Bilingualism: Language and Cognition

cambridge.org/bil

Consolidation as a mechanism for word learning in sequential bilinguals

Giacomo Tartaro[1], Atsuko Takashima[1] and James M. McQueen[1,2]

[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands and [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Research Article

Cite this article: Tartaro G, Takashima A, McQueen JM (2021). Consolidation as a mechanism for word learning in sequential bilinguals. *Bilingualism: Language and Cognition* **24**, 864–878. https://doi.org/10.1017/S1366728921000286

Received: 2 March 2020
Revised: 28 April 2021
Accepted: 15 May 2021
First published online: 30 June 2021

Keywords:
word learning; memory consolidation; lexical integration; Italian–English sequential bilinguals

Address for correspondence:
James M. McQueen
Donders Centre for Cognition,
Donders Institute for Brain, Cognition and Behaviour,
Radboud University,
Thomas van Aquinostraat 4,
6525 GD Nijmegen
The Netherlands
Email: james.mcqueen@donders.ru.nl

Abstract

First-language research suggests that new words, after initial episodic-memory encoding, are consolidated and hence become lexically integrated. We asked here if lexical consolidation, about word forms and meanings, occurs in a second language. Italian–English sequential bilinguals learned novel English-like words (e.g., *apricon*, taught to mean "stapler"). fMRI analyses failed to reveal a predicted shift, after consolidation time, from hippocampal to temporal neocortical activity. In a pause-detection task, responses to existing phonological competitors of learned words (e.g., *apricot* for *apricon*) were slowed down if the words had been learned two days earlier (i.e., after consolidation time) but not if they had been learned the same day. In a lexical-decision task, new words primed responses to semantically-related existing words (e.g., *apricon-paper*) whether the words were learned that day or two days earlier. Consolidation appears to support integration of words into the bilingual lexicon, possibly more rapidly for meanings than for forms.

Introduction

Learning vocabulary is an essential part of language acquisition. Vocabulary is especially important when sequential bilinguals learn a second language (L2), since knowing more words is the best way to improve communication. This is why there is so much focus on vocabulary in the L2 classroom. But what are the cognitive mechanisms that support word learning? While this ability rests on many sub-processes, a key one is that memories for new words need to be consolidated. In essence, a new word starts off as an episodic experience (e.g., a foreign-language teacher providing for the first time the L2 label for an existing concept) but a memory of this experience is not enough for the learner to be able to use the new word efficiently and appropriately. Through consolidation, however, the new word becomes integrated into the mental lexicon such that the learner can use it as they communicate in the L2. Little is known about memory consolidation in L2 learners (but see Qiao & Forster, 2017; Nakayama & Lupker, 2018). In this study, we examine if and how consolidation processes support lexical integration, at form and meaning levels, as sequential bilinguals learn new words.

It seems plausible, a priori, that word-learning mechanisms are likely to be the same in the L2 as in the first language (L1). It is the most parsimonious account; language learners have a set of cognitive mechanisms that they can use to learn words, and they use them irrespective of the language the words are spoken in. The lack of evidence for age of acquisition effects for vocabulary is consistent with this hypothesis. While there is convincing evidence that there are sensitive periods for the acquisition of phonology (Flege, Yeni-Komshian & Liu, 1999; Granena & Long, 2012) and grammar (Granena & Long, 2012; Hartshorne, Tanenbaum & Pinker, 2018), this does not appear to be the case for vocabulary (Hartshorne & Germine, 2015; Snow & Hoefnagel-Höhle, 1978). Furthermore, L1 users need to be able to continue learning new L1 words throughout their lives, and indeed do so without any apparent decrease in that ability (Hartshorne & Germine, 2015). It thus seems likely that the same mechanisms for word learning are available across the lifespan, and hence also that they can be applied when learning words in a late-acquired L2. Neuropsychological and neuroimaging evidence further supports this view. While brain damage can differentially affect L1 and L2 grammar, this appears not to be the case for L1 versus L2 vocabulary (Ullman, 2001; Ullman & Lovelett, 2018). Similarly, the brain networks active in L1 word processing tasks appear to overlap with those used in L2 word processing (Ullman, 2001).

There are, however, critical differences between L1 and L2 vocabulary acquisition. The late L2 learner may be confronted with problems with phonology (e.g., difficulties hearing non-native speech segments correctly may interfere with learning the new word's form; Pajak, Creel & Levy, 2016) and/or syntax (e.g., difficulties with non-native grammar may interfere with learning a new word's grammatical role). Furthermore, L2 vocabulary builds on L1 vocabulary, such that (at least in most cases in late L2 acquisition) new L2 words refer to

concepts that are already known. Translation is thus an important means to support L2 vocabulary learning and processing (Kroll & Stewart, 1994; Dijkstra et al., 2019) but is something that is obviously not available in L1 acquisition. It would therefore be a mistake to assume that L2 word learning is identical to L1 word learning. For exactly this reason it is essential to seek to specify which mechanisms are or are not shared across L1 and L2 learning contexts. We therefore asked in this study whether memory consolidation processes are shared between L1 and L2 vocabulary acquisition.

Consolidation is a central feature of language learning and of other domains of cognition. Across cognitive domains, after new experiences have been encoded in episodic memory, they are consolidated through transfer into semantic memory (McGaugh, 2000; Alvarez & Squire, 1994; Winocur & Moscovitch, 2011). In the Complementary Learning Systems (CLS) account (McClelland, McNaughton & O'Reilly, 1995), integration of new knowledge into semantic memory requires slow consolidation because the new knowledge must be interleaved with old knowledge without catastrophic interference (i.e., overwriting of old knowledge by new knowledge). Fast learning in the episodic memory system complements this slow consolidation process by providing initial storage of the new knowledge. Episodic memory is supported by activity in medial temporal structures centered on the hippocampus, whereas semantic memory is supported by activity in neocortical structures, and consolidation entails a shift in the relative engagement of these two sets of structures (Takashima, Petersson, Rutters, Tendolkar, Jensen, Zwarts, McNaughton & Fernández, 2006; Frankland & Bontempi, 2005).

New words need to be consolidated and integrated into the semantic network to function as stable lexical representations: after initial episodic experience with them, they must be integrated into the lexicon such that they can be used in communication (e.g., speaking and listening to speech). But this again needs to be done so that the integration of new words does not disrupt existing lexical knowledge (i.e., without catastrophic interference; McClelland et al., 1995). Beginning with the work of Gaskell and Dumay (2003), there is now a substantial body of evidence, consistent with the CLS account, for consolidation in L1 word learning (for reviews, see Davis & Gaskell, 2009; James, Gaskell, Weighall & Henderson, 2017). Behavioral studies have shown that, after a period of consolidation, new words start to function like existing words and thus appear to be integrated into the lexicon at both the form and meaning levels. At the form level, new words start to engage in competition with other words, as shown by the emergence, after consolidation, of inhibition of responses to similar-sounding existing words (e.g., decisions to *cathedral* become slower after the new word *cathedruke* has been consolidated and hence starts to compete with *cathedral*; Gaskell & Dumay, 2003). This effect has been shown in multiple studies using the pause-detection task, where participants have to decide whether they heard an artificially inserted pause in the base words (e.g., in *cathedr…al)*. The competition appears to involve abstract (i.e., non-episodic) representations of the new words (Bakker, Takashima, van Hell, Janzen & McQueen, 2014) and does not require the new words to be associated with meanings (Gaskell & Dumay, 2003). When novel-word meanings are taught to participants, the emergence of semantic priming effects after consolidation suggests lexical integration of the meanings of the new words (e.g., facilitation of responses in a primed lexical decision task to the new word *pamat*, which was taught to be a kind of cat, when primed by *dog*; Bakker, Takashima, van Hell, Janzen & McQueen, 2015a; Bakker, Takashima, van Hell, Janzen & McQueen, 2015b; see also van der Ven, Takashima, Segers & Verhoeven, 2017, where the new words served as primes rather than targets).

Lexical consolidation is enhanced by sleep. Although some studies have shown consolidation effects without any sleep in the time interval between learning and testing (Kapnoula, Gupta, Packard & McMurray, 2015; Lindsay & Gaskell, 2013; Szmalec, Page & Duyck, 2012), others have shown that sleep in that interval enhances consolidation (Dumay & Gaskell, 2007; Gaskell & Dumay, 2003) and that electrophysiological activity during sleep is associated with behavioral indicators of consolidation (Tamminen, Payne, Stickgold, Wamsley & Gaskell, 2010). Thus, while sleep appears not to be necessary for consolidation to occur, it supports what is a gradual and ongoing process.

Further evidence for consolidation in L1 word learning comes from neuroscientific studies. In such studies, comparisons are usually made between the neural activity measured in response to words learned either on the day of test (less-consolidated novel words) or on a previous day (i.e., after at least one night of sleep; more-consolidated novel words). Results from fMRI (Davis, Di Betta, Macdonald & Gaskell, 2009; Takashima, Bakker, van Hell, Janzen & McQueen, 2014), EEG (Bakker et al., 2015a, 2015b) and MEG (Bakker-Marshall et al., 2018) suggest, in keeping with the CLS model, a shift from greater engagement of the hippocampus and medial temporal structures for less-consolidated words to greater engagement of neocortical structures for more-consolidated words. A key neocortical structure showing this greater engagement after consolidation is the left posterior Medial Temporal Gyrus (pMTG); the pMTG appears to be a lexical hub in the perisylvian language network, one that plays a key role in binding together semantic, phonological, and orthographic knowledge about words (Bakker-Marshall et al., 2018).

There is also evidence in children of lexical integration after sleep from behavioral studies (Henderson, Weighall, Brown & Gaskell, 2012, 2013; van der Ven et al., 2017); evidence that electrophysiological activity as children sleep is associated with memory consolidation of words (Smith et al., 2018); and fMRI evidence from children of decreasing hippocampal activity for novel words as the interval increases between learning and testing (Takashima, Bakker-Marshall, van Hell, McQueen & Janzen, 2019). Although there seem to be differences in the details of the consolidation process with respect to which types of new words are more strongly consolidated (James, Gaskell & Henderson, 2019), the general picture emerging from this developmental work is that consolidation processes in children appear to be equivalent to those in adults (James et al., 2017). These developmental data thus support the idea that the same consolidation mechanisms support word learning throughout the lifespan and hence the hypothesis that these mechanisms will also be used by sequential bilinguals as they learn L2 words.

We test this hypothesis here. We thus test for evidence that would challenge the recent claim that there are differences in lexical consolidation between L1 word learning in English native speakers (Qiao & Forster, 2013) and L2 word learning in Chinese–English sequential bilinguals (Qiao & Forster, 2017). Those studies used a masked priming technique in visual lexical decision. A Prime Lexicality Effect (PLE) on the form level was observed in the native English participants' data (i.e., after training, no significant priming from a novel word prime, e.g., *baltery*,

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

on responses to target words differing in only one letter, e.g., *battery*). The novel word primes thus behaved like real words and not like nonwords (which do produce facilitatory priming). In the bilingual participants' data, however, there was no PLE even after multiple days of training on the novel words, and instead there was significant facilitatory priming (e.g., *baltery* primed responses to *battery*). The PLE is usually interpreted as an indicator of lexical competition, so Qiao and Forster (2017) took the absence of the PLE in L2 participants as evidence that the novel words had not been integrated into the lexicon in the same way as L1 words are integrated, and hence that the L2 lexicon may be different in kind from the L1 lexicon. Another study, with L1 Japanese participants tested on L2 English, showed a similar facilitation effect on known English words (Nakayama & Lupker, 2018). The difference between L1 and L2 participants found in Qiao and Forster (2017) may thus be driven by the difference in orthography between L1 (logographic Chinese) and L2 (alphabetic English). However, because the Qiao and Forster (2017) findings conflict with our earlier arguments that word learning (and more specifically lexical consolidation) is likely to be the same across ages and languages, it is important to test for lexical consolidation in L2 speakers using other methods, specifically in a situation where there are no differences in orthography between L1 and L2. We took that approach here.

Our design was modelled on the seminal L1 study on lexical consolidation (Gaskell & Dumay, 2003). We taught Italian–English sequential bilinguals novel English-like words (e.g., *apricon*, taught to mean "stapler") and then tested them on these novel words in a free recall task. A variety of tasks were used during training to try to ensure that the new words (presented in both written and spoken form) were associated with the concepts which were represented as pictures of objects (e.g., for *apricon*, a picture of a stapler). Two days later, the bilinguals were taught a matched second set of words in the same way. They were again tested on those words in free recall, and then they underwent a final test phase with three tasks. We used multiple test tasks, which examined different aspects of consolidation, and fMRI, in order to obtain a more complete picture of lexical integration in L2 word learning.

The first task tested recognition-memory in the fMRI scanner. Participants decided which of four words was the correct referent of a picture shown in the middle of the display. This task was designed to test non-episodic memory of the words through the use of pictures that had not been presented during training (e.g., for *apricon*, a different picture of a stapler). This task, which requires retrieval of the association of the new word with the abstract concept it refers to, enabled us to focus on neural activity in the networks responsible for linking word forms to word meanings, and the pMTG in particular (Bakker-Marshall et al., 2018; Takashima et al., 2014). The second test task was pause detection, which was used to measure integration of the novel words into the lexicon at the form level. Participants were asked to decide whether silent pauses had been inserted into base words – that is, real English words from which the novel words had been derived (e.g., for *apricon*, whether there was a pause in *apricot*). Following, for example, Gaskell and Dumay (2003), decisions to the base words should be slowed down (relative to matched base words for novel words that had not been learned) because of competition between the base word and the novel word. Such competition would indicate lexical integration of the novel words at the form level. The third test task was primed lexical decision. The novel words served as semantically related primes for existing target words (e.g., *apricon-paper*). Speeding of responses to targets (relative to a semantically unrelated condition) would again indicate lexical integration of the novel word, but now at the meaning level (Bakker et al., 2015a, 2015b; van der Ven et al., 2017). The constraints imposed on the choice of materials that could be used in all these tasks (e.g., that the base words were known to the participants) made it impossible to use actual English words for training. In keeping with many prior studies (e.g., Gaskell & Dumay, 2003), we therefore used pseudowords for training – in this case, those with English phonological and orthographic features.

We also collected data on the participants' English proficiency, including their scores on standardized tests, their length of stay outside Italy, and a measure of their English vocabulary size. These secondary, individual-differences measures made it possible to ask, through correlational analyses, whether there was any association between L2 proficiency and degree of consolidation as estimated by the primary measures.

For all three primary measures, the main comparison was between the words learned on the day of test ("Recent" words), which we expected to be less consolidated, and the words learned two days earlier ("Remote" words), which we expected to be more consolidated. In the fMRI data, we predicted greater activity in the hippocampus and related medial temporal structures for Recent than for Remote words, and greater activity in the pMTG for Remote than for Recent words. In the pause-detection task, we predicted stronger competition (i.e., more slowing of responses relative to those for base words of untrained words) for the base words of the Remote words than for those of the Recent words. In the primed lexical-decision task, we predicted more facilitation of responses to targets primed by Remote words than to those primed by Recent words. Across all three measures, we were thus able to ask whether memory consolidation is a mechanism that operates as sequential bilinguals learn new L2 words, and whether there are consolidation effects for word forms and/or meanings.

Method

Participants

Fifty-two right-handed native Italians (30 female; average age 25 years, range 18–41) living in Nijmegen, the Netherlands, were recruited. No participants had a history of neurological or language-related disorders and all reported having normal or corrected-to-normal vision and hearing. One participant was excluded from the analysis because she was not able to conduct the experiment in the MRI scanner due to an attack of claustrophobia. Three more participants were excluded because their Reaction Times (RTs) were very slow (greater than the threshold of 2.5 standard deviations (SDs) above the group mean) in one or more of the primary behavioral tasks. One other participant's data were discarded from the MRI analyses due to technical problems with the imaging data. The analysis was therefore based on 48 participants for the behavioral data and 47 participants for the imaging data. All participants gave informed consent; the study was conducted according to a protocol approved by the regional ethical review board (CMO Region Arnhem-Nijmegen, The Netherlands).

Design

The study is summarized in Figure 1. Prior to the experiment, participants completed a pre-test (explained in the procedure

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

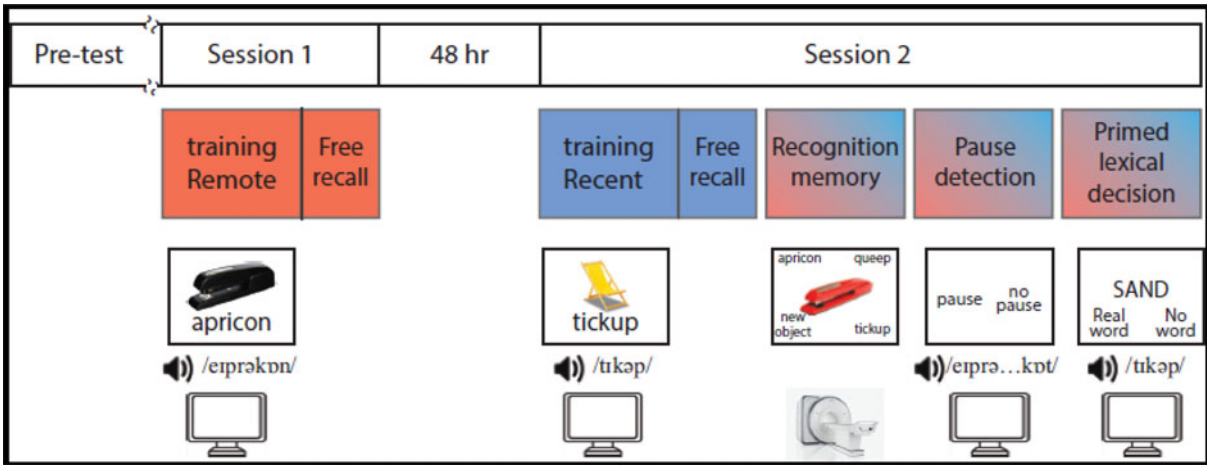

Figure 1. Overall study design.

section). On the first day (session 1), participants were trained on one set of novel words (Remote words) and a free-recall memory test. Two days later, hence with two nights allowing for sleep-related memory consolidation of the words learned on the first day, participants took part in the second training session with another set of novel words (Recent words) and a second free-recall task. After the training session on the second day, they took part in the recognition-memory task in the MRI scanner, followed by the pause-detection task, and finally the primed lexical-decision task. The latter two tasks were done in a behavioral laboratory (i.e., not in the scanner).

Materials

Words

One hundred and twenty English-like pseudowords were created and divided into three sets of 40 (hereafter called novel words). The novel words in each set were matched for orthographic length (range 4–11 letters), phonological length (number of phonemes, range 3–10, and number of syllables, range 2-5).

The novel words were derived from existing English words (hereafter called base words) using Wuggy (a multilingual pseudoword generator; Keuleers & Brysbaert, 2010). For every base word, one to three phonemes of the last syllable were changed, in order to create a novel word that did not exist in English or Italian (e.g., *apricon* derived from *apricot*). In some cases of monosyllabic existing words, two or three phonemes were added at the end of the word (e.g., *cliffon* from *cliff*). The base words were not English-Italian cognates (e.g., we avoided *cathedral* - *cattedrale*). The selected base words had a frequency of occurrence in the range of 1.15 to 4.35 log10 per million according to the data from SubtlexUS (Brysbaert & New, 2009) and CLEARPOND (Marian, Bartolotti, Chabal & Shook, 2012). Mean frequency of occurrence of the base words was balanced across the three lists. In addition, the base words had small phonological neighborhoods (80% with less than three neighbors, min. 0, max. 6) and small orthographic neighborhoods (86% with less than three neighbors, min. 0, max. 7). In this way, we enhanced the competition effect in the pause-detection task, which was the only task where the participants heard the base words (with or without pauses inserted). Forty high-frequency English words were selected for use as fillers in the pause-detection task.

The novel words were used in the training sessions, the explicit memory tasks (the free-recall and the recognition-memory task) and as primes in the primed lexical-decision task. For each novel word, two high-frequency existing English words were selected for use as targets in the primed lexical-decision task. One word in each pair was semantically related to the concept expressed by the novel word (as taught during the training sessions); the other was not. A further 240 pseudowords derived from these 240 target words were created with Wuggy, also for use as targets in the primed lexical-decision task. The critical stimuli are listed in the Supplementary Material (Table S1, Supplementary Materials).

Pictures

One hundred and twenty objects whose English names were unfamiliar to Italians were selected and two photos depicting the objects were selected from the internet. All pictures represented meaningful objects and were selected from different semantic categories (tools, animals, furniture, etc.). The pictures were divided into three sets with the categories equally distributed and paired with the three sets of 40 novel words. Any letters or words present in the pictures were blurred. The selected images were checked by 12 people who did not take part in the main experiment. They were asked if the object looked meaningful (i.e., if the object was recognizable), and whether each pair of pictures could be considered to be the same object. Picture pairs were chosen only if they met these criteria. The two pictures could differ in the angle of the photo, the color of the object, or other features (see Figure 2). The more distinct member of each pair was placed in Picture List 1 (used during training); the other one was placed in Picture List 2 (used during the recognition-memory test).

Speech recordings and speech editing

All 120 base words, the 120 novel words, and the 40 filler words for the pause-detection task were each recorded three times by a young female native speaker of American English. In each case, the best of the three recordings was selected. A silent pause of 200 ms was then inserted immediately before the point of divergence between the base word and the novel word (e.g., before the [t] in *apricot*) using the audio editor Audacity. Fade-in and

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

Figure 2. Example of a pair of pictures (for the novel word *tickup*) used in the training session (left panel) and the recognition-memory test (right panel).

fade-out manipulations were applied in order to prevent any clipping sounds generated by the insertion of the pause.

Procedure

Pre-test

Participants first underwent a preliminary screening in order to check their eligibility. They had to be right-handed Italian native speakers, without any neurological or language disorders, to have no metal parts in their body (for safety during the scanner session) and no signs of claustrophobia. If they were eligible, they were scheduled for the two days of training and testing.

Furthermore, participants filled out a questionnaire. In the first part, knowledge of all English base words used in the main experiment was tested (e.g., "Do you know the meaning of apricot?"). The total number of known words was used as a proxy measurement of the participant's vocabulary size and performance on this part of the questionnaire was also used to filter the pause-detection data (and, for consistency, the primed lexical-decision data), such that the data was limited, per participant, to only those base words that they knew.

The second part of the questionnaire included information regarding the participant's proficiency in English. We asked for their IELTS score when available, and if not, scores on other tests, such as TOEFL, TOEIC or Cambridge CELA. The latter scores were converted to the equivalent IELTS score. We choose the IELTS as measurement of proficiency since it is widely considered to provide a reliable estimation (Charge & Taylor, 1997). We also asked the length of stay of the participant in English-speaking countries. Given the generally high level of English proficiency in the sample, the international English-speaking environment they all lived in, and the fact that none of the participants spoke Dutch, the Netherlands was considered to be an English-speaking country with respect to this length-of-stay measure.

Training sessions

The participants were assigned to one of six groups. Each group was trained on one of the three sets of novel words on the first day (i.e., for the Remote training session tasks) and on another set two days later (i.e., for the Recent training session tasks). The remaining set was used as the New condition in the recognition-memory task (see below). Assignment of novel word sets to session and tasks was counterbalanced over groups.

At the beginning of the first training session, instructions about the purpose of the experiment and the procedure were given. All instructions (i.e., for all training and test sessions) were in English. After initial instruction, participants underwent ten different training tasks. They were trained on the spelling of the words, the sound of the words and their pronunciation (i.e., reading, listening and speaking). For each trial, they were always informed if the given answer was correct or not and told what the correct answer was if they were incorrect, and, at the end of each task, they were shown the total number of correct answers. All tasks were presented using Presentation (www.neurobs.com) and used a response button box and/or a computer keyboard. See the Supplementary Material for details on the ten tasks (Table S2 and accompanying text, Supplementary Materials).

Testing sessions

After the training session on each day, participants performed a FREE-RECALL memory test. They had five minutes to type all the trained novel words they could remember. After the free-recall test on the second day of testing, the three main tests were conducted.

First, the RECOGNITION-MEMORY TASK was performed in the MRI scanner. Participants saw a picture on a computer screen together with four printed response options. Their task was to decide which printed word was depicted by the picture. As noted above, these pictures had not been presented during training. At the beginning of the trial, a black fixation cross appeared for a jittered inter-trial interval (ITI) of 1-7 seconds. Then, the fixation cross turned blue for one second, prompting the participant that the next picture would be presented. Subsequently a picture appeared in the middle of the screen with four different word options in lower-case in the four corners of the screen. The picture was an image of an object from one of the following three conditions: the object had been studied the same day (Recent condition) or two days before (Remote condition), or it was novel

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

(New condition). One of the four response options was the correct novel word, one was another novel word from the same training set (e.g., a word from the Remote set if the correct word was a word from the Remote set), one was a novel word taken from the other training set (e.g., a word from the Recent set if the correct word was a word from the Remote set), and the fourth option was the phrase "new object". Participants were instructed to choose the option "new object" if they judged that the object was not in either of the studied lists. The positions of the four options were randomized on every trial. Participants were instructed to respond as soon as they knew the answer. If the participant did not respond within 2500 ms after the picture onset, the picture disappeared and the word "Respond" appeared in the center of the screen, to further prompt participants for a response. After the response, a black bar appeared under the chosen option for 500 ms, but no corrective feedback was given. This procedure was repeated for all 120 pictures, one at a time in random order.

Second, the PAUSE-DETECTION TASK was administered. In this task, the participant had to identify as quickly as possible if a pause was present or not in the spoken words presented through a loudspeaker. Of the 160 trials, there were 40 trials with the base words from which the Remote novel words had been created, 40 with the base words from which the Recent novel words had been created, 40 with the base words from which the untrained set of novel words had been created, and 40 with filler words. Overall, half of the trials had a pause inserted in the base word and half did not. The order of the trials (Remote/Recent/Untrained/Filler, Pause+/Pause-) was randomized for every participant.

At the beginning of the trial, a fixation cross appeared in the middle of the screen for 1000 ms, then it disappeared and the spoken word was played at the same time as two response options appeared on the screen: "Pause" on the left, "No pause" on the right. The participants were instructed to press the corresponding button (i.e., the left one when there was a pause and the right one when there was not) as fast and as accurately as possible. All responses given before the onset of the pause (or in the equivalent location in stimuli without a pause) were treated as errors. When a button response was made, a black bar appeared under the chosen option for 500 ms, but no corrective feedback was given. If the participant did not reply within 3 seconds after the offset of the sound stimulus "Too late" appeared on the screen and the trial was coded as a missing response.

Finally, participants took part in the PRIMED LEXICAL-DECISION TASK. In this task, they heard a novel word from the trained sessions, and then saw a word on the computer screen. Their task was to decide whether the word on the screen was an existing English word or not. This task was composed of 320 trials, half of them with the Remote novel words as the prime and the other half with Recent novel words as primes. Each prime word was played four times, once with a related word target, once with an unrelated word target, and twice with two different pseudoword targets. Order of primes was pseudo-randomized such that every sequence of four trials contained one related-word, one unrelated-word and 2 pseudoword targets. After trials 80, 160 and 240 there was a short break. At the beginning of each trial, a fixation cross appeared at the center of the screen for 1 second, then it disappeared and the prime novel word was presented auditorily. Participants were instructed to think of the meaning of the prime word. At acoustic offset of the prime, the target word appeared at the center of the screen together with the response options, on the left of the screen "Real word" and on the right "No word". The participants' task was to press the left button when the word on the screen was an existing word, and the right one when the word was a pseudo-word, as fast and as accurately as possible. When the answer was given, a black bar appeared under the chosen option for 500 ms, but no corrective feedback was given. If the subject did not reply within 4 seconds of target word onset, the trial was categorized as a miss.

At the end of the experiment, the participants were informed about the purpose of the experiment and told that the novel words were pseudowords rather than actual English words.

fMRI scanning and analysis procedures

fMRI data were recorded in a 3 T Prisma scanner (Siemens Healthcare, Erlangen, Germany) using a 32-channel head coil. For functional images, we used a multiband sequence with the following parameters: acceleration factor of six, repetition time (TR): 1000 ms, echo time: TE 34 ms, 66 slices, ascending slice order, 2.0 mm slice thickness, inplane field of view (FOV): 210 x 210 mm, flip angle: 60°. Slices were angulated in an oblique axial manner to reach whole-brain coverage. In addition, an inverted EPI with the same parameters was collected. Also, for field-map images, we used a multiband sequence with the following parameters: TR: 620.0 ms, echo time: TE 1 of 4.70 ms and TE 2 of 7.16 ms, 66 slices, ascending slice order, 2.0 mm slice thickness, inplane FOV: 210 x 210 mm, flip angle: 60°. T1-weighted anatomical scans at 1 mm isotropic resolution were acquired with TR 2300 ms, TE 3.03 ms, flip angle 8°, 1.0 mm slice thickness, and inplane FOV 256 × 256 mm.

Image pre-processing and statistical analysis was performed using SPM12 (www.fil.ion.ucl.ac.uk). The first ten volumes of each participant's functional scan were discarded to allow for T1 equilibration. Field-map images, T1 structural images and functional images were then converted from DICOM files to nifti files to allow the use of standard fMRI preprocessing tools in SPM12. Successively, the field-map deformation was calculated and the functional images were realigned and un-warped. Then the subject-mean image of the functional run was co-registered to the corresponding structural MRI and applied to all functional scans. Consecutively, T1 structural images were segmented and the functional images were normalized using the normalization parameter estimated in this segmentation step. Lastly, the functional scans were smoothed at 4 mm full-width half maximum.

The fMRI data were analyzed statistically using a general linear model (GLM) and statistical parametric mapping (www.fil.ion.ucl.ac.uk). Four explanatory variables were included in the model: correct trials for remote condition, recent condition, new condition (untrained) and other trials. Trials in the "other" category included all trials that were responded to incorrectly and trials with overly long RTs (> 2 SD above the mean RT for each participant). We took this approach of discarding trials with long RTs from the correct response regressors, as these trials may not be well modelled when the onset time is set to the picture onset. For 4 participants whose RTs (due to a technical error) were not logged, no trials were labeled as overly long. These explanatory variables were temporally convolved with the canonical Hemodynamic Response Function (HRF) provided by SPM12. Each event was time-locked to the onset of the picture. The design matrix also included the six head-motion regressors (three translations, three rotations). A high pass filter was implemented using a cut-off period of 128 s to remove low-frequency effects from the time series.

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

Table 1. Explicit memory tasks: free recall and recognition memory (mean ± SD)

	Recognition Test					
	Number of Trials (max 40)			RT (ms)	Free Recall (max 40 words)	
Conditions	Correct	Incorrect	Correct but too long	Correct	Number of words typed	Number of correct responses
Remote	37.3 ± 2.7	2.2 ± 2.4	0.6 ± 0.7	2423 ± 623	20.3 ± 7.6	13.4 ± 6.4
Recent	36.9 ± 2.3	2.3 ± 1.9	0.8 ± 1.0	2320 ± 561	22.6 ± 6.6	15.1 ± 6.3
New	35.7 ± 2.5	2.0 ± 2.1	2.3 ± 1.8	2933 ± 696		

Note. Due to failure in logging, RTs for the recognition test for 4 participants and free-recall data in the Recent condition for 2 participants are missing.

Two different Regions-of-Interest (ROI) masks were created using MarsBaR (Brett, Anton, Valabregue & Poline, 2002). One of the masks covered the anatomical area of the hippocampus defined by the HIP AAL template (Tzourio-Mazoyer, Landeau, Papathanassiou, Crivello, Etard, Delcroix, Mazoyer & Joliot, 2002), the other the pMTG area. The latter was created by combining two different masks, an anatomical mask of the whole left MTG defined by the T2 AAL template (Tzourio-Mazoyer et al., 2002) and a sphere of 2 cm diameter with the center at [-58 -60 0] (MNI coordinates). These coordinates were taken from the peak of activation in the left pMTG as observed in Takashima et al. (2014).

Results

The data from the training sessions are given in the Supplementary Materials (Table S3, Supplementary Materials).

Explicit memory tests

Behavioural results from the free recall and fMRI recognition-memory tasks are presented in Table 1. Due to technical problems, RTs of 4 participants during the recognition-memory task were not logged. Thus the RT data are based on 43 participants.

Recognition-memory task

A repeated measures ANOVA with the factor Time (Remote, Recent, New) on the number of correct responses revealed that there was no difference in performance across these three levels ($p = .621$). RTs for correct responses were, however, significantly different from each other ($F(2,84) = 32.7$, $p < .001$, $\eta p^2 = .438$). The New condition was responded to more slowly than the Remote or Recent conditions (both $ps < .001$). There was a trend for the Recent condition to be faster than the Remote condition ($p = .089$).

Free recall test

There was a trend for more words to be typed overall, and to be typed correctly in the (Recent) second session than in the (Remote) first session (number of words typed $p = .095$, number of correctly typed words $p = .083$, see Table 1). This may be because participants anticipated the free recall test during the training phase in the second session based on their experience with this test in the first session.

We also asked if performance on the free recall task on both days was correlated with our measurements of individual differences in English proficiency by testing for the partial correlation coefficients. The descriptive data for the three individual differences measures are given in Table 2. Contrary to our hypothesis, length of stay and proficiency did not correlate with the number of correctly remembered words. Interestingly, we found different results for the two days with the respect to vocabulary size. On the second day, no significant correlation was found, but on the first day we found a correlation of vocabulary size with the number of correctly remembered words ($r = .372$, $p = .009$). Participants who knew more base words in the experimental word list were also the ones who were able to recall more novel words from the training list in session 1. Overall, these data contradict our hypothesis of an influence of individual differences in the memorization process except for vocabulary size, which correlated with the number of remembered words but only on the first day. Probably the advantages of a bigger vocabulary can be easily overwhelmed by better familiarization with the task (i.e., what and how well the words should be learned).

fMRI analyses

Whole brain analysis

We predicted that words learnt 2 days before (Remote condition) would show a consolidation effect compared to words learnt just before the test (Recent condition), and thus show increased activity in the left pMTG and decreased activity in the hippocampus. To test this hypothesis, we contrasted the Remote and Recent conditions using a one-sample t-test. The Remote > Recent comparison showed two clusters in the left inferior frontal gyrus (pars triangularis peak [-34 30 6] cluster size 152; pars orbitalis peak [-34 30 -14], cluster size 80). No significant clusters were found in the predicted left pMTG. The reverse contrast Recent > Remote showed significant clusters in the bilateral occipito/parietal areas (right middle occipital gyrus peak [44 -74 28] cluster size 208, left angular gyrus peak [-52 -62 30] cluster size 124, right precuneus peak [8 -60 26] cluster size 85, left cuneus peak [-16 -56 26] cluster size 61, and right supramarginal gyrus peak [52 -38 38] cluster size 57). See Figure 3 (panels A and B) and Table 3.

The Remote and Recent conditions both showed increased activity in the bilateral hippocampus compared to the New condition but the difference between Remote and Recent condition did not show a significant cluster in the hippocampus. See Figure 3 (panels C and D) and Table 3. Moreover, no increase in activity with consolidation was found within the left pMTG. Our hypothesis was only partially confirmed. Hippocampal activation was found for the Recent condition as expected, but the

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

Table 2. Individual difference measures (IELTS, length of stay, word knowledge)

	IELTS	duration stay (months)	word knowledge (max 396)
average	7.0	17.9	362.0
SD	1.2	18.9	25.8
min	4	0	286
max	9	101	396

Remote condition also showed a heightened activation pattern in this region. Furthermore, we did not observe the expected pattern in the left pMTG.

ROI analysis

We also looked specifically at activity levels within the bilateral hippocampi and the left pMTG and, within these ROIs, tested for differences between the Recent and Remote conditions (see Figure 3E). The beta values extracted within these ROI were investigated with two repeated measures ANOVAs, with factor Time (Remote/Recent/New). One ANOVA was used to investigate the values for the left pMTG and the other the values for the bilateral hippocampus. For the comparison of mean beta values of the left pMTG (Remote (mean = 0.022, SD = 1.540), Recent (mean = 0.104, SD = 1.383), and New (mean = 0.137, SD = 1.526) no significant difference was found (p = .366). However, for the hippocampus we observed a significant difference ($F(2,92) = 49.532$, $p < .001$, $\eta p^2 = .518$): Remote (mean = 0.195, SD = 0.747), Recent (mean = 0.233, SD = 0.687), and New (mean = -0.354, SD = 0.697). This difference was driven by the values of the hippocampus in the New condition which were significantly different from the values in the other two conditions (both $ps < .001$; see Figure 3E). The difference between the Remote and Recent conditions, however, was not statistically significant ($p = .890$). The ROI analysis thus confirmed the findings that we observed in the whole-brain analysis.

Influence of different levels of English across participants

We assumed that there may be inter-participant variability in the brain responses to the consolidation trajectory, with longer experience or better knowledge of English speeding up the consolidation process. To test this hypothesis, we included extra regressors to the one-sample t-test comparing the Remote and Recent conditions. Although highly correlated with each other, IELTS, duration of stay, and vocabulary size differed slightly in different ways, and thus we tested each parameter separately. For IELTS, one cluster in the right posterior cingulate cortex (peak [2 -50 24] cluster size 47) was significant. More activity for the Recent condition relative to Remote condition was found if the IELTS score was higher. The reverse correlation did not show any significant clusters. For duration of stay in an English speaking environment, 2 clusters were found to be significant, one in the left inferior parietal cortex (peak [-48 -44 36] cluster size 52) and another in the right posterior middle temporal gyrus (peak [52 -42 -2] cluster size 56). More activity was found for the Recent condition if the duration of stay was longer. For vocabulary size, no significant clusters were observed for either direction. In sum, the correlation analyses with English proficiency measures did not show any effect in either the hippocampus or the pMTG.

Pause-detection task

Responses were excluded from this analysis for three reasons. First, all responses faster than 100 ms were removed from analysis and from calculation of mean RTs and SDs. Second, all remaining responses more than 2.5 SDs from the subject's mean were excluded. Third, we excluded all trials whose words were responded to as unknown during pre-test questionnaire on knowledge of English words. The results for the remaining data in this task are shown in Table 4 and Figure 4A.

With a repeated-measure design ANOVA, with two factors Time (Remote/Recent/Untrained), and Pause (With Pause/Without Pause), we investigated if the base words in the remote condition showed slower RTs compared to base words in the other conditions. As expected, a main effect of Time was observed ($F(2,47) = 4.182$, $p = .018$, $\eta p^2 = .082$). Planned t-tests comparing the Remote condition with the other conditions showed that RTs in the Remote condition were slower than those in either the Recent condition ($t = 2.383$, $p = .021$) or the Untrained condition ($t = 2.503$, $p = .016$, corrected $\alpha = .025$). Thus, slower pause detection RTs were observed for the base words whose related novel words were trained 2 days before. There was a main effect of Pause, where participants responded faster when there was no pause inserted ($p < .001$) but the presence/absence of pause did not interact with Time ($p = .107$). In order to investigate if the three measures on English affinity (IELTS, duration stay, knowledge of English words) had any influence on the size of the effect (i.e., the difference of RTs between the Remote and Recent conditions), we correlated the size of the effect with the three covariates. None of the covariates significantly correlated with the consolidation effect (all $ps > .184$).

We observed the expected pattern for this task. That is, the interference effect was observed in the Remote condition, but not in the other two conditions. This suggests word-form based integration was observed after a period of consolidation.

Primed lexical-decision task

As with the pause-detection task, responses were excluded from this analysis for three reasons. First, all responses faster than 100 ms were removed from the analysis and hence from the calculation of mean RTs and SDs. Second, all remaining responses more than 2.5 SD from the subject's mean were excluded. Third, for consistency with the pause-detection analysis, we excluded all trials on a by-participant basis for all novel words whose base words were responded to as unknown by the participant during the English word pre-test. See Table 5 for a summary of the number of trials entered into the analysis and the RTs.

We performed a repeated-measure design ANOVA, with two factors: Time (Remote/Recent) and Priming (Related/Unrelated) (see also Figure 4B). The analysis revealed a main effect of Priming, where related targets were responded faster than the unrelated targets ($F(1,47) = 16.757$, $p < .001$, $\eta p^2 = .263$), but we did not observe an effect of Time or an interaction of Time by Priming ($p = .856$, $p = .253$, respectively), although numerically the priming effect (the Unrelated–Related difference) was larger in the Remote condition (36.5 ± 64.8 ms) than in the Recent condition (17.9 ± 78.8 ms). Given this numerical trend, we performed an additional analysis in which we did not exclude trials on items with base words that the participants did not know in the English word pre-test (note that because the base words were not used in the primed lexical-decision task, data from their

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

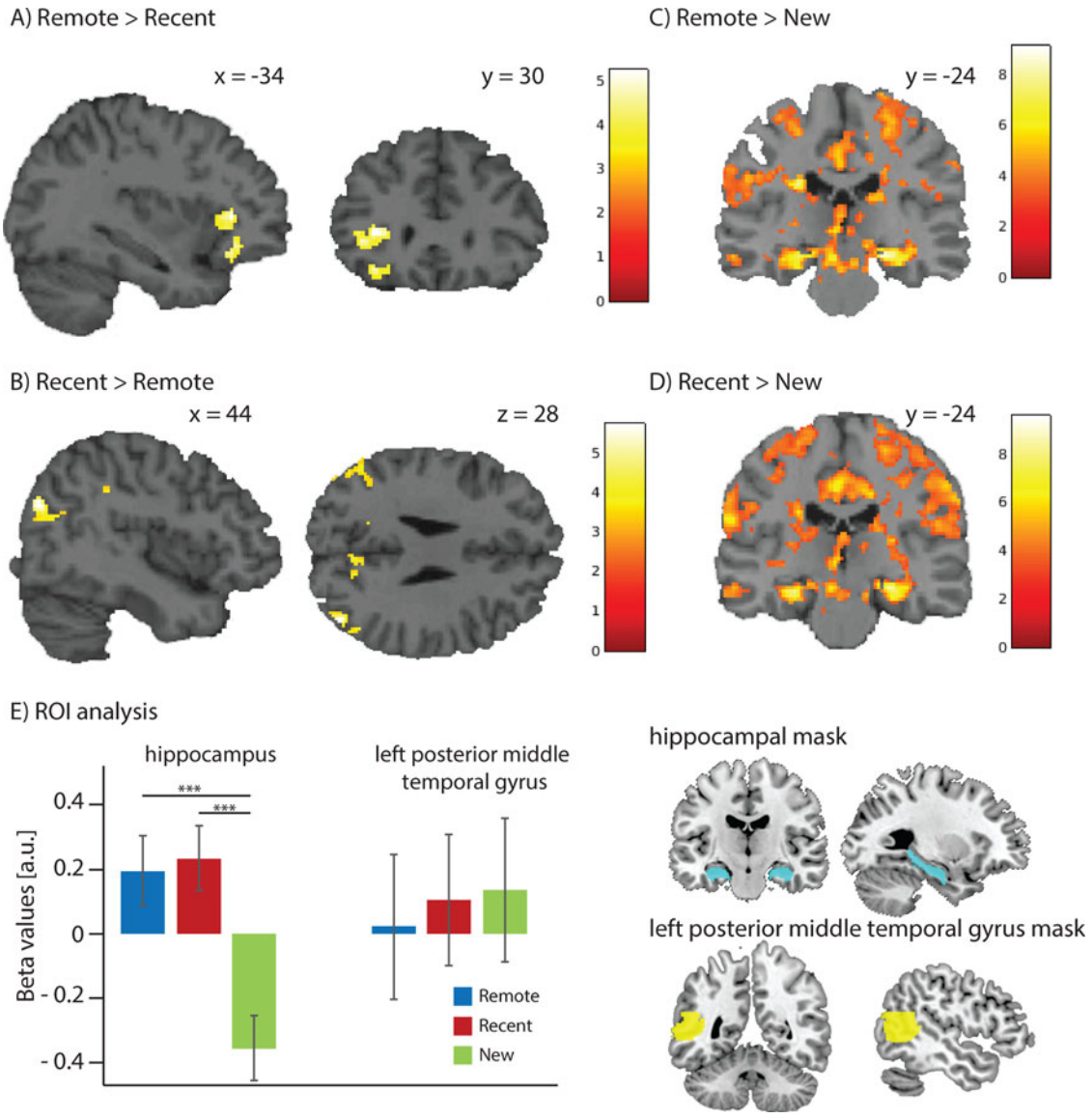

Figure 3. fMRI results. A) Significant clusters found for the Recent > Remote contrast superimposed on a template brain. B) Significant clusters found for the Remote > Recent contrast. C) Significant clusters found for the Remote > New contrast. D) Significant clusters found for the Recent > New contrast. E) The left sub-panel shows the mean and SD of the extracted beta-values for the regions-of-interest (ROIs): the bilateral hippocampus and the left posterior middle temporal gyrus (pMTG). The right sub-panel shows the ROIs from which the beta-values were extracted.

corresponding primes and targets, reflecting the meaning of the novel word rather than of the base word, is valid). This analysis once again revealed a main effect of Priming ($F(1,47) = 11.397$, $p = .001$, $\eta p^2 = .195$) and no effect of time ($p = .626$). The interaction of Time by Priming was again not significant ($F(1,47) = 3.227$, $p = .079$, $\eta p^2 = .064$). As with the pause-detection task, we also ran repeated measures ANOVAs with the English affinity covariates (IELTS, duration stay, knowledge of English). We did not find any significant effects with any of the covariates ($p > .149$).

Here, our hypothesis was partially confirmed. We found a main effect of priming: that is, the novel words primed semantically related target words, such that RTs were facilitated. Although this effect was numerically larger in the Remote condition than in the Recent condition, as we had expected, this difference was not significant.

Discussion

In this study, we asked whether memory consolidation supports the lexical integration (at the form and meaning levels) of novel L2 words in sequential bilinguals. Young adult native speakers of Italian with English as their L2 were taught two sets of English-like novel words, one set on the day of final testing (the Recent set) and one set two days earlier (the Remote set). We predicted that neural activity, as measured with fMRI, would shift from greater involvement of the hippocampus (and associated medial temporal structures) during the processing of Recent novel words (which should be less consolidated) to greater involvement of neocortical structures (the pMTG in particular) during the processing of Remote novel words (which should be more consolidated). This fMRI prediction was not confirmed. We also predicted two behavioral signatures of consolidation. The first was that there would be stronger form-based

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

Table 3. fMRI: significant clusters

	Cluster		Peak MNI coordinates		
Brain Regions	p(FWE-corr)	Voxel size	X	Y	z
Recent > Remote					
R middle occipital	< 0.001	209	44	−74	28
R middle temporal			48	−66	20
R middle temporal			56	−56	16
R precuneus	0.001	85	8	−60	26
R precuneus			12	−66	38
R precuneus			16	−64	26
R precuneus	0.047	43	18	−70	48
L cuneus	0.008	61	−16	−56	26
L calcarine			−16	−62	12
R supramarginal	0.01	58	52	−38	38
R inferior parietal			56	−42	46
L angular	< 0.001	124	−52	−62	30
L angular			−40	−54	26
L middle occipital			−46	−74	32
Remote > Recent					
L inferior frontal (pars Triangularis)	<0.001	152	−34	30	6
L inferior frontal (pars Triangularis)			−40	20	−2
L orbitofrontal	0.001	80	−34	30	−14
Recent > New					
L anterior cingulate	< 0.001	32294	−6	46	0
L cuneus			−4	−64	22
L hippocampus			−28	−34	−6
L middle temporal	0.001	90	−38	−56	0
L superior temporal			−40	−40	8
L inferior frontal (pars Triangularis)	< 0.001	202	−44	36	24
L inferior frontal (pars Triangularis)			−32	36	6
L middle frontal			−34	34	18
L middle frontal	< 0.001	304	−24	30	50
L superior frontal			−22	26	38
L superior frontal			−18	32	30
L middle temporal (anterior)	< 0.001	165	−60	−18	−20
L middle temporal (anterior)			−60	−28	−16
L middle temporal (anterior)			−64	−32	−10
L middle frontal orbital	0.001	93	−26	42	−6
L middle frontal orbital			−26	34	−16
L middle frontal orbital			−32	50	−12
R inferior temporal	0.001	91	56	−58	−4
R middle temporal			50	−52	−2

(*Continued*)

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

Table 3. (*Continued.*)

	Cluster		Peak MNI coordinates		
Brain Regions	p(FWE-corr)	Voxel size	X	Y	z
Remote > New					
L Cuneus	< 0.001	24001	−8	−62	24
thalamus			0	−6	6
L cerebellum			−26	−46	−24
L mid frontal orbital	< 0.001	141	−28	34	−16
L mid frontal orbital			−26	42	−6
L superior frontal	< 0.001	130	−18	26	46
L middle frontal			−26	28	50
L supramarginal	< 0.001	566	−52	−26	30
L posterior cingulate			−58	−20	20
L supramarginal			−64	−22	26
L cerebellum	0.001	101	−32	−62	−38
L cerebellum			−26	−46	−40
L cerebellum			−42	−72	−36
R. hippocampus	0.014	62	20	−6	−10
R. amygdala			24	4	−12
R. hippocampus			24	−4	−20
R middle occipital	< 0.001	123	38	−76	38
R angular			36	−70	46
R middle occipital			44	−76	28
L middle temporal	0.002	89	−54	−28	−16
L middle temporal			−64	−18	−18
L middle temporal			−64	−32	−10
L post central	0.05	48	−48	−8	50
L post central			−46	−10	42

L: left, R: right

Table 4. Pause-detection task: Number of analysed trials and Reaction Times [mean (SD)]

Conditions		num Trials	Reaction Times [ms]	
Remote	Pause	16 (4)	727 (171)	693 (147)
	No Pause	16 (3)	659 (144)	
Recent	Pause	16 (4)	698 (146)	672 (124)
	No Pause	16 (3)	646 (120)	
New	Pause	15 (3)	691 (140)	675 (139)
	No Pause	17 (4)	658 (153)	

competition (as measured by slower pause-detection latencies) between novel words and existing phonologically related English words for the Remote words (more consolidated) than for the Recent words (less consolidated). The second was that there would be stronger semantic priming for the Remote words than for the Recent words (again because the former are more consolidated than the latter). The first of these behavioral predictions was confirmed while the second was not (there was a numerical trend in the predicted direction, but the interaction was not statistically significant). The behavioral findings thus suggest that consolidation does indeed appear to support word learning in the L2, but that integration at the level of lexical form may have a longer time-course than that at the level of lexical meaning.

The lack of neural evidence for consolidation does not undermine this conclusion, simply because it is a null result. The behavioral measures may be more sensitive than the fMRI measures. This may be for technical reasons, such as that the hypothesized differences in neural activity due to the effects of consolidation processes may not be adequately captured by the BOLD signal (i.e., it may have insufficient spatial or temporal resolution). It is unlikely, however, that this reason is the whole story, given the successful imaging of consolidation effects in prior fMRI word-learning studies (Davis et al., 2009; Takashima et al., 2014). An additional or alternative reason for the present null result is that the recognition-memory task that we used in the scanner may have been sub-optimal. Remember that the displays

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

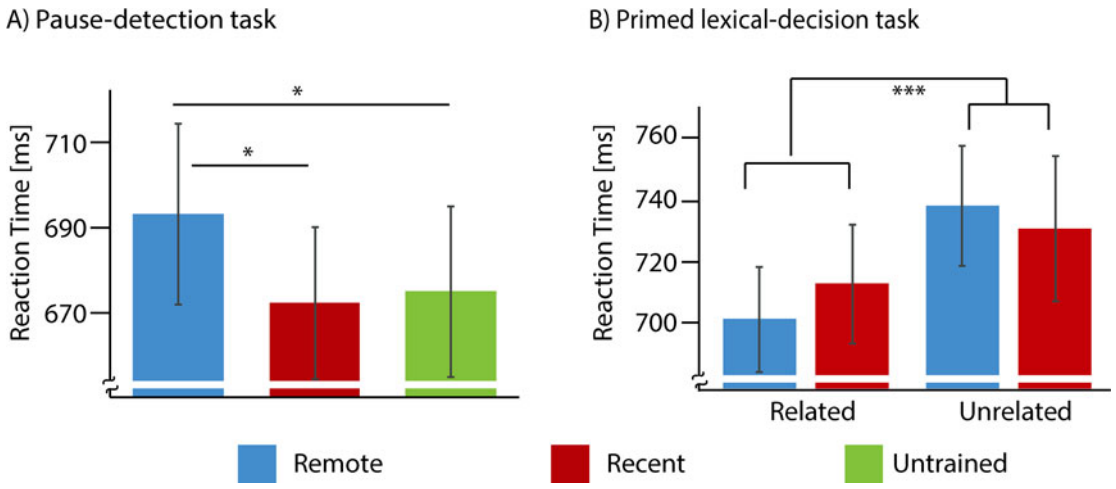

Figure 4. Mean Reaction Times (mean ± S.E.) in the pause-detection task (left panel) and the primed lexical-decision task (right panel). * $p < .05$, *** $p < .001$.

Table 5. Primed lexical-decision task: Mean RTs and mean error rates.

		Mean num Trials (SD)	RT (SD)
Remote	Related	28 (4)	702 (117)
	Unrelated	36 (3)	739 (133)
Recent	Related	27 (4)	714 (133)
	Unrelated	36 (2)	732 (162)

involved the presentation of multiple printed words as response options. The act of reading these response options may have triggered episodic memories of the presentation of those words during training and thus increased the amount of hippocampal activity, making it harder to see the predicted difference between the Recent and Remote conditions. The difference observed between activity in the New condition and that in the two trained conditions (see Figure 3E) is consistent with the idea that, even though we used novel pictures in the scanner, both Remote and Recent words called up episodic memories from the training phase. An alternative task (e.g., a cued recall task in which the participants are asked to produce the learnt words when cued with the picture, rather than choose one from the options on the screen) could avoid this problem. Another possible interpretation is that the successful retrieval of remotely learned novel words still made use of the hippocampal episodic memory system. Both episodic and semantic memory representations can be activated upon retrieval if both memory representations are present (Takashima & Bakker, 2017), and the episodic memory network was strongly engaged when retrieving the word meaning in both conditions. Future research is required to investigate further the neural signatures of memory consolidation in L2 word learning.

The current study failed to reveal effects of differences in English proficiency on either the degree of competition found in the pause-detection task or the amount of priming in the lexical-decision task. On the one hand, these null effects may again reflect lack of sensitivity (in this case, with respect to the individual differences measures): length of stay in an English-speaking country, standard test scores, and knowledge of the vocabulary that was tested in the study. These measures may not be precise enough indicators of proficiency and/or there may have been insufficient variance in the measures for associations with the consolidation effects to be detected. It may also be possible that the majority of participants were proficient enough for differences in amount of consolidation to go undetected. For example, differences in consolidation could arise if participants varied in their ability to hear non-native phonological categories accurately (e.g., the /ð/ in the novel word *brothon*) but perhaps there were few (or no) participants with such phonological difficulties. On the other hand, and with the usual provisos about null effects, one might argue that the lack of effects of proficiency is the expected outcome. If lexical integration is achieved through the same mechanisms as are used in L1 vocabulary acquisition (and, hence, are available to every L2 learner whatever their proficiency level), then lexical integration should not vary with language proficiency.

The results in the pause-detection and primed lexical-decision tasks are indeed consistent with the hypothesis that word learning in L2 is based on the same consolidation mechanisms as have been shown to affect L1 word learning in adults (Gaskell & Dumay, 2003; Dumay & Gaskell, 2007) and children (Henderson et al., 2012, 2013; van der Ven et al., 2017). The lack of an effect of learning time (i.e., no interaction between Priming and Time) in the primed lexical-decision task may reflect insufficient power. If a study with greater statistical power were to detect this interaction, it would suggest that integration at the form and meaning levels has the same time course. Alternatively, however, if such a study were to confirm that the numerical trend observed here is not a real effect, it would suggest that meaning-level integration is indeed more rapid than form-level integration.

This difference in the time-course of lexical integration at the two levels would be consistent with evidence suggesting that meaning-level integration with new words can be very rapid (possibly due to "fast mapping"; Borovsky, Kutas & Elman, 2010; Borovsky, Elman & Kutas, 2012; Carey & Bartlett, 1978). Furthermore, it is plausible that meaning-level integration may be quicker than form-level integration because there is more new knowledge to integrate in the latter case. Indeed, when a new L2 word refers to an existing concept (as in this study), no new concept needs to be acquired, and hence no reorganization

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

of the semantic lexicon is required. All that is required is the integration of the knowledge that the concept has a new label. In contrast, the new form (unless it is a cognate, which we avoided here) is a novel package of phonological and orthographic knowledge that will overlap with (and thus need integration with) potentially many phonological and orthographic neighbors. While further research is required to examine possible differences between form- and meaning-level integration, we tentatively conclude, based on the present non-significant interaction and on these empirical and theoretical arguments, that consolidation processes may take more time for phonological than for semantic knowledge to be integrated in L2 learning. It is an interesting possibility that a novel L2 word might be lexically integrated at the meaning level while it still exists only as an episodic trace and thus not (fully) integrated at the form level.

One might consider that the order in which the tests were given might have had an effect. The participants were always tested for recognition memory in the scanner first, followed by the pause detection task, and finally the primed lexical decision task. It is thus possible that the recognition memory task (on Remote and Recent novel words) could have weakened potential differences between these two conditions in the other two tasks. Such an effect seems unlikely in the pause detection task because it was done on the base words of the novel words, which had not previously been presented. Furthermore, there was a difference between the Recent and Remote conditions in this task. This difference thus arose in spite of any effect of the preceding recognition memory task. In the lexical decision task, however, although the target words had also never appeared in any other task, it is possible that retrieval of the novel words' meanings during the recognition memory task could have acted to remove differences in the efficacy of the Recent versus the Remote primes. The order of tests may thus be another reason (beyond insufficient power, as noted above) why the priming effect was not significantly modulated by the recent/remote training manipulation. This underscores the need for further research (looking e.g., at test order) to examine whether lexical integration has the same time course at the form and meaning levels.

The evidence of integration of novel words at both levels suggests that L2 vocabulary acquisition may be different from the acquisition of L2 grammar (Hartshorne et al., 2018) or L2 phonology (Flege et al., 1999). Vocabulary acquisition appears to rely on mechanisms that are basically the same across ages and languages (i.e., lexical integration is effectively the same throughout one's lifetime whichever language one is learning). Acquisition of grammar and phonology, in contrast, appears to depend on mechanisms which change with age, and thus (if the L2 is learned later in life than the L1) across languages.

Our claim that mechanisms of word learning are shared across L1 and L2 stands in sharp contrast to the recent suggestion that word learning is fundamentally different in L1 and L2 (Qiao & Forster, 2017). Qiao and Forster suggest that this qualitative difference reflects the closure of a critical period for language acquisition. As just discussed, there are grounds to question this suggestion: there is evidence that word-learning ability (unlike the abilities to learn other aspects of language) does not appear to depend on a critical period (Hartshorne & Germine, 2015; Snow & Hoefnagel-Höhle, 1978).

It remains the case, however, that our positive evidence of lexical consolidation in Italian–English sequential bilinguals is inconsistent with the negative evidence presented in Qiao and Forster's (2017) masked priming experiment with Chinese–English sequential bilinguals. There are several possible explanations for this discrepancy. First, there may be global differences between learning words in both the spoken and printed modalities (as here) versus learning words only in the print modality (as in Qiao & Forster, 2017). This seems unlikely, given other evidence of lexical consolidation in the print modality (e.g., Bakker et al., 2014; Qiao & Forster, 2013). Second, there are other procedural differences between the studies, including whether the primes were masked (as in Qiao & Forster) or unmasked (as in the current priming task).

Third, as discussed in the study by Nakayama and Lupker (2018), where they also found a facilitation effect for form-priming in an L2 English experiment with L1 Japanese participants, the unexpected findings in the L2 English materials with L2 Chinese participants in the Qiao and Forster study (2017) may be due to the differences between the L1 and L2 orthographies. It may be that an alphabetic orthographic lexicon (for L2 English) is different (and engages different processes) in someone who has an L1 logographic orthographic lexicon (as in the case of L1 Chinese participants) than in someone whose L1 uses an alphabetic orthography. The L1 visual masked priming study by Qiao and Forster (2013) with English L1 participants indeed revealed interference effects which parallel those observed here in the remote condition of the pause detection task and those in previous studies investigating novel word learning in L1 (e.g., Bakker et al., 2014; Gaskell & Dumay, 2003). These interference effects, we have argued, reflect lexical integration effects after consolidation. It may be that lexical consolidation in L2 may be slower (or even fundamentally different) when it entails a different kind of writing system from that used in the participant's L1. One way to test this possibility would be to rerun Qiao and Forster's masked priming experiment in English with L2 participants who speak an L1 that uses the same orthographic writing system. Another way would be to test Chinese–English bilinguals but with a longer consolidation period than in Qiao and Forster's study (e.g., more days of training and/or a longer delay between training and final test). Without further evidence from such experiments for differences in lexicalization processes between L1 and L2, and given the present results and the data showing similarities in consolidation processes between children and adults, it appears unfounded to argue, as Qiao and Forster (2017) do, that all L2 words (irrespective of language and orthography) are represented and processed in a qualitatively different way from L1 words.

Our data suggest instead that consolidation processes are similar across languages and are consistent with the idea that those processes are also similar across ages. This similarity, however, does not mean that there are no age-related changes in vocabulary acquisition whatsoever. In fact, there appear to be subtle differences between adults and children in the efficacy of consolidation processes. In particular, while degree of consolidation in children is modulated by the lexical neighborhood of the new word (i.e., new words are consolidated better if they have fewer neighbors), no such difference is observed in adults (James et al., 2019). Further research is required to investigate these age-related differences in greater detail. It is possible that effects of neural maturation may modulate the consolidation process (Takashima et al., 2019), potentially through age-related changes in the engagement of schema-based memory formation processes (van Kesteren, Beul, Takashima, Henson, Ruiter & Fernández, 2013; van Kesteren, Ruiter, Fernández & Henson, 2012). In spite of these potential age-related effects, however, it appears that integration of new word forms has the same basic consolidation trajectory

https://doi.org/10.1017/S1366728921000286 Published online by Cambridge University Press

(i.e., the forms of recently learned words are less likely to be integrated than those of more remotely learned words) in monolinguals and bilinguals.

It is important to emphasize that there is much more to word learning than lexical consolidation. Many other cognitive mechanisms are also involved. As with consolidation, in many cases those mechanisms are likely to be the same in the L2 as in the L1. They include, for example, the ability to hold novel word-forms in phonological short-term memory (Baddeley, Gathercole & Papagno, 1998; McQueen, Eisner, Burgering & Vroomen, 2020) and the ability to form label-meaning associations (which as we have already noted can be very fast; Borovsky et al., 2010, 2012; Carey & Bartlett, 1978). But other mechanisms are necessarily specific to the learning of L2 words. In particular, translation to the corresponding L1 word is an obvious way to learn an L2 word (and indeed traditional vocabulary learning in the language classroom involves rote learning of lists of L1-L2 word pairs). We should note that, in order to discourage a translation-based learning strategy, we tried to select referents of the new words that did not have readily available Italian translations (e.g., because they were low in frequency, which was of course also necessary to ensure that the participants did not already know the English words). Nevertheless, we cannot rule out that our participants used translation as a strategy. More generally, it is clear that there will always be differences between learning an L2 word (through translation to the L1 word or otherwise) when the referent is an existing concept than when (as is often the case, especially early in life, in the L1) the referent is a new concept.

Our claim is thus that lexical consolidation is an essential mechanism of word learning that can be found across ages and languages but neither that it is the only mechanism underlying word learning nor that those other mechanisms are always the same in L2 as in L1 learning. As we have also argued, our results across the two behavioral tasks suggest that consolidation may have a different time course at the meaning and form levels. It is possible that the consolidation process is more rapid for the integration of word meanings than for the integration of word forms, and this may especially be the case when the L2 word is a new, non-cognate label for an existing concept. Our findings suggest that, most clearly with respect to word forms, memory consolidation appears to support the gradual integration of new L2 words into the bilingual lexicon.

Supplementary Material. The supplementary materials comprise one file containing:

1. Materials: Critical stimuli (Table S1).
2. Procedure: Timeline of tasks (Table S2) and descriptions of each training-phase task.
3. Results: Error rates during the training phase (Table S3).

For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728921000286

Acknowledgments. This article is based on the MSc thesis of the first author, under the supervision of the other two authors, carried out as part of the Cognitive Neuroscience Research Master program at Radboud University, Nijmegen. We thank Tobia Spampatti for his invaluable assistance with preparing and running the experiment.

References

Alvarez, P, & Squire, LR (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences of the United States of America* **91**(15), 7041–7045.

Baddeley, A, Gathercole, S, & Papagno, C (1998). The phonological loop as a language learning device. *Psychological Review* **105**(1), 158–173.

Bakker, I, Takashima, A, van Hell, JG, Janzen, G, & McQueen, JM (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language* **73**(1), 116–130.

Bakker, I, Takashima, A, van Hell, JG, Janzen, G, & McQueen, JM (2015a). Changes in theta and beta oscillations as signatures of novel word consolidation. *Journal of Cognitive Neuroscience* **1–12**. doi:10.1162/jocn_a_00801

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

Bakker-Marshall, I, Takashima, A, Schoffelen, J.-M, van Hell, JG, Janzen, G, & McQueen, JM (2018). Theta-band oscillations in the Middle Temporal Gyrus reflect novel word consolidation. *Journal of Cognitive Neuroscience* **30**(5), 621–633. doi:10.1162/jocn_a_01240.

Borovsky, A, Kutas, M, & Elman, J (2010). Learning to use words: Event related potentials index single-shot contextual word learning. *Cognition* **116**(2), 289–296. doi: 10.1016/j.cognition.2010.05.004

Borovsky, A, Elman, J, & Kutas, M (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development* **8**(3), 278–302. doi: 10.1080/15475441.2011.614893

Brett, M, Anton, J.-L, Valabregue, R, & Poline, J.-B (2002). Region of interest analysis using an SPM toolbox. *Neuroimage* **16**(2, Supplement 1), xvii-xx. doi: 10.1016/S1053-8119(02)90013-3

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. *Behavior Research Methods* **41**(4), 977–990.

Carey, S, & Bartlett, E (1978). Acquiring a single new word. *Papers and Reports on Child Language Development* **15**, 17–29.

Charge, N, & Taylor, LB (1997). Recent developments in IELTS. *ELT Journal* **51**(4), 374–380.

Davis, MH, Di Betta, AM, Macdonald, MJE, & Gaskell, MG (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience* **21**(4), 803–820.

Davis, MH, & Gaskell, MG (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1536), 3773–3800. doi:10.1098/rstb.2009.0111

Dijkstra, T, Wahl, A, Buytenhuijs, F, van Halem, N, Al-Jibouri, Z, De Korte, M, & Rekké, S (2019). Multilink: A computational model for bilingual word recognition and word translation. *Bilingualism: Language and Cognition* **22**(4), 657–679.

Dumay, N, & Gaskell, MG (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

Flege, JE, Yeni-Komshian, GH, & Liu, S (1999). Age constraints on second-language acquisition. *Journal of Memory and Language* **41**, 78–104.

Frankland, PW, & Bontempi, B (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience* **6**(2), 119–130.

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

Granena, G, & Long, MH (2012). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* **29**(3), 311–343.

Hartshorne, JK, & Germine, L.T (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science* **26**, 433–443.

Hartshorne, JK, Tanenbaum, JB, & Pinker, S (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* **177**, 263–277.

Henderson, LM, Weighall, AR, Brown, H, & Gaskell, MG (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science* **15**(5), 674–687. doi:10.1111/j.1467-7687.2012.01172.x

Henderson, LM, Weighall, AR, Brown, H, & Gaskell, MG (2013). Online lexical competition during spoken word recognition and word learning in children and adults. *Child Development* **84**(5), 1668–1685. doi:10.1111/cdev.12067

James, E, Gaskell, MG, Weighall, A, & Henderson, L (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience & Biobehavioral Reviews* **77**, 1–13. doi://doi.org/10.1016/j.neubiorev.2017.01.054

James, E, Gaskell, MG, & Henderson, LM (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science* **22**(3), e12776. doi:10.1111/desc.12776

Kapnoula, E, Gupta, P, Packard, S, & McMurray, B (2015). Immediate lexical integration of novel word forms. *Cognition* **134**, 85–99.

Keuleers, E, & Brysbaert, M (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods* **42**(3), 627–633.

Kroll, JF, & Stewart, E (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations. *Journal of Memory and Language* **33**, 149–174. doi://doi.org/10.1006/jmla.1994.1008

Lindsay, S, & Gaskell, MG (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **39**, 608–622.

Marian, V, Bartolotti, J, Chabal, S, & Shook, A (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE* **7**(8): e43230. doi:10.1371/journal.pone.0043230

McClelland, JL, McNaughton, BL, & O'Reilly, RC (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.

McGaugh, JL (2000). Memory--a Century of Consolidation. *Science* **287** (5451), 248–251. doi:10.1126/science.287.5451.248

McQueen, JM, Eisner, F, Burgering, M, & Vroomen, J (2020). Specialized memory systems for learning spoken words. *Journal of Experimental Psychology: Learning, Memory and Cognition* **46**, 189–199.

Nakayama, M, & Lupker, SJ (2018). Is there lexical competition in the recognition of L2 words for different-script bilinguals? An examination using masked priming with Japanese–English bilinguals. *Journal of Experimental Psychology: Human Perception and Performance* **44**(8), 1168–1185. doi:10.1037/xhp0000525

Pajak, B, Creel, SC, & Levy, R (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **42**(9), 1377–1399.

Qiao, X, & Forster, KI (2013). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory and Cognition* **39**(4), 1064–1074.

Qiao, X, & Forster, KI (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition* **158**, 147–152.

Smith, FRH, Gaskell, MG, Weighall, AR, Warmington, M, Reid, AM, & Henderson, LM (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science* **21**(5), e12639. doi:10.1111/desc.12639

Snow, CE, & Hoefnagel-Höhle, M (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development* **49**, 1114–1128.

Szmalec, A, Page, MPA, & Duyck, W (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language* **67**, 342–354.

Takashima, A, Petersson, KM, Rutters, F, Tendolkar, I, Jensen, O, Zwarts, MJ, McNaughton, BL, & Fernández, G (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America* **103**(3), 756–761. doi:10.1073/pnas.0507774103

Takashima, A, Bakker, I, van Hell, JG, Janzen, G, & McQueen, JM (2014). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *Neuroimage* **84**, 265–278.

Takashima, A, & Bakker, I (2017). Memory consolidation. In *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Boston, MA, US: De Gruyter Mouton. pp. 177–200.

Takashima, A, Bakker-Marshall, I, van Hell, JG, McQueen, JM, & Janzen, G (2019). Neural correlates of word learning in children. *Developmental Cognitive Neuroscience* **37**, 100649. doi: 10.1016/j.dcn.2019.100649

Tamminen, J, Payne, JD, Stickgold, R, Wamsley, EJ, & Gaskell, MG (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience* **30**, 14356–14360.

Tzourio-Mazoyer, N, Landeau, B, Papathanassiou, D, Crivello, F, Etard, O, Delcroix, N, Mazoyer, B, & Joliot, M (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**(1), 273–289.

Ullman, MT (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition* **4**, 105–122.

Ullman, MT, & Lovelett, JT (2018). Implications of the declarative/procedural model for improving second language learning: The role of memory enhancement techniques. *Second Language Research* **34**(1), 39–65. doi:10.1177/0267658316675195

van der Ven, F, Takashima, A, Segers, E, & Verhoeven, L (2017). Semantic priming in Dutch children: Word meaning integration and study modality effects. *Language Learning* **67**(3), 546–568. doi:10.1111/lang.12235

van Kesteren, MTR, Beul, SF, Takashima, A, Henson, RN, Ruiter, DJ, & Fernández, G (2013). Differential roles for medial prefrontal and medial temporal cortices in schema-dependent encoding: from congruent to incongruent. *Neuropsychologia* **51**(12), 2352–2359.

van Kesteren, MTR, Ruiter, DJ, Fernández, G, & Henson, RN (2012). How schema and novelty augment memory formation. *Trends in Neurosciences* **35**(4), 211–219. doi://dx.doi.org/10.1016/j.tins.2012.02.001

Winocur, G, & Moscovitch, M (2011). Memory transformation and systems consolidation. *Journal of the International Neuropsychological Society* **17** (5), 766–780. doi:10.1017/S1355617711000683