How Do We Learn New Meanings for Words Already Known? Evidence from EEG and MEG Studies

by

Xiaoping Fang

BS, Beijing Normal University, 2010

MS, Beijing Normal University, 2013

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Xiaoping Fang

It was defended on

August 30, 2019

and approved by

Marc N. Coutanche, Assistant Professor, Department of Psychology

Michael W. Dickey, Associate Professor, Department of Communication Science & Disorders

Julie A. Fiez, Professor, Department of Psychology

Thesis Advisor: Charles Perfetti, Distinguished University Professor, Department of Psychology

How Do We Learn New Meanings for Words Already Known? Evidence from EEG and MEG Studies

Xiaoping Fang, PhD

University of Pittsburgh, 2019

In addition to learning new words, people often learn new meanings for words they already know. For example, one might learn that the word 'skate' refers to a type of fish long after knowing its more common roller- or ice-skating meaning. Different from learning a new word, this type of learning requires updating the lexical knowledge by associating new information with an existing word and involves the co-activation of new and prior word knowledge. This dissertation investigates the mechanisms underlying the learning of new meanings for known words. In particular, it focuses on the influence of overnight consolidation on the learning of new meanings for known words and the role of the left posterior middle temporal gyrus (pMTG) in binding new meanings to known words. Study 1 showed that the processing of both new and original meanings became faster after overnight sleep. This indicated reduced interference between new and original meanings over time, especially after overnight consolidation occurred. However, the event-related potential (ERP) data showed that accessing the new meanings was still mainly supported by episodic retrieval even 24 hours after learning. To investigate how new meanings are associated with known words, Study 2a first demonstrated that presenting word meanings as verbal definitions were sufficient to drive a semantic category effect. Based on this result, Study 2b further showed that the left pMTG, in addition to sensorimotor cortices relevant to the representation of new meanings, was involved in binding new meanings to known words. Combined with the previous findings on learning novel words, the results suggest that the co-

activation of new and prior knowledge is essential to the integration of new word knowledge into the mental lexicon. The left pMTG not only supports the formation of novel form-meaning associations, but also the associations between new meanings and previously known words.

Table of Contents

List of Tables

List of Figures

Preface

In the past six years, there are so many people I am greatly indebted to.

First of all, I would like to thank my advisor, Dr. Charles Perfetti, who is always supportive of my growth as a scholar at my own pace in every way he can. I have been extremely lucky that I have been able to explore without being rushed.

Next, I wanted to thank former and current Perfetti lab members: Adeetee Bhide, Elena Cimino, Wendy Chang, Lin Chen, Regina Calloway, Maxwell Helfrich, Anne Helder, Erika Laing, Adrian Maries, Kimberley Muth, Ben Rickles, Joseph Stafura, and Zoe Zhou. I greatly enjoyed the discussions in and out of the lab. I wanted to thank the research assistants and undergraduates that I have worked with, especially Alexia Cardiges, Joseph Jaros, and Paula Pascual. I cannot imagine how my projects would be finished without their hard work.

I also wanted to thank the faculty who served on my mentoring and milestone committees and have been guiding and supporting me on my professional and personal development: Drs. Marc Coutanche, Michael Dickey, Julie Fiez, Scott Fraundorf, Melissa Liberturs, Ben Rottman, Natasha Tokowicz, and Tessa Warren.

In addition, I received a lot of support from my climbing/hot-pot/…/study buddies: Josh Bow, Qin Gu, Ruizhe Liu, Jing Tian, Aria Wang, Charles Wu, Yimeng Zhang, Zoe Zhou; and my "twalking" buddies: Susan Kuo, Sweyta Lohani, and Krystina Teoh. I know I would never make it to the end of the program without their company. I also wanted to thank Veronique Schreurs who welcomed me to her home and greatly supported my transition to the new culture.

Finally, I wanted to thank my parents, sisters, brothers-in-law, and Zhifeng Xu who are always there for me.

1.0 Introduction

1.1 The complementary learning systems model of word learning

Even native adult speakers continuously update their vocabulary knowledge by learning new words through reading and communication (Brysbaert, Stevens, Mandera, & Keuleers, 2016). Although new words can be successfully recognized shortly after minimal exposure, they are not necessarily represented in the mental lexicon in the same way as existing words. For example, existing words with similar pronunciations compete with each other during word recognition (Marslen-Wilson, 1987). Behavioral studies have shown that new words (e.g., “cathedruke”) do not usually compete with existing words with similar pronunciations (e.g., “cathedral”) until the occurrence of offline consolidation, which typically involves overnight sleep (e.g., Bakker, Takashima, van Hell, Janzen, & McQueen, 2014; e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003).

To account for the time-dependent change in the processing of new words and existing words, Davis and Gaskell (2009) applied the complementary learning systems model (CLS; Frankland & Bontempi, 2005; Mcclelland, Mcnaughton, & Oreilly, 1995) to the field of word learning. According to the CLS model, the hippocampal learning system quickly encodes episodic memories of learning events. These memories are gradually transformed to neocortex-based semantic memory through repeated memory “replay”, especially during overnight consolidation. Within this framework, Davis and Gaskell (2009) proposed that episodic memories of a new word are quickly formed with the support of the hippocampal learning system. With these initial memories, learners are able to recognize or recall the new word by activating relevant episodic memories. After memory transformation through offline consolidation, a new word is gradually integrated into the mental lexicon. Only after becoming part of the mental lexicon is a novel word represented and processed like an existing word.

Based on the CLS model, the processing of new words becomes less hippocampus-dependent and relies more on the neocortex as they are integrated over time. Recent learning studies have shown that the left posterior middle temporal gyrus (pMTG) is associated with the formation of new lexical representations for newly learned words (Landi et al., 2018; Takashima, Bakker, van Hell, Janzen, & McQueen, 2014, 2017). The activation in the left pMTG during the processing of novel words becomes stronger after offline sleep occurred, especially when novel words have been paired with meanings or when word meanings contain richer semantic features (Ferreira, Gobel, Hymers, & Ellis, 2015; Takashima et al., 2014, 2017). One proposal is that this region simultaneously activates the neocortical areas associated with representations of different lexical constitutes and binds the activated information together through theta oscillations (Bakker-Marshall et al., 2018).

The importance of the left pMTG in word processing has been widely acknowledged. In the dual-stream model, the left pMTG is the lexical interface – mapping word forms onto their meanings (Gow, 2012; Hickok & Poeppel, 2004, 2007). In the Memory, Unification, and Control (MUC) model, this region is relevant to the storage and retrieval of lexical knowledge including lexical-semantics (Hagoort, 2005). A related argument is that the left pMTG is associated with the representations of underspecified but core semantic features and activates more specific representation in relevant neocortical areas (Papeo et al., 2015).

Overall, the left pMTG plays a crucial role in the storage and processing of word knowledge. It is possible that the left pMTG slowly takes over the binding role of the hippocampus over time. Before the shift is completed, both the left pMTG and the hippocampus support word processing (Takashima et al., 2014, 2017).

1.2 Learning new meanings for known words

In addition to learning novel words, people often add new meanings to words they already know, thus resulting in the update of word knowledge. For example, one may learn that the word *skate* can refer to a type of fish after knowing its roller (or ice)-skating meaning. Different from learning novel words, learning new meanings for known words requires meaning learning only, without creating new lexical representations. In addition, this type of learning involves an intensive interaction between new and prior knowledge, because prior word knowledge is automatically activated upon word presentation (Humphreys, Evett, & Taylor, 1982; Lesch & Pollatsek, 1993; Perfetti, Bell, & Delaney, 1988). The reactivation of prior word knowledge can affect the encoding and integration of new information (Schlichting & Frankland, 2017). In particular, neurons that are relevant to the representation of prior word knowledge, including word forms and original meanings, are likely to be reactivated and become part of the representation of new meanings during the initial encoding. During offline replay of the newly encoded memory, these neurons are likely to be reactivated, further enhancing the co-activation of new and prior knowledge. The boosted co-activation may facilitate the integration of new meanings into the mental lexicon and make learning less reliant on overnight consolidation.

However, when new and original meanings are semantically unrelated, the co-activation of new and prior knowledge is likely to involve interference among meanings, possibly slowing down integration during some stages of learning. Previous studies on the learning of new meanings for known words have shown that interference from original meanings can hinder the initial acquisition of new meanings, especially when new meanings are semantically unrelated to the original one (Fang & Perfetti, 2019; Rodd et al., 2012). However, over time, the interference seemed reduced and did not necessarily pose a disadvantage one day or one week following learning (Fang & Perfetti, 2019). It is possible that more distinctive or less overlapped representations for different meanings are created over time. However, given the characteristic differences relative to the learning of novel words, it remains unclear whether learning new meanings for known words involves the shift from episodic- to semantic-based processing over time and whether the left pMTG is the functional area for the binding of new meanings to known words.

1.3 Aims and studies

This dissertation examines the mechanisms underlying the learning of new meanings for known words. In particular, I focused on: (1) the role of overnight consolidation in learning new meanings for known words (Chapter 2); (2) the role of left pMTG in binding new meanings to known words (Chapter 3).

Chapter 2 (Study 1) investigates how overnight consolidation affected the learning of new meanings for known words using event-related potentials (ERPs). Participants studied new meanings for two sets of words on two consecutive days – one set before overnight sleep occurred and the other after. Behavioral and ERP responses to the words were compared. The findings showed faster semantic judgments on the new meanings after overnight consolidation, suggesting faster meaning access over time. While learning new meanings did slow access to original meanings on the day of learning, such effect was absent 24 hours later. The ERP data showed that overnight consolidation did not affect lexical-semantic processing significantly. Instead, episodic retrieval seems to play an equally important role in meaning access within the 24 hours. Overall, overnight sleep does play a beneficial role in the learning of new meanings for known words, even though new meanings are not fully integrated within the 24 hours.

Chapter 3 reports two studies and investigates the role of left pMTG in the learning of new meanings for known words. Study 2a first demonstrated that presenting word meanings as verbal definitions is sufficient to drive the activation of relevant sensorimotor features associated. In particular, after multiple-session training, ERPs evoked by novel spoken words paired with definitions describing action and non-action features started showing a difference (i.e., semantic category effect) around 100 ms when a novel word can be uniquely identified. This suggests rapid and (arguably) automatic activation of semantic features during word recognition, possibly involving relevant sensorimotor cortices.

With these findings, Study 2b further used Magnetoencephalography (MEG) to examine the role of the left pMTG and that of the sensorimotor cortices in learning new action meanings for known words and for novel words. The MEG data showed different parts of the sensorimotor circuits are involved in accessing the new action meanings for novel words and for known words. Enhanced involvement of the left MT+, the human homologue of the monkey motion complex, was observed in the processing of new action meanings of novel words. In contrast, enhanced involvement of the left frontal motor regions (BA44 and lateral precentral gyrus) that are relevant to the representation of abstract action meanings, was found for known words. In addition, enhanced source activation in the left pMTG was observed when participants were accessing the new meanings of known words, while there was only a trend of enhanced source activation for meaning access in novel words. Overall, the findings suggest new meanings of known words are more integrated into the mental lexicon than those of novel words and that the left pMTG may play an important role in binding the new meanings to known words.

In Chapter 4, based on the reported studies and previous work, I proposed how the co-activation of new and prior word knowledge is involved in the learning new meanings for known words within the standard model of system consolidation (Frankland & Bontempi, 2005). Furthermore, the findings from the learning of new meanings provide implications for the mechanisms underlying word learning in general.

2.0 Study 1: The Role of Overnight Consolidation in Learning New meanings for Known Words

2.1 Introduction

Based on a large-scale study, native speakers of American English aged 20 to 60 learn, on average, a new word every other day (Brysbaert et al., 2016). While a new word may be recognized shortly after minimal exposure, learning is not completed when the initial learning event ends. Instead, to integrate a new word into the mental lexicon, offline memory consolidation during overnight sleep is usually needed (Davis & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). However, most of the evidence for the importance of overnight consolidation comes from the learning of novel spoken or written words, when new lexical representation is established. Another type of word learning that occurs very often, however, is less studied – the learning of new meanings for previously known words (Fang, Perfetti, & Stafura, 2017; Fang & Perfetti, 2017, 2019; Maciejewski, Rodd, Mon-Williams, & Klepousniotou, 2019; Rodd et al., 2012). For example, one might learn that the word ‘skate’ refers to a type of fish long after knowing its more common roller- or ice-skating meaning. Different from learning a novel word, this type of learning requires updating the lexical knowledge by associating new information with an existing word. The current study examined the role overnight consolidation may play in the learning of new meanings for known words.

Within the framework of complementary learning systems (Kumaran, Hassabis, & McClelland, 2016; Mcclelland et al., 1995), Davis and Gaskell (2009) proposed that learning a novel word involves two learning and memory systems. The initial memories about a novel word are encoded as episodic memories of specific learning events. Over time, those memories are transformed to semantic memories, especially during overnight sleep. The core mechanism for memory transformation is memory replay or reactivation (Diekelmann & Born, 2010; Dudai, Karni, & Born, 2015). In particular, the hippocampus and surrounding areas replay the encoded episodic memories repeatedly, leading to re-activation of neocortical areas relevant to new memories and also those relevant to older and similar memories. Such repeated co-activation gradually establishes direct connections among neocortical areas, eventually leading to the formation of hippocampus-independent semantic memories and integration of new information (Frankland & Bontempi, 2005). In the case of word learning, the formation of a semantic memory indicates the integration of a novel word into the mental lexicon and allows it to be represented and processed like a previously acquired word.

To investigate the role of overnight consolidation in word learning, a study examined behavioral and ERP responses to novel words that were studied either before or after overnight sleep (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015). After learning both sets of words, participants made semantic judgments on the studied words while EEG was recorded. Behavioral data showed that participants were faster in responding to the words studied prior to overnight sleep than those after. Furthermore, neural responses to novel words were also influenced by overnight sleep, as indicated by the difference in the N400 component. The N400 component, a negative going component peaking around 400 ms after word onset at the central midline site of scalp, is a classical indicator of semantic processing (Kutas & Federmeier, 2011). A more negative N400 usually means more effortful meaning access or more extensive search of semantic memory. Bakker et al. (2015) found a reduced N400 within 300-500 ms for words that were studied before overnight sleep than those after. Furthermore, the N400 became comparable with that for existing words after overnight sleep. A series of behavioral studies also argues for the importance of overnight consolidation in integrating new words into the mental lexicon (e.g., Bakker et al., 2014; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003; Wang et al., 2016).

Different from learning novel words, learning new meanings for known words involves an intensive interaction between new and prior knowledge during the encoding of new meanings, because prior word knowledge is automatically activated upon word presentation (Humphreys et al., 1982; Lesch & Pollatsek, 1993; Perfetti et al., 1988). This provides opportunity for the co-activation of neocortical areas that are relevant to the representation of new meanings and those relevant to the representation of prior word knowledge, even before overnight consolidation occurs. As a result, integration of new meanings may occur faster and rely less on overnight consolidation. However, because of the interference between new and original meanings, integration may need more overnight consolidation so that the learning of new meanings does not affect the previously acquired word-meaning mappings.

A recent behavioral study examined the time-dependent changes in the processing of both new and original meanings (Fang & Perfetti, 2019). In the study, participants learned new meanings for high and low frequency words, and the processing of both new and original meanings were tested at three time points: immediately following learning, one day later, and one week later. The results showed that participants remembered new meanings of high frequency words better one week following the initial learning, even though immediate retention of new meanings suffered more from the interference from the original meanings. In addition, participants became slower in making semantic relatedness judgments based on original meanings on the day of learning, but not one day or one week later. The time-dependent changes suggest that, as in the learning of novel words, offline consolidation may also be essential to the long-term retention and integration of new meanings.

The current study examined the role of overnight consolidation in the learning of new meanings for known words by comparing words that were studied on two different days, as in Bakker et al. (2015). New meanings for one set of words were studied prior to overnight sleep while the other set was studied after. Additionally, the same number of words was presented without new meanings on each day, serving as exposure controls. Behavioral and ERP measurements for the processing of new meanings were taken. In addition, the processing of original meanings was tested with a semantic category judgment task, aiming to replicate and extend the previous findings by Fang and Perfetti (2019) by using a task that required participants to make judgments only on the original meanings and including only one test point.

If overnight consolidation benefited the integration of new meanings, we would expect more automatic and faster processing of new meanings following overnight consolidation, as found in the learning of novel words (Bakker et al., 2015). If more automatic processing was a result of enhanced involvement of semantic memory and reduced episodic memory, we expected this would be reflected by the ERP components relevant to the processing of episodic and semantic memory. If new meanings are integrated over time and more semantic processing is involved in the processing of new meanings, a difference in N400 amplitude between words studied before and after overnight sleep occurred was expected. Previous studies have shown that semantically ambiguous words evoke a more negative N400 than unambiguous words, reflecting that more information is accessed from the semantic memory and possibly also competition among meanings (Lee & Federmeier, 2006). In the current study, if new meanings were integrated into the mental lexicon following overnight consolidation and were accessed from the lexico-semantic memory, a more negative N400 was expected.

In terms of episodic memory processing, the left parietal positivity, typically observed after 500 ms, is an ERP indicator of memory recollection (Maratos, Allan, & Rugg, 2000; Rugg & Curran, 2007; Tsivilis et al., 2015). If the involvement of episodic memory reduced over time, we would expect to see a reduced left parietal positivity. Another ERP component that is often related to episodic memory processing is the mid-frontal negativity. This component is often associated with memory recognition based on subjective familiarity rather than memory recollection (Rugg & Curran, 2007; Tsivilis et al., 2015). Previous studies suggest that stimuli presented within the preceding 40 minutes and those presented one day or four weeks ago did not differ in the mid-frontal negativity (Curran & Friedman, 2004; Tsivilis et al., 2015). Therefore, we did not expect any differences between words that were studied 24 hours apart in this component. Furthermore, if any of the above effects truly reflected the changes in meaning processing over time, we would see that overnight sleep affected words with new meanings and exposure controls differently.

2.2 Methods

2.2.1 Participants

Thirty right-handed native English speakers (15 females, 18.33 ± 0.61 years old) participated in the study. They had normal or corrected-to-normal vision and none reported any learning or language disabilities. Based on self-reports, the participants had 7.29 ± 1.25 hours of sleep (range: 4.5-11 hours) the night after studying Day 1 words and most of them reported that the sleep quality was the same as the past month (4.00 ± 1.02 in a 7-point scale, ranged from 2 to 6; 1 = worse, 7 = better). Participants provided written informed consent before the experiment and received course credits for their participation. The procedure of the study was approved by the institutional review board at the University of Pittsburgh.

2.2.2 Stimuli

Words

Word stimuli included 64 concrete nouns selected from a database by Medler, Arnoldussen, Binder, and Seidenberg (2005). On a scale of 1-6, all of the words had low ratings for the attributes of motion (0.64 ± 0.44), manipulation (2.13 ± 0.68), sound (0.78 ± 0.56), and emotion (0.64 ± 0.94). The words had 4.42 ± 0.97 letters (range: 3-6) and were rated as highly concrete nouns (4.83 ± 0.17, from Brysbaert, Warriner, & Kuperman, 2013). Each word had only one meaning but may have had multiple senses according to the Wordsmyth English Dictionary (Parks, Ray, & Bland, 1998). The words were separated into four groups, one group for each of the four conditions (Remote/Recent * Meaning/Control; see Table 1 for examples and Appendix B.1 for full list), with the assignment of words to the conditions counterbalanced across participants. Within each group, half of the words refer to man-made objects and the other half referred to natural objects. Additionally, 64 filler words, half referring to man-made objects and the other half referring to natural objects, were selected and used in the semantic category judgment task. Filler words and trained words were matched in word frequency and number of letters (both $ps > .26$).

Meanings

Thirty-two new meanings were created and separated into two groups (see Appendix B.2 for full list). Within each group, half of the meanings described specific actions involving hand, arm, or finger movement, while the other half described actions involving foot, leg, or toe movement. One group was used for each day, with the assignment of meanings to Remote and Recent conditions counterbalanced across participants. For exposure controls, they were paired with a string of asterisks.

Table 1. Stimulus examples (trained words)

Condition	Word	Meaning
Remote (Day 1)		
Meaning	cloud	lifting with one hand
Control	stone	******************
Recent (Day 2)		
Meaning	grass	walking backwards
Control	lamp	******************

Notes: Sixteen items for each condition. The assignment of words and meanings to conditions was counterbalanced across participants.

2.2.3 Procedure

The overview of experimental procedure is presented in Figure 1. Participants learned new meanings for one group of words and were exposed to the same number of exposure controls on

Day 1 (i..e, remote words). On the second day, they learned a second set of words with the same procedure (i.e., recent words). Then participants were tested on both recent and remote words before they performed a semantic relatedness judgment task tapping the processing of new meanings while EEG was recorded. Following that, they performed a semantic category judgment task on original meanings of both sets of words, and finally completed a vocabulary test and some questionnaires (results of vocabulary test and questionnaires are not reported here). The procedure of each task is described below.

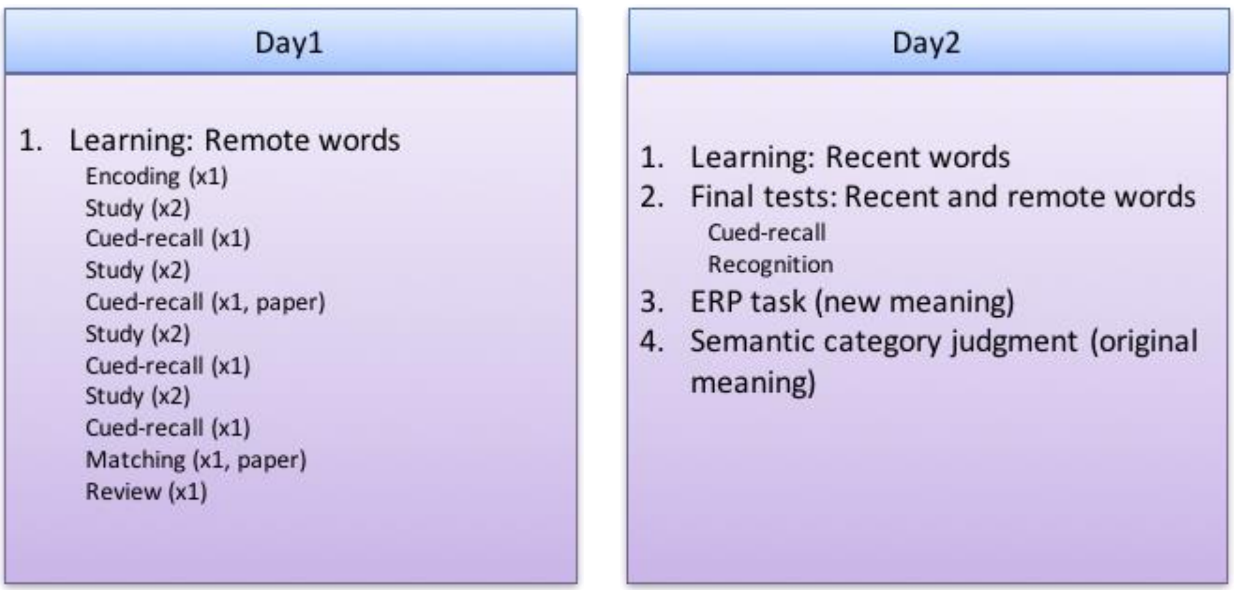

Figure 1. Tasks participants performed on each day.

The procedure of learning of recent words (Day 2 words) is the same as that of learning remote words (Day 1 words). Number in parentheses indicates the number of exposures to the trained words in different stages of the learning phase.

Learning

Participants first encoded words with new meanings and the exposure controls. They were presented with a word for a second, and then either a definition or a row of asterisks was presented for four seconds below the word and. Each word was presented once in the encoding stage. To facilitate their learning following the initial encoding, participants studied the words eight more times and performed a cued-recall test after every two times of studying. In the study trials, participants were presented with a word and were instructed to recall what had been paired with it before pressing the space bar to see the answer and restudy the meaning. After that, they pressed the space bar to continue to the next word. In the first, third and the fourth tests, participants typed out the new meanings for each word, and typed "n" if a word did not have a new meaning. Following each response, the correct answer was presented for them to study. In the second test, participants were provided with a sheet of paper listing all the words and wrote the new meanings for each word. Again, they wrote "n" for exposure controls. Participants were then given a sheet with the correct answers and instructed to score their answers and reported their accuracy to the experimenter. Following the last cued-recall test, participants performed a matching test, in which words and meanings were presented in two separate columns. The meanings were numbered from 1 to 16. Participants wrote the number of the meaning behind the corresponding words and wrote "n" for exposure controls. They then scored and reported their performance to the experimenter. As the last task of Day 1, participants reviewed all the words and their new meanings before they left and the procedure was the same as in the study trials. On Day 2, participants learned another set of words (i.e., recent words) with the same procedure.

Final tests on recent and remote words

On Day 2, following learning the second set of words, participants were tested on both sets of words with a cued-recall (typing) test and a recognition (multiple-choice) test. The procedure of the typing test was the same as described above except both recent and remote words were presented and no feedback was provided. In the multiple-choice test, one word and four options were presented on each page, and participants were instructed to select what has been paired with each word. Among the four options, the fourth option was always a string of asterisks. The overall accuracy and correct answers for all the words were provided upon completion of the task.

ERP task (new meanings)

Participants read the meaning-taught words and the exposure controls, which were presented at the center of the screen one by one (see Figure 2). They were instructed to read each word carefully. Following 18.75% of the words, they saw an underlined phrase (e.g., "<u>hand movement</u>") and needed to judge whether the new meaning of the preceding word was semantically related to the phrase. Sometimes the phrase "has a new meaning" or "without a new meaning" was presented and participants needed to judge whether a word had been paired with a new meaning. Participants were not aware of when a phrase would be presented and what the content would be so that expectation and motor preparation was minimized. Each trial began with a fixation (600-700 ms), followed by a blank screen (200 ms). Then a word was presented for 1 000 ms and followed by 1 500 ms of either a blank screen or a phrase. If a phrase was presented, it stayed on the screen until participants responded and the next trial began after 1 500 ms of blank screen. Participants were instructed to stay still during the task and to blink only after a word disappeared but before the next fixation was presented. There were four blocks, and each of the trained words were presented once in each block. Therefore, there were 64 trials per condition.

Participants took a break of at least 20 seconds between blocks and each block lasted around four minutes. A short practice session with 12 additional words was administered twice at the beginning of the task to familiarize participants with the task procedure.

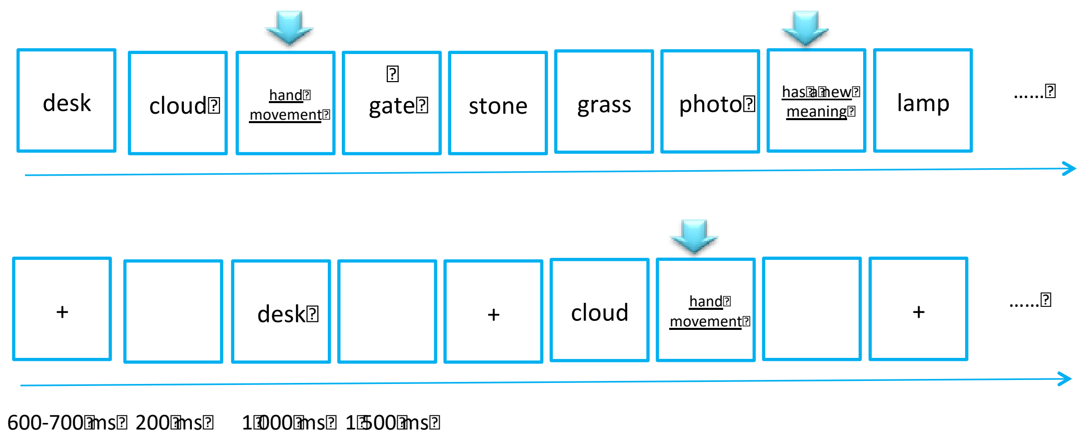

Figure 2. Procedure for ERP task.

Fixations and inter-trial intervals are omitted in the upper panel but shown in the lower panel.

Semantic category judgment task (original meanings)

Participants made semantic category judgments on the 64 trained words and 64 filler words. The trial procedure was largely the same as in previous studies using the same task (Bowers, Davis, & Hanley, 2005; Coutanche & Koch, 2017; Coutanche & Thompson-Schill, 2014; Wang et al., 2016). Each trial began with a central fixation (800 ms), followed by a blank screen (350 ms) and a word (500 ms). Participants were asked to judge whether the original meaning of a presented word referred to a man-made object or a natural object as quickly and as accurately as possible. They indicated their judgments by pressing one button for man-made objects and another button for natural objects. Following each response or 1500 ms after the onset of a word, feedback (i.e., correct, incorrect, or no response detected) was presented for 1000 ms before the next trial began. The task lasted around seven minutes and participants took a break in the middle of the task. Participants received practice trials with 12 additional words before they performed the task.

2.2.4 EEG data acquisition and preprocessing

Participants were fitted with a Geodesic Sensor Net with a 128 Ag/AgCl electrode array and data were recorded using the associated NetStation acquisition software (Electrical Geodesics, Inc., Eugene, OR) with a sampling rate of 1 kHz and a hardware bandpass filter of 0.01–200 Hz. Data were preprocessed with NetStation Tool software. A bandpass filter of 0.1–30 Hz was applied, and then data were segmented into 900-ms epochs, starting 100 ms before the onset of word stimuli. Epochs with artifacts were rejected for further analysis, including eye blink (exceeding±140 μV), eye-movement (exceeding 55 μV), and extreme variance (larger than 100 μV). Channels with artifacts on more than 20% of epochs were marked as bad channels, and data from surrounding channels were used for interpolation. After artifact rejection, on average, there were 59.09 ± 4.78 out of 64 valid epochs per condition (ranged from 42 to 64). For each trial, data were referenced to the average of whole scalp and then baseline correction was performed (-100-0 ms).

2.2.5 Data analysis

Missing data statement

Data from three participants were not complete. For two participants, data from the final cued-recall tests on both recently and remotely learned words in Session 2 were missing. We did include both participants in the analysis of all the other tasks where their data were available. The third participant completed all the tasks except the EEG task, because of the malfunction of the

EEG equipment. We excluded the participant from the analysis of the semantic judgment task that was administered right after the EEG task for all the other participants. Not performing the EEG task that required the retrieval of new meanings four times may influence the access to the original meanings.

Behavioral data

Behavioral data were analyzed using linear mixed effects modeling with the lme4 package in R (Baayen, Davidson, & Bates, 2008). Participants' responses in the cued-recall tests were scored from 0 to 5 independently by two trained research assistants who were blind to conditions, based on how close their responses were to the studied meanings (see Appendix B.3 for rubric). Responses with differences between the two scores larger than 1 were discussed before a final score was assigned. For other responses, the average scores were the final scores. In the recognition tests, participants reached an accuracy of 100% in the one of the conditions; therefore empirical logit was calculated and by-subject and by-item analyses were performed separately (Donnelly & Verkuilen, 2017). Mixed effects logistic regression was conducted to analyze the accuracy data from the ERP and semantic category judgment tasks. For response time data, incorrect trials and trials with response times 2.5 SD beyond the mean were excluded and then inverse transformation was performed before they entered linear mixed effects modeling.

For all of the tasks, fixed effects included Day (Remote vs. Recent) and Type (Meaning vs. Control). However, the fixed factors were coded differently to better capture the effects of most interest in different tasks. In the tasks on the new meanings (cued recall, recognition, and ERP task), Day was effect-coded and Type was treatment-coded to detect the effect of Day in the meaning conditions and to learn whether the effect of Day is different there from that in the control conditions. In the semantic category judgment task, which was designed to examine the effect of meaning learning compared to mere exposure, Type was effect-coded and Day was treatment-coded, to show the effect of Type on the day of learning and how it changed over time, as in our previous study (Fang & Perfetti, 2019). Random effect terms included intercepts of subject and item, except that only intercept of subject or intercept of item was included in the by subject- and by item analysis of empirical logits in the recognition test. A by-subject or by-item slope was added if model comparisons showed a significant contribution and models converged.

EEG data

The electrodes were grouped into 11 clusters (See Figure 3). Two time-windows were of most interest: 300-500 ms and 500-800 ms. Three clusters were of most interest: the central midline cluster (Cz), the middle frontal cluster (Fz), and the left parietal cluster (P3). For each cluster within each time window, a linear mixed effects modeling on mean amplitude within the time window and across channels for single trials was conducted. As in other tasks tapping new meanings, Day was effect-coded while Type was treatment-coded. Therefore, three effects are examined; simple effect of Day in the two Meaning conditions, main effect of Type, and the interaction between Day and Type. To avoid missing results that were not anticipated, for each of the three effects, the spatio-temporal data were searched for any significant clusters and multiple-comparison was corrected with cluster-based permutation tests (Maris & Oostenveld, 2007) using MNE-Python (Gramfort et al., 2013; Gramfort et al., 2014). For the sake of computational efficiency, the analysis was performed on ERP data that were generated by averaging valid trials under the same conditions for each participant rather than on single trial data.

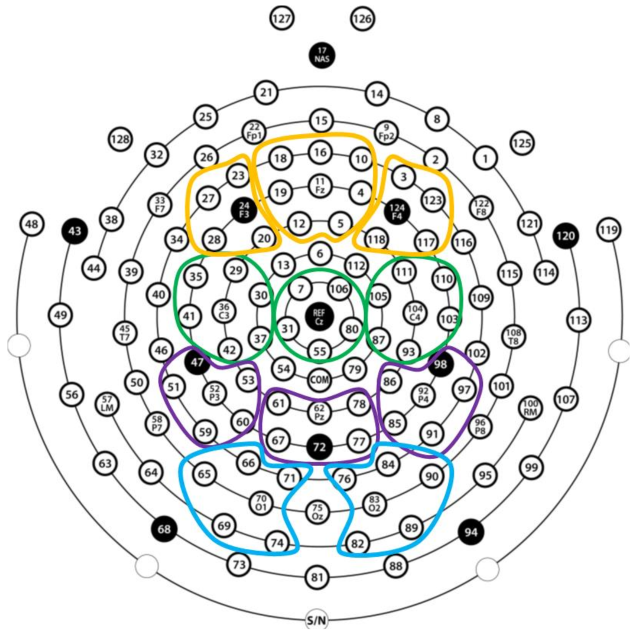

Figure 3. The definition of clusters.

Orange: F3, Fz, and F4; Green: C3, Cz, and C4. Purple: P3, Pz, and P4. Blue: O1 and O2.

2.3 Results

2.3.1 New meanings: Cued-recall and recognition tests

As shown in Table 2 and Figure 4, participants recalled more information about the new meanings of recently presented (Day 2) words than those of remotely presented words (Day 1; $t = -5.875$, $p < .001$), suggesting memory decay over 24 hours. This difference was larger than in the control words, as indicated by a significant interaction between Day and Type ($t = -6.755$, $p < .001$). The same pattern was found in the recognition or multiple-choice tests, as revealed by both by-subject and by-item analyses (*p*s < .01). Additionally, in both tests, participants performed better in the control conditions than in the meaning conditions (*p*s < .001).

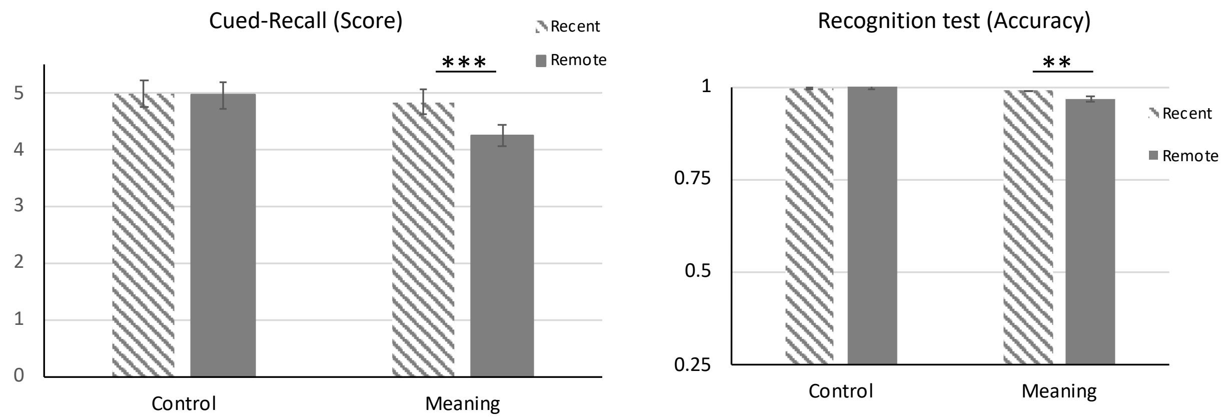

Figure 4. Participants' performance in the cued-recall (left) and recognition (right) tests.

Error bar represents 1 SEM with between-participant variance removed (Franz & Loftus, 2012; same for other figures). ***: $p < .001$; **: $p < .01$.

Table 2. Fixed effect estimates for mixed effects models of behavioral tests on new meanings

Fixed effect	***beta***	***SE***	***t***	***p***	
Cued-recall (typing)					
Intercept	4.555	0.077	58.884	< .001	***
Day	-0.590	0.101	-5.875	< .001	***
Type	-0.434	0.082	-5.305	< .001	***
Day:Type	-0.568	0.084	-6.755	< .001	***
Recognition (multiple-choice)					
by-subject					
Intercept	3.177	0.058	54.632	< .001	***
Day	-0.367	0.106	-3.451	0.001	**
Type	-0.280	0.075	-3.724	< .001	***
Day:Type	-0.445	0.151	-2.954	0.003	**
by-item					
Intercept	2.601	0.030	87.082	< .001	***
Day	-0.186	0.060	-3.106	0.002	**
Type	-0.150	0.042	-3.548	< .001	***
Day:Type	-0.224	0.084	-2.656	0.008	**

Notes: The intercept represents averaged performance in meaning conditions across recent and remote words; Day represents difference between recent and remote words in the meaning conditions (i.e., simple main effect); Type represents overall difference between control and meaning conditions. Model in the cued-recall tests: lmer (Score ~ 1 + Day*Type + (1+Day+Type|Subject) + (1|Item)); models in the recognition test: lmer (EmpLogit ~ Day * Type + (1 | Subject)) for by-subject analysis and EmpLogit ~ Day * Type + (1 | Item) for by-item analysis. EmpLogit = empirical logit of accuracy. ***: $p < .001$; **: $p < .01$; *: $p < .05$.

2.3.2 New meanings: ERP task

Behavioral performance

As shown in Figure 5 and Table 3, participants performed the ERP task after learning the recent words on Day 2. They overall were more accurate and faster at making judgments on control words than on words with new meanings (both *p*s < .001). For the words with new meanings, remote words were responded to more accurately ($z = 2.154$, $p = .012$) and faster ($t = 3.082$, $p = .002$). In contrast, for the control words, accuracy was marginally higher for recent condition ($z = -1.912$, $p = .056$), and the difference in response times was not significant ($t = -0.106$, $p = .916$).

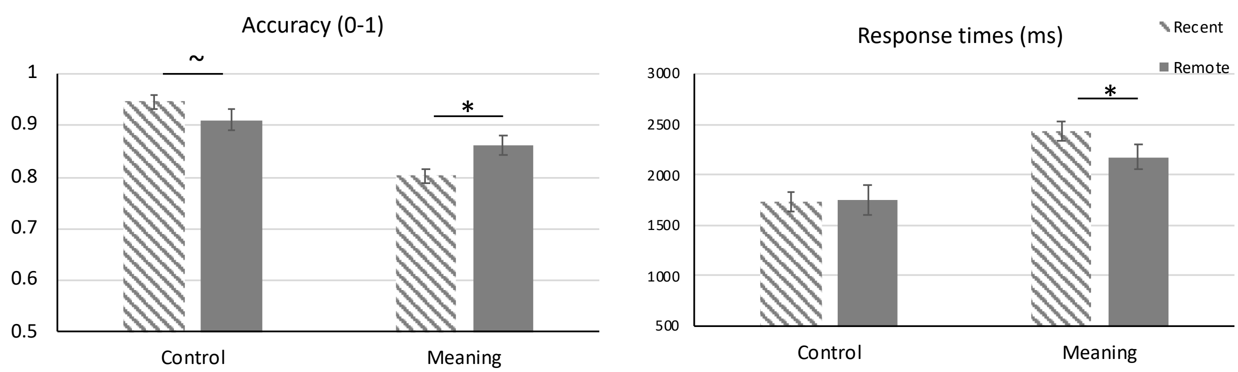

Figure 5. Accuracy (left) and response times (right) of semantic judgments in the ERP task.

Table 3. Fixed effect estimates for mixed effects models of behavioral performance in the ERP task

Fixed effect	*beta*	*SE*	*t or z*	*p*	
Accuracy					
Intercept	2.091	0.228	9.187	< .001	***
Day	0.595	0.237	2.514	0.012	*
Type	1.233	0.330	3.738	< .001	***
Day:Type	-1.231	0.409	-3.013	0.003	**
Response times					
Intercept	0.578	0.045	12.764	< .001	***
Day	0.058	0.019	3.082	0.002	**
Type	0.159	0.029	5.060	< .001	***
Day:Type	-0.060	0.026	-2.297	0.022	*

Notes: The intercept is averaged performance in meaning conditions across recent and remote words; Day is difference between recent and remote words in the meaning conditions (i.e., simple main effect); Type represents overall difference between control and meaning conditions. Model in the accuracy: glmer (Accuracy ~ Day * Type + (1 + Type | Subject) + (1 | Item)); models in the response times: lmer (Inversed response times in seconds ~ Day * Type + (1 + Type | Subject) + (1 | Item)). ***: $p < .001$; **: $p < .01$; *: $p < .05$.

ERP results

Figure 6 shows the ERPs evoked by the four types of words in the 11 predefined clusters when participants were making semantic judgments based on the new meanings. Findings from the clusters and time windows of interest are reported first, followed by the whole scalp analysis.

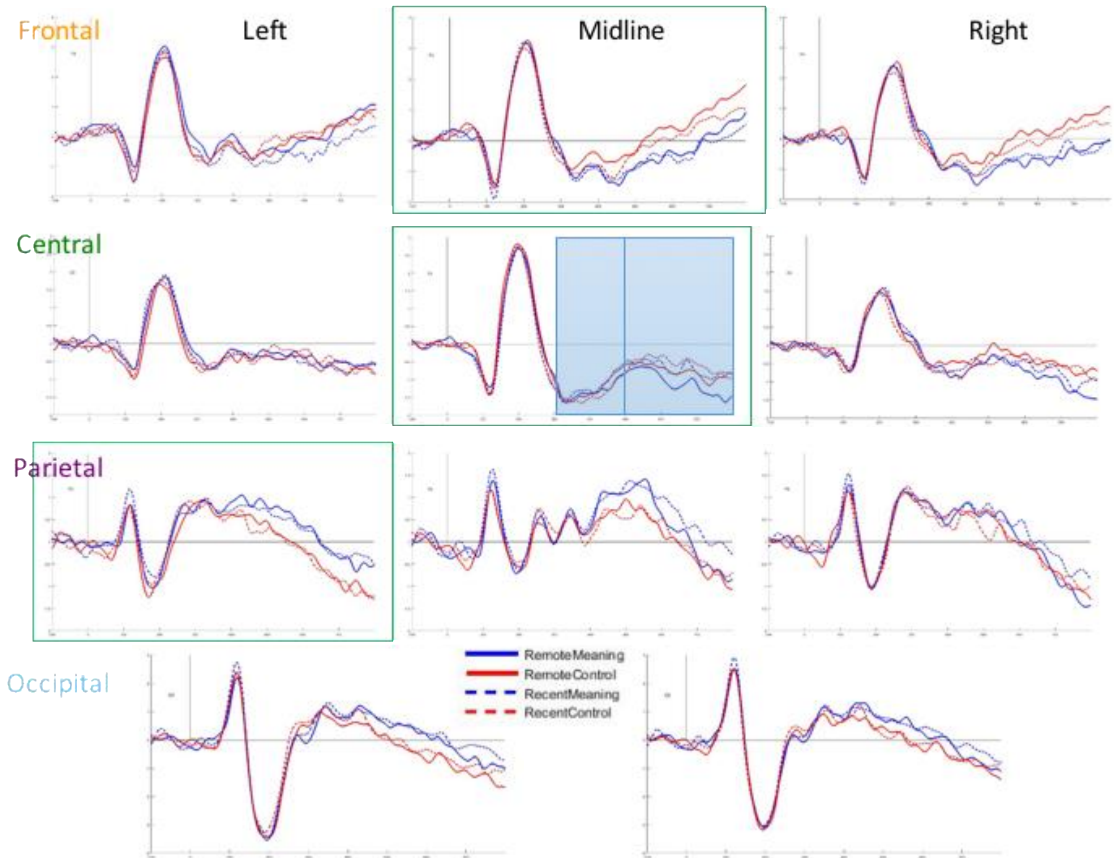

Figure 6. ERP waveforms evoked by the four different types of words.

Clusters of interest (Fz, Cz, and P3) are highlighted with green outline and time windows of interest (300-500 ms and 500-800 ms) are highlighted with blue boxes at the Cz cluster.

At the Cz cluster, among the words with new meanings, amplitudes in response to remote words were comparable to those of recent words within 300-500 ms (p = .742; see Table 4 for details). Within 500-800 ms, remote words yielded a marginally larger negativity than recent words (p = .086). In addition, combining words with new meanings and exposure controls together, remote words overall evoked a marginally larger negativity than recent words within 500-800 ms (p = .071) but not within 300-500 ms (p = .871). However, there was no interaction between Day and Type in either time window (both ps > .30).

At the Fz and P3 clusters, we did not observe a significant difference between recent and remote words that had been paired with new meanings in either time window (all *p*s > .48). Furthermore, there was no interaction between Day and Type (all *p*s > .32). Instead, we observed a robust main effect of Type. Compared to exposure controls, words with new meanings evoked a larger negativity in both time windows at the frontal site (500-700 ms: $p = .036$; 500-800 ms: $p < .001$), and a larger positivity in the left parietal site especially in the later time window (500-700 ms: $p = .105$; 500-800 ms: $p < .001$).

Table 4. Fixed effect estimates for mixed effects models of amplitudes in the ERP task

Fixed effect	*beta*	*SE*	*t*	*p*	
Cz: 300-500 ms					
Intercept	-1.304	0.267	-4.882	< .001	***
Day	-0.056	0.171	-0.329	0.742	
Type	-0.024	0.146	-0.162	0.871	
Day: Type	0.249	0.243	1.025	0.305	
Cz: 500-800 ms					
Intercept	-0.929	0.259	-3.585	< .001	***
Day	-0.311	0.181	-1.716	0.086	~
Type	0.298	0.165	1.804	0.071	~
Day: Type	0.197	0.257	0.767	0.443	
Fz: 300-500 ms					
Intercept	-1.059	0.313	-3.386	< .001	***
Day	0.027	0.192	0.144	0.886	
Type	0.293	0.140	2.091	0.036	*
Day: Type	0.267	0.271	0.984	0.325	
Fz: 500-800 ms					
Intercept	-0.209	0.331	-0.632	0.527	
Day	0.145	0.209	0.692	0.489	
Type	0.756	0.150	5.045	< .001	***
Day: Type	0.216	0.296	0.729	0.466	
P3: 300-500 ms					
Intercept	0.801	0.250	3.203	0.001	**
Day	0.067	0.157	0.424	0.672	

Type	-0.215	0.132	-1.622	0.105	
Day: Type	0.016	0.222	0.070	0.944	
P3: 500-800 ms					
Intercept	0.046	0.236	0.196	0.844	
Day	0.000	0.170	-0.002	0.999	
Type	-0.567	0.139	-4.086	< .001	***
Day: Type	0.026	0.241	0.110	0.913	

Notes: The intercept is averaged amplitude in meaning conditions across recent and remote words; Day represents difference between recent and remote words in the meaning conditions (i.e., simple main effect); Type represents overall difference between control and meaning conditions. Final model: lmer (Amplitude ~ Day * Type + (1 + Type | Subject) + (1 | Item)). ***: $p < .001$; **: $p < .01$; *: $p < .05$.

The whole scalp analysis yielded two spatio-temporal clusters for the main effects of Type: A frontal negativity (cluster $p = .003$) and a left lateralized parietal positivity (cluster $p = .006$) for words with new meanings than exposure controls (see Figure 7). None of the other effects were significant after multiple-comparison correction (*p*s > .06 for the main effect of Day in the Meaning conditions and *p*s > .30 for interaction).

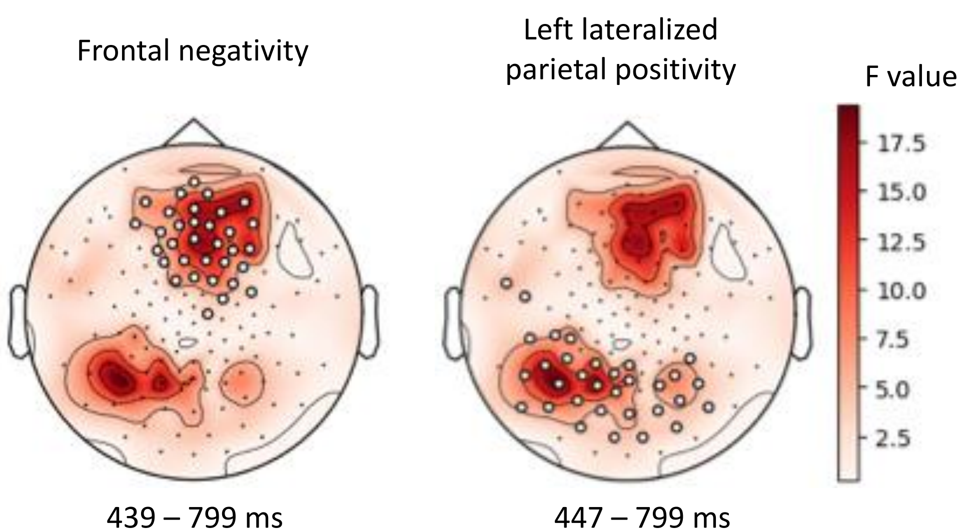

Figure 7. Two spatio-temporal clusters (F maps) for the main effect of Type (Meaning vs. Control).

Left: frontal negativity; right: left lateralized parietal positivity. White circles in topography indicate the channels in each cluster and the time windows are noted at the bottom of each map. Statistical significance at cluster level was determined through spatio-temporal cluster permutation test (1 000 permutations and initial threshold of $p < .005$).

2.3.3 Original meanings: Semantic category judgment

In this task, participants made semantic category judgments (i.e., man-made or natural) based on the original meanings. The contrast between meaning and control conditions and the change over time were of most interest. As shown in Figure 8 and Table 5, for recent words, participants were slower making judgments on the original meanings when words had been paired with new meanings than when they had merely been exposed (t = -2.147, p = .032). However, the difference was reduced over time as indicated by the significant interaction between Type and Day (t = 2.056, p = .040). An additional analysis on the remote conditions further showed the difference between meaning and control conditions was absent (t = 0.769, p = .442). In contrast, neither meaning learning nor time affected the accuracy in making judgment on original meanings (ps > .12).

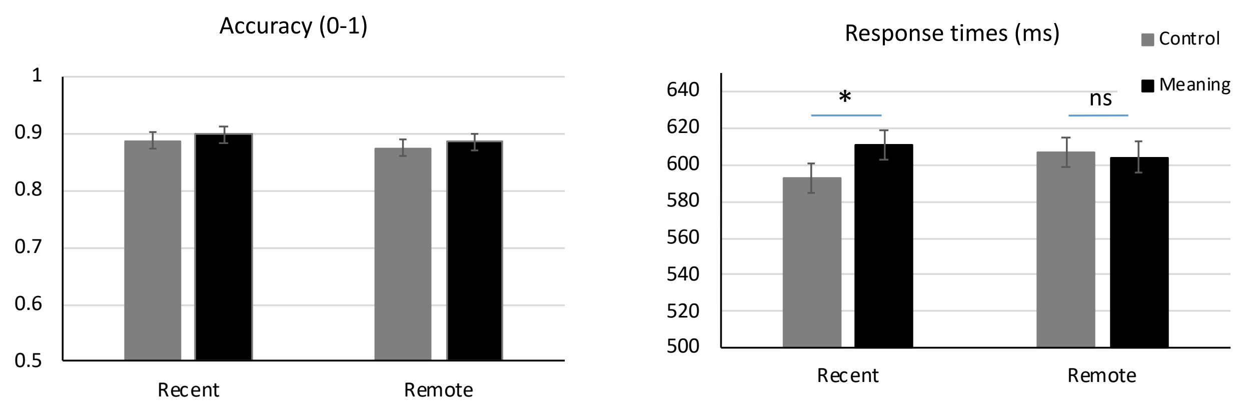

Figure 8. Accuracy (left) and response times (right) in the semantic category judgment task.

Table 5. Fixed effect estimates for mixed effects models of semantic category judgments

Fixed effect	*beta*	*SE*	*t or z*	*p*	
Accuracy					
Intercept	2.441	0.181	13.473	< .001	***
Type	-0.111	0.152	-0.732	0.464	
Day	-0.386	0.254	-1.522	0.128	
Type: Day	0.000	0.305	0.001	0.999	
Response times					
Intercept	1.777	0.053	33.629	< .001	***
Type	-0.048	0.022	-2.147	0.032	*
Day	-0.009	0.016	-0.590	0.55	
Type: Day	0.066	0.032	2.056	0.040	*

Notes: The intercept is averaged performance in recent words across meaning and control conditions; Type is difference between recent control and meaning conditions (i.e., simple main effect); Day represents overall difference between recent and remote conditions. Model in the accuracy: glmer (Accuracy ~ Type*Day + (1 | Subject) + (1 +Day | Item)); models in the response times: lmer (Inversed response times in seconds ~ Type*Day + (1 | Subject) + (1 | Item)). ***: $p < .001$; **: $p < .01$; *: $p < .05$.

2.4 Discussion

The current study examined the role of overnight consolidation in the learning of new meanings for known words, by comparing words that were studied on two consecutive days. Behavioral data revealed that access to both new and original meanings was influenced by overnight consolidation. In particular, while explicit memory of new meanings decayed over time, semantic judgments of new meanings became faster. Meanwhile, meaning learning slowed down the processing of original meanings on the day of learning but not 24 hours later. The ERP data showed a larger frontal negativity and a larger left lateralized parietal positivity for words with new meanings than their exposure controls, regardless of when they were tested. This suggests that episodic retrieval continues playing an important role in accessing the new meanings even after overnight sleep occurred. In contrast, the evidence for increased involvement of semantic memory over time was minimal. A larger negativity of borderline significance was found at the central midline site within 500-800 ms when new meanings were accessed 24 hours later than on the day of learning. However, the time-dependent change was not significantly different from that observed on exposure controls.

2.4.1 Beneficial role of overnight consolidation

Following overnight consolidation, semantic judgments of new meanings became more accurate and faster. This is consistent with previous findings on the learning of novel words. For example, participants were faster making semantic judgments on words studied prior to overnight sleep than those studied after (Bakker et al., 2015). It has also been reported that free recall of studied words was better after overnight sleep had occurred (Dumay & Gaskell, 2007). The facilitative effect could reflect strengthening of connections between word forms and the new meanings, possibly resulted from repeated memory replay or reactivation during overnight sleep.

The processing of original meanings is also affected by overnight consolidation. While meaning learning slowed down the access to the original meaning when words were tested on the day of learning, the effect was absent on words studied one day earlier. The finding is consistent with one of our previous studies (Fang & Perfetti, 2019). The slower access to original meanings on the day of learning could be a result of repeated suppression of original meaning during the encoding of new meanings (Fang & Perfetti, 2017, 2019). However, in Fang and Perfetti (2019), the same set of words was tested three times with the same task over a week, making it difficult to conclude whether the absence of slow-down effect one day or one week later was a result of faster meaning selection when the task was repeated. The current study was able to exclude this possibility by having two sets of studied words and only one test point. Furthermore, the previous study used a semantic relatedness judgment task, a task requiring not only accessing the original meanings but also evaluating the semantic relationship between words. In contrast, the semantic category judgment task used in the current study required participants to make judgments only on the original meanings, providing a more direct evidence for how original meanings are accessed.

Taking the time-dependent changes in the processing of both new and original meanings into consideration, it is possible that the function of overnight consolidation is to provide the opportunity to reduce the interference between new and original meanings. In learning new meanings for known words, resolution of such interference is essential, as typically only one of the meanings is appropriate in a language context. Being able to select the wanted meaning and suppress the unwanted meaning efficiently is essential to reading and communication. During overnight consolidation, the replay of both new and original meanings may lead to the formation of even more distinctive representations for different meanings, even though they share the same word form. One possible solution is to create different context nodes for different meanings of the same word (Armstrong & Plaut, 2008). The context nodes may not be established until learners have sufficient opportunities to encounter each meaning in variant contexts. However, once the connections between context nodes and meanings are created, these context nodes would allow for the efficient selection of meanings based on the matching with language input.

As in many previous studies (e.g., Bakker et al., 2015; Gaskell & Dumay, 2003), the current design does not allow us to confidentially disentangle the effect of overnight sleep from that of the time passing or any other activities that occurred within the 24 hours. However, studies that controlled the length of interval time between study and test and manipulated whether overnight sleep was involved can be informative (Dumay & Gaskell, 2007; Wang et al., 2016). For example, in Dumay and Gaskell (2007), novel words were studied at 8pm and tested at 8am on the next day (overnight sleep involved), in contrast to that words were studied at 8am and tested at 8pm on the same day (without overnight sleep). They were able to show that it was overnight sleep rather than just the time that led to the change in the processing of novel words[1]. Although the effect of time passing by has not yet been directly addressed in the learning of new meanings for known words, it is likely that overnight consolidation at least contributes to the observed time-dependent changes in the processing of new and original meanings.

2.4.2 Sustained involvement of episodic retrieval

While behavioral data suggest that new meanings are integrated over time, the ERP data do not seem to speak for that conclusion. To be specific, the left lateralized parietal positivity evoked by words with new meanings was not modulated by overnight consolidation, indicating recollection of relevant learning episodes still supports meaning access even one day after learning. In addition, the larger positivity for words with new meanings than for their exposure controls could reflect the difference in the amount of information associated with learning episodes (Vilberg, Moosavi, & Rugg, 2006). If new meanings are integrated after overnight sleep occurs so that richer semantic information is accessed from the mental lexicon, then we would see an increased N400 over time, as observed in the contrast between ambiguous words and unambiguous

[1] However, such design also confounds the effect of time-of-day.

words (Lee & Federmeier, 2006). Different from our expectation, the typical N400 was not modulated by meaning learning or overnight sleep. We did observe a marginally larger negativity at the central midline cluster, but in a later time window (500-800 ms) when new meanings were accessed 24 hours later than on the day of learning. This suggests that original meanings are processed primarily within the typical N400 time window. New meanings are accessed faster after 24 hours, but still slower than original meanings. It is possible that the integration of new meanings has started but is not completed within 24 hours.

In addition to a larger left lateralized parietal positivity, words with new meanings also evoked a larger mid-frontal negativity than their exposure controls, regardless of whether overnight sleep had occurred when tested. The frontal negativity has been commonly associated with familiarity-based recognition (Rugg & Curran, 2007). Previous studies typically report a larger frontal negativity for new stimuli than recently exposed stimuli when participants are making “old/new” judgments (Maratos et al., 2000; Tsivilis et al., 2015). However, it is unlikely that the subjective familiarity with the words in the current study was different, given that the relatively short interval time between study and test, and all the words were presented in the final cued-recall and recognition tests right before the ERP task. Furthermore, one would expect words with new meanings were more familiar to participants because they were likely to be more attended to during learning, rather than the other way around. Given the very similar patterns in the frontal negativity and in the left parietal positivity, the mid-frontal negativity may be relevant to memory recollection. It is possible that when a task that requires the retrieval of specific content of episodic memory, familiarity-based recognition and recollection are more tightly coupled than when participants only need to be able to recognize a stimulus to make old/new judgments.

In Bakker et al. (2015), novel words evoked a word-like N400 and meaning access became faster after overnight sleep. Although the access of new meanings of known words became faster 24 hours later, we did not find evidence for reduced involvement of episodic memory or enhanced involvement of semantic memory. The slower integration and more sustained involvement of episodic memory may be essential to the learning of new meanings for known words when new meanings are unrelated to original meanings. According to recent updates of the complementary learning systems model, when new information is congruent with prior knowledge, integration can occur faster or even immediately (Kumaran et al., 2016; van Kesteren, Ruiter, Fernandez, & Henson, 2012). If new meanings are semantically related to the original meanings, new meanings may be integrated into the mental lexicon faster and the modulation of overnight sleep on the N400 and left parietal positivity may be observed.

Another explanation for the null effect of overnight sleep on the N400 is that the relative involvement of different memory systems is influenced by the specific task participants are performing. A previous study found a larger negativity for words with new meanings than their exposure controls at the central midline site within 300-700 ms, even on the day of learning when participants were performing a one-back task (Fang & Perfetti, 2017). While a one-back task does not require meaning access, the current study asked participants to be always prepared for semantic judgment on the new meanings. According to the complementary learning systems model (Davis & Gaskell, 2009), episodic memory is involved to support efficient lexical access when a lexical representation is not established yet. Participants may have to take advantage of the still-fresh episodic memory when the lexical representation of new meanings is not well established, so that they can access the new meaning as quickly as possible. Future research is needed to address the trade-off of the two learning and memory systems for the efficiency of task completion and how the trade-off changes over time.

2.5 Conclusion

Following overnight consolidation, new meanings of known words were accessed faster. The access to original meanings was slow down by meaning learning right after learning but not 24 hours later. The ERP data further showed that accessing new meanings was associated with a larger mid-frontal negativity and a larger left lateralized parietal positivity regardless of whether the words were tested on the day of learning or 24 hours later, suggesting continued involvement of episodic retrieval. In contrast, the N400, a component relevant to lexical semantic processing, was not modulated by meaning learning or overnight consolidation.

Overall, the findings indicate that the processing of both new and original meanings benefits from a study-test interval involving overnight sleep, even though new meanings are not fully integrated within that period. Episodic memory seems to play an important role in accessing the new meanings even 24 hours later, at least in a task requiring efficient meaning access.

3.0 Studies 2a and 2b: Neural Mechanisms underlying the Binding of New Meanings to Known Words

Recent learning studies suggest that the left pMTG suppoprts the formation of new lexical representations for newly learned words by mapping word forms and meanings (Bakker-Marshall et al., 2018; Ferreira et al., 2015; Landi et al., 2018; Takashima et al., 2014, 2017). It is possible that the left pMTG slowly takes over the binding role of the hippocampus over time. If the left pMTG also serves a binding role in the learning of new meanings for known words, we expected that the processing of new meanings would be associated with the activation in the left pMTG, in addition to the neocortical areas relevant to the representations of more specific semantic features.

According to the embodied account of language comprehension, word meanings, especially those referring to perception and actions, are partially represented in sensorimotor cortex as a result of perceptual experience (Barsalou, 2008; Binder & Desai, 2011; Pulvermüller, 2001; Pulvermuller & Fadiga, 2010). For example, the contrast between action and non-action words often reveals stronger brain activation for action words in the left ventral premotor cortex and the pars opercularis of the left inferior frontal gyrus (IFG) that extends into the premotor cortex (Moseley & Pulvermuller, 2014; Raposo, Moss, Stamatakis, & Tyler, 2009). It has been found that these regions are involved in the processing of action meanings regardless of the involved body part or effector (MacGregor, Pulvermuller, van Casteren, & Shtyrov, 2012; Moseley, Pulvermuller, & Shtyrov, 2013; Tettamanti et al., 2005). Based on this, the regions have been proposed to be associated with the abstract representation of actions.

In addition to the frontal areas, the human homologue of the monkey motion complex, or the human middle temporal plus near adjacent motion sensitive areas (i.e., the MT+), has also been found relevant to the representation of action meanings. Functional magnetic resonance imaging (fMRI) studies often reported stronger activation in the bilateral MT+ in the processing of visual motion, as revealed by contrasting moving and stationary stimuli (James & Gauthier, 2003; Tootell et al., 1995; Watson et al., 1993). In language comprehension, the left MT+ has been found to show stronger activation for sentences describing actions than describing non-action content (Saygin, McCullough, Alac, & Emmorey, 2010).

Recent word learning studies using fMRI and transcranial magnetic stimulation (TMS) have shown the involvement of the left motor cortex and the MT+ in the learning and processing of novel action words compared to those with non-action meanings (Liuzzi et al., 2010; Revill, Aslin, Tanenhaus, & Bavelier, 2008). However, because of the low temporal resolution of fMRI and TMS, it is unclear whether the involvement of the sensorimotor circuits in the processing of new action meanings results from post-activation processes (e.g. imagery) or from automatic activation of relevant semantic features as would be the case of existing words. If the involvement of sensorimotor circuits is part of word meaning comprehension, then the regions would be activated within the time window when lexical-semantic processing typically occurs.

In an ERP study particularly relevant for this time course question, Fargier et al. (2014) associated novel spoken words with videos of hand movement or animated artificial images and tested the processing of novel words multiple times over two days. They found that ERPs within 100-400 ms of auditory onset were able to distinguish novel words paired with different types of meanings immediately after training and on the next day before further training. However, such differences seemed to diminish with further training. While this study suggested a relatively early difference between words associated with different types of videos, it remains unclear whether the differences reflect stable associations between novel spoken word forms and the meanings. It is possible that one- or two-day learning may not be sufficient to form stable connections between words and their referents. In addition, because direct sensorimotor input was provided during learning by presenting word meanings through videos or pictures, the question of whether such differences emerge when meanings are presented verbally remains unclear. Much word learning occurs through verbal contexts that may or may not associate with established word meanings that are partially grounded in sensorimotor connections. These verbally mediated connections were described as "secondary grounding" in Pulvermuller (2013).

In this chapter, both of the reported studies spanned learning over four days to provide more opportunity for memory consolidation and knowledge integration to occur. Study 2a first demonstrated that presenting word meanings as verbal definitions was sufficient to drive the semantic category effect that has been observed in the processing of existing words (action vs. non-action words). Based on the findings of Study 2a, Study 2b further examined the role of the left pMTG and that of the brain regions relevant to the representation of action meanings (left frontal motor areas and MT+) in the learning of new action meanings for known words and for novel words using MEG.

3.1 Study 2a: Learning action meanings presented as verbal definitions

3.1.1 Research design and expectations

Study 2a aimed to examine whether word meanings presented as verbal definitions were sufficient to drive the semantic category effect and how early the semantic category effect would emerge relative to the point when words can be uniquely identified (i.e., recognition point). To do this, we took advantage of coarse-grain semantic distinctions. Participants learned novel words that were associated with either action or non-action (i.e., static visual) definitions and were instructed to visualize the meanings. Spoken words were used in this study, so that recognition point can be identified. Learning spanned over three consecutive days to provide sufficient opportunity for the integration of novel word meanings into semantic memory. On the fourth day, participants performed semantic judgments on novel words and also previously known words while EEGs was recorded.

High behavioral performance in the tests on novel words after the multiple-session training paradigm was expected. Of most interest is evidence for early discrimination of semantic categories: whether novel words associated with action meanings described by definitions would become distinguishable from those associated with non-action meanings in early processing stage, as examined with ERP that has high temporal resolution. If presenting meanings as verbal definitions was sufficient to drive difference in meaning representation, then divergence between novel action and non-action words would be predicted. If this divergence is driven by automatic meaning access to novel words after three nights following the initial learning, we expected it to emerge in an early time window close to recognition point of the spoken words. If meaning access required a slower episodic retrieval of the learning event, then a later separation beyond the recognition point would be expected.

Central electrode clusters were the focus of interest, because they have been the locus of differences between action words and non-action words in previous ERP studies (Fargier et al., 2014; Pulvermuller, Lutzenberger, & Preissl, 1999; Vanhoutte et al., 2015). However, if meaning access results from episodic retrieval, then the difference may emerge much later, 500 ms following word onset, when the recollection of episodic memory is typically observed (Rugg & Curran, 2007). Bakker, Takashima, van Hell, Janzen, & McQueen (2015) found a larger frontal and central negativity within 500 -700 ms for novel words than for existing words one day following learning, suggesting that meaning access to novel words relies partially on episodic retrieval. It is possible that such a difference may not be observed when more a longer learning phase occurs, as it does in the present study, to support integration of novel words into semantic memory.

3.1.2 Methods

3.1.2.1 Participants

Twenty-seven right-handed native English speakers participated in the study (10 females, 19.07 ±1.14 years old). Data from four additional participants were excluded from analysis because of not completing all the sessions (N = 3) or, in one case, due to an error in the assignment of experimental materials across sessions (N = 1). Participants provided written informed consent prior to the experiment and received course credits for their participation as an option to fulfill part of a course requirement. The study was approved by the University of Pittsburgh Institutional Review Board.

3.1.2.2 Materials

Novel words included forty English pseudowords (See Table 6 for examples and Appendix B.4 for the full set of stimuli). The novel words have 4.4 ± 0.71 phonemes (ranged from 3 to 6). The novel words were separated into two sets (20 in each set), matched in the number of phonemes and the number of letters. Forty existing words were selected and used in the ERP task: half refer to action meanings (typically used as verbs) and the other half refer to color, shape or pattern (typically used as nouns and/or adjectives). None of the existing words had a homophone. It is difficult to match grammatical category between the two sets of existing words; however, semantic information plays the dominant role when words are presented in isolation according to a meta-analysis study (Vigliocco, Vinson, Druks, Barber, & Cappa, 2011). While the two sets of existing words differ in terms of word category and semantic features, they were carefully matched on word frequency, number of phonemes, orthographic and phonological neighbor sizes, valence and arousal (See Table S1-1 for descriptive statistics and Appendix B.4 for the full set of stimuli). To match existing words with action and non-action meanings, we selected existing words with a small number of phonemes. However, because novel words with the same small number of phonemes tended to remind native speakers of one or more than one similar sounding real words, the final set of novel words was longer than existing words by 0.65 phonemes on average.

Table 6. Stimulus examples

Condition	Word	Meaning	Task
Novel/Action	bloosh	lifting with one hand	Learning/Test/ERP
Novel/Non-action	bropt	with a dark blue surface	Learning/Test/ERP
Existing/Action	kick		ERP
Existing/Non-action	green		ERP

Notes: Sixteen items in each condition; the assignment of novel words to action and non-action conditions was counterbalanced across participants.

All the word stimuli were spoken by a female native English speaker with a neutral tone and recorded at the sampling rate of 44.1 kHz. The acoustic intensity of each word was normalized to 70dB using Pratt. Because acoustic information unfolds over time during spoken word recognition, we further marked the recognition points for each word. For novel words, the recognition point was defined as the onset of the phoneme starting at which the phonemes of a novel word diverged from those of all the other novel words. For existing words, it was the onset of the phoneme starting which a word diverged from all other words according to the Carnegie Mellon University (CMU) pronunciation dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict/)[2]. On average, a novel word can be distinguished from the other novel words after 1.12 phonemes (103.07 ms), and an existing word can be identified after 3.78 phonemes (408.00 ms). However, if the novel words are integrated into the real word lexicon, then both novel and existing words need to be considered when recognition point is decided. In the selected novel

[2] Based the SUBTLEX (US) corpus developed by Brysbaert and New (2009) and words with word frequency lower than 1 per million are not considered in the CMU pronunciation dictionary.

words, the recognition point was 2.56 phonemes (328.42 ms) when all the words in the CMU pronunciation dictionary were considered; the recognition point of the 40 selected existing words was not affected by the including of the 40 novel words. For most of the novel words, the last phoneme is the divergence point.

Forty descriptive definitions were created for the study (see Appendix B.5 for the full list). Twenty of them described specific actions typically involving movement of finger/hand/arm, toe/foot/leg, or whole body. The other twenty described visual features including color, shape, pattern, or a combination of them. The pairing between definitions and novel words was counterbalanced across participants, such that a novel word was paired with an action meaning for half of the participants and with a non-action meaning for the other half of participants. The meanings of the words were spoken by the same speaker who recorded the words.

3.1.2.3 Procedure

The study had four sessions occurring on four consecutive days. The tasks participants performed on each day are shown in Figure 9A. On Day 1, they first learned the novel words by listening to and saying the novel words out loud and then seeing the written word forms to confirm which word they heard (i.e., form encoding). In this part, each word was encountered twice and participants were allowed to listen to and see the novel words once more if they wanted to. Then they started encoding the meanings for the novel words (i.e., meaning encoding). In each trial, participants heard a novel word and then its meaning. They were instructed to visualize the meanings for as long as they needed. When they were ready, they need to evaluate how hard it was to visualize the word meaning on a scale of 1 (easy) to 6 (hard). The evaluation was to make sure participants followed the instruction to visualize the word meanings. Following participants' rating, written forms of both word and their meaning were presented on the screen until they were ready to learn the next word. Following the initial encoding of word meanings, they experienced multiple study and test cycles to improve their learning. During the study trials, participants were presented with novel spoken words and instructed to recall the meanings in their head. The written meaning was presented at the center of the screen following participants' attempt to recall. After encountering each word twice, participants performed a test where they typed the meanings out for each novel word. The study and test cycle was repeated three times and participants were allowed to have one more study cycle if they wanted to. After that, participants performed a multiple-choice test where they chose the correct meaning from four options for each word they heard. After participants completed the task, they were presented with the overall accuracy and the correct responses for each word. Learning on Day 1 ended with a refresh of all the novel words and their meanings. On Days 2 and 3, participants repeated some of the tasks from Day 1 to further improve their learning or maintain their performance.

On Day 4, following the final tests on the novel words, participants performed a semantic judgment task while EEG was recorded. In the task, participants heard novel words and existing words in two separate blocks, with the order counterbalanced across participants. Each word was presented twice to increase the number of trials per condition, and participants had a break in the middle of a block and between blocks. They were told what type of words (i.e., novel words or previously known words) they would hear at the beginning of each block. During the task, participants were instructed to look directly at at the center of the screen and listen to the words carefully. Following about 20%[3] of the words, they saw a phrase (e.g., leg movement, dark color)

[3] The actual percentage was reduced 19.375% because the question for the last novel word was not presented due to an error in the experiment run script.

on the screen and judged whether it was related to the word that immediately preceded the phase (see Figure 9B). In each trial, a fixation cross was presented on the screen for 500-700 ms before word onset, during word presentation, and 800-1500 ms following word presentation. If a phrase was presented, participants had up to 5000 ms to respond. The next trial began following 1500 ms of blank screen. The jittering of duration of fixation before and after word presentation was to minimize participants' expectations about stimulus presentation (both spoken words and phrases). Participants were instructed not to blink when a fixation across was on the screen. A practice session with another 10 words was administered to familiarize participants with the task procedure. Following the ERP task, measurements of individual differences in vocabulary size and learning strategies were taken but results are not reported here. Participants' vocabulary was assessed with the Nelson-Denny Vocabulary Test (Brown, 1960), and learning strategies were surveyed through the Verbalizer-Visualizer Questionnaire (Kirby, Moore, & Schofield, 1988). They were asked questions about their strategies about the learning and retention of the novel words and then provided with debriefing.

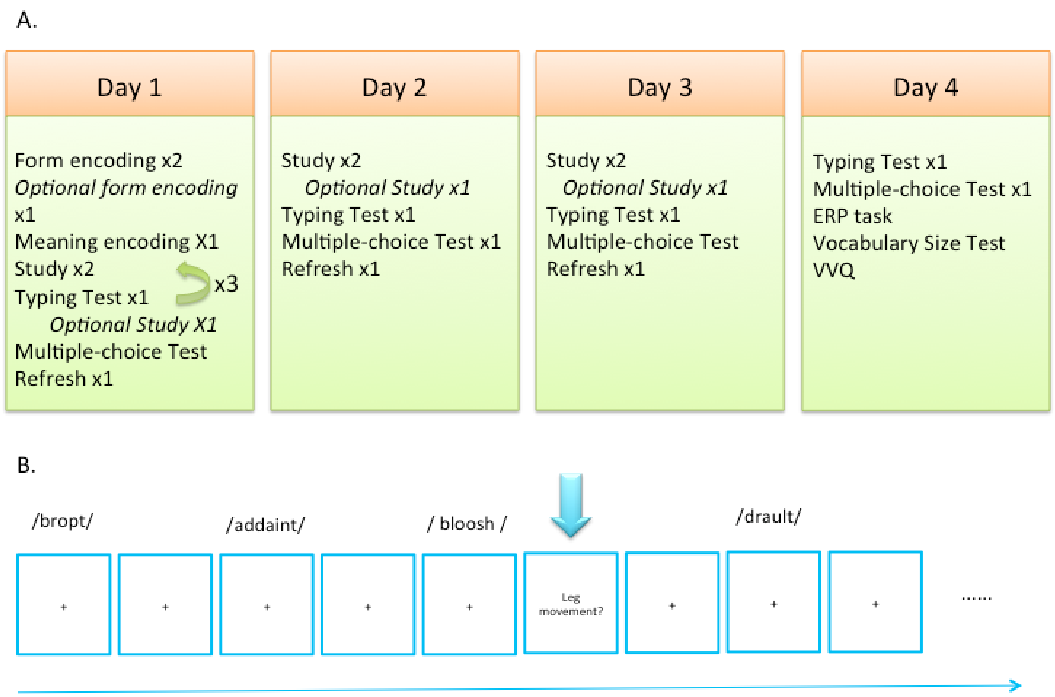

Figure 9. Task procedure.

A. Tasks participants performed on each day. B. The procedure of the ERP task (inter-trial intervals with blank screen are not presented for the sake of space). VVQ = verbalizer-visualizer questionnaire.

3.1.2.4 EEG data recording and preprocessing

The procedure for EEG data recording and preprocessing is the same as in Study 1. Participants were fitted with a Geodesic Sensor Net with a 128 Ag/AgCl electrode array and data were recorded using the associated NetStation acquisition software (Electrical Geodesics, Inc., Eugene, OR) with a sampling rate of 1,000 HZ and a hardware bandpass filter of 0.01–200 Hz. Data were preprocessed with NetStation Tool software. A bandpass filter of 0.1–30HZ was applied, and then data were segmented into 900-ms epochs, starting 100 ms before the onset of spoken words. Epochs with artifacts were excluded from further analysis, including eye blink (exceeding±140 µV), eye-movement (exceeding 55 µV), and extreme variance (larger than 100 μV). As a result, data from one participant with fewer than 20 (out of 40) valid epochs in one of the conditions were excluded. For participants whose data entered statistical analysis, on average there were 36.02 ± 3.77 out of 40 epochs per condition. Channels with artifacts on more than 20% of epochs were marked as bad channels, and data from surrounding channels within the cluster were used for interpolation. For each participant, ERPs for each type of words were acquired through averaging epochs under the same condition. The averaged waveforms were referenced to the average of the whole scalp before baseline correction (100 ms before onset of spoken words).

3.1.2.5 Data analysis

Behavioral data

Data were analyzed with mixed effects modeling (Baayen et al., 2008). Participants' typed responses in the cued-recall tests were scored from 0 (no response or unrelated response) to 5 (the exact meaning) by two trained research assistants independently. Responses with inconsistent ratings larger than 1 were discussed before a final score was assigned. The scores were analyzed with Meaning Type (Action/Non-action) and Day (1/2/3/4) as the fixed factors. The Meaning Type factor was effect-coded (Action vs. Non-action) and the Day factor was coded with backward difference to reflect changes between consecutive days to capture the change of performance over time (i.e., Day 2 vs. Day 1, Day 3 vs. Day 2, and Day 4 vs. Day 3). The accuracy data from the multiple-choice tests were analyzed with logistic mixed effects modeling, using the same fixed factors. For behavioral data from the ERP task, we analyzed the response times and accuracy with similar methods and included Meaning Type and Lexicality (Novel words vs. Existing words) as fixed factors. In all the models, the intercepts of subjects, word form, and definitions were included. Random slopes were included if a random effect significantly contributed to the model according to model comparison and did not cause perfect correlation among random effects or failure of model convergence. The final models are reported in the table notes in Results section.

ERP data

Repeated-measures ANOVAs (by-subject analysis) were conducted in the analysis of the preprocessed ERP data. ERPs at the central clusters (C3, Cz, and C4, see Figure 10) are of most interest. We focus on ERP components that are typically observed in sequence following word presentation – N1, P2, N400 and late positive complex (LPC). The divergence between words with different semantic features has been observed when N1 and P2 are typically observed (Kiefer, Sim, Herrnberger, Grothe, & Hoenig, 2008; Vanhoutte et al., 2015). The specific time windows for N1 and P2 were determined based on the peaks of the averaged waveforms of all the four conditions at the central clusters (102 ms for N1 and 186 ms for P2). The final time windows are 82-122 ms for N1 and 136-236 ms for P2. Because all the non-action words described visual features and difference between existing action words and non-action words have been found in the occipital clusters (Pulvermuller et al., 1999), we also included occipital clusters (O1 and O2) as another site of interest. However, it is important to note that action meanings also contain visual information, which could make the difference in the occipital clusters subtle or even absent. There are three contrasts of interest: novel action words vs. novel non-action words; existing action words vs. existing non-action words; and novel words vs. existing words. We first reported the effects of interest in the central and occipital clusters, and then the effects in the lateral clusters and midline clusters. For lateral clusters (F3, F4, C3, C4, P3, P4, O1, and O2), Anteriority (frontal, central, parietal, and occipital) and Hemisphere (left and right) were included as the additional variables; for midline clusters (Fz, Cz, and Pz), Anteriority (frontal, central, and parietal) was the additional fixed factor.

In addition to the predefined time windows, we also examined the time course of semantic category effect in the central and occipital clusters in novel words and in existing words by running permutation test over time. To reduce the number of tests, we average data from every 10 consecutive time points without overlap, and then run a permutation test (1000 iterations) at each time step. Only differences larger than 97.5% or smaller than 2.5% of the permutation results in the same time point were considered as significant (e.g., two-sided test). The distribution of maximal clusters size was generated from the same 1000 permutations and for positive (more positive for Action condition) and negative (more negative for Action condition) clusters separately. Here the cluster size is defined as the sum of the absolute values of the differences over time points that significant differences were found. Clusters were only considered significant when a cluster was larger than 95% within the distribution. The results are presented in *Figure S1-1*.

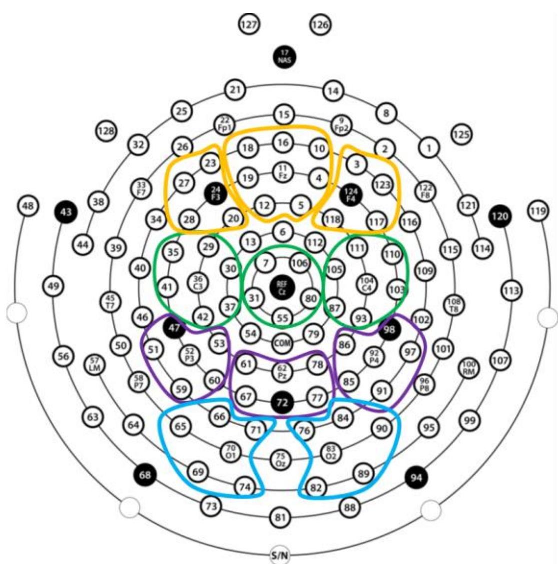

Figure 10. Layout of electrodes and clusters.

Orange: F3, Fz, and F4; green: C3, Cz, and C4; purple: P3, Pz, and P4; blue: O1 and O2. (same as in Study 1)

3.1.3 Results

3.1.3.1 Tests on novel words

Participants' performance in both the cued-recall (i.e., typed responses) and recognition (i.e., multiple-choice) tests improved across four days, as indicated by significant improvement between consecutive days (ps < .001, see Figure 11 and Table 7). However, the improvement on the last two days, especially in the multiple-choice tests (increase of less than 0.4%), was very small. The significant effects seemed driven by the very small variance when the participants' performances reached ceiling. The main effect of Meaning Type was not significant in either task (ps > .58). Additionally, in the cued-recall tests, the improvement of performance for non-action words was larger than that for action words from Day 1 to Day 2 (p = .015) and marginally larger from Day 2 to Day 3 (p = .085). However, the performance for the two types of novel words was comparable on each of the days (Day 1: β = -0.182, SE = 0.126, t = -1.443, p = .149; ps > .42 for Days 2, 3, and 4), suggesting again the significant interaction is driven by small variance of the data. In the multiple-choice tests, the improvement of performance over days was comparable between action and non-action words (all ps > .31 for all the interaction terms).

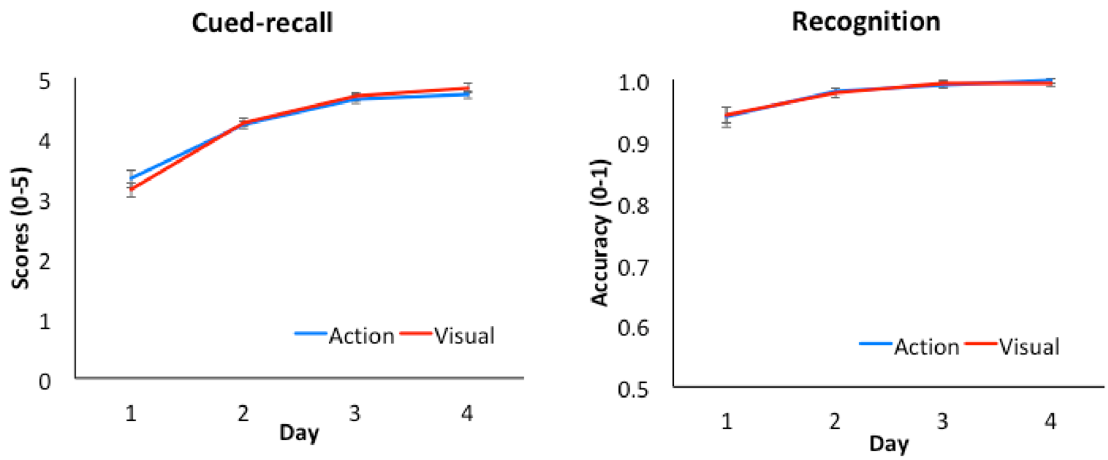

Figure 11. Participants' performance in the cued-recall and recognition tests.

Error bars represent ±1 SEM.

Table 7. Fixed effect estimates for mixed effects of tests on novel words

Fixed effect	*β*	*SE*	*t or z*	*p*	
Cued-recall (Typing test)					
Intercept	4.221	0.127	33.14	< .001	***
MeaningType	0.003	0.102	0.03	.976	
Day2 (vs. Day1)	2.004	0.074	27.25	< .001	***
Day3 (vs. Day2)	1.974	0.085	23.25	< .001	***
Day4 (vs. Day3)	1.091	0.074	14.83	< .001	***
MeaningType:Day2(vs. Day1)	-0.357	0.147	-2.43	.015	*
MeaningType:Day3(vs. Day2)	-0.293	0.170	-1.72	.085	~
MeaningType:Day4(vs. Day3)	-0.209	0.147	-1.42	.156	
Recognition (Multiple-choice test)					
Intercept	5.538	0.496	11.169	< .001	***
MeaningType	0.212	0.384	0.553	.580	
Day2 (vs. Day1)	3.350	0.412	8.140	< .001	***
Day3 (vs. Day2)	4.211	0.734	5.736	< .001	***
Day4 (vs. Day3)	2.922	0.898	3.252	.001	**
MeaningType:Day2(vs. Day1)	0.344	0.818	0.421	.674	
MeaningType:Day3(vs. Day2)	0.369	1.466	0.252	.801	
MeaningType:Day4(vs. Day3)	1.817	1.796	1.012	.312	

Notes: Intercept represents mean performance across two types of novel words over four days; Meaning Type represents the overall difference between action words and non-action words; interaction terms represent the change of the effect of Meaning Type on consecutive days. Final model for typing test: Score ~ 1 + MeaningType * Day + (1 + MeaningType | Subject) + (1 | Word) + (1 | Meaning); final model for multiple-choice test: log(ACC) ~ 1 + MeaningType * Day + (1 | Subject) + (1 | Meaning). ***: $p < .001$, **: $p < .01$, *: $p < .05$, ~: $p < .10$.

3.1.3.2 ERP semantic judgment task

Accuracy and response times

As shown in Figure 12 and Table 8, participants made semantic judgments about novel words as accurately as existing words (p = .137). However, the overall response times were faster for existing words than for novel words (p = .007). In addition, the accuracy and response times for words with different meaning types were comparable and there was no interaction between MeaningType and Lexicality (both ps > .36).

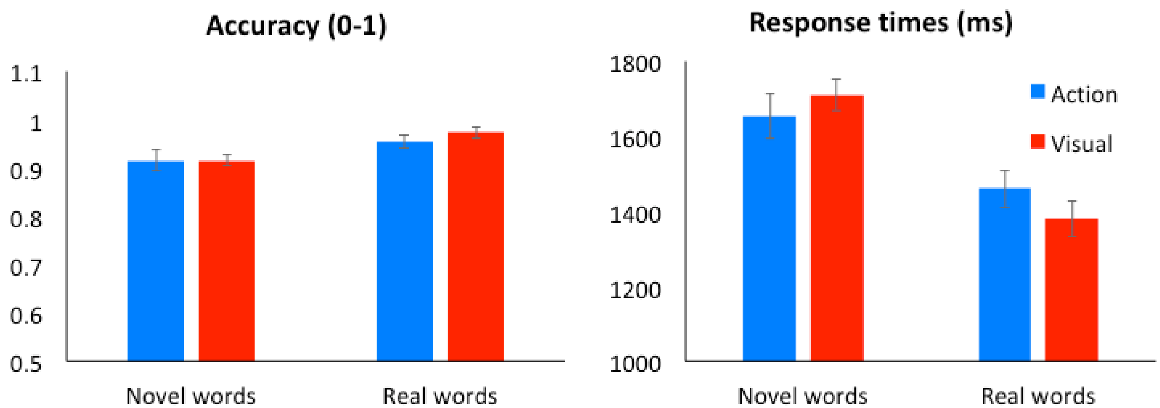

Figure 12. Participants' accuracy and response times in the ERP task.

Error bars represent ±1 SEM.

Table 8. Fixed effect estimates for mixed effects of performance in the ERP task

Fixed effect	*β*	*SE*	*t or z*	*p*	
Accuracy					
Intercept	4.068	0.559	7.274	< .001	***
Meaning Type	-0.398	0.5	0.885	.376	
Lexicality	-1.453	0.978	-1.487	.137	
Meaning Type * Lexicality	0.431	0.901	0.478	.633	

Response times					
Intercept	1534.3371	81.6197	18.799	< .001	***
Meaning Type	-0.7346	75.5265	-0.01	.992	
Lexicality	244.6536	91.0319	2.688	.007	**
Meaning Type * Lexicality	-124.4564	137.0526	-0.908	.364	

Notes: Intercept represents mean performance across all four conditions; Meaning Type represents the overall difference between action and non-action words; Lexicality represents the overall difference between novel and existing words. Final model for the accuracy data: log(ACC) ~ Meaning Type * Lexicality + (1 + Lexicality | Subject) + (1 | Word); Final model for the response times: RT ~ Meaning Type * Lexicality + (1 + Meaning Type * Lexicality | Subject) + (1 | Word).

ERPs data

Waveforms for each type of words are presented in Figure 13 and the topography of the semantic category effect within the time windows of interest is shown in Figure 14. For each of the contrasts of interest, results at the central and occipital clusters are reported first, followed by analyses of lateral and midline clusters to reveal the scalp distribution of difference.

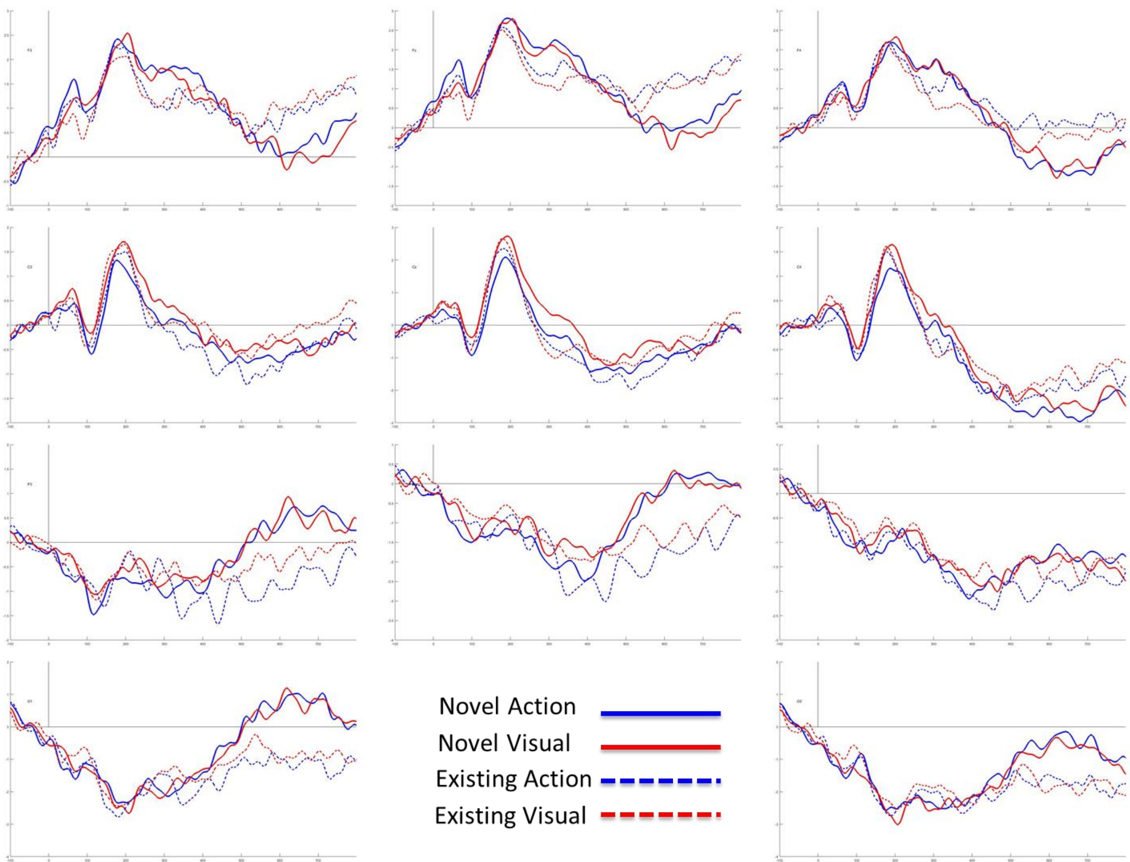

Figure 13. ERP waveforms for the four types of words.

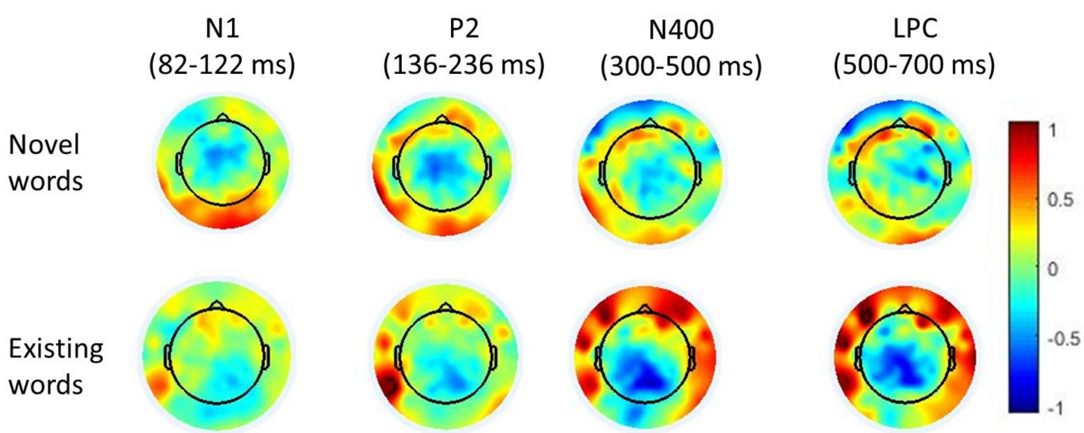

Figure 14. The topography of the Meaning Type effect (action vs. non-action) in novel words and existing words within the time windows of interest.

Novel action words vs. novel non-action words

At the central clusters, a larger negativity for novel action words than for non-action words was observed within the N1 and P2 time windows, whereas the difference was not statistically significant in the N400 or LPC time window (N1: $F(1,25) = 5.475$, $p = .028$, $\eta_p^2 = .074$; P2: $F(1,25) = 11.990$, $p = .002$, $\eta_p^2 = .080$; N400: $F(1,25) = 2.771$, $p = .108$, $\eta_p^2 = .039$; LPC: $F(1,25) = 2.765$, $p = .109$, $\eta_p^2 = .026$). At the occipital clusters, there was no difference between the two types of words in any of the time windows (all ps > .30).

In the lateral clusters, the main effect of Meaning Type was marginally significant within the P2 time window ($F(1,25) = 3.289$, $p = .082$, $\eta_p^2 = .138$), reflecting a larger negativity for novel action words. However, the effect was not significant in any other time windows or in the midline clusters, and none of the interactions involving Meaning Type within any time window in the lateral or midline clusters was significant (all ps > .13).

Existing action words vs. existing non-action words

At the central clusters, a larger negativity for action words than for non-action words was found in the N400 (marginally significant) and LPC time windows, but not in N1 and P2 time windows (N1: $F(1,25) = 0.136$, $p = .716$, $\eta_p^2 = .001$; P2: $F(1,25) = 0.835$, $p = .370$, $\eta_p^2 = .003$; N400: $F(1,25) = 3.287$, $p = .082$, $\eta_p^2 = .014$; LPC: $F(1,25) = 6.161$, $p = .020$, $\eta_p^2 = .026$). At the occipital clusters, there was no difference between the two types of words in any of the time windows (all ps > .18).

In the distribution analysis, neither the effect of Meaning Type nor any interaction involving Meaning Type was significant within the N1 and P2 time windows (all ps > .20). The effect of Meaning Type in the lateral clusters was significant in the N400 time window ($F(1,25) = 5.893$, $p = .023$, $\eta_p^2 = .003$) and marginally significant in the LPC time window ($F(1,25) = 4.136$,

p = .053, η_p^2= .003). In addition, in the midline clusters the effect of Meaning Type and its interaction with Anteriority within the N400 time window was marginally significant (Meaning Type: $F(1,25) = 2.949$, $p = .098$, η_p^2= .003; interaction: : $F(2,50) = 1.891$, $p = .058$, η_p^2= .006). The interaction was driven by larger negativity for action words than for non-action words in Pz than in Cz and Fz clusters. None of the other effects involving Meaning Type was significant (all ps > 10).

Novel words vs. existing words

At the central clusters, there was no significant difference between novel and existing words in any of the four time-windows (all ps > .20). In the occipital clusters, there was larger positivity for novel words than for existing words within 500-700 ms, but not in the other time windows (LPC: $F(1,25) = 24.540$, $p < .001$, η_p^2= .288; other ps > .13).

The overall difference between novel words and existing words was significant in both lateral and midline clusters within the N400 time window (lateral: $F(1,25) = 4.935$, $p = .040$, η_p^2= .017; midline: $F(1,25) = 5.992$, $p = .022$, η_p^2= .042), with larger negativity for existing words. In the later LPC time window, the larger negativity was significant in the lateral clusters but not in the midline clusters (lateral: $F(1,25) = 20.160$, $p < .001$, η_p^2= .039; midline: $F(1,25) = 1.928$, p = .177, η_p^2= .019). Within 500-700 ms, there was also a significant interaction between Lexicality with anteriority in both lateral and midline clusters, driven by a larger frontal negativity and a larger parietal/occipital positivity for novel words than for existing words (lateral: $F(3,75)$ = 15.210, $p < .001$, η_p^2= .258; midline: $F(2,50) = 18.110$, $p < .001$, η_p^2= .292). None of the other effects involving Meaning Type was significant within the N1 or P2 time window (all ps > .10).

3.1.4 Discussion

Following multiple-day training, participants learned both types of novel words very well. The effect of Meaning Type observed in the ERPs evoked by novel words indicated that word meanings presented as verbal definitions are sufficient to drive a semantic category effect. Furthermore, the very early effect of Meaning Type (within N1 and P2 time windows) suggests that the meanings of novel words can be accessed very fast and (arguably) automatically following intensive training spanning over days. However, a clear difference between novel words and existing words in ERP after 500 ms suggests that novel words are not fully integrated yet even days after learning. Episodic retrieval seems to continue playing an important role in recollecting relevant information in support of semantic judgments on the newly learned words.

Especially interesting is that the semantic effect emerged approximately when one novel word could be distinguished from another or around the point of recognition relative to only the novel words (~103 ms). This recognition point is early compared to the recognition point that would be defined for the entire vocabulary of English spoken words (~328 ms). This suggests that novel words are tagged separately from existing words, and represented in functionally separable memories. It seems the presentation of a word cues the relevant memory set: An existing word cues the relevant memory set of all existing words, whereas a novel word cues the relevant memory set of the novel words or a mini-lexicon consisting of the 40 newly learned words. A likely explanation for this result is in the procedure of testing novel words and existing words in separate blocks. This allows the novel word “lexicon” to be activated throughout the presentation of novel words, functioning as the vocabulary within which the novel word point of recognition functions[4].

To summarize, Study 2a demonstrated that presenting word meanings as verbal definitions without providing direct sensorimotor input is sufficient to drive a semantic category effect on novel words. The very early semantic effect indicated rapid and (arguably) automatic meaning access and the integration of novel words. Although the ERP findings do not directly argue for the involvement of the sensorimotor cortices, they do suggest the rapid activation of semantic features following multiple-session training. With these findings, Study 2b examined the role of left pMTG and neocortical areas relevant to the representation of action meanings in the binding of new action meanings to known words and to novel words using MEG that has better spatial resolution.

[4] The argument is supported by a follow-up study where novel words and existing words were presented in the same blocks. The effect of Meaning Type was absent in both novel words and existing words, while the difference between novel words and existing words was again observed.

3.2 Study 2b: The role of pMTG in learning new meanings for known words

3.2.1 Research design and expectations

In Study 2b, we examined the role of the left pMTG in the learning of new meanings for known words. The new meanings described actions for words whose original meanings have minimal motor or action component. We compared the learning of new meanings for known words with that for novel words. To account for the effect of mere exposure, we again included exposure controls. As in Study 2a, learning spanned four days and the MEG session occurred after participants had learned all the words very well and had more opportunity to integrate the new information compared to Study 1. Participants performed one meaning judgment and one meaning-unrelated (i.e., letter judgment) task on the studied words. In addition, a functional localizer task was included to localize specific regions relevant to the processing of action meanings using familiar words that either had action or non-action meanings.

We expected that words with new meanings would yield stronger source activation than exposure controls (i.e., the meaning learning effect) in the left pMTG for both novel words and known words, if the region is associated with the binding of lexical constitutes in general. Because the new meanings are all relevant to actions, stronger activation for words with new meanings than exposure controls was predicted in the left frontal areas involved in the representation of action meanings and the left MT+, which is sensitive to visual motion perception.

The learning of new meanings for known words involves co-activation of new and prior word knowledge from the very beginning of learning. With more time for offline consolidation, the new meanings may be integrated better than those of novel words. If this is the case, we expected the effect of meaning learning to be stronger in known words than in novel words. If meanings were automatically activated following multiple-session learning, we expected the meaning learning effects to be observed when lexical-semantic processing typically occurs in both meaning and letter judgment tasks (i.e., 300-500 ms). However, the effects may be more robust in the meaning judgment when meaning information is relevant to task completion.

3.2.2 Methods

3.2.2.1 Participants

Seventeen native English speakers participated in the study (six females, 23.12 ± 4.95 years old). They were all right-handed and had normal or corrected-to-normal vision and none reported any learning or language disabilities. Two of the participants (one female) did not complete the MRI session and were excluded from the MEG data analysis. The procedure of the study was approved by the institutional review board at the University of Pittsburgh and that at the Carnegie Mellon University. Participants provided written informed consent prior to the experiment and received course credits and/or monetary compensation for their participation.

3.2.2.2 Stimuli

Trained words

Thirty-two familiar words were selected from Wisconsin Perceptual Attribute Rating Database (Medler et al., 2005) and they were also part of stimuli in Study 1. The words had low ratings in the attribute of motion and low-to-medium ratings in the attributes of manipulate, emotion, and sound in the database (see Table 9 for examples and Appendix B.6 for full list). The words were separated into two groups, and within each group half of the words referred to man-made objects while the other half referred to natural objects. In addition, thirty-two novel words that were partially overlapped with those in Study 2a were separated into two groups. Novel words and known words were not statistically different in word length (p = .255) or bigram frequency (p = .172). One group of known words and one group of novel words were paired with new action meanings (i.e., meaning condition), and the other group served as exposure controls (i.e., control condition), with the assignment of words to the conditions counterbalanced across participants. The descriptive statistics of lexical and sub-lexical characteristics of word stimuli are presented in Table S2-1.

Untrained words

Sixteen action words and sixteen non-action words were presented in a localizer task. Action words described actions involving movement of upper or lower part of body. Non-action words described color or shape. The two types of words were matched for word frequency, word length, orthographic neighborhood size, ratings of valence and arousal, number of senses (see Table 9 for examples, Table S2-1 for lexical characteristics, and Appendix B.6 for full list).

Table 9. Stimulus examples

Condition	Word	Meaning	Task
Trained words			
Known/Meaning	cloud	typing rapidly	Learning/Test/MEG
Known/Control	stone	****************	Learning/Test/MEG
Novel/Meaning	trebe	lifting with one hand	Learning/Test/MEG
Novel/Control	bape	****************	Learning/Test/MEG
Untrained words			
Action	pull		MEG

Non-action	green	MEG

Notes: Sixteen items per condition. The assignment of trained words to the Meaning and Control conditions was counterbalanced across participants. The untrained words were used in the localizer task in the MEG session.

Meanings

The same 32 definitions from Study 1 were used (see Appendix B.2 for full list). The meanings were separated into two groups. Within each group, half of the meanings described actions typically involving finger, hand, or arm movement, while the other half described actions involving toe, foot, or leg movement. One group of meanings was paired with known words, while the other group was paired with novel words, with the assignment counterbalanced across participants.

3.2.2.3 Procedure

Over three days, as shown in Figure 15, participants learned new meanings for known words and novel words, which were presented along with exposure controls. On Day 4, they performed the final tests on the studied words and then the MEG tasks. Following that, they completed some questionnaires. Participants' T1-weighted anatomical images were acquired in a separate session 1-3 weeks following the MEG session.

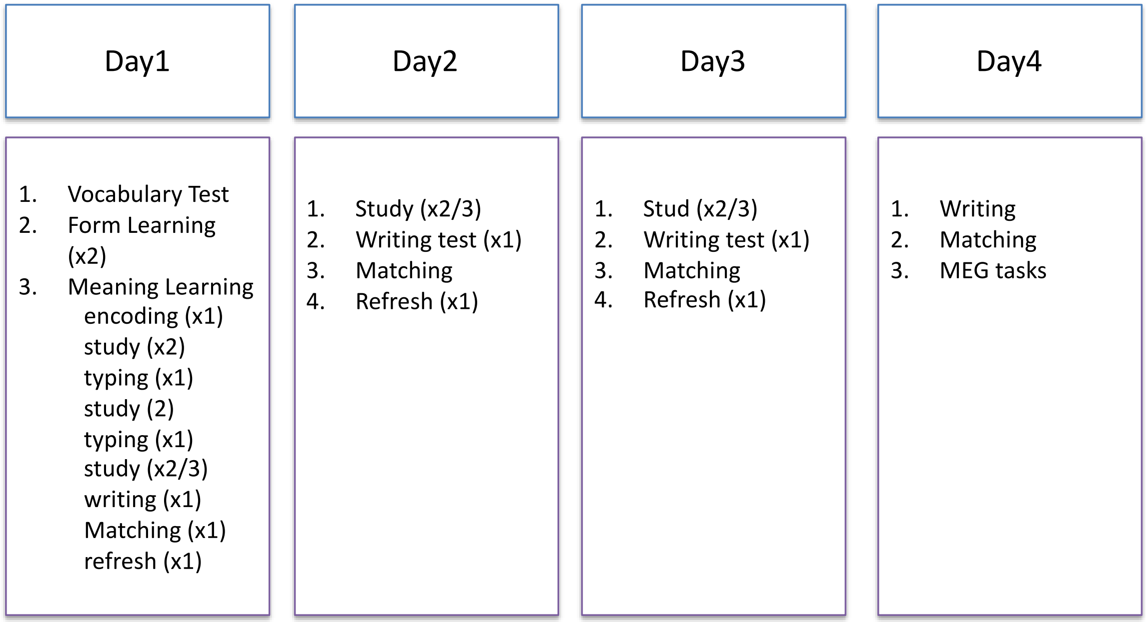

Figure 15. Overview of tasks on each day.

Numbers in parentheses indicate the number of exposures to the trained words in each task.

On Day 1, participants' English vocabulary was assessed with the Nelson-Denny vocabulary test (Brown, 1960). Following that, they started studying the word forms (i.e., form learning). In this part, no meanings were presented and participants saw each word twice. In each trial, a word was presented on the screen for five seconds and then the next word was presented automatically. Participants were instructed to pay attention to the spelling of the words. This was designed to reduce the difference between known words and novel words in familiarity, although it is unlikely the difference diminished after two exposures (Fang et al., 2017; Fang & Perfetti, 2017). Following learning the word forms, participants studied the meanings. In each trial, participants were presented with a word for one second, and then its meaning or a string of asterisks below the word for eight seconds. Participants were instructed to visualize the meanings. The next word was presented automatically and each word was presented once. Following meaning encoding, participants experienced three cycles of study and cued-recall test. In each study block, each word was presented twice in a self-paced manner. Participants were presented with a word, and then instructed to recall what had been paired with it. Following the attempt to recall, the correct answer was presented on the screen for them to study. In the cued-recall test, participants typed out what had been paired with each word and the correct answer was provided as feedback following each response. Each word was presented once within each test. For exposure controls, participants typed “n” for none (no meaning associated with the word). After each test, participants estimated how many words they got correct and reported their estimation to experimenter.

Following the three study-and-test cycles, participants performed a multiple-choice test to assess the recognition of the associations between the words and meanings. In each trial, one word and four response options were presented on the screen. The first three options were studied meanings, including one correct answer and two foils – one had been paired with a known word and the other had been paired with a novel word. The fourth option was always a string of asterisks. Participants selected what had been paired with each word. Following that, participants reviewed the words once before they left. On both Day 2 and Day 3, participants experienced one study-test cycle, followed by a recognition or multiple-choice test. They reviewed all the words once at the end of each session.

On Day 4, participants first performed a cued-recall and a recognition test. The procedure for each test was the same as described above. Following that, participants performed the MEG tasks, including one meaning judgment task and one letter judgment task on trained words, and then a localizer task on untrained words (see Figure 16 for diagram). The order of the meaning and letter judgment tasks was counterbalanced across participants. The localizer task was always the last task. In each of the first two tasks, participants were presented with each studied word four times, once in each of the four blocks. Following 18.75% of the words, a phrase or a letter was presented. In the meaning judgment task, participants judged whether the phrase was semantically related to the studied meaning of the preceding word. On some trials, the phrase presented was "has a new meaning" or "without a new meaning" and participants made decisions based on whether or not a word had been paired with a new meaning. In the letter judgment task, participants judged whether the presented letter occurred in the preceding word or not. Participants' responses were recorded. Participants could not predict when a phrase or a letter would be presented or what the content would be, thus motor preparation was minimized.

The procedure of the localizer task was the same as the meaning judgment task, except that it has only two blocks. In the localizer task, untrained but familiar action and non-action words were presented, and each word was presented twice within each block (i.e., four repetitions for each word). Following the MEG tasks, they completed the Pittsburgh Sleep Index questionnaire (Buysse, Reynolds, Monk, Berman, & Kupfer, 1989), and the Epworth Sleepiness Scale (Johns, 1991).

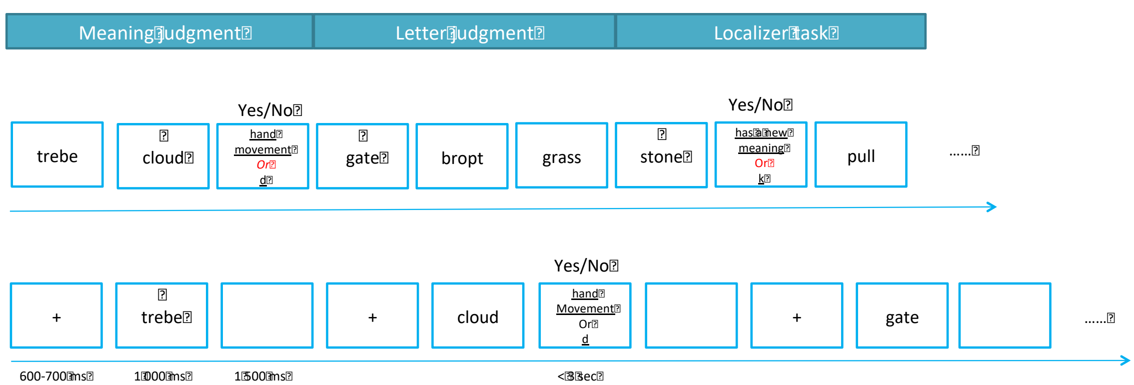

Figure 16. MEG tasks and trial procedure.

Upper panel: MEG tasks that participants performed. Middle panel: task procedure for with fixation and inter-trial-interval not shown. Bottom panel: Task procedure with full details. Notes: The order of meaning judgment and letter judgment tasks on trained words was counterbalanced across participants; the procedure of the three tasks are the same except letters are presented as questions in the letter judgment task and phrase are presented in the meaning judgment task and the localizer task.

3.2.2.4 MEG and MRI data acquisition

The MEG data were recorded with the 306-channel scanner (Elekta Neuromag, Helsinki, Finland) located at the Brain Mapping Center, University of Pittsburgh Medical Center (UPMC). The sampling rate is 1 kHz, with a band pass filter at 0.1-330 Hz. Participants' head position was tracked with four Head Position Indicator (HPI) coils that were attached to participants' head. Additionally, participants' vertical and horizontal eye movement (EOG), and heartbeat (ECG) were recorded using additional electrodes. Prior to MEG scan, at least 100 points on participants' heads were digitized in addition to the three fiducial points and four HPI coils. An empty room measurement was taken within the same session.

Participants' high-resolution anatomical images were acquired using a T1-weighted MPRAGE sequence (1 mm × 1 mm × 1 mm, 176 sagittal slices, TR = 2 300 ms, TI = 900 ms, Flip angle = 9°) with a Siemens Verio 3T Scanner located at the Scientific Imaging and Brain Research Center at the Carnegie Mellon University.

3.2.2.5 Data analysis

Behavioral data

For the cued-recall tests, two trained research assistants independently scored participants' responses from 0 (i.e., no response or unrelated meaning) to 5 (i.e., exact meaning, see Appendix B.3 for rubric). Inconsistency in rating larger than 1 was resolved through discussion and a final score was assigned. For the recognition test, participants' responses were scored either as correct or incorrect. Behavioral data in the tests on the studied words and those from the MEG tasks were analyzed with linear mixed effects modeling using the lme4 package in R (Baayen et al., 2008). In both the cued-recall and recognition tests, the fixed factors included Lexicality (known words vs. novel words), Type (control vs. meaning), and Session (Session 1/2/3/4). For accuracy and response times in the letter and meaning judgment tasks, the fixed factors included Lexicality, Type, and Task. For localizer task, MeaningType (Action vs. Visual) was the only fixed factor. For accuracy data, logistic regression was used. However, because participants reached 100% accuracy in at least one of the conditions in the recognition tests, empirical logits were calculated for by-subject and by-item analyses separately before linear mixed modeling was performed (Donnelly & Verkuilen, 2017). Effect coding was used for Type and Lexicality in all the behavioral tasks. For cued-recall and recognition tests, Session was coded to capture the changes between consecutive days (i.e., Session 2 vs. Session 1, Session 3 vs. Session 2, and Session 4 vs. Session 3).

MEG data preprocessing

The preprocessing and analysis of MEG data were performed primarily using MNE-Python and followed the typical workflow (Gramfort et al., 2013; Gramfort et al., 2014). Using empty-room measurement, five projectors for gradiometer sensors and five for magnetometer sensors were generated to capture external noise and then applied to the raw MEG data. The raw data were then low-pass filtered at 40 HZ, and bad sensors were detected and removed. Stereotypical artifacts related to blink and heartbeat were removed using independent component analysis (ICA; 25 components). The bad sensors were then interpolated using data from surrounding sensors. The artifact-compensated data were then segmented into epochs, including 200 ms before word presentation and 800 ms after. An epoch was excluded from further analysis if at least one MEG channel had extreme values (gradiometers: 4000e-13 T/m, magnetometers: 4e-12 T). On average, there were 63.40 ± 1.71 out of 64 valid trials per condition per task.

To facilitate source localization analysis, the MEG data and the structural MRI data were co-registered based on the three fiducial points and the additional digitized points using *mne* *analyze*. Brain surfaces were created by segmenting individual anatomical images using Freesurfer (http://surfer.nmr.mgh.harvard.edu/). A source space with spacing of approximately 4.9 mm between vertices (i.e., 4096 sources per hemisphere) was created. An elementary boundary model (BEM) was generated based on a single layer model (inner skull, conductivity = 0.3 S/m). Noise covariance matrices were created with the baseline data (200 ms before word onset) from all the valid epochs from the same blocks and were then used to make inverse operators for individual blocks. Source activation was estimated by applying inverse operators to the single trial data, using dynamic statistical parametric mapping (dSPM; Dale et al., 2000) and therefore the unit for source activation is z score. The source estimation was then averaged over trials from the same condition across all four blocks within each task. Prior to group analysis, source activation in individual surfaces were transformed to the common surface. For the sake of computational efficiency, data were down-sampled to 250 HZ and analysis were limited to time points after word onset.

MEG data analysis

Functional localizer. The localizer task was designed to localize brain regions involved in the processing of action meaning. However, the contrast between action and non-action words did not yield any significant difference between the two types of words across participants. Therefore, the main part of the analysis was based on the anatomically defined regions of interest.

Regions of interest (ROIs) analysis. Source activities in the left pMTG, left pre-motor cortex and the left IFG (frontal motor ROI), and left MT+/V5 were most relevant to the research hypotheses. As in previous MEG studies (MacGregor et al., 2012; Moseley et al., 2013), the frontal motor ROI included the left BA44 and the ventral part of precentral gyrus based on the "Desikan-Killiany" cortical atlas (Desikan et al., 2006). The MT+ was based on the atlas by Fischl et al. (2008). The left pMTG cortex was defined as the posterior part of the left MTG (Desikan et al., 2006). Similar to Bakker-Marshall et al. (2018), a vertical line was drawn in the middle of the left MTG on the inflated surface along the anterior-posterior axis. As in Study 1, for each ROI, we focused on two time-windows: 300-500 ms and 500-800 ms. For each ROI, within each time window, we first examined the effect of meaning learning (Meaning vs. Control) in known words and novel words separately, and then compared the effect of meaning learning on known words and that on novel words. For the effect of task, we compared the meaning learning effect in the meaning judgment and that in the letter judgment task, for known words and novel words separately. One-tail paired t tests were used to test the differences in the ROI analysis, because we have strong research hypotheses about the direction of the differences.

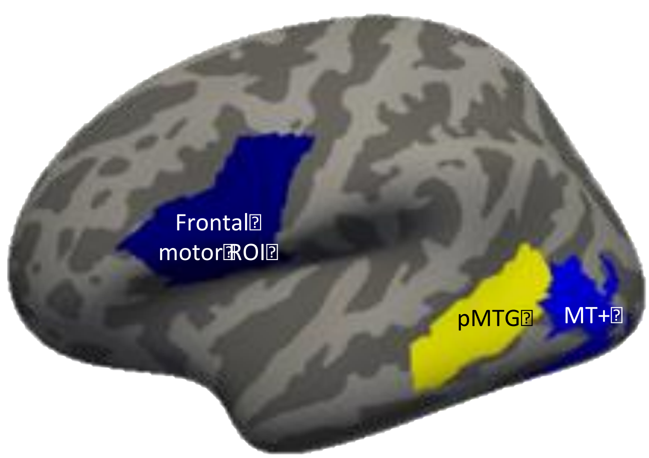

Figure 17. Regions of interest.

The MNI coordinates for the center of mass are: left frontal motor ROI (-46.54, 2.89, 17.32); left MT+ (-39.82, -74.46, 1.31); left pMTG (-58.81, -51.66, 1.98).

Whole-brain analysis. To avoid missing any meaningful effects outside of the ROIs, differences between conditions were searched in both spatial (vertices) and temporal (time points) dimensions and a cluster-based permutation test was applied to correct for multiple comparisons

(Maris & Oostenveld, 2007). Specifically, the significance of each cluster was tested against the distribution of 1024 permutations with an original threshold of $p < .05$[5].

3.2.3 Results

3.2.3.1 Performance in cued-recall and recognition tests on trained words

Overall, participants performed well in both cued-recall and recognition tests: For all types of words across the four sessions, the mean score of the cued-recall tests is 4.955 out of 5, and the mean accuracy in the recognition tests is 99.8% (see Table 10 for means). This very good performance was not surprising, as participants had the opportunity to study the words before the tests, except in the final session. Statistical analysis showed that participants overall recall of information about the studied words increased steadily over sessions (ps < .01); however, the improvement between consecutive sessions was small (0.018-0.023 out of 5). In addition, the change of performance across sessions was different among words under different conditions. For example, the increase of performance from Session 1 to Session 2 was larger for novel words than

---

[5] Because no clusters yielded a significant effect when all the vertices were included, we ran additional analyses on the vertices falling within a language mask to reduce the number of vertices. The mask was defined as a combination of brain regions that are typically involved in language processing and were based on the "Desikan-Killiany" cortical atlas (Desikan et al., 2006). As in Kocagoncu, Clarke, Devereux, and Tyler (2017), the language mask included bilateral IFG, MTG, superior temporal gyrus (STG), inferior temporal gyrus (ITG), supramarginal gyrus (SMG), and angular gyrus (AG). We additionally included precentral gyrus as the regions have been reported relevant to the processing of action words (MacGregor et al., 2012; Moseley et al., 2013). However, we did not find any significant difference within the mask.

for known words (interaction of Lexicality and Session2vs1: $\beta = 0.118$, $SE = 0.045$, $z = 2.615$, $p = .009$).

Table 10. Performance in cued-recall and recognition tests

Test	Type	Lexicality	Session 1	Session 2	Session 3	Session 4
Cued-recall (score)	Control	Known words	5.000	4.945	4.982	4.982
		Novel words	4.982	5.000	4.982	5.000
	Meaning	Known words	4.912	4.938	4.945	4.963
		Novel words	4.800	4.903	4.956	4.993
Recognition (accuracy)	Control	Known words	1.000	1.000	1.000	1.000
		Novel words	0.996	0.996	1.000	1.000
	Meaning	Known words	1.000	0.996	1.000	1.000
		Novel words	0.985	1.000	1.000	0.996

Notes: Mean scores (out of 5) in the cued-recall tests and mean accuracy in the recognition tests are reported.

In the recognition tests, again participants reached performance ceiling even though statistically some differences among conditions emerged because of the small variance in participants' performance. In the final tests in Session 4, participants reached performance ceiling in both tests, indicating that they had a good knowledge of the studied words before they performed the MEG tasks. The complete statistical results were presented in Table S2-2.

3.2.3.2 Performance in the MEG tasks

As shown in Figure 18 and Table 11, participants had a high accuracy in both letter and meaning judgment tasks (above 95% in all of the conditions). Across the two tasks, they responded more accurately to known words with new meanings than novel words with meanings, while no difference between exposure controls of novel words and those of known words was found (interaction of Type and Lexicality: z = 2.929, p = .022; see Table 11). Response time data showed that participants were overall faster in making letter judgments than meaning judgments (main effect of Task: t = -8.327, p < .001). None of the other effects in the accuracy or response time data was significant (all ps > .19). In the localizer task, the accuracy and response times were comparable across action and visual words (both ps > .53).

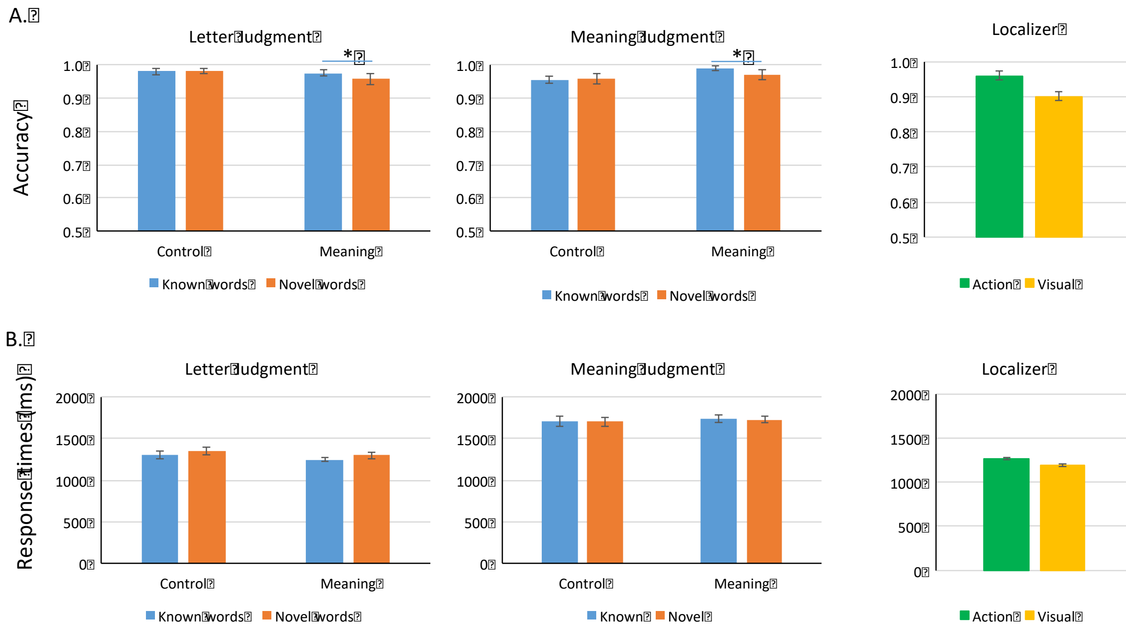

Figure 18. Accuracy (panel A) and response times (Panel B) in the MEG tasks.

Error bars represent 1 SEM with between-participant variance removed (Franz & Loftus, 2012).

Table 11. Fixed effect estimates for mixed effects models of performance in the MEG tasks

	beta	*SE*	*t or z*	*p*	
Letter and meaning judgment: Accuracy					
Intercept	3.613	0.165	21.913	< .001	***
Task (Meaning vs. Letter)	-0.417	0.330	-1.263	0.207	
Type (Meaning vs. Control)	0.224	0.330	0.679	0.497	
Lexicality (Novel vs. Known)	-0.080	0.330	-0.243	0.808	
Task:Type	-0.863	0.660	-1.309	0.191	
Task:Lexicality	-0.255	0.660	-0.387	0.699	
Type:Lexicality	1.512	0.660	2.292	0.022	*
Task:Type:Lexicality	-0.511	1.320	-0.387	0.699	
Letter and meaning judgment: Response times					
Intercept	0.844	0.040	21.175	< .001	***
Task	-0.189	0.023	-8.327	< .001	***
Type	0.042	0.021	1.997	0.046	*
Lexicality	-0.015	0.024	-0.634	0.526	
Task:Type	-0.033	0.031	-1.060	0.289	
Task:Lexicality	0.014	0.032	0.447	0.655	
Type:Lexicality	0.000	0.030	-0.004	0.997	
Task:Type:Lexicality	0.006	0.045	0.143	0.886	
Localizer: Accuracy					
Intercept	5.446	1.493	3.647	< .001	***
MeaningType (Action vs. Visual)	0.304	1.489	0.204	0.838	
Localizer: Response times					
Intercept	0.875	0.049	17.972	< .001	***
MeaningType	-0.035	0.056	-0.618	0.537	

Notes: Intercept is the mean performance across all the conditions in each task. Final model for accuracy data in letter and meaning judgment tasks: log(ACC) ~ Task * Type * Lexicality + (1| Subject); final model for response times in letter and meaning judgment: Inverted response times in seconds ~ Task * Type * Lexicality + (1|Subject) + (1|Word); final model of accuracy data in localizer task: log(ACC) ~ Condition + (1|Subject) + (1|Word); final model for response times: inverted response times in seconds ~ Condition + (1|Subject) + (1|Word). ***: $p < .001$, **: $p < .01$, *: $p < .05$.

3.2.3.3 Source activation in the localizer task

Different from our expectation, no statistically significant difference between action and non-action words was observed in the localizer task. In the ROI analysis, no difference between the two types of words was observed in any of the ROIs in either time window (all *p*s > .25, see

Figure 19A). In the whole-brain analysis, we did see stronger source activation in the left superior temporal sulcus (STS) and STG including the auditory cortex within 300-500 ms, although the differences did not survive multiple-comparison correction (cluster *p*s > .38, see Figure 19B). We return to discussion the null findings later.

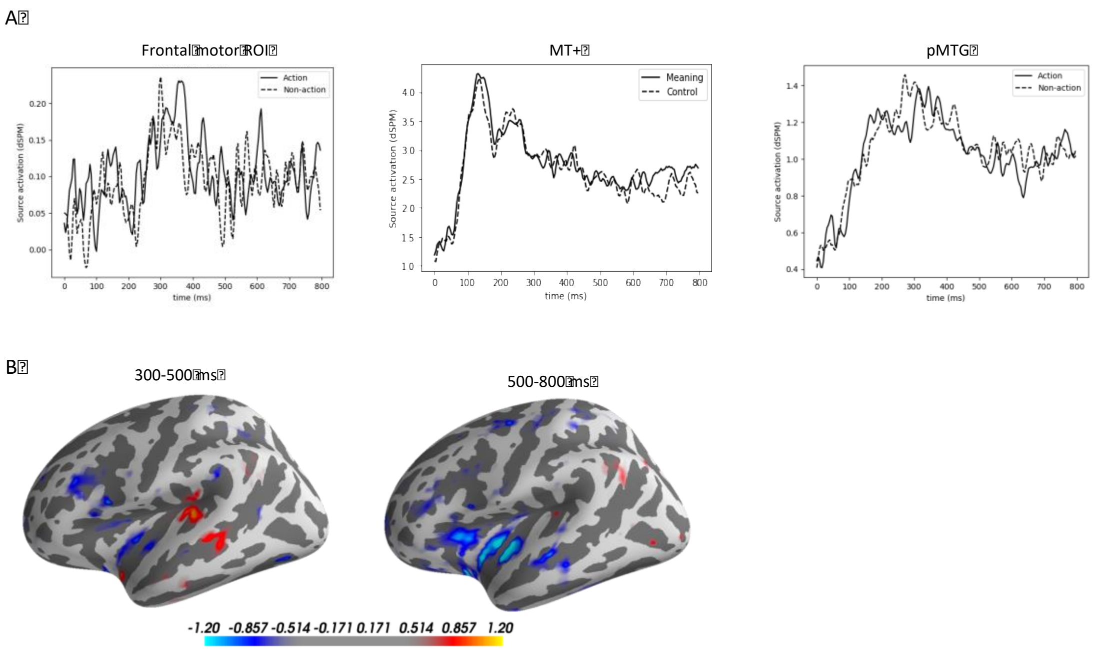

Figure 19. Source activation in the localizer task.

A. Time course of source activation within the ROIs. B. Activation map for the difference between action and non-action words within 300-500 ms and within 500-800 ms.

3.2.3.4 Source activation in the meaning judgment tasks

As shown in Figure 20, for known words, source activation in the frontal motor ROI was stronger for words with new meanings than exposure controls. The difference was marginally significant within 300-500 ms ($t(14) = 1.708$, $p = .055$) and became significant within 500-800 ms ($t(14) = 2.280$, $p = 0.019$). No meaning learning effect was observed in the left MT+ in either time window (both *p*s > .15). In the left pMTG, we also observed stronger source activation for words with new meanings in both time windows ($t(14) = 1.937$, $p = .037$ within 300-500 ms; $t(14) = 2.724$, $p = .008$ within 500-800 ms).

For novel words, words with new meanings evoked stronger source activation in the left MT+ within 300-500ms ($t(14) = 1.801$, $p = .047$) and marginally stronger within 500-800 ms ($t(14) = 1.468$, $p = .082$). For novel words, although not statistically significant (all ps > .11), stronger source activation for words with new meanings was observed in the left frontal motor ROI and the left pMTG, especially within 300-500 ms.

We also directly compared the meaning learning effect (Meaning – Control) in known words and that in novel words. A marginally larger effect of meaning learning on known words was found in the left frontal motor ROI within 500-800 ms ($t(14) = 1.681$, $p = .057$), while a marginally larger effect for novel words was observed in the left MT+ within 300-500 ms ($t(14) = 1.568$, $p = .070$). The difference in the meaning learning effect between novel words and known words was not statistically significant in the left pMTG (ps > .13 for both time windows), or in the other time window in the left frontal motor ROI or in the left MT+ (ps > .47).

The whole-brain analysis did not show any significant cluster (cluster ps > .18). This could be a result of a large number of comparisons when the number of vertices and the number of time points are both considered. However, we sill presented the activation map for the meaning learning effect in known words and novel words within the time windows of interest (Figure 20B). Overall, in the meaning judgment task, stronger activation for the meaning condition than the control condition was observed in known words in the left posterior STG/STS, IFG, and central sulcus. For novel words, the difference was mainly observed in the left MT+, ITG, and IFG, and precentral gyrus within 300-500 ms, and in the left STG and MT+ within 500-800 ms.

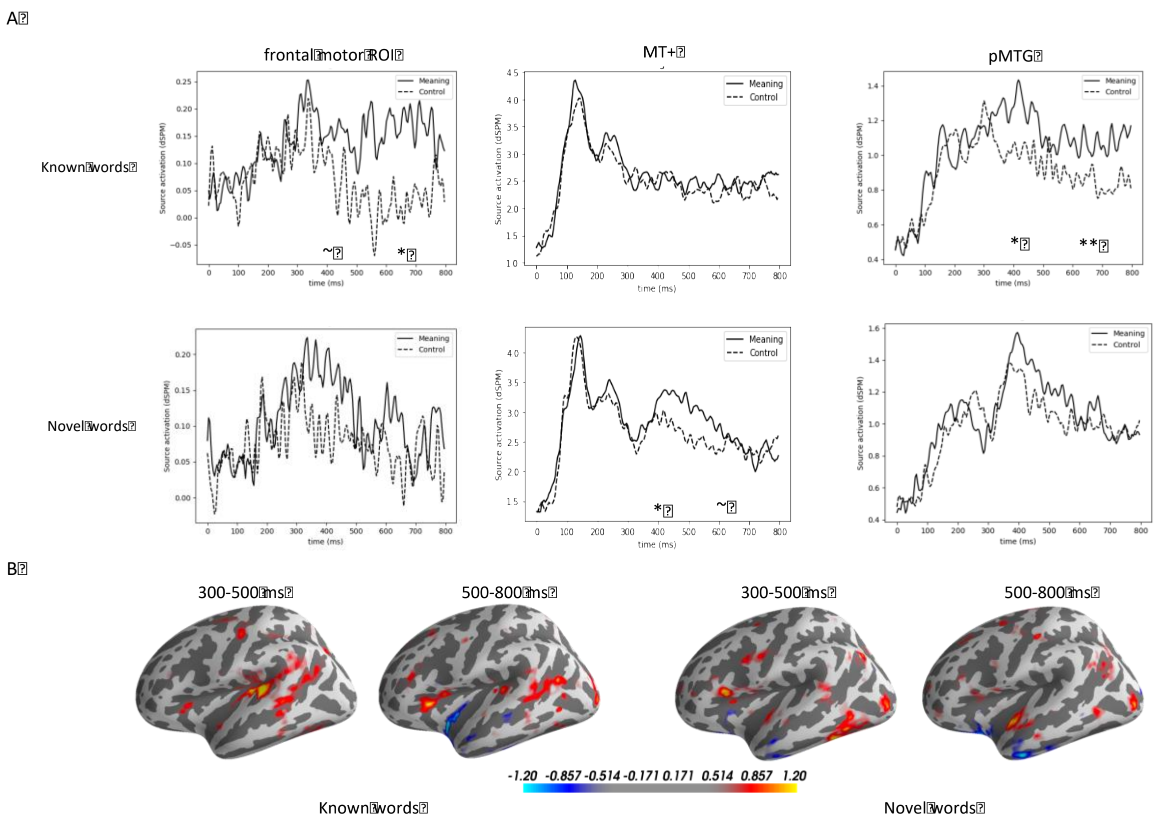

Figure 20. Source action in the meaning judgment task.

A. Source estimates in the regions of interest. B. Activation maps for the meaning learning effect within 300-500 ms and within 500-800 ms (B) in the meaning judgment task. **: $p < .01$, *: $p < .05$, ~: $p < .10$.

3.2.3.5 Source activation in the letter judgment

For known words, meaning learning yielded a reserved meaning learning effect in the left pMTG or left MT+, with reduced activation for words with new meanings, although the difference was not significant (all *p*s > .10, see Figure 21A). In the frontal motor ROI, we observed stronger activation for exposure controls within 500-800 ms ($t(14) = -1.982$, $p = .040$), again a pattern opposite to what we expected. For novel words, source activation was stronger for words with new meanings than exposure controls in the left MT+ within 300-500 ms ($t(14) = 2.112$, $p = .027$) but not within 500-800 ms ($t(14) = 0.403$, $p = .346$). As in the meaning judgment task, there was no significant effect of meaning learning in the frontal motor ROI or pMTG in either time window (all *p*s > .24).

Direct comparison of meaning learning effects between known words and novel words did not yield significant difference (all *p*s > .10), except for a larger meaning learning effect in novel words than in known words in the left MT+ within 300-500 ms ($t(14) = 2.539$, $p = .012$). As in the meaning judgment task, the whole-brain analysis did not show any significant cluster for the meaning learning effect in known words or novel words (cluster *p*s > .87). As shown in Figure 21B, we mainly observed reduced source activation for the meaning condition in the left ITG and insular within both time windows. Reduced source activation in the left STS and MT+ was additionally found within 500-800 ms. For novel words, we observed stronger activation for the meaning condition in the left MT+ and post-central gyrus.

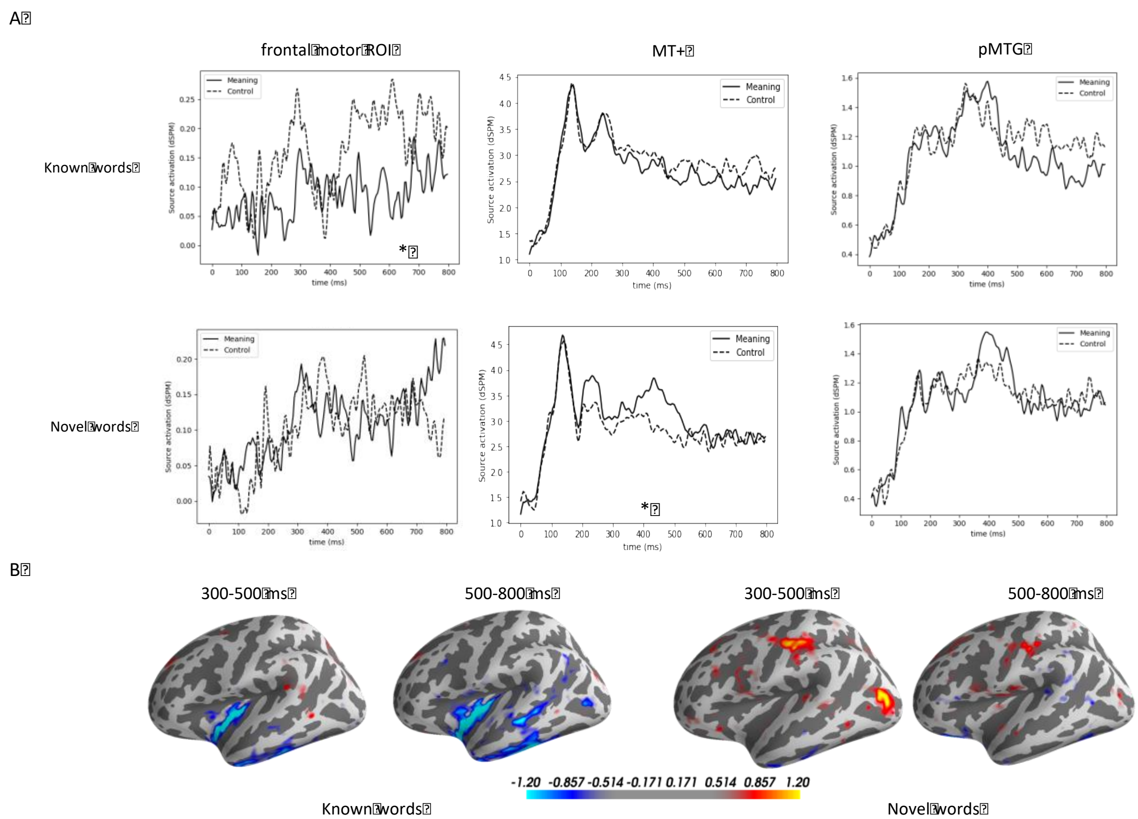

Figure 21. Source action in the letter judgment task.

A. Source estimates in the regions of interest. B. Activation maps for the meaning learning effect within 300-500 ms and within 500-800 ms (B) in the meaning judgment task. *: $p < .05$.

3.2.3.6 Cross-task comparison of source activation in the ROIs

The findings from the letter and meaning judgment tasks suggest that task modulated the meaning learning effect. We provide more direct evidence for the task effect by comparing the effect of meaning learning in the letter and meaning judgment tasks. For known words, the cross-task difference was observed in the frontal motor ROI in both time windows ($t(14) = -1.467$, $p = .082$ within 300-500 ms; $t(14) = -2.640$, $p = .009$ within 500-800 ms), in the pMTG within 500-800 ms ($t(14) = -2.650$, $p = .009$; $t(14) = -1.138$, $p = .137$ within 300-500 ms). A significant task effect was also observed in the left MT+ within 500-800 ms ($t(14) = -1.953$, $p = .036$; $t(14) = -$

0.579, p = .286 within 300-500 ms). The observed differences were driven by opposite patterns in the letter and meaning judgment tasks: source activation in the left pMTG, left frontal motor ROI, and the left MT+ for words with new meanings was stronger than for exposure controls in the meaning judgment task; a reversed pattern was found in the letter judgment task. In contrast, source activation for novel words in the ROIs was not affected by task (all ps > .26).

3.2.4 Discussion

The current study examined the role of the left pMTG in learning new meanings for known words. We compared the learning of new meanings for known words with that for novel words. Through a four-session training paradigm, participants reached performance ceiling in both cued-recall and recognition tests on the studied words. Behavioral data from the MEG task showed that after four-day training, participants were not slower in accessing the new meanings of known words than those of novel words. Instead, they were slightly more accurate in making judgments on known words in both letter and meaning judgment tasks. The result suggests reduced interference between new and original meanings over time, a finding consistent with Study 1 and one of our previous studies (Fang & Perfetti, 2019).

In the MEG data, we focused on source activation in the left pMTG and brain areas relevant to the processing of action meaning presented as verbal definitions (e.g., “lifting with one hand”). Although not statistically significant, we did see a trend of stronger activation for novel words with meanings than exposure controls in the left pMTG. In addition, regardless of task requirement, stronger source activation for novel words with action meanings than exposure controls was observed in the left MT+, a brain region sensitive to visual motion. For known words, the IFG (BA44) and lateral precentral gyrus and the left pMTG but not the left MT+ were more involved in the meaning judgment task for known words with new action meanings than exposure controls. However, when participants were performing a meaning-unrelated letter detection task, the activation was reduced when words had been paired with new meanings.

Overall, the MEG findings suggest that the new meanings of novel and known words are processed through the involvement of different parts of the sensorimotor circuits. The left pMTG was more reliably involved in the learning of new meanings for known words than for novel words. Furthermore, while source activation for novel words seems unaffected by task, the processing of new meanings of known words is modulated by whether new meanings are needed for a task.

3.2.4.1 The role of sensorimotor circuits in learning new action meanings

We focused on two ROIs that are relevant to the processing of action meanings. Previous studies comparing action and non-action words found stronger activation for action words in the left lateral precentral gyrus including ventral premotor cortex and the pars opercularis of the left IFG that extends into the premotor cortex (i.e., the frontal motor ROI here). The regions have been proposed to be associated with the abstract representation of actions as they are involved in the processing of action words regardless of involved body part (MacGregor et al., 2012; Moseley et al., 2013; Tettamanti et al., 2005). In addition to the frontal areas, the MT+ has also been found associated with the processing of action meanings. In particular, the region is sensitive to visual motion features of actions even when actions are verbally described (Saygin et al., 2010).

In our study, we presented action meanings in the format of verbal definitions and participants were instructed to visualize the meanings during the initial encoding. We observed enhanced activation of the left MT+ within 300-500 ms for novel words with action meanings than their exposure controls, suggesting fast reactivation of mental image of actions that were generated previously. Given the relatively early effect, it is unlikely that the visual motion features resulted from on-site motor imaginary during the MEG tasks. The retrieval of mental images suggests that the encoded episodic memory may still be important in accessing the meanings of novel words, even though meaning access has become faster over time. This is also consistent with Study 2a that found continuing involvement of episodic memory in the meaning access of novel words three days after initial learning.

Interestingly, accessing the new action meanings of known words seems less reliant on the left MT+. Instead, the left frontal motor ROI was more involved when participants were making judgments on the new meanings. Given that the frontal motor ROI is relevant to the processing of abstract action meaning while the left MT+ is sensitive to visual motion, it is possible that the new action meanings of known words have been more integrated or semanticized. This would be consistent with our argument that the integration of new meanings of known words benefits from the co-activation of new and prior word knowledge in the long run, even though interference from the original meanings may hinder the learning or access of new meanings before overnight sleep occurs (Fang et al., 2017; Fang & Perfetti, 2019). Alternatively, the reduced reliance on the left MT+ could reflect a lingering effect of original meanings. Maintaining the visual motion features of action meanings could potentially lead to the interference between new and original meanings, as the objects that the selected words originally refer to have dominant visual features such as color or shape (e.g., “bench”, “snow”). In contrast, this is not an issue for novel words as the action meanings are the only meanings.

3.2.4.2 The role of the left pMTG in meaning learning

One of the hypotheses about the function of the left pMTG in word processing is that the region serves as the lexical hub and maps lexical forms and word meanings (Hickok & Poeppel, 2004, 2007). Recent word learning studies suggest that this region binds new lexical constituents, replacing the role of the hippocampus when a new word is integrated into the mental lexicon (Bakker-Marshall et al., 2018; Ferreira et al., 2015; Landi et al., 2018; Takashima et al., 2014, 2017). In our study, we aimed to examine whether the left pMTG is also involved in binding new meanings to previously known words.

Our results showed when participants were making meaning judgments, source activation in the left pMTG was nonsignificantly stronger for novel words with new meanings than exposure controls, while the meaning learning effect in known words was significant within 300-800ms. This suggests that left pMTG was more reliably involved in accessing the new meanings of known words than those of novel words. For known words, the left pMTG is likely to be activated when a word is presented (Hagoort, 2005; Hickok & Poeppel, 2004, 2007), making it easier to create new connections with the rest of the brain (Schlichting & Frankland, 2017). Meanwhile, the left pMTG may interact with the hippocampal learning system, facilitating the takeover of its binding role. In contrast, novel words are assumed to be first represented in the hippocampal learning system and the involvement of the left pMTG is minimal during the initial learning. The hippocampus serves the main role of form-to-meaning mapping until lexical representation for novel words are established. Although the current analysis approach does not allow for the estimation of source activation in the hippocampus in a reliable way, it is possible that the hippocampus still supports the form-meaning mapping in the meaning access of novel words, as found in Study 2a.

Overall, the left pMTG seems to support the binding of lexical constitutes in general, a role the hippocampus plays when words are initially represented in the format of episodic memory. A connectivity analysis would be able to provide more direct evidence for this.

3.2.4.3 The modulation effect of task in processing new meanings

Another interesting finding in our study is that tasks modulate the meaning learning effects in novel words and known words differently. Regardless of task, a significant meaning learning effect in novel words was observed in the left MT+ within 300-500 ms, suggesting the mental imaginary of the action meanings were rapidly activated within the time window when lexico-semantic processing typically occurs. Such fast activation indicates that meaning access has become automatic four days following learning, even though those meanings may not be fully integrated yet, as discussed above.

In the processing of known words, we observed reserved patterns in both the left pMTG and left frontal motor ROI for the meaning learning effect in the meaning and letter judgment tasks. Specifically, source activation was stronger for words with new meanings in the meaning judgment task, while the activation was stronger for exposure controls in the letter judgment task. Although we expected a smaller meaning learning effect when meaning access is not needed for task completion, the reserved patterns were not anticipated.

In the letter judgment task, participants need to maintain information about word forms for later letter detection, which is likely to involve phonological processing. The frontal motor ROI including the pars opercularis of the left IFG and the precentral gyrus is associated with phonological processing (Hickok & Poeppel, 2004, 2007; Poldrack et al., 1999; Roskies, Fiez, Balota, Raichle, & Petersen, 2001). The suppression of action meanings may serve letter detection by making the regions available for phonological processing. It is unlikely that meaning suppression is always needed in a meaning-irrelevant task, Instead, suppression is needed here because all the tasks that participants had been performing right before the letter judgment task required them to access the new meanings, making the new actions meanings very accessible or even more dominant (Rodd, Lopez Cutrin, Kirsch, Millar, & Davis, 2013). The reduced source activation in the left frontal motor ROI may be associated with reduced source activation in the left pMTG. However, again, connectivity analysis is needed to test this hypothesis. The absence of meaning suppression for novel words in the letter judgment task could be because the processing of new action meanings is mainly supported by the left MT+, rather than the frontal areas. Therefore, meaning activation caused little interference with the letter judgment.

3.2.4.4 Null findings in the localizer task

The null finding from the localizer task is unexpected. The words presented in the localizer task were selected to maximize the difference in the motor involvement. Furthermore, Study 2a, using basically the same stimuli, showed a clear difference in ERPs between action and non-action words. One concern is that the localizer task was the final MEG task and most participants performed the task 2-2.5 hours after the beginning of the session. As a result, alpha activities were more dominant towards the end of the session when participants were experiencing more fatigue, leading to the relative low signal-noise ratio of the data overall. One additional concern was that, the anatomically defined ROIs are likely to include vertices that are not necessarily involved in the cognitive processes of interest. Therefore, future analysis using functionally constrained and individually defined ROIs may increase the power of detecting the difference between conditions. In addition, the relatively small sample size (N=15) could lead to less robust effects overall including the effect in the localizer task.

3.2.5 Conclusion

The learning of new action meanings presented as verbal definitions is supported by sensorimotor cortices relevant to the representation of actions. While the left MT+ is associated with the processing of new action meanings of novel words, the left inferior frontal gyrus (BA44) and precentral gyrus are associated with the processing of new actions meanings of known words. Such difference suggests new meanings of known words are more semanticized than those of novel words. Meanwhile, the left posterior middle temporal gyrus (pMTG) seems more reliably involved in the learning of new meanings for known words than for novel words. Overall, the findings suggest that the left pMTG is involved in binding new meanings to previously known words, possibly by interacting with neocortical areas relevant to more specific representation of new meanings.

4.0 General Discussion

The presented studies are part of a larger research program that aims to uncover the mechanisms underlying the learning of new meanings for known words. Study 1 showed that learning new meanings benefits from the study-test interval involving overnight sleep. However, within the first 24 hours, new meanings seem not yet fully integrated and episodic retrieval still plays a crucial role in accessing the new meanings. When learning spanned over days as in Study 2b, new meanings are more integrated through neocortical learning. In particular, following four-day learning, the left pMTG and the left frontal areas were involved in accessing the new action meanings of known words, supporting the important role of the left pMTG in binding new meanings to known words. For novel words, accessing the new action meanings is associated with the left MT+, a region sensitive to visual motion. While the integration of new meanings is slower than that of novel words within the first 24 hours possibly because of interference, it seems to catch up later when more time and more learning opportunities are given. In this chapter, I explain different stages of learning based on the studies reported in the dissertation and our previous work and then provide implications for word learning in general.

4.1 The co-activation model for the learning of new meanings for known words

One unique feature of learning new meanings for known words is that strong co-activation of new and prior word knowledge is involved from the very beginning of learning. Here I emphasize the role of such knowledge co-activation in different stages of learning, based on the standard model of system consolidation (Frankland & Bontempi, 2005). When a word is presented, the knowledge about the word form, meaning, and mapping between them is automatically activated (Humphreys et al., 1982; Lesch & Pollatsek, 1993; Perfetti et al., 1988). As shown in Figure 22, during the initial encoding, the hippocampus receives input from neocortical areas representing prior knowledge and also those relevant to the representation of new meanings. The initial connections between new and prior knowledge are created and represented in the hippocampus, and the connections among neocortical areas are very weak during the encoding phase. At this point, these hippocampus-dependent connections are sufficient to support the recognition and recall of new meanings. During memory consolidation, the hippocampus replays the memories to the neocortex. It is likely that new and original meanings are both reactivated during overnight consolidation. This again provides opportunity for the interaction between hippocampal learning and neocortical learning to occur. In addition to the interaction among neocortical areas relevant to the representation of word forms and original meanings and that of new meanings, the left pMTG and the neocortical areas relevant to the new meanings are also co-activated. Over time, these neocortical connections are stronger and stronger while the hippocampus-dependent connections become weaker and weaker. Eventually the mapping between new meanings and words is represented in the left pMTG.

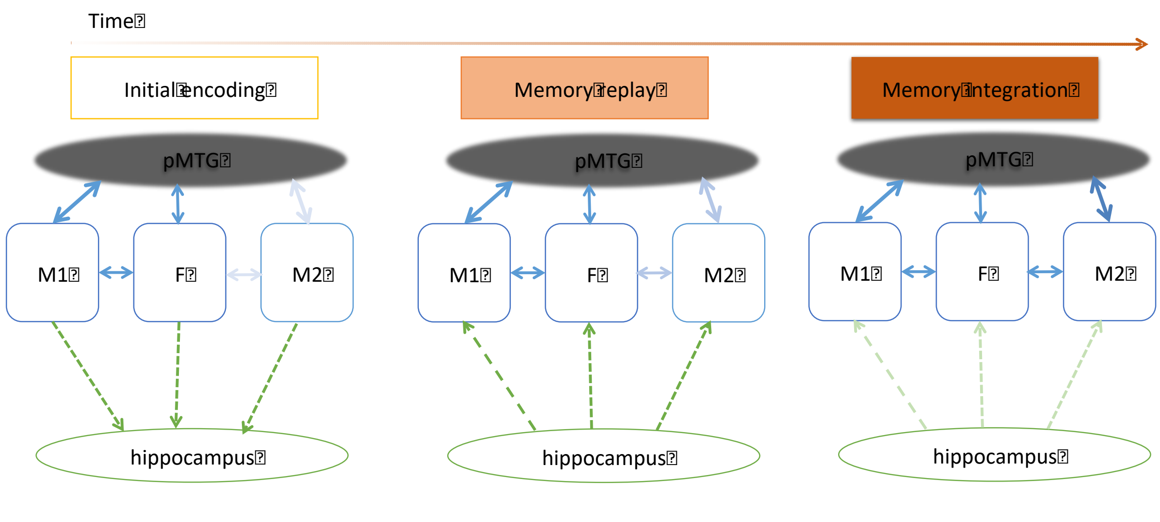

Figure 22. Model for time-dependent changes in the learning of new meanings for known words.

Memory replay mainly occurs during overnight consolidation. The left pMTG represents existing form-meaning mappings and supports the formation and updating of lexical representations. When new and original meanings are semantically unrelated, there are inhibitory connections between them. Solid lines represent connections among neocortical areas; dash lines represent connections between the hippocampus and neocortical areas; darkness of lines represents strength of connections.

While knowledge co-activation facilitates the integration of new meanings in the long run, it can sometimes be a disadvantage. The reactivation of original meanings can interfere with the initial learning of new meanings, especially when new and original meanings are semantically unrelated (Rodd et al., 2012) or when participants know the original meanings very well (Fang et al., 2017; Fang & Perfetti, 2019). Meanwhile, such interference slows down the access of original meanings on the day of learning, as found in Study 1 and previous studies (Fang & Perfetti, 2019). It is likely that inhibitory connections between new and original meanings are formed during the initial learning. Such connections are important in performing tasks that only need to access one of the meanings.

The most interesting finding is that the disadvantage of knowledge co-activation wanes over time, especially after overnight consolidation occurs. New meanings of known words are more accessible after overnight consolidation occurs (Study 1). What's more, although strong interference between new and original meanings hinders the initial learning, it benefits long-term retention of new meanings (Fang & Perfetti, 2019). Meanwhile, the access of original meanings also benefits from overnight sleep so that the speed of meaning access recovers after 24 hours (Study 1; Fang & Perfetti, 2019). Therefore, both new and original meanings benefit from overnight consolidation.

The faster access to both new and original meanings over time suggests that more distinctive representations for different meanings of a word are formed. In the encoding of similar events, the dentate gyrus, a subfield of the hippocampus, is associated with forming more distinctive hippocampal representations (Duncan & Schlichting, 2018; Favila, Chanales, & Kuhl, 2016). Establishing distinctive representations for new and original meanings may rely on a different mechanism, because prior word knowledge is represented in the neocortex rather than in the hippocampus. As mentioned in the discussion of Study 1, one possible solution to differentiate meanings is to include certain contextual information or context nodes for different meanings (Armstrong & Plaut, 2008). For example, the context nodes for new meanings may be about experimental sessions and are initially represented in the format of episodic memory. When new meanings are encountered in different language contexts, the context nodes become more meaning-relevant. Such information may also need repeated exposure and overnight consolidation to become part of the meaning representations. Once the context nodes are established, meaning selection can be achieved by matching language input with context nodes.

Another scenario is that new meanings are semantically related to original meanings. In this case, the interference among meanings is much weaker, if there is any. In the absence of strong inhibitory connections between them, new meanings may become integrated much faster and rely less on overnight consolidation. These hypotheses are consistent with recent updates of the complementary learning system models where the influence of prior knowledge on the speed of integration is acknowledged (Kumaran et al., 2016; McClelland, 2013).

Overall, the co-activation of new and prior word knowledge benefits the long-term retention and integration of new meanings, even though it may hinder initial learning. Similar to the learning of novel words, overnight sleep plays an important role in the learning of new meanings for known words. The role of overnight consolidation may involve reducing the interference between new and original meanings and facilitating selective meaning access, possibly by establishing context nodes for different meanings.

4.2 Implications for word learning in general

Establishing the associations between word forms and word meanings is an essential part of building vocabulary knowledge. With well-established form-meaning connections, one can efficiently retrieve the most appropriate words to express ideas (i.e., meaning-to-form) or comprehend what is heard or seen (i.e., form-to-meaning). While learning new meanings for known words is a special case of word learning, the relevant findings can inform the research on word learning in general. In particular, the knowledge co-activation can be applied to the learning of novel words. Previous studies have shown that under certain learning conditions when knowledge co-activation is enhanced by making the link between new and prior knowledge more available, integration of novel words can occur faster. For example, when novel words and existing words that have similar pronunciations are presented in an interleaved way, novel words can be integrated right away (Lindsay & Gaskell, 2013). In a fast-mapping paradigm where the picture referents of novel words are presented along with familiar and relevant pictures, immediate integration has also been reported (Coutanche & Thompson-Schill, 2014). Enhancing knowledge co-activation is essentially engaging the neocortical learning system and also facilitating the interaction between hippocampal and neocortical learning systems during initial learning. The connections formed during the encoding of new information can further facilitate knowledge co-activation during post-learning memory replay and benefits integration of new word knowledge in the long run.

In terms of neural mechanisms for word learning, the left pMTG supports the establishment of form-meaning mappings for novel words (Bakker-Marshall et al., 2018; Landi et al., 2018; Takashima et al., 2014, 2017). The same region seems to function in the binding of new meanings to known words, suggesting that the left pMTG is involved in the establishment of new form-meaning mappings in general. This is consistent with the argument that the left pMTG is the lexical hub (Hagoort, 2005; Hickok & Poeppel, 2004, 2007). Anatomically, the left pMTG is connected with other brain regions relevant to the language processing through fiber tracks (Friederici, 2011). In addition, this region has also been proposed to represent modality-independent or under-specific semantic information (Binder & Desai, 2011; Papeo et al., 2015). While the specific role of the left pMTG remains controversial, the region is an ideal replacement for the hippocampus in the long run given its functional and anatomical properties. Overall, the left pMTG may connect the regions relevant to the specific representation of word meanings and those relevant to word form representation (e.g., visual word form area or VWFA in Dehaene and Cohen (2011)), although more direct evidence is needed.

4.3 Summary and conclusions

The thesis examined how overnight consolidation affected the learning new meanings for known words and the role of the left pMTG in binding new meanings to known words. Study 1 showed that the processing of both new and original meanings became faster after overnight sleep. This indicates reduced interference between new and original meanings over time, especially after overnight consolidation occurs. However, the ERP data showed that accessing the new meanings was still mainly supported by episodic retrieval even 24 hours after learning. To investigate how new meanings are associated with known words, Study 2a first demonstrated that presenting word meanings as verbal definitions is sufficient to drive semantic category effect. Based on this, Study 2b further showed that the left pMTG, in addition to other neocortical areas relevant to the specific representation of new meanings, is involved in binding new meanings to known words.

Combined with the previous findings on learning novel words, the dissertation results suggest that the co-activation of new and prior knowledge is essential to the integration of new word knowledge into the mental lexicon. The interactions between the hippocampal and neocortical learning systems are likely to be part of the mechanisms. The left pMTG not only supports the formation of novel form-meaning associations, but also the associations between new meanings and previously known words.

Appendix A Supplementary Tables and Figures

A.1 Study 2a

Table S1-1. Descriptive statistics of lexical and sub-lexical characteristics of existing and novel words (Study 2a)

	Condition	Phoneme	Letter	NB(O)	NB(P)	Frequency	Valence	Arousal
Existing words	Action	3.7 (0.66)	4.7 (0.86)	5.95 (3.94)	12.05 (7.49)	38.86 (52.83)	5.59 (0.79)	4.10 (0.68)
	Non-action	3.8 (0.62)	7.8 (1.00)	5.95 (5.81)	10.25 (8.16)	44.39 (56.22)	5.74 (0.69)	3.83 (0.64)
Novel words		4.35 (0.75)	5.45 (0.69)					

Notes: Number of phonemes, letters and orthographic neighbor size are reported for both novel words and existing words. Phonological neighbor size, word frequency (based on the SUBTLEX(US) corpus, Brysbaert & New, 2009), ratings for valence and arousal (Warriner, Kuperman, & Brysbaert, 2013) are reported for existing words additionally. Except ratings for valence and arousal, all the other data were extracted from the English Lexicon Project (Balota et al., 2007).

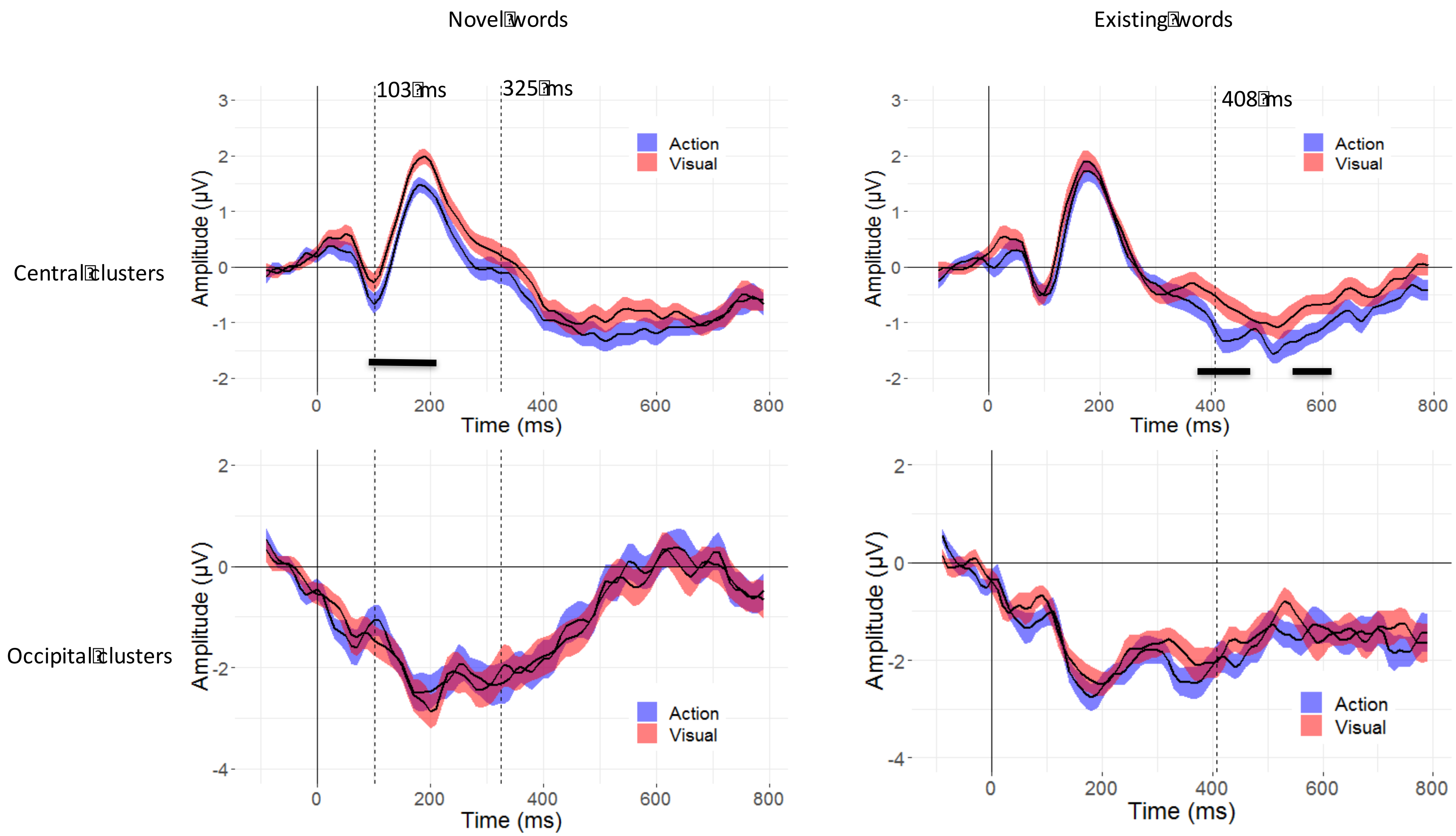

Figure S1-1. Permutation test results on central and occipital clusters for novel words and existing words (Study 2a).

Left: Novel words; Right: Existing words; Upper: Central clusters; Lower: Occipital clusters. Black lines represent time points showing significant difference between action and non-action conditions, based on 10,000 permutations. Vertical line represents recognition points.

A.2 Study 2b

Table S2-1. Lexical and sublexical characteristics of word stimuli (Study 2b)

Condition	Number of Letters	Mean Bigram Frequency (log)	Word Frequency (SUBTL)	Number of Senses	Valence	Arousal	Concreteness	Orthographic Neighborhood Size	Phonologic Neighborhood Size
Known words	5.13 ± 0.71	3.24 ± 0.19	12.77 ± 12.31	4.63 ± 2.32	5.94 ± 0.61	3.40 ± 0.57	4.84 ± 0.19	4.35 ± 4.22	10.16 ± 7.37
Novel words	5.31 ± 0.59	3.18 ± 0.17							
Localizer words									
Action	4.69 ± 0.95	3.14 ± 0.21	36.87 ± 53.96	11.31 ± 5.19	5.54 ± 0.79	4.15 ± 0.71	4.06 ± 0.26	6.38 ± 4.26	12.13 ± 8.29
Non-action	4.88 ± 1.02	3.26 ± 0.20	51.79 ± 60.57	8.88 ± 5.21	5.85 ± 0.58	3.90 ± 0.64	4.33 ± 0.33	5.25 ± 6.01	10.81 ± 8.89

Notes: Word frequency is based on the SUBTL (US) corpus (Brysbaert & New, 2009); ratings for valence and arousal are from the database by Warriner et al. (2013); Concreteness is based on the database by Brysbaert et al. (2013); number of senses is from the Wordsmith English Dictionary (Parks et al., 1998). Word frequency, orthographic and phonological neighborhood size are retrieved from the English Lexicon Project (Balota et al., 2007).

Table S2- 2. Fixed effect estimates for models of performance in cued-recall and recognition tests (Study 2b)

	beta	***SE***	***t or z***	***p***	
Cued-recall (score)					
Intercept	4.955	0.020	245.293	< .001	***
Type (Meaning vs. Control)	-0.058	0.013	-4.478	< .001	***
Lexicality (Novel vs. Known)	-0.006	0.014	-0.468	0.640	
Session2vs1	0.063	0.022	2.819	0.005	**
Session3vs2	-0.081	0.026	-3.114	0.002	**
Session4vs3	0.059	0.022	2.615	0.009	**
Type: Lexicality	-0.041	0.026	-1.562	0.118	
Type:Session2vs1	0.154	0.045	3.432	< .001	***
Type:Session3vs2	-0.143	0.052	-2.760	0.006	**
Type:Session4vs3	0.090	0.045	2.002	0.045	*
Lexicality:Session2vs1	0.118	0.045	2.615	0.009	***
Lexicality:Session3vs2	-0.085	0.052	-1.628	0.104	
Lexicality:Session4vs3	0.061	0.045	1.348	0.178	
Type:Lexicality:Session2vs1	0.107	0.090	1.185	0.236	
Type:Lexicality:Session3vs2	-0.206	0.104	-1.981	0.048	*
Type:Lexicality:Session4vs3	0.103	0.090	1.144	0.253	
Recognition (accuracy, by-subject)					
Intercept	3.462	0.020	177.553	< .001	***
Type	-0.034	0.021	-1.642	0.101	
Lexicality	-0.051	0.030	-1.730	0.084	~
Session2vs1	0.103	0.036	2.844	0.004	**
Session3vs2	0.103	0.042	2.463	0.014	*
Session4vs3	0.034	0.036	0.948	0.343	
Type: Lexicality	-0.034	0.042	-0.821	0.412	
Type:Session2vs1	0.137	0.072	1.896	0.058	~
Type:Session3vs2	0.068	0.083	0.821	0.412	
Type:Session4vs3	0.000	0.072	0.000	1.000	
Lexicality:Session2vs1	0.239	0.072	3.318	< .001	***
Lexicality:Session3vs2	0.137	0.083	1.642	0.101	
Lexicality:Session4vs3	0.034	0.072	0.474	0.636	
Type:Lexicality:Session2vs1	0.342	0.144	2.370	0.018	*
Type:Lexicality:Session3vs2	0.000	0.166	0.000	1.000	
Type:Lexicality:Session4vs3	-0.068	0.144	-0.474	0.636	
Recognition (accuracy, by-item)					
Intercept	2.841	0.015	191.381	< .001	***
Type	-0.057	0.029	-1.971	0.049	*
Lexicality	-0.096	0.030	-3.220	< .001	***
Session2vs1	0.029	0.021	1.394	0.163	
Session3vs2	0.048	0.021	2.323	0.020	*
Session4vs3	0.038	0.021	1.858	0.063	~
Type: Lexicality	-0.115	0.058	-1.971	0.049	*
Type:Session2vs1	0.057	0.041	1.394	0.163	
Type:Session3vs2	0.057	0.041	1.394	0.163	

Type:Session4vs3	0.038	0.041	0.929	0.353	
Lexicality:Session2vs1	0.096	0.041	2.323	0.020	*
Lexicality:Session3vs2	0.096	0.041	2.323	0.020	*
Lexicality:Session4vs3	0.076	0.041	1.858	0.063	~
Type:Lexicality:Session2vs1	0.191	0.082	2.323	0.020	*
Type:Lexicality:Session3vs2	0.115	0.082	1.394	0.163	
Type:Lexicality:Session4vs3	0.076	0.082	0.929	0.353	

Notes: Intercept represents mean performance across all the conditions. Final model for cued-recall test: Score ~ Type * Lexicality * Session + (1 | Subject) + (1 | Word); final model for by-subject analysis of recognition test: EmpLogit(ACC) ~ Type *Lexicality * Session + (1 + Lexicality | Subject); final model for by-item analysis of recognition test: EmpLogit (ACC) ~ Type *Lexicality * Session + (1 | Word). ***: $p < .001$, **: $p < .01$, *: $p < .05$, ~: $p < .10$.

Appendix B Experimental Stimuli and Rubric

Table S3- 1. Word stimuli (Study 1)

Category	**Word**	**Category**	**Word**	**Category**	**Word**	**Category**	**Word**
man-made	bench	man-made	bread	man-made	boot	man-made	brick
man-made	bucket	man-made	desk	man-made	cereal	man-made	dime
man-made	fence	man-made	hat	man-made	glass	man-made	hut
man-made	ink	man-made	linen	man-made	lamp	man-made	map
man-made	maze	man-made	pole	man-made	mirror	man-made	porch
man-made	roof	man-made	seat	man-made	rug	man-made	sleeve
man-made	soup	man-made	tape	man-made	stair	man-made	tent
man-made	tray	man-made	vase	man-made	tube	man-made	vest
natural	banana	natural	carrot	natural	birch	natural	moss
natural	dew	natural	grape	natural	garlic	natural	snow
natural	ice	natural	leaf	natural	ivy	natural	cloud
natural	moon	natural	mud	natural	onion	natural	grass
natural	oak	natural	pea	natural	pearl	natural	nest
natural	pear	natural	pepper	natural	wheat	natural	peach
natural	shell	natural	stone	natural	lemon	natural	tomato
natural	walnut	natural	wood	natural	plum	atural	yam

Table S3- 2. Definitions (Study 1 and Study 2b)

Group	Type	Verb	Definition
1	hand	reach	reaching upward
1	hand	kneed	kneading forcefully
1	hand	wave	waving side to side
1	hand	type	typing rapidly
1	leg	tip-toe	tip-toeing stealthily
1	leg	hike	hiking up an incline
1	leg	kneel	kneeling slowly
1	leg	run	running quickly
1	hand	smash	smashing instantly
1	hand	write	writing from right to left
1	hand	slide	sliding fingers to the right
1	hand	push	pushing towards the left
1	leg	skate	skating smoothly
1	leg	jump	jumping high
1	leg	skip	skipping very fast
1	leg	walk	walking backwards
2	hand	pat	patting sporadically
2	hand	tug	tugging back and forth
2	hand	grasp	grasping firmly
2	hand	rub	rubbing in circles
2	leg	dig	digging with toes
2	leg	stomp	stomping intensely
2	leg	kick	kicking with force
2	leg	pedal	pedaling with effort
2	hand	lift	lifting with one hand
2	hand	squeeze	squeezing repeatedly
2	hand	poke	poking with both hands
2	hand	throw	throwing underhand
2	leg	march	marching in place
2	leg	hop	hopping on one leg
2	leg	limp	limping stiffly
2	leg	tap	tapping one foot constantly

Table S3- 3. Rubric for the scoring of responses in the cued-recall tests (Study 1 and Study 2b)

Score	Response
0	No response or unknown indicated
1	Relevant modifier + irrelevant/missing verb
2	Correct modifier + irrelevant/missing verb
3	Correct verb + irrelevant/missing modifier
4	Correct verb + relevant modifier
5	Correct verb + correct modifier

Table S3- 4. Novel words (Study 2a)

Set	Written form	Pronunciation	Set	Written form	Pronunciation
1	adaint	ədeint	2	attave	ərev
1	blauge	blaʊʤ	2	bloosh	bluʃ
1	bropt	bɹɒpt	2	chalph	ʧælf
1	crulge	kɹɔlʤ	2	drault	drɑlt
1	dwock	dwak	2	flerp	flɚp
1	fralt	fɹælt	2	fruch	fɹuʧ
1	gaiph	gef	2	gelb	gɛlb
1	glerg	glɚg	2	grelve	grɛlv
1	knisp	nɪsp	2	knurt	nɝːɹt
1	larsk	laɹsk	2	lootch	luʧ
1	maldge	malʤ	2	moip	mɔɪp
1	plauve	plaʊv	2	plisk	plɪsk
1	praff	pɹæf	2	relsh	rɛlʃ
1	rhonge	rɔnʤ	2	smange	smænʤ
1	snalve	snalv	2	spronk	spɹɑːnk
1	strimph	stɹɪmf	2	swulch	swɑlʧ
1	thalp	θælp	2	thralk	θɹalk
1	trithy	tɹɪθi	2	tweche	twɛtʃ
1	twoom	twum	2	twult	twɔːlt
1	vanty	vænti	2	vorsh	vɔɹʃ

Table S3- 5. Existing words (Study 2a)

Condition	Word	Condition	Word
Action	swim	Visual	tan
Action	crawl	Visual	purple
Action	spin	Visual	red
Action	slide	Visual	green
Action	dance	Visual	brown
Action	dive	Visual	yellow
Action	throw	Visual	pink
Action	scratch	Visual	white
Action	scoop	Visual	black
Action	pull	Visual	stripe
Action	carve	Visual	circle
Action	dip	Visual	curve
Action	draw	Visual	cube
Action	bend	Visual	square
Action	nudge	Visual	cone
Action	catch	Visual	flat
Action	swing	Visual	sphere
Action	tread	Visual	star
Action	leap	Visual	arrow
Action	stride	Visual	oval

Table S3- 6. Definitions (Study 2a)

Condition	Definition	Condition	Definition
action	twisting one's body around	visual	with a dark blue surface
action	bouncing from side to side	visual	with golden flowers
action	trembling without stopping	visual	with silver hair
action	rotating to the right	visual	with a maroon outline
action	lifting with one hand	visual	with a bright teal color
action	pushing toward the left	visual	with rust-colored spots
action	reaching upward	visual	resembling an orange ellipse
action	grasping firmly	visual	covered in beige dots
action	tugging back and forth	visual	appearing like a rainbow
action	typing rapidly	visual	with a wood grain pattern
action	squeezing repeatedly	visual	covered in crossing lines
action	smashing instantly	visual	with many thin cracks
action	shoving in a different direction	visual	speckled with turquoise
action	hitting with force	visual	covered in gray zig-zags
action	stomping intensely	visual	with a triangle head
action	kicking mightily	visual	with a long tail
action	walking backward	visual	with a wide cylindrical body
action	running quickly	visual	in a heart shape
action	hopping up and down	visual	in the shape of an octagon
action	digging with one's toes	visual	with tiny ears

Table S3- 7. Known words (Study 2b)

Group	Word	Sound	Color	Manipulation	Motion	Emotion
1	bench	0.67	1.45	1.83	0.15	-0.15
1	brick	1.23	3.82	1.38	0	-0.55
1	cereal	1.67	3.7	2.58	0.42	1.69
1	glass	2.92	1.67	2.86	0.36	0.91
1	linen	0.77	3.73	1.91	0.45	0.55
1	porch	1.27	1.33	1.54	0.64	1.67
1	soup	1.4	3	2.92	0.82	1.83
1	vase	0.31	1.73	2.67	0.08	1.08
1	banana	0.2	5.26	2.89	0.26	1.9
1	carrot	1.07	5.27	2.07	0.16	0.84
1	garlic	0.46	2	2	0.25	-0.46
1	moss	0.46	5.1	0.69	0.33	0.75
1	pearl	0.54	3.9	2.62	0.09	2.85
1	shell	1.75	2.54	1.46	0.23	1.75
1	tomato	0.21	5.54	2	0.18	0.63
1	grape	0.42	4.73	2.64	0	1.92
2	bread	0.42	2.33	3.17	0	1.36
2	bucket	1	1	2.17	0.45	-0.82
2	fence	0.3	2	2.21	0	-0.54
2	lamp	1.33	2.77	2.69	0.1	0.62
2	mirror	0.64	1.5	2.82	0.31	0.08
2	sleeve	0.3	0.45	2.69	0.38	0
2	tent	1.27	2.27	2.69	0.55	1.58
2	vest	0.08	2	2.64	0.1	0
2	grass	0.64	4.85	2.25	0.67	1
2	birch	1.11	2.38	1	0.22	1
2	cloud	0.5	3.45	0.42	2	1.82
2	lemon	0.58	4.91	2.67	0.08	0.58
2	onion	0.58	2.91	2.85	0.31	-0.91
2	pepper	0.45	3.98	1.98	0.53	-0.17
2	stone	1.23	2.36	2.67	0.27	-0.58
2	walnut	0.64	3.31	2	0.29	0.25

Notes: Ratings of five attributes from the Wisconsin perceptual attribute ratings database (Medler et al., 2005) are presented.

Table S3- 8. Novel words (Study 2b)

Group	Word
1	adaint
1	bloosh
1	chalph
1	drault
1	flerp
1	frert
1	gelb
1	knisp
1	larsk
1	maldge
1	plisk
1	praff
1	slere
1	spronk
1	trithy
1	vanty
2	attave
2	bropt
2	criph
2	dakle
2	fraine
2	garck
2	glerg
2	knart
2	loatch
2	moip
2	plauve
2	relsh
2	snalve
2	thalp
2	twalt
2	vorsh

Table S3- 9. Untrained words (Study 2b)

Condition	Word
Action	swim
Action	crawl
Action	dance
Action	dive
Action	scratch
Action	scoop
Action	pull
Action	carve
Action	dip
Action	draw
Action	bend
Action	nudge
Action	catch
Action	tread
Action	leap
Action	stride
Visual	tan
Visual	purple
Visual	red
Visual	green
Visual	brown
Visual	yellow
Visual	white
Visual	black
Visual	circle
Visual	curve
Visual	square
Visual	cone
Visual	sphere
Visual	star
Visual	arrow
Visual	oval

Bibliography

Armstrong, B.C., & Plaut, D.C. (2008). *Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence.* Paper presented at the Proceedings of the Cognitive Science Society.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

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*(0), 116-130. doi: http://dx.doi.org/10.1016/j.jml.2014.03.002

Bakker, I., Takashima, A., van Hell, J.G., Janzen, G., & McQueen, J.M. (2015). Tracking lexical consolidation with erps: Lexical and semantic-priming effects on n400 and lpc responses to newly-learned words. *Neuropsychologia, 79*(Pt A), 33-41. doi: 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. *J Cogn Neurosci, 30*(5), 621-633. doi: 10.1162/jocn_a_01240

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The english lexicon project. *Behav Res Methods, 39*(3), 445-459.

Barsalou, L.W. (2008). Grounded cognition. *Annu Rev Psychol, 59*, 617-645. doi: 10.1146/annurev.psych.59.103006.093639

Binder, J.R., & Desai, R.H. (2011). The neurobiology of semantic memory. *Trends Cogn Sci, 15*(11), 527-536. doi: 10.1016/j.tics.2011.10.001

Bowers, J.S., Davis, C.J., & Hanley, D.A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition, 97*(3), B45-54. doi: 10.1016/j.cognition.2005.02.002

Brown, J.I. (1960). The nelson-denny reading test.

Brysbaert, M., & New, B. (2009). Moving beyond kucera and francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. *Behav Res Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in psychology*, *7*, 1116.

Brysbaert, M., Warriner, A.B., & Kuperman, V. (2013). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior research methods*, 1-8.

Buysse, D.J., Reynolds, C.F., 3rd, Monk, T.H., Berman, S.R., & Kupfer, D.J. (1989). The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res, 28*(2), 193-213.

Coutanche, M.N., & Koch, G.E. (2017). Variation across individuals and items determine learning outcomes from fast mapping. *Neuropsychologia*, *106*, 187-193. doi: 10.1016/j.neuropsychologia.2017.09.029

Coutanche, M.N., & Thompson-Schill, S.L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology-General, 143*(6), 2296-2303. doi: 10.1037/xge0000020

Curran, T., & Friedman, W.J. (2004). Erp old/new effects at different retention intervals in recency discrimination tasks. *Cognitive Brain Research, 18*(2), 107-120. doi: 10.1016/j.cogbrainres.2003.09.006

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., & Halgren, E. (2000). Dynamic statistical parametric mapping: Combining fmri and meg for high-resolution imaging of cortical activity. *Neuron, 26*(1), 55-67.

Davis, M.H., & Gaskell, M.G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philos Trans R Soc Lond B Biol Sci, 364*(1536), 3773-3800. doi: 10.1098/rstb.2009.0111

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends Cogn Sci, 15*(6), 254-262. doi: 10.1016/j.tics.2011.04.003

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., . . . Killiany, R.J. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage, 31*(3), 968-980. doi: 10.1016/j.neuroimage.2006.01.021

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nat Rev Neurosci, 11*(2), 114-126. doi: 10.1038/nrn2762

Donnelly, S., & Verkuilen, J. (2017). Empirical logit analysis is not logistic regression. *Journal of Memory and Language, 94*, 28-42. doi: 10.1016/j.jml.2016.10.005

Dudai, Y., Karni, A., & Born, J. (2015). The consolidation and transformation of memory. *Neuron, 88*(1), 20-32. doi: 10.1016/j.neuron.2015.09.004

Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39. doi: 10.1111/j.1467-9280.2007.01845.x

Duncan, K.D., & Schlichting, M.L. (2018). Hippocampal representations as a function of time, subregion, and brain state. *Neurobiol Learn Mem, 153*(Pt A), 40-56. doi: 10.1016/j.nlm.2018.03.006

Fang, X., Perfetti, C., & Stafura, J. (2017). Learning new meanings for known words: Biphasic effects of prior knowledge. *Lang Cogn Neurosci, 32*(5), 637-649. doi: 10.1080/23273798.2016.1252050

Fang, X., & Perfetti, C.A. (2017). Perturbation of old knowledge precedes integration of new knowledge. *Neuropsychologia, 99*, 270-278. doi: 10.1016/j.neuropsychologia.2017.03.015

Fang, X., & Perfetti, C.A. (2019). Learning new meanings for known words: Perturbation of original meanings and retention of new meanings. *Mem Cognit, 47*(1), 130-144. doi: 10.3758/s13421-018-0855-z

Fargier, R., Ploux, S., Cheylus, A., Reboul, A., Paulignan, Y., & Nazir, T.A. (2014). Differentiating semantic categories during the acquisition of novel words: Correspondence analysis applied to event-related potentials. *J Cogn Neurosci, 26*(11), 2552-2563. doi: 10.1162/jocn_a_00669

Favila, S.E., Chanales, A.J.H., & Kuhl, B.A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications, 7*. doi: ARTN 1106610.1038/ncomms11066

Ferreira, R.A., Gobel, S.M., Hymers, M., & Ellis, A.W. (2015). The neural correlates of semantic richness: Evidence from an fmri study of word learning. *Brain Lang, 143*, 69-80. doi: 10.1016/j.bandl.2015.02.005

Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., . . . Zilles, K. (2008). Cortical folding patterns and predicting cytoarchitecture. *Cereb Cortex, 18*(8), 1973-1980. doi: 10.1093/cercor/bhm225

Frankland, P.W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nat Rev Neurosci, 6*(2), 119-130. doi: 10.1038/nrn1607

Franz, V.H., & Loftus, G.R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing loftus and masson (1994) and avoiding the biases of alternative accounts. *Psychon Bull Rev, 19*(3), 395-404. doi: 10.3758/s13423-012-0230-1

Friederici, A.D. (2011). The brain basis of language processing: From structure to function. *Physiol Rev, 91*(4), 1357-1392. doi: 10.1152/physrev.00006.2011

Gaskell, M.G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*(2), 105-132. doi: 10.1016/s0010-0277(03)00070-2

Gow, D.W. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain and Language, 121*(3), 273-288. doi: 10.1016/j.bandl.2012.03.005

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., . . . Hamalainen, M. (2013). Meg and eeg data analysis with mne-python. *Front Neurosci, 7*, 267. doi: 10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., . . . Hamalainen, M.S. (2014). Mne software for processing meg and eeg data. *Neuroimage, 86*, 446-460. doi: 10.1016/j.neuroimage.2013.10.027

Hagoort, P. (2005). On broca, brain, and binding: A new framework. *Trends Cogn Sci, 9*(9), 416-423. doi: 10.1016/j.tics.2005.07.004

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1-2), 67-99. doi: 10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nat Rev Neurosci, 8*(5), 393-402. doi: 10.1038/nrn2113

Humphreys, G.W., Evett, L.J., & Taylor, D.E. (1982). Automatic phonological priming in visual word recognition. *Memory & Cognition, 10*(6), 576-590.

James, T.W., & Gauthier, I. (2003). Auditory and action semantic features activate sensory-specific perceptual brain regions. *Curr Biol, 13*(20), 1792-1796. doi: 10.1016/j.cub.2003.09.039

Johns, M.W. (1991). A new method for measuring daytime sleepiness: The epworth sleepiness scale. *Sleep, 14*(6), 540-545. doi: 10.1093/sleep/14.6.540

Kiefer, M., Sim, E.J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *J Neurosci, 28*(47), 12224-12230. doi: 10.1523/JNEUROSCI.3579-08.2008

Kirby, J.R., Moore, P.J., & Schofield, N.J. (1988). Verbal and visual learning styles. *Contemporary educational psychology, 13*(2), 169-184.

Kocagoncu, E., Clarke, A., Devereux, B.J., & Tyler, L.K. (2017). Decoding the cortical dynamics of sound-meaning mapping. *J Neurosci, 37*(5), 1312-1319. doi: 10.1523/JNEUROSCI.2858-16.2016

Kumaran, D., Hassabis, D., & McClelland, J.L. (2016). What learning systems do intelligent agents need? Complementary learning systems theory updated. *Trends Cogn Sci, 20*(7), 512-534. doi: 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. doi: 10.1146/annurev.psych.093008.131123

Landi, N., Malins, J.G., Frost, S.J., Magnuson, J.S., Molfese, P., Ryherd, K., . . . Pugh, K.R. (2018). Neural representations for newly learned words are modulated by overnight consolidation, reading skill, and age. *Neuropsychologia, 111*, 133-144. doi: 10.1016/j.neuropsychologia.2018.01.011

Lee, C.L., & Federmeier, K.D. (2006). To mind the mind: An event-related potential study of word class and semantic ambiguity. *Brain Res, 1081*(1), 191-202. doi: 10.1016/j.brainres.2006.01.058

Lesch, M.F., & Pollatsek, A. (1993). Automatic access of semantic information by phonological codes in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 285.

Lindsay, S., & Gaskell, M.G. (2013). Lexical integration of novel words without sleep. *J Exp Psychol Learn Mem Cogn, 39*(2), 608-622. doi: 10.1037/a0029243

Liuzzi, G., Freundlieb, N., Ridder, V., Hoppe, J., Heise, K., Zimerman, M., . . . Hummel, F.C. (2010). The involvement of the left motor cortex in learning of a novel action word lexicon. *Curr Biol, 20*(19), 1745-1751. doi: 10.1016/j.cub.2010.08.034 S0960-9822(10)01037-7 [pii]

MacGregor, L.J., Pulvermuller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nat Commun, 3*, 711. doi: 10.1038/ncomms1715

Maciejewski, G., Rodd, J.M., Mon-Williams, M., & Klepousniotou, E. (2019). The cost of learning new meanings for familiar words. *Language, Cognition and Neuroscience*, 1-23. doi: 10.1080/23273798.2019.1642500

Maratos, E.J., Allan, K., & Rugg, M.D. (2000). Recognition memory for emotionally negative and neutral words: An erp study. *Neuropsychologia, 38*(11), 1452-1465. doi: Doi 10.1016/S0028-3932(00)00061-0

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg- and meg-data. *J Neurosci Methods, 164*(1), 177-190. doi: 10.1016/j.jneumeth.2007.03.024

Marslen-Wilson, W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1-2), 71-102.

McClelland, J.L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology-General, 142*(4), 1190-1210. doi: 10.1037/a0033812

Mcclelland, J.L., Mcnaughton, B.L., & Oreilly, 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*(3), 419-457. doi: Doi 10.1037/0033-295x.102.3.419

Medler, D., Arnoldussen, A., Binder, J., & Seidenberg, M. (2005). The wisconsin perceptual attribute ratings database: Retrieved from www. neuro. mcw. edu/ ratings.

Moseley, R.L., & Pulvermuller, F. (2014). Nouns, verbs, objects, actions, and abstractions: Local fmri activity indexes semantics, not lexical categories. *Brain Lang, 132*, 28-42. doi: 10.1016/j.bandl.2014.03.001

Moseley, R.L., Pulvermuller, F., & Shtyrov, Y. (2013). Sensorimotor semantics on the spot: Brain activity dissociates between conceptual categories within 150 ms. *Sci Rep, 3*, 1928. doi: 10.1038/srep01928

Papeo, L., Lingnau, A., Agosta, S., Pascual-Leone, A., Battelli, L., & Caramazza, A. (2015). The origin of word-related motor activity. *Cereb Cortex, 25*(6), 1668-1675. doi: 10.1093/cercor/bht423

Parks, R., Ray, J., & Bland, S. (1998). Wordsmyth english dictionary--thesaurus. University of chicago.

Perfetti, C.A., Bell, L.C., & Delaney, S.M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27*(1), 59-70.

Poldrack, R.A., Wagner, A.D., Prull, M.W., Desmond, J.E., Glover, G.H., & Gabrieli, J.D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage, 10*(1), 15-35. doi: 10.1006/nimg.1999.0441

Pulvermuller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458-470.

Pulvermüller, F. (2001). Brain reflections of words and their meaning. *Trends in Cognitive Sciences, 5*(12), 517-524. doi: 10.1016/s1364-6613(00)01803-9

Pulvermuller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*(5), 351-360. doi: 10.1038/nrn2811

Pulvermuller, F., Lutzenberger, W., & Preissl, H. (1999). Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cereb Cortex, 9*(5), 497-506. doi: 10.1093/cercor/9.5.497

Raposo, A., Moss, H.E., Stamatakis, E.A., & Tyler, L.K. (2009). Modulation of motor and premotor cortices by actions, action words and action sentences. *Neuropsychologia, 47*(2), 388-396. doi: 10.1016/j.neuropsychologia.2008.09.017

Revill, K.P., Aslin, R.N., Tanenhaus, M.K., & Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proc Natl Acad Sci U S A, 105*(35), 13111-13115. doi: 10.1073/pnas.0807054105

Rodd, J.M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M.G., & Davis, M.H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Mem Cognit, 40*(7), 1095-1108. doi: 10.3758/s13421-012-0209-1

Rodd, J.M., Lopez Cutrin, B., Kirsch, H., Millar, A., & Davis, M.H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*(2), 180-198. doi: 10.1016/j.jml.2012.08.002

Roskies, A.L., Fiez, J.A., Balota, D.A., Raichle, M.E., & Petersen, S.E. (2001). Task-dependent modulation of regions in the left inferior frontal cortex during semantic processing. *J Cogn Neurosci, 13*(6), 829-843. doi: 10.1162/08989290152541485

Rugg, M.D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn Sci, 11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Saygin, A.P., McCullough, S., Alac, M., & Emmorey, K. (2010). Modulation of bold response in motion-sensitive lateral temporal cortex by real and fictive motion sentences. *J Cogn Neurosci, 22*(11), 2480-2490. doi: 10.1162/jocn.2009.21388

Schlichting, M.L., & Frankland, P.W. (2017). Memory allocation and integration in rodents and humans. *Current Opinion in Behavioral Sciences, 17*, 90-98. doi: 10.1016/j.cobeha.2017.07.013

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

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

Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., Scifo, P., . . . Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *J Cogn Neurosci, 17*(2), 273-281. doi: 10.1162/0898929053124965

Tootell, R.B., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Malach, R., . . . Rosen, B.R. (1995). Visual motion aftereffect in human cortical area mt revealed by functional magnetic resonance imaging. *Nature, 375*(6527), 139-141. doi: 10.1038/375139a0

Tsivilis, D., Allan, K., Roberts, J., Williams, N., Downes, J.J., & El-Deredy, W. (2015). Old-new erp effects and remote memories: The late parietal effect is absent as recollection fails whereas the early mid-frontal effect persists as familiarity is retained. *Frontiers in Human Neuroscience, 9*. doi: 10.3389/fnhum.2015.00532

van Kesteren, M.T., Ruiter, D.J., Fernandez, G., & Henson, R.N. (2012). How schema and novelty augment memory formation. *Trends Neurosci, 35*(4), 211-219. doi: 10.1016/j.tins.2012.02.001

Vanhoutte, S., Strobbe, G., van Mierlo, P., Cosyns, M., Batens, K., Corthals, P., . . . Santens, P. (2015). Early lexico-semantic modulation of motor related areas during action and non-action verb processing. *Journal of Neurolinguistics, 34*, 65-82. doi: 10.1016/j.jneuroling.2015.01.001

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H., & Cappa, S.F. (2011). Nouns and verbs in the brain: A review of behavioural, electrophysiological, neuropsychological and imaging studies. *Neurosci Biobehav Rev, 35*(3), 407-426. doi: 10.1016/j.neubiorev.2010.04.007

Vilberg, K.L., Moosavi, R.F., & Rugg, M.D. (2006). The relationship between electrophysiological correlates of recollection and amount of information retrieved. *Brain Research, 1122*, 161-170. doi: 10.1016/j.brainres.2006.09.023

Wang, H.C., Savage, G., Gaskell, M.G., Paulin, T., Robidoux, S., & Castles, A. (2016). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychon Bull Rev*. doi: 10.3758/s13423-016-1182-7

Warriner, A.B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 english lemmas. *Behav Res Methods, 45*(4), 1191-1207. doi: 10.3758/s13428-012-0314-x

Watson, J.D., Myers, R., Frackowiak, R.S., Hajnal, J.V., Woods, R.P., Mazziotta, J.C., . . . Zeki, S. (1993). Area v5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb Cortex, 3*(2), 79-94. doi: 10.1093/cercor/3.2.79