Applied Psycholinguistics (2026), **47**, e50, 1–27
doi:10.1017/S0142716426100927

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Second language vocabulary consolidation through incidental learning

Rosa Tanja Padt[1], Katherine Kerschen[2], Zofia Wodniecka[3] and Carrie N. Jackson[4]

[1]Penn State University, USA, [2]Germanic Languages and Literatures, Washington University in St Louis, USA, [3]Institute of Psychology, Jagiellonian University, Poland and [4]Department of Germanic and Slavic Languages and Literatures, Penn State University, USA
Corresponding author: Carrie N. Jackson; Email: cnj1@psu.edu

(Received 13 August 2024 UTC; revised 26 June 2026 UTC; accepted 17 July 2026 UTC)

Abstract

Previous first (L1) and second language (L2) vocabulary acquisition research suggests that newly learned words are initially encoded episodically and are only fully integrated into the lexicon after a period of offline consolidation. While prior L2 consolidation research has utilized intentional vocabulary learning tasks, L2 vocabulary learning often occurs incidentally during meaning-focused tasks. To investigate consolidation under more incidental learning conditions, adult L1 Polish-L2 English learners read and listened to a story that contained 12 novel English-like words with novel semantic meanings. Lexical integration was measured using a pause detection task in which response latencies were compared between competitor words similar in form to the newly learned English-like words (e.g., *daffodil*) and control words. Overall, there was no increase in response latency for competitor words after a 24-hour consolidation period. However, higher word recognition accuracy immediately after the learning task led to increased response latencies for competitor words relative to control words after a 24-hour consolidation period. These results suggest that initial learning of the L2 form—as measured by recognition accuracy—is critical for L2 lexical integration and that consolidation effects can extend to adult L2 word learning under more incidental learning conditions.

Keywords: Incidental vocabulary learning; second language word learning; vocabulary consolidation

At the core of language acquisition lies vocabulary acquisition, which provides a key building block for communication. Vocabulary acquisition is also of paramount importance for second language (L2) learning (Webb et al., 2023). In L2 learning, learners often encounter new vocabulary within contextualized input, such as stories and other forms of connected discourse; for example, through reading (Godfroid et al., 2018; Pellicer-Sánchez, 2016), listening (Chang & Ma, 2018;

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Feng & Webb, 2020), or a combination of the two (Brown et al., 2008; Chen, 2021; Teng, 2018; Webb & Chang, 2012, 2015). Learning words from L2 input while engaging in a meaning-focused task, in which vocabulary learning is not the primary aim, is termed incidental learning (Webb et al., 2023). Such incidental exposure plays an important role in shaping L2 vocabulary learning (Teng, 2018). Even though research shows that incidental learning via exposure to L2 input is effective (for a review, see Webb et al., 2023), open questions remain regarding the mechanisms that underlie this form of L2 vocabulary learning and the quality of word knowledge that emerges. One area of interest is the process by which vocabulary becomes consolidated and lexicalized; that is, the mechanisms by which newly learned words become integrated into the learners' L2 lexical-semantic system. In the field of first language (L1) vocabulary acquisition, the complementary learning systems (CLS) model has been used to explain L1 novel word learning and consolidation in both school-aged (e.g., Henderson et al., 2021) and adult L1 users (for review, see Gore et al., 2022). While the CLS model is well established in the L1 word learning literature, few studies have explored how consolidation and lexical integration proceed among L2 learners when learning new L2 words. Further, those studies that have investigated L2 consolidation have all relied on intentional word learning tasks in which learners' attention is specifically directed towards learning new L2 words, such as memorizing novel L1-L2 word translation pairs (e.g., Bakker-Marshall et al., 2021; Qiao & Forster, 2017; Tartaro et al., 2021; for reviews, see Ekerdt et al., 2023, Palma & Titone, 2021). However, this neglects the critical role that incidental word learning plays in L2 vocabulary acquisition. Applying the basic principles of the CLS model and exploiting the rich experimental methodology developed to examine L1 word learning (e.g., Davis & Gaskell, 2009; Gore et al., 2022), the present study examines the process of consolidation and lexical integration via exposure to stories in adult L2 learners. As such, we seek to advance our understanding of how novel L2 words are integrated and consolidated in adult L2 learners' lexical-semantic memory under an incidental word learning condition, in which learners are engaged in a meaning-focused task (i.e., reading and listening to a story for comprehension). In so doing, the present study provides greater insight into the underlying cognitive mechanisms that support incidental L2 vocabulary learning in adults.

