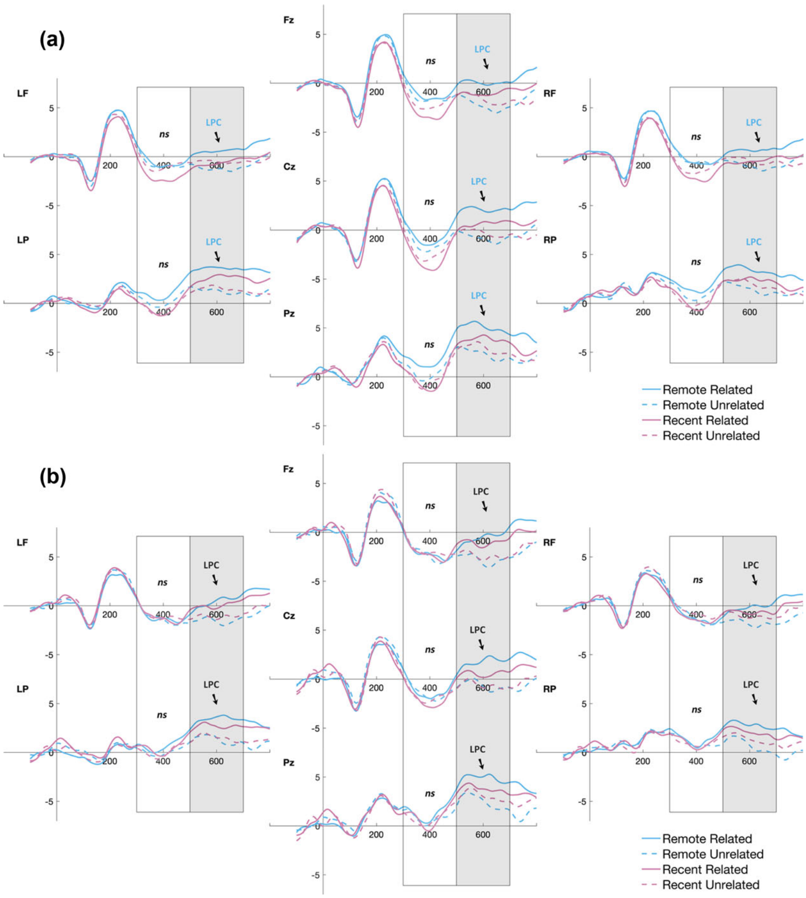

Figure 4 Semantic priming effects for novel words tested on Day 2 (Panel A) and Day 8 (Panel B). Event-related potentials time-locked to novel words preceded by a semantically related or unrelated prime word. Black lettered annotations indicate the labeled effect (N400 or late positive component) was significant for both remote and recent learning conditions. Blue lettered annotations indicate the effect was significant only for words in the remote condition. A 15 Hz low-pass filter was applied to the waveforms for presentation purposes. Fz, Cz, and Pz are midline electrodes. LF = left frontal; LP = left posterior; RF = right frontal; RP = right posterior; LPC = late positive component; *ns* = not significant.

Finally, planned tests to examine relatedness effects for novel words in the remote and recent conditions separately revealed no N400 semantic priming effects.

Laterality ANOVA. In the N400 window, the laterality ANOVA revealed significant interactions of Condition × Word-type × Relatedness × Laterality, $F(1, 30) = 5.74$, $p = .02$, $\eta_p^2 = .16$, and Relatedness × Anteriority, $F(1, 30) = 13.09$, $p < .01$, $\eta_p^2 = .30$.

To follow up on the Condition × Word-type × Relatedness × Laterality interaction, we conducted Condition × Relatedness × Laterality ANOVAs for each word type separately. For existing words, there was a significant effect of relatedness, $F(1, 30) = 88.38$, $p < .001$, $\eta_p^2 = .75$, with unrelated trials eliciting more negative-going mean amplitudes than related trials (see Figure 3). For novel words, however, there was no relatedness effect (see Figure 4).

To follow up on the Relatedness × Anteriority interaction, we conducted separate relatedness ANOVAs for the anterior and posterior regions. For both the anterior and posterior regions, there was a significant effect of relatedness, $F(1, 30) = 32.83$, $p < .001$, $\eta_p^2 = .52$; $F(1, 30) = 83.19$, $p < .001$, $\eta_p^2 = .74$, respectively. For both regions, the mean amplitude for unrelated trials (anterior: $M = -1.41$ μV, $SD = 2.06$, 95% CI [−2.16, −0.65]; posterior: $M = 0.06$ μV, $SD = 1.72$, 95% CI [−0.57, −0.69]) was significantly more negative-going than for related trials (anterior: $M = -0.17$ μV, $SD = 2.14$, 95% CI [−0.61, 0.95]; posterior: $M = 2.27$ μV, $SD = 1.81$, 95% CI [1.61, 2.93]).

Finally, planned tests to examine the main effect of relatedness for novel words, separately for the remote and recent conditions, revealed no N400 semantic priming effects.

In sum, on Day 2, we found an N400 semantic priming effect for existing words in both midline and lateral ANOVAs. Novel words, however, did not yield N400 semantic priming effects.

Late Positive Component Semantic Priming Effect on Day 2

Midline ANOVA. In the LPC time window, the omnibus midline ANOVA revealed a significant Condition × Relatedness interaction, $F(1, 30) = 16.10$, $p < .001$, $\eta_p^2 = .35$. Subsequent relatedness ANOVAs for the remote and recent conditions showed a significant effect of relatedness for remote words, $F(1, 30) = 50.37$, $p < .001$, $\eta_p^2 = .63$, and recent words, $F(1, 30) = 13.50$, $p < .01$, $\eta_p^2 = .31$. A relatedness effect in the LPC window reflected more positive amplitudes for related than for unrelated trials. For both remote and recent conditions, related trials (remote: $M = 1.78$ μV, $SD = 2.06$, 95% CI [1.03, 0.12]; recent: $M = 0.65$ μV, $SD = 1.76$, 95% CI [0.01, 1.30]) elicited significantly greater deflections than did unrelated trials (remote: $M = -0.52$ μV, $SD = 1.75$, 95% CI [−1.17, 2.54]; recent: $M = -0.82$ μV, $SD = 1.72$, 95% CI [−1.45, −0.19]).

For the novel words, planned tests of the main effect of relatedness in remote and recent learning conditions of novel words revealed a significant effect of relatedness for remote novel words, $F(1, 30) = 29.87$, $p < .001$, $\eta_p^2 = .50$, with related trials eliciting more positive-going amplitudes than unrelated trials, but not for recent novel words.

Laterality ANOVA. The omnibus laterality ANOVA revealed a significant Condition × Relatedness interaction, $F(1, 30) = 16.43$, $p < .001$, $\eta_p^2 = .35$. Subsequent relatedness ANOVAs revealed significant relatedness effects for the remote condition, $F(1, 30) = 45.19$, $p < .001$, $\eta_p^2 = .60$, and the recent condition, $F(1, 30) = 9.67$, $p < .01$, $\eta_p^2 = .24$.

Focused on the novel words, planned tests conducted to examine the main effect of relatedness of novel words in the remote and recent learning conditions revealed a significant relatedness effect in the remote condition, $F(1, 30) = 26.82$, $p < .001$, $\eta_p^2 = .47$, with related trials eliciting more positive-going amplitudes than unrelated trials, but not in the recent condition.

In sum, on Day 2, both midline and laterality ANOVAs showed LPC semantic priming effects for all existing words and for novel words trained on Day 1, but not for novel words trained on Day 2.

N400 Semantic Priming Effect on Day 8

Midline ANOVA. In the N400 window, the midline ANOVA revealed a significant Word-type × Relatedness interaction, $F(1, 22) = 23.32$, $p < .001$, $\eta_p^2 = .51$, and a Midline × Relatedness interaction, $F(1.24, 27.31) = 5.70$, $p = .02$, $\eta_p^2 = .21$.

To examine the Word-type × Relatedness interaction, we conducted separate relatedness ANOVAs for existing and novel words. The relatedness effect was significant for existing words, $F(1, 22) = 31.55$, $p < .001$, $\eta_p^2 = .59$ (with unrelated trials eliciting more negative-going mean amplitudes than did related trials, indexing an N400 semantic priming effect), but not for novel words.

To examine the Midline × Relatedness interaction, we conducted separate relatedness ANOVAs for the midline channels (Fz, Cz, Pz). All the ANOVAs showed significant N400 relatedness effects (unrelated trials eliciting more negative-going mean amplitudes than did related trials): Fz, $F(1, 22) = 11.09$, $p = .01$, $\eta_p^2 = .34$; Cz, $F(1, 22) = 25.23$, $p < .001$, $\eta_p^2 = .53$; Pz, $F(1, 22) = 29.46$, $p < .001$, $\eta_p^2 = .57$. Figures 3 and 4 show that these significant relatedness effects stemmed from existing words and not from novel words.

Laterality ANOVA. The laterality ANOVA for the N400 window revealed a significant Word-type × Relatedness × Laterality interaction, $F(1, 22) = 5.57$, $p = .03$, $\eta_p^2 = .20$, and a significant Relatedness × Anteriority interaction, $F(1, 22) = 11.95$, $p < .01$, $\eta_p^2 = .35$.

To examine the Word-type × Relatedness × Laterality interaction, we conducted separate Laterality × Relatedness ANOVAs for existing words and for novel words. For existing words, but not for novel words, there was a significant relatedness effect (unrelated trials eliciting more negative-going mean amplitudes than did related trials), $F(1, 22) = 27.42$, $p < .001$, $\eta_p^2 = .55$.

Post-hoc relatedness ANOVAs, conducted separately for the anterior and posterior regions, showed significant relatedness effects for both the anterior region, $F(1, 22) = 7.78$, $p = .02$, $\eta_p^2 = .26$, and the posterior region, $F(1, 22) = 36.84$, $p < .001$, $\eta_p^2 = .63$, that is, unrelated trials elicited more negative-going mean amplitudes than did related trials. Figures 3 and 4 show that these relatedness effects stemmed from existing words and not from novel words.

In sum, on Day 8, both midline and laterality ANOVAs showed N400 semantic priming effects for existing words but not for novel words.

Late Positive Component Semantic Priming Effect on Day 8

Midline ANOVA. For the LPC time window, the midline ANOVA revealed only a significant main effect of relatedness, $F(1, 22) = 26.70$, $p < .001$, $\eta_p^2 = .55$, with related trials eliciting more positive-going mean amplitudes than unrelated trials.

Laterality ANOVA. The omnibus laterality ANOVA for the LPC time window also revealed only a significant main effect of relatedness, $F(1, 22) = 28.12$, $p < .001$, $\eta_p^2 = .56$, and no significant interactions with relatedness.

Thus, on Day 8, midline and laterality ANOVAs showed LPC semantic priming effects for both existing and novel words, regardless of learning condition (recent or remote).

Training-Type: Comparing Definition-Only and Definition-Image Training (Question 2). Behavioral Measurements

We have reported only the results related to the training-type variable (in the online Supporting Information, see Appendix S10 for the complete results tables and Appendix S11 for behavioral performance figures that directly compare training-type).

Definition Recall

For Day 2, the ANOVA with training-type as a between-subjects variable revealed a significant main effect of training-type, $F(1, 55) = 8.72$, $p = .01$, $\eta_p^2 = .14$. Accuracy was 10.12% higher for the definition-image ($M = 82.84\%$, $SD = 26.02$) than for the definition-only ($M = 72.72\%$, $SD = 30.04$) training.

For Day 8, the effect of training-type was also significant, $F(1, 55) = 8.72$, $p = .005$, $\eta_p^2 = .16$. Accuracy was 12.32% higher for the definition-image training ($M = 73.48\%$, $SD = 32.82$) than for the definition-only training ($M = 61.16\%$, $SD = 36.78$).

Four-Alternative-Forced-Choice Word Recognition Task

For Day 2, the ANOVA revealed a significant Training-type × Condition interaction, $F(1, 53) = 6.52$, $p = .01$. Post-hoc analyses conducted separately for each condition revealed a trend toward a marginally significant effect of training-type for the remote condition, $F(1, 53) = 4.28$, $p = .09$, $\eta_p^2 = .07$. Accuracy was slightly higher for the definition-image training ($M = 97.58\%$, $SD = 4.59$) than for the definition-only training ($M = 95.29\%$, $SD = 6.56$). No training effect was present for the recent condition. The Training-type × Word-type interaction was also significant, $F(1, 53) = 9.56$, $p = .003$. Post-hoc analyses conducted separately for each word-type revealed an effect trending toward significance of training-type for novel words, $F(1, 53) = 4.27$, $p = .09$, $\eta_p^2 = .07$. Accuracy was slightly higher for the definition-image training ($M = 96.69\%$, $SD = 5.04$) than for the definition-only training ($M = 93.53\%$, $SD = 7.70$). No effect for existing words was present.

For Day 8, the ANOVA with training-type as a between-subjects variable revealed a significant Training-type × Word-type interaction, $F(1, 43) = 10.75$, $p = .02$, $\eta_p^2 = .20$. Post-hoc analyses conducted separately for word type revealed a marginal significant effect of training-type for novel words, $F(1, 43) = 5.35$, $p = .05$, $\eta_p^2 = .11$. Recognition accuracy of novel words was 6.09% higher for the definition-image training ($M = 96.52\%$, $SD = 5.26$) than for the definition-only training ($M = 90.43\%$, $SD = 13.26$). There was also a borderline significant effect of training-type for existing words, $F(1, 43) = 5.16$, $p = .06$, $\eta_p^2 = .11$. Accuracy was slightly lower for the definition-image training ($M = 97.50\%$, $SD = 3.91$) than for the definition-only training ($M = 99.19\%$, $SD = 2.42$).

Semantic Relatedness Judgment Task

For Day 2 semantic relatedness judgments, the ANOVA including training-type revealed a significant Training-type × Condition interaction, $F(1, 55) = 6.01$,

$p = .02$, $\eta_p^2 = .10$, and a significant Training-type × Word-type interaction, $F(1, 55) = 7.51$, $p = .01$, $\eta_p^2 = .12$. Post-hoc analyses for the Training-type × Condition interaction conducted separately for the remote and recent conditions revealed a significant training-type effect only for recent words, $F(1, 55) = 11.31$, $p = .003$, $\eta_p^2 = .17$. Accuracy was 6.86% higher for the definition-image training (M = 86.09%, SD = 15.41) than for the definition-only training (M = 79.23%, SD = 19.92). Post-hoc analyses for the Training-type × Word-type interaction conducted separately for existing and novel words showed a significant training-type effect only for novel words, $F(1, 55) = 10.49$, $p = .004$, $\eta_p^2 = .16$. Accuracy was 7.94% higher for the definition-image training (M = 78.21%, SD = 17.73) than for the definition-only training (M = 70.27%, SD = 19.35).

For Day 8, the ANOVA revealed a significant Training-type × Condition interaction, $F(1, 44) = 4.30$, $p = .04$, $\eta_p^2 = .09$, and a significant Training-type × Word-type interaction, $F(1, 44) = 8.03$, $p = .01$, $\eta_p^2 = .15$. Post-hoc analyses for the Training-type × Condition interaction conducted separately for remote and recent conditions revealed no training-type effect for remote words but a trend toward a marginally significant effect for recent words, $F(1, 44) = 4.14$, $p = .10$, $\eta_p^2 = .09$. Post-hoc analyses for the Training-type × Word-type interaction revealed no training-type effect for existing words, but a borderline significant effect for novel words, $F(1, 44) = 5.00$, $p = .06$, $\eta_p^2 = .10$ (definition-image: M = 78.12%, SD = 20.00; definition-only; M = 70.68%, SD = 20.84).

In conclusion, Day 2 definition recall, 4AFC word recognition, and semantic relatedness judgment scores for novel word meanings were higher for the definition-image training than for the definition-only training. This definition-image training advantage for novel word meanings was still present on Day 8 for definition recall and 4AFC word recognition performance.

Event-Related Potential Measurements

ANOVAs with training-type as a between-subjects variable yielded no N400 or LPC semantic priming effects involving the variable training-type for novel words tested on Day 2 and Day 8 (see Appendices S12 and S13 in the online Supporting Information for complete results tables and the report, respectively). The absence of training-type effects in the ERP data contrasted with the consistent definition-image training advantage for novel word meanings in all behavioral measures.

Table 1 provides a summary of the ERP results for definition-image group and the training-type results for novel words.

Table 1 Summary of event-related potential results for definition-image training and the behavioral and event-related potential training-type results for novel words

		Test on Day 2	Test on Day 8
Definition-image group semantic priming ERP results			
Existing	Remote	N400, LPC	N400, LPC
	Recent	N400, LPC	N400, LPC
Novel	Remote	LPC	LPC
	Recent	*ns*	LPC
Training-type (definition-image vs. definition-only) novel words results			
Definition recall		Definition-image > Definition-only	Definition-image > Definition-only
4AFC word recognition		Marginal; Definition-image > Definition-only	Marginal; Definition-image > Definition-only
Semantic relatedness judgment		Definition-image > Definition-only	Marginal; Definition-image > Definition-only
Semantic priming (ERPs)		*ns*	*ns*

Note. ERPs = event related potentials; LPC = late positive component; 4AFC = four-alternative-forced-choice; *ns* = not significant.

Discussion

This study examined how verbalizations and visualizations impact the learning and consolidation of novel words and novel meanings in monolingual speakers. Day 2 ERP results showed that existing words elicited the classic N400 semantic priming effect followed by an LPC semantic priming effect. Novel words trained the previous day, with one night of offline consolidation, elicited an LPC semantic priming effect but not a N400 semantic priming effect. Novel words trained shortly before testing, with no overnight offline consolidation, did not demonstrate any semantic priming effects. On Day 8 of testing, novel words of both recent and remote training conditions demonstrated LPC semantic priming effects, but no N400 effects. We have presented overview tables summarizing the study manipulation and results in Appendix S14 in the online Supporting Information.

These findings align with the complementary learning systems account of word learning (e.g., Davis & Gaskell, 2009) in which a period of offline consolidation is required for novel words to be consolidated and integrated in lexical-semantic memory. In our study, trained novel words needed a period of time—overnight offline consolidation—to demonstrate LPC effects. LPC semantic priming effects are taken to index a controlled, conscious, and strategic process of semantic access (e.g., Hoshino & Thierry, 2012; Juottonen, Revonsuo, & Lang, 1996; Rohaut et al., 2015), which suggests that meaning retrieval for these novel words was more controlled and less invariable and automatic compared to existing words that did demonstrate N400 semantic priming effects. Thus, this study shows that novel word meaning retrieval becomes progressively more word-like over time, evidenced by the LPC semantic priming effect on Day 2 for novel words learned the previous day and on Day 8 for words learned a week earlier. But it appears that one training session followed by one day and one week of offline consolidation was not enough for novel words to become fully lexicalized (eliciting N400 effects).[1]

To what extent does encoding novel words with meaning definitions and visualizations strengthen encoding and consolidation over time relative to encoding with verbal definitions only? We conducted direct comparisons of two groups of learners (definition-only group from Liu and Van Hell, 2020; definition-image group from our present study) with training-type as a between-subjects variable. Both learner groups consisted of undergraduate students at the same public university with similar language backgrounds, ages, and gender ratios. For both the N400 and LPC time windows on both Day 2 and Day 8, the ERP signatures of the definition-image and definition-only trainings did not differ. Both trainings showed LPC semantic priming effects on Day 2 for only novel words learned the previous day (remote condition), whereas one week after training on Day 8, we obtained (for both trainings) LPC priming effects for novel words learned on both Day 1 and Day 2. This pattern of results parallels Takashima et al.'s (2017) fMRI word learning study in which there were similar patterns of neocortical activation patterns for words learned with a picture only (visual modality) and words learned with a definition only (verbal modality). Our behavioral data, however, did show an effect of training; definition recall and semantic judgment performance for novel words were better for the definition-image training than for the definition-only training.

Based on the dual coding theory, we had predicted that images combined with definitions (imagens combined with logogens) would facilitate the encoding and consolidation of novel word meanings relative to the definition-only condition. Specifically, the definition-image training would elicit an N400 followed by an LPC effect in the semantic priming task, and behavioral recall and recognition performance would be higher for novel word meanings trained with definitions and images relative to the definition-only training. Our ERP results indicate otherwise. The presence of LPC semantic priming effects but the lack of N400 semantic priming effects even one week post learning, for both the definition-image and the definition-only trainings, suggests that these novel words had not obtained the status of existing words and that the addition of visualizations to verbal definitions did not expedite the lexicalization of novel word meanings.

Interestingly, the behavioral data did show that definition recall and semantic judgment performance for novel words was better for the definition-image training than for the definition-only training. This suggests that images do facilitate the encoding and retrieval of word meaning, possibly due to dual encoding. Alternatively, the behavioral recall advantage of the definition-image training can be explained by the depth-of-processing hypothesis (e.g., Craik & Lockhart, 1972; see Nassaji & Hu, 2012, for a L2 incidental word learning study that manipulated the depth of processing of novel words), which postulates stronger memory traces for information that is more deeply processed. Specifically, studying images alongside definitions during training may involve deeper processing than does just definitions alone. Despite using the same 5-s exposure time for each word in the definition-image training (our study) and the definition-only training (Liu & Van Hell, 2020), the participants' effort and depth of encoding may have been more profound when meaning was presented verbally (in writing) and illustrated visually, facilitating word meaning retrieval. Particularly, visualizations that are colorful and attractive depictions of novel word meanings (as was the case for the images that we used, see Figure 1) may serve as effective retrieval cues and facilitate the retrieval of word meaning during behavioral recall and recognition tasks. Previous research has shown that memory for colored images is better than for black-and-white images (e.g., Wichmann, Sharpe, & Gegenfurtner, 2002). By providing colorful images of the novel-word meanings during training, the three components of each novel meaning (i.e., object category with two novel features) were encapsulated in one visual representation that could act as an effective retrieval cue. We propose that the images merely served as a visual memory cue that supported the retrieval of the novel word meanings, rather than further expediting the consolidation and lexical-semantic integration of novel word meanings in memory (and yielding N400 semantic priming effects on Day 2 and Day 8 as initially predicted).

The ERP analyses yielded LPC semantic priming effects, but no N400 semantic priming effects, for both the definition-image and definition-only training, suggesting that word meaning retrieval becomes progressively more word-like over time, but that one training is not sufficient for novel words to become fully lexicalized (and to elicit N400 effects). That novel word meanings were not yet fully lexicalized (not even when images were added to visualize the verbal definitions) may be explained by the inconsistency of the novel word meanings with the existing lexical network (i.e., prior knowledge, McClelland, 2013; Van Kesteren, Ruiter, Fernández, & Henson, 2012). In the CLS model, the integration of new information in memory networks may be facilitated if it is consistent with existing representations, and inconsistent information may slow down this process (McClelland, 2013; Palma & Titone, 2020; Palma, Marin, Onishi, & Titone, 2022). The stimuli that we used were inconsistent at the semantic level—we designed the novel meanings to consist of one object category with two novel features typically not associated with this noun (e.g., *an apple that is square and grows on the ground*). In order to integrate these novel word meanings, the preexisting semantic network representations would need to be disrupted as new representations and connections are formed. The slower integration of inconsistent information has been proposed to be in place to prevent catastrophic interference with existing representations in the neocortex (McClelland, 2013).

The lack of lexicalization of novel words may also have been due to the fact that training was constrained to one session for each set of 20 novel items and may not have been sufficient for the participants to fully form strong associations between the novel word forms and their meanings and to integrate these into lexical-semantic memory. The participants' at-ceiling test performance for the 4AFC word recognition task (all *M*s $>$ 96%, see Figure S4.1, Panel B, in

Appendix S4 in the online Supporting Information) on both test days suggests that they were learning the words and associated concepts, but their definition recall performance, however, varied (*M* range: 41–77%, see Figure S4.1, Panel A, Appendix S4 in the online Supporting Information). Better definition recall performance might be needed before N400 lexicalization effects can be observed in ERP data. Future studies may increase the intensity of training (number of trials, number of training sessions; e.g., Walker et al., 2019), the nature of the training, or set a high definition recall performance threshold to examine whether the lack of lexicalization in our study can be ascribed to insufficient training. Our study materials are available on OSF (https://osf.io/gkd2h/) and IRIS (iris-database.org) for future studies.

The finding that LPC semantic priming effects were found for words learned one day previously (remote condition), but not immediately prior to (recent condition) Day 2 testing, is in line with the CLS model but not with alternative perspectives proposing a faster scenario for memory consolidation and rapid integration into the lexicon (e.g., Borovsky, Elman, & Kutas, 2012; Coutanche & Thompson-Schill, 2014; Kapnoula, Packard, Gupta, & McMurray, 2015; Lindsay & Gaskell, 2013), such as the fast mapping account (for critical discussion, see Cooper, Greve, & Henson, 2019). For example, Kapnoula et al. (2015) observed lexical competition effects immediately after training in a visual world paradigm and Lindsay and Gaskell (2013) observed competition effects within a single day when novel and existing words were interleaved during the training phase. These rapid lexicalization effects have been related to the ease with which prior information can be assimilated to prior knowledge or a schema, suggesting that under certain circumstances learning may involve only limited, or no, hippocampal contribution and that lexical integration of newly learned words does not require (an extended period of) offline consolidation.

An interesting pattern observed in the behavioral definition recall task, 4AFC recognition task, and semantic relatedness judgment d' scores is that, on Day 8, novel remote words (learned on Day 1) were better recalled and had higher d' scores than recent novel words (learned on Day 2; see Appendix S4 in the online Supporting Information for the full report of this analysis). One possible interpretation of this finding is related to the classical definition of consolidation as the development of increasing resistance to interference over time. Newly encoded information is malleable and fragile and requires gradual stabilization and integration into long-term memory to become resistant to forgetting (e.g., McGaugh, 2000). Newly encoded memories are subject to interference and forgetting, and learners recall less when the interval between learning and recall is filled with cognitively demanding activities (e.g., Martini, Martini, Maran, & Sachse, 2018). In our study, unlike learning words on Day 1, learning words on Day 2 (recent condition) was followed by the Day 2 testing tasks. This Day 2 testing may have created interference for words encoded on Day 2—interference that was absent for words encoded on Day 1—that may have led to better memory performance on Day 8 for words learned on Day 1 versus words learned on Day 2. Alternatively, the first test session (on Day 2) may have served as a restudy session, albeit without feedback as in typical restudy training regimes, because retrieval during testing has been shown to enhance retention (e.g., Kim & Webb, 2022; Roediger & Karpicke, 2006). During Day 2 testing, the participants were exposed to words learned the previous day (remote words) and words learned immediately before testing (recent words). Re-exposure to words that had already benefited from offline consolidation (remote words) may have further solidified the memory for these words relative to re-exposure to words that had been learned immediately before testing, yielding the observed behavioral recall advantage for remote words on Day 8.

Conclusion

In conclusion, we examined the impact of (written) verbal and visual information on the encoding and consolidation of novel word meanings. Monolingual learners demonstrated LPC (but no N400) semantic priming effects for novel words after one overnight offline consolidation period and one week later as was the case for Liu and Van Hell's (2020) learners who were presented with verbal definitions only. Neurocognitively, these findings suggest that lexicalization of novel word meanings is a gradual process and that one training session (even when verbal definitions are combined with visualizations) followed by one day and one week of offline consolidation is not enough for novel word meanings to become fully lexicalized. In the behavioral definition recall and semantic relatedness judgment tasks, however, the learners in the definition-image group (present study) performed significantly better than did those in the definition-only group (Liu & Van Hell's, 2020, study). The behavioral and neurocognitive findings combined suggest that images visualizing the verbal definition of word meanings are an efficient memory retrieval cue that facilitates the recall of novel word meanings (as measured with behavioral memory tasks), but adding images to definitions does not significantly enhance the lexical-semantic integration of novel word meanings.

Final revised version accepted 9 February 2022

Note

1 The prediction here is not that novel words will elicit N400 effects identical to existing words. Rather, the presence of an N400 effect is predicted, likely with a smaller magnitude than for the N400 effect for existing words.

References

Angwin, A. J., Phua, B., & Copland, D. A. (2014). Using semantics to enhance new word learning: An ERP investigation. *Neuropsychologia*, *59*, 169–178. https://doi.org/10.1016/j.neuropsychologia.2014.05.002

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*, 116–130. https://doi.org/10.1016/j.jml.2014.03.002

Bakker, I., Takashima, A., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2015a). Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience*, *27*, 1286–1297. https://doi.org/10.1162/jocn_a_00801

Bakker, I., Takashima, A., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2015b). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Bakker-Marshall, I., Takashima, A., Schoffelen, J. M., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2018). Theta-band oscillations in the middle temporal gyrus reflect novel word consolidation. *Journal of Cognitive Neuroscience*, *30*, 621–633. https://doi.org/10.1162/jocn_a_01240

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*, 126–140. https://doi.org/10.1016/j.cedpsych.2010.04.001

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, *23*, 3181–3196. https://doi.org/10.1162/jocn_a_00013

Bermúdez-Margaretto, B., Beltrán, D., Cuetos, F., & Domínguez, A. (2018). Brain signatures of new (pseudo-) words: Visual repetition in associative and non-associative contexts. *Frontiers in Human Neuroscience*, *12*, Article 354. https://doi.org/10.3389/fnhum.2018.00354

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research*, *76*, 192–203. https://doi.org/10.1007/s00426-011-0335-6

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meaning from a single exposure in context.

Language Learning and Development, *8*, 278–302. https://doi.org/10.1080/15475441.2011.614893

Brain Products GmbH. (n.d.). actiCAP [Apparatus]. https://www.brainproducts.com/productdetails.php?id=4

Buysse, D. J., Reynolds, C. F., III, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*, 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Compumedics Neuroscan. (n.d.). SynAmps RT [Apparatus]. https://compumedicsneuroscan.com/product/synamps-rt-64-channel-eeg-erp-ep-amplifier/

Cooper, E., Greve, A., & Henson, R. N. (2019). Investigating fast mapping task components: No evidence for the role of semantic referent nor semantic inference in healthy adults. *Frontiers in Psychology*, *10*, Article 394. https://doi.org/10.3389/fpsyg.2019.00394

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*, 2296–2303. https://doi.org/10.1037/xge0000020

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*, 3773–3800. https://doi.org/10.1098/rstb.2009.0111

Deacon, D., Dynowska, A., Ritter, W., & Grose-Fifer, J. (2004). Repetition and semantic priming of nonwords: Implications for theories of N400 and word recognition. *Psychophysiology*, *41*, 60–74. https://doi.org/10.1111/1469-8986.00120

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: On the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*, 1251–1261. https://doi.org/10.1162/jocn.2009.21297

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*, 35–39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Federmeier, K. D. (2021). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, *59*, Article e13940. https://doi.org/10.1111/psyp.13940

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory*, *13*, 259–262. https://doi.org/10.1101/lm.132106

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences*, *95*, 2703–2708. https://doi.org/10.1073/pnas.95.5.2703

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology*, *10*, 431–436. https://doi.org/10.1111/j.1469-8986.1973.tb00801.x

Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, *14*, 938–950. https://doi.org/10.1162/089892902760191153

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*, 97–110.

Hoshino, N., & Thierry, G. (2012). Do Spanish–English bilinguals have their fingers in two pies–or is it their toes? An electrophysiological investigation of semantic access in bilinguals. *Frontiers in Psychology*, *3*, Article 9. https://doi.org/10.3389/fpsyg.2012.00009

Johns, M. W. (1991). A new method for measuring daytime sleepiness: The Epworth sleepiness scale. *Sleep*, *14*, 540–545. https://doi.org/10.1093/sleep/14.6.540

Juottonen, K., Revonsuo, A., & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, *4*, 99–107. https://doi.org/10.1016/0926-6410(96)00022-5

Kaczer, L., Bavassi, L., Petroni, A., Fernández, R. S., Laurino, J., Degiorgi, S., … Pedreira, M. E. (2018). Contrasting dynamics of memory consolidation for novel word forms and meanings revealed by behavioral and neurophysiological markers. *Neuropsychologia*, *117*, 472–482. https://doi.org/10.1016/j.neuropsychologia.2018.07.001

Kandhadai, P., & Federmeier, K. D. (2010). Automatic and controlled aspects of lexical associative processing in the two cerebral hemispheres. *Psychophysiology*, *47*, 774–785. https://doi.org/10.1111/j.1469-8986.2009.00969.x

Kapnoula, E. C., Packard, S., Gupta, P., & McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, *134*, 85–99. https://doi.org/10.1016/j.cognition.2014.09.007

Kim, S., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning*, *72*. https://doi.org/10.1111/lang.12479

Kumaran, D., Hassabis, D., & McClelland, J.L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends in Cognitive Sciences*, *20*, 512–534. https://doi.org/10.1016/j.tics.2016.05.004

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

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology*, *55*, 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001

Lei, D., Liu, Y., & Van Hell, J. G. (2022a). *Experimental scripts. Materials from "Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures"* [Software: Experimental script]. IRIS Database, University of York, UK. https://doi.org/10.48316/23af-hy33

Lei, D., Liu, Y., & Van Hell, J. G. (2022b). *Questionnaires. Materials from "Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures*" [Questionnaire]. IRIS Database, University of York, UK. https://doi.org/10.48316/59kt-ah37

Lei, D., Liu, Y., & Van Hell, J. G. (2022c). *Visual stimuli. Materials from "Novel word learning with verbal definitions and images: Tracking consolidation with behavioral and event-related potential measures*" [Collection: Stimuli]. IRIS Database, University of York, UK. https://doi.org/10.48316/w01w-7835

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 608–622. https://doi.org/10.1037/a0029243

Liu, Y., & Van Hell, J. G. (2020). Learning novel word meanings: An ERP study on lexical consolidation in monolingual, inexperienced foreign language learners. *Language Learning*, *70*(S2), 45–74. https://doi.org/10.1111/lang.12403

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, Article 213. https://doi.org/10.3389/fnhum.2014.00213

Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, MA: MIT press.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. https://doi.org/10.1371/journal.pone.0043230

Martini, M., Martini, C., Maran, T., & Sachse, P. (2018). Effects of post-encoding wakeful rest and study time on long-term memory performance. *Journal of Cognitive Psychology*, *30*, 558–569. https://doi.org/10.1080/20445911.2018.1506457

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General*, *142*, 1190–1210. https://doi.org/10.1037/a0033812

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457. https://doi.org/10.1037/0033-295X.102.3.419

McGaugh, J. L. (2000). Memory – A century of consolidation. *Science*, *287*, 248–251. https://doi.org/10.1126/science.287.5451.248

Nassaji, H., & Hu, H. C. M. (2012). The relationship between task-induced involvement load and learning new words from context. *International Review of Applied Linguistics in Language Teaching*, *50*, 69–86. https://doi.org/10.1515/iral-2012-0003

Nelson, D. L., McEvoy, C .L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*, 402–407. https://doi.org/10.3758/BF03195588

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia*, *9*, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Paivio, A. (2014). *Mind and its evolution: A dual coding theoretical approach*. New York, NY: Psychology Press. https://doi.org/10.4324/9781315785233

Palma, P., Marin, M.-F., Onishi, K. H., & Titone, D. (2022). Learning, inside and out: Prior linguistic knowledge and learning environment impact word learning in bilingual individuals. *Language Learning*, 72. https://doi.org/10.1111/lang.12501

Palma, P., & Titone, D. (2020). Something old, something new: A review of the literature on sleep-related lexicalization of novel words in adults. *Psychonomic Bulletin & Review*, *28*, 96–121. https://doi.org/10.3758/s13423-020-01809-5

R Core Team. (2020). R: A language and environment for statistical computing (Version 4.0.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. www.r-project.org.

Rice, C. A., & Tokowicz, N. (2020). A review of laboratory studies of adult second language vocabulary training. *Studies in Second Language Acquisition*, *42*, 439–470. https://doi.org/10.1017/S0272263119000500

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Rohaut, B., Faugeras, F., Chausson, N., King, J. R., El Karoui, I., Cohen, L., & Naccache, L. (2015). Probing ERP correlates of verbal semantic processing in patients with impaired consciousness. *Neuropsychologia*, *66*, 279–292. https://doi.org/10.1016/j.neuropsychologia.2014.10.014

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*, 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Smith, B. D., Stahl, N., & Neil, J. (1987). The effect of imagery instruction on vocabulary development. *Journal of College Reading and Learning*, *20*, 131–137. https://doi.org/10.1080/10790195.1987.10849909

Smith, B. D., Miller, C., Grossman, F., & Valeri-Gold, M. (1994). Vocabulary retention: Effects of using spatial imaging on hemispheric-preference thinkers. *Journal of Research & Development in Education*, *27*, 244–252.
Stickgold, R., & Walker, M. P. (2007). Sleep-dependent memory consolidation and reconsolidation. *Sleep Medicine*, *8*, 331–343. https://doi.org/10.1016/j.sleep.2007.03.011
Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Richness of information about novel words determines how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, *84*, 265–78. https://doi.org/10.1016/j.neuroimage.2013.08.023
Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and Language*, *167*, 44–60. https://doi.org/10.1016/j.bandl.2016.05.009
Van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neuroscience*, *35*, 211–219. https://doi.org/10.1016/j.tins.2012.02.001
Walker, S., Henderson, L. M., Fletcher, F. E., Knowland, V. C. P., Cairney, S. A., & Gaskell, M. G. (2019). Learning to live with interfering neighbours: The influence of time of learning and level of encoding on word learning. *Royal Society Open Science*, *6*, Article 181842. https://doi.org/10.1098/rsos.181842
Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 509–520. https://doi.org/10.1037/0278-

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Stimuli List.
Appendix S2. Event-Related Potential Data Number of Trials per Condition.
Appendix S3. Event-Related Potential Results With Time Variable.
Appendix S4. Behavioral Results Summary.
Appendix S5. Behavioral Data Descriptive Statistics.
Appendix S6. Behavioral Results.
Appendix S7. Event-Related Potential Results.
Appendix S8. Descriptive Statistics for Event-Related Potentials.
Appendix S9. Lexicality Event-Related Potential Results Summary.
Appendix S10. Training-Type Behavioral Results.
Appendix S11. Training-Type Behavioral Plots.

Appendix S12. Training-Type Event-Related Potential Results.
Appendix S13. Training-Type Event-Related Potential Results Summary.
Appendix S14. Study Overview.
Appendix S15. Sleep Questionnaire.
Appendix S16. Language History Questionnaire.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Novel words learned better with definition and image than definition only, but no brain activity differences between the two types of training

What This Research Was About and Why It Is Important

Research has shown that when individuals learn new words, a period of consolidation (e.g., sleep) aids the association of the new words and their meaning with known word meanings in the brain. Given that adults continue to learn new words even after they have gained fluency in a language, it is important to examine how different training regimes impact novel word learning in adults. In this study, the impact of images on novel word learning was examined by training young monolingual adults on one set of novel words paired with novel definitions and images on Day 1 and another set on Day 2. These participants (the definition-image group) were tested on all the words immediately after learning (Day 2) and one week later (Day 8). Tests included a definition recall test, a multiple-choice test, and a relatedness judgment test. Brain activity was recorded during the relatedness judgment test. A previous study (by Liu & Van Hell, 2020) had trained and tested participants with the same design, but their participants were only exposed to definitions during the learning (the definition-only group). We compared the definition-image and definition-only groups and found that the definition-image group performed better in the definition recall and semantic judgment test, but there was no brain activity difference between the two groups.

What the Researchers Did

- Thirty-three participants (the definition-image group) were trained on novel (made up) words (e.g., ‘jabary’) paired with novel definitions (e.g., a car with two wheels that can move under water) and matching images.
- Participants were trained on one set of new words on Day 1 and a different set of new words on Day 2.
- Participants were tested on all trained words on Day 2 and on Day 8.

- Brain activity was also recorded during the testing sessions on Day 2 and Day 8.
- Test outcomes of the definition-image group were compared with the test outcomes of the definition-only group from a previous study (Liu & Van Hell, 2020).

What the Researchers Found

- After one-night of sleep, brain activity of novel word meanings learned on Day 1 showed brain activity reflecting association of novel words with existing words, but that was not the case for words learned on Day 2.
- The definition-image group and the definition-only group did not show differing brain activity when tested on Day 2 or Day 8.
- The definition-image group showed better performance in the definition recall and the relatedness judgment test than the definition-only group.

Things to Consider

- Images appear to have facilitated encoding and might be an efficient memory retrieval cue that facilities the recall of novel word meanings.
- Additional training might be needed to see differences in brain activity between the two training paradigms.

Material, data, open access article: Materials are available from IRIS (www.iris-database.org) and OSF (https://osf.io/gkd2h/).

How to cite this summary: Lei, D., Liu, Y., & Van Hell, J. G. (2022). Novel words learned better with definition and image than definition only, but no brain activity differences between the two types of training. *OASIS Summary* of Lei, Liu, & Van Hell (2022) in *Language Learning*. https://oasis-database.org