Complementary learning systems model

Consolidation is a critical step in word learning during vocabulary acquisition, including incidental learning. Consolidation involves establishing connections between newly acquired and previously acquired words in semantic memory, e.g., semantically (*vehicle* and *car*) or phonologically (e.g., *fallow* and *fellow*) related words. Memory traces that are consolidated are resistant to forgetting. Building on the complementary learning systems theory of learning and memory (McClelland et al., 1995), the complementary learning systems model for word learning (CLS; Davis & Gaskell, 2009) posits that the learning and retention of novel words depend on the dynamic interaction of distinct neural systems within the brain. This account distinguishes two processes: rapid initial learning and slow lexical consolidation.

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

In the initial stages of word learning, newly acquired words are stored as episodic memory traces in the hippocampal region of the brain. Only after a period of time are newly acquired words integrated into the mental lexicon and stored as stable, long-term representations in the neocortex. Consolidation thus entails a shift in the relative engagement of these two complementary neural structures; novel words are assumed to undergo a consolidation process during which hippocampal connections decay and neocortical connections are strengthened. A consequence of consolidation is that the newly learned words interact with existing words, at which point the novel word representation is thought to have reached the state of lexicalization, i.e., it is lexically integrated (for a review, see Ekerdt et al., 2023).

The CLS model further posits that during the rapid initial learning stage, learners can recall the newly learned word information, including formal properties and semantic information, immediately after exposure. However, this newly learned information is not yet integrated into the mental lexicon (i.e., lexicalized) and thus cannot interact with existing word knowledge. Only after a period of offline consolidation, during which novel word information becomes integrated into the existing semantic network, can newly learned word information interact with existing word representations (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010). Furthermore, research within the CLS paradigm suggests that not just time, but sleep, promotes memory consolidation (Bakker-Marshall et al., 2018; Batterink et al., 2017; Brodt et al., 2023; Dumay & Gaskell, 2007; Tamminen et al., 2010), although some studies have observed consolidation effects without any sleep between learning and testing (Kapnoula et al., 2015; Lindsay & Gaskell, 2013).

Measuring successful memory consolidation and lexicalization involves tasks that go beyond the recognition/recall of novel words, as such measurements may only tap into the retrieval of purely episodic memory traces. It is possible for a novel word form to be successfully retrieved but to lack connections to other word forms and semantic information that are crucial for language use and continued language development. Evidence that novel words have been successfully integrated into semantic memory (i.e., have become lexicalized) and have developed fully-established lexical-semantic representations come from tasks that assess the ability of novel words to impact the processing of existing words. Such tasks include semantic priming (e.g., Bakker et al., 2015; Liu & Van Hell, 2020) to test semantic integration and pause detection (e.g., Davis & Gaskell, 2009; Gaskell & Dumay, 2003) to test form-based integration. For example, Gaskell and Dumay (2003) familiarized participants through repeated exposure to spoken pseudowords (e.g., *cathedruke*) that exhibit strong phonological overlap with existing words (e.g., *cathedral)*. Then participants listened to recordings of words that could contain an artificially manipulated pause (here: *cathe_dral*), and they had to identify whether the recording contained a pause. The pause detection task (for details, see Gaskell & Dumay, 2003; Mattys & Clark, 2002) exploits a phenomenon central to theories of word recognition: lexical competition. In lexical competition, multiple candidates that match an incoming signal compete for selection, thereby slowing down recognition of a given target word. The critical assumption is that pauses are detected more quickly when there is only one candidate for completion (e.g., *cathe_dral*) than when there is competition between several alternatives (e.g., *cathe_dral/cathe_druke*). Thus, pause detection latencies for the target

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

cathedral increase only when the newly learned pseudoword *cathedruke* has been lexicalized and thus functions as a lexical competitor to the existing word *cathedral*. Indeed, Gaskell and Dumay (2003) observed that exposure to newly learned pseudowords (*cathedruke*) that overlapped with existing words impacted pause detection performance on existing words (*cathedral*) one week after exposure, but not immediately after exposure. They took this delayed competition effect to imply that lexical integration of newly learned words does not occur immediately but rather requires a period of consolidation.

Lexical consolidation in a second language

While much research has examined the CLS model in L1 vocabulary learning (Davis & Gaskell, 2009; Gore et al., 2022), exploring the CLS model for L2 vocabulary learning is understudied (Ekerdt et al., 2023; Palma & Titone, 2021). On the one hand, it seems reasonable to assume that consolidation also emerges as a pivotal mechanism in L2 vocabulary learning, facilitating the seamless interaction between newly learned L2 words and pre-existing L2 knowledge (Tartaro et al., 2021). In fact, L1 word learning continues across the lifespan (Gore et al., 2022). Thus, there is no a priori reason to assume that the neurocognitive mechanisms associated with L1 word learning should be fundamentally different from (adult) L2 word learning (Lindsey & Gaskell, 2010).

At the same time, critical differences between L1 and L2 vocabulary learning may differentially impact the nature of L1 and L2 word learning and consolidation. Particularly among L2 learners who acquire a second language after childhood, L2 vocabulary learning builds on an already existing L1 lexicon, with well-established connections between L1 words and concepts. For example, bilingual lexical-semantic memory models (e.g., the Revised Hierarchical Model (Kroll & Stewart, 1994; Kroll et al., 2014) and the developmental bilingual interactive activation model (Grainger et al., 2010)) posit that less-proficient L2 learners access the semantic representations of L2 words via connections between the L2 and L1 lexical forms, essentially relying on translating between the L2 and the L1 because the direct connections between L2 lexical forms and concepts are weak. Only at more advanced proficiency levels do learners build direct connections between L2 lexical forms and their associated concepts. Although later research has shown that even less-proficient learners can directly access semantic information via their L2, particularly for comprehension-based tasks (e.g., Sunderman & Kroll, 2006), asymmetries in direct conceptual access for L1 words versus L2 words remain (Kroll et al., 2010), which could affect lexical integration.

In recent years, a few studies have applied the principles of the CLS model to investigate L2 word learning and consolidation (Bakker-Marshall et al., 2021; Qiao & Forster, 2017; Tartaro et al., 2021; for a review, see Ekerdt et al., 2023). For example, Bakker-Marshall et al. (2021) trained monolingual English speakers and bilingual Spanish-English speakers on Swahili words. Participants learned two sets of L2 words: remote words on Day 1 and recent words on Day 2. The training procedure involved four blocks, in which participants saw the Swahili words and their English translations interspersed with four training tasks: matching a Swahili

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

word to its English translation in a four-alternative forced-choice (4AFC) task, typing the English translation for a Swahili word, matching an English word to its Swahili translation in a 4AFC task, and typing the Swahili translation for an English word. For all tasks, participants saw the correct answer after each response. The researchers collected both behavioral and fMRI brain imaging data; we will focus here on the behavioral outcome measures. In testing on Day 2, participants completed a primed lexical decision task. In this task, participants had to decide if a presented letter string was a real English word or not. The target items were preceded by semantically related or unrelated Swahili primes. When remote novel words served as semantically related primes (e.g., *farasi* (horse)—SADDLE), participants responded faster than for unrelated primes. This priming effect was not found for recent words learned on Day 2, showing evidence of meaning-based lexical integration after a 24-hour period but not immediately after training, in line with the CLS model (Davis & Gaskell, 2009). Note that this effect was found only with the monolingual English speakers but not the bilingual Spanish-English speakers, who showed no priming effects for remote or recently learned words (see Bakker-Marshall et al., 2021, for further discussion).

In another L2 study, Tartaro et al. (2021) trained L1 Italian-L2 English speakers on pseudowords that were similar in phonological form to existing English words and referred to real-world objects (e.g., the word *apricon* was paired with a picture of a stapler). The training included 10 different tasks that intentionally focused on learning different aspects of the target words, including spelling and pronunciation practice. As in Bakker-Marshall et al. (2021), participants learned two sets of target words: remote novel words on Day 1 and recent novel words on Day 2. At the end of training on both Day 1 and Day 2, participants completed a free recall test in which they typed the new words upon seeing their corresponding pictures, to measure their knowledge of the new word forms and their meanings. Results from testing on Day 2 revealed that participants recalled more words from Day 1 (remote novel words) than Day 2 (recent novel words). Results from a pause detection task on Day 2 revealed that participants' response times (RTs) for existing English words (e.g., *apricot*) were the same for recent words learned on Day 2 and untrained words, showing that no consolidation had occurred for recently learned novel words. In contrast, pause detection RTs were slower for remote words learned on Day 1 compared to untrained words, showing that consolidation had occurred for remotely learned novel words from Day 1. These results suggest that the novel words learned 48 hours prior to testing on Day 2 had undergone lexical integration, exhibiting form-based competition with existing words, while those learned just before the test had not, in line with the CLS model (Davis & Gaskell, 2009). On Day 2, participants also completed a primed lexical decision task, similar to Bakker-Marshall et al. (2021). When both remotely and recently learned novel words were used as primes, participants responded faster to related targets than unrelated targets. The evidence of a semantic priming effect for both remotely and recently learned words diverges from the results from Bakker-Marshall et al. (2021) and the predictions of the CLS model (Davis & Gaskell, 2009). This suggests that L2 consolidation and lexicalization follow a similar process as in the L1, but that the integration of L2 semantic information may occur more quickly than the integration of L2 forms, and possibly not require a period of offline consolidation.

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Further, Tartaro et al. found that participants' knowledge of the L2 base words—used as a proxy for L2 vocabulary size—and overall L2 proficiency did not correlate with the emergence of lexical integration effects in either pause detection or semantic priming tasks. However, they did report limited evidence that vocabulary size positively predicted performance on a free-recall task of the novel words immediately after training, but not on subsequent days of testing. Overall, despite some deviations, the outcomes of these studies are in line with the CLS model (Davis & Gaskell, 2009) and Lindsay and Gaskell's (2010) proposal that L1 and L2 lexical integration would proceed in similar ways (but see Qiao & Forster, 2017, for counterevidence[1]).

Consolidation during incidental vocabulary learning

In previous L2 consolidation studies, novel words were learned by association with existing concepts via pictures or direct L1 translations (Bakker-Marshall et al., 2021; Tartaro et al., 2021). These represent intentional learning tasks. However, L2 learners often learn L2 vocabulary under incidental learning conditions, such as while reading (Reynolds, 2025) or listening to a story or conversation (Webb & Nation, 2017). Incidental vocabulary acquisition can also occur through multimodal learning tasks, such as reading while listening. Reading while listening is an effective means for L2 learners to learn both the form and meaning of new words, with learning rates on par with or sometimes even greater than reading or listening alone (e.g., Bisson et al., 2014; Brown et al., 2008; Chen, 2021; Teng, 2018; Webb & Chang, 2012; for meta-analysis see Webb et al., 2023). Learning rates during incidental vocabulary tasks are higher for form and meaning recognition and lower for meaning recall (Webb et al., 2023). While L2 incidental vocabulary learning studies have investigated word learning outcomes, using traditional word recognition and recall tasks, how newly learned word forms and meanings become integrated into the mental lexicon (cf. lexical network formation, see Meara, 2009; Reynolds, 2025) and interact with existing lexical-semantic knowledge has rarely been studied. This question is particularly relevant for adult L2 learners in the context of reading-based incidental learning: Adults not only often acquire L2 lexical-semantic knowledge incidentally through reading, as skilled readers they retrieve word knowledge from memory to serve as contextual cues to derive the meaning of unknown words encountered during L2 reading. Therefore, the present study examines how consolidation occurs during incidental L2 vocabulary learning during L2 reading-while-listening, and individual factors that facilitate or constrain the integration of new L2 lexical-semantic knowledge into the existing lexicon.

In a seminal incidental word learning study from the L1 consolidation literature, Henderson et al. (2015) examined how children and adults integrate newly and previously acquired words when they are exposed to new words by listening to a story. Using a pause detection task to measure lexical integration (e.g., Dumay & Gaskell, 2007; Gaskell & Dumay, 2003), they found that children's and adults' ability to recognize and recall new word forms and integrate these new words with their existing lexical knowledge increased after 24 hours, compared to immediately after exposure. Among the children, but not the adults, those with larger expressive

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

vocabulary knowledge showed a larger lexical integration effect, suggesting that children with a larger vocabulary consolidate new words more effectively. Children with larger expressive vocabulary also demonstrated greater explicit memory of the novel word forms, as measured by a cued recall task 24 hours after exposure to the story. These findings highlight the reciprocal relationship between lexical integration and vocabulary knowledge during L1 acquisition.

Paralleling the results from Henderson et al. (2015), several studies on incidental L2 vocabulary learning have found a positive relationship between learning gains and prior L2 vocabulary knowledge among adolescent and adult L2 learners (e.g., Feng & Webb, 2020; Webb & Chang, 2015). In their meta-analysis of L2 incidental vocabulary learning, Webb et al. (2023) also reported that increased L2 proficiency was associated with greater incidental word learning, with L2 proficiency broadly defined to include measures of prior L2 vocabulary knowledge, standardized L2 proficiency scores (e.g., TOEFL scores) and instructor judgments. Increased vocabulary knowledge and increased L2 proficiency likely make the linguistic context that surrounds unknown words more accessible and facilitate comprehension, which in turn may aid the learning of these words (e.g., Na & Nation, 1985). Better comprehension also likely leaves more cognitive resources available to attend to and learn unknown words (Godfroid et al., 2018).

The fact that previous studies find that vocabulary knowledge and proficiency have an impact on incidental word learning among L1 children and L2 learners—and that, by extension, not all learners are equally able to learn and integrate new words into the mental lexicon while reading and/or listening to a story—highlights the variable nature of incidental vocabulary learning. Given the variability in word learning outcomes from incidental learning tasks, the question arises whether lexical integration in the L2 is similarly variable and whether the same factors that predict successful L2 word learning under incidental conditions also predict successful lexical integration. As such, it behooves researchers to go beyond traditional measures of explicit word knowledge, like word recognition and recall, to investigate how lexical integration and consolidation proceed among L2 learners during an incidental word learning task. In addition to deepening our understanding of L2 vocabulary consolidation, this may help shed light on why L2 incidental vocabulary learning is more variable in the first place.

The present study

To investigate outstanding questions on consolidation and lexical integration in L2 incidental vocabulary learning, the present study examines adults' incidental L2 word learning in a story context. We adapted Henderson et al.'s (2015) story-based task and testing materials to study the learning and consolidation of novel L2 English words among adult L1 Polish-L2 English bilinguals. Similar to research on L2 vocabulary learning (e.g., Webb et al., 2023), here we define word learning as the explicit knowledge of novel L2 words, as measured by a form recognition task, a cued recall task to measure learners' recall of novel word forms, and a definition task to measure learners' recall of novel word meanings. Consolidation and lexical integration, as defined by the CLS model (e.g., Davis & Gaskell, 2009), refer to how

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

novel words are integrated into the mental lexicon. This was measured by a pause detection task to identify whether novel word forms compete with already existing entries in the lexicon during lexical processing. In contrast to previous L2 consolidation research that focused on intentional word learning (Bakker-Marshall et al., 2021; Tartaro et al., 2021; Qiao & Forster, 2017), the present study examined L2 word learning and the consolidation and lexicalization of newly learned words in an incidental learning task. Grounded in the CLS model, we asked the following research questions:

1) Do word learning and lexical integration occur after a 24-hour period of offline consolidation for L2 words learned through an incidental, story-based learning task?
2) Are individual factors (L2 proficiency and L2 vocabulary size) and task-based performance factors (story comprehension and novel word recognition) associated with word learning and lexical integration after a 24-hour period of offline consolidation?

If L2 novel word learning under more incidental conditions (i.e., through reading and listening to a story) is driven by similar mechanisms of consolidation and lexicalization as L2 word learning during more intentional vocabulary learning tasks, such as paired-associate learning (e.g., Bakker-Marshall et al., 2021; Tartaro et al., 2021), we expect to replicate these earlier L2 studies and observe lexical integration of newly learned L2 words after a 24-hour consolidation period, but not immediately after exposure to the story. If this prediction is upheld, it would also indicate that the mechanisms underlying consolidation and lexicalization of novel L1 words learned in the context of a story (as in Henderson et al., 2015) extend to L2 novel word learning through an incidental task.

For individual factors, Henderson et al. (2015) found that children with a larger existing (L1) vocabulary exhibited better explicit recall of newly learned words and larger lexical integration effects after a 24-hour period of offline consolidation. L2 incidental word learning studies have also found improved word learning among adolescent and adult L2 learners with greater prior L2 vocabulary knowledge (Feng & Webb, 2020; Webb & Chang, 2015) and greater L2 proficiency (Webb et al., 2023). Therefore, we expect that L2 vocabulary size and L2 proficiency will positively predict performance on explicit measures of new word knowledge, including word recognition and recall. Further, if consolidation and lexical integration under incidental L2 word learning conditions among L2 adults proceed similarly as consolidation among L1 children, then we also expect that L2 vocabulary size, and possibly L2 proficiency, will positively predict the size of lexical integration effects after a 24-hour period of offline consolidation.

Method

Participants

Sixty-four L1 Polish-L2 English speakers living in Poland participated in the study in exchange for payment. Participants were recruited from a population of

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Table 1. Participant demographic information

	M(SD)	95% CI	Range
Age (years)	23.0 (4.8)	21.5, 24.5	19–41
Age of first exposure to English (years)	7.6 (3.9)	6.7, 8.6	4–16
LexTALE (% correct)	77.3 (0.1)	74.4, 80.3	62–98
Cambridge Assessment (out of 25)	21.8 (2.1)	21.1, 22.4	18–25

university students and the surrounding community. Three participants were excluded because they had learned another language in early childhood (2 English; 1 Greek). Two participants were excluded because they did not achieve B2-level proficiency on the Cambridge English Assessment ("Cambridge University Press and Assessment," n.d.). Eleven participants were excluded because they did not complete one or more experimental tasks. Six participants were excluded because their responses to the definitions task were not recorded due to technical difficulties. Two participants were excluded because they responded incorrectly to over 25% of the experimental items in the pause detection task. Therefore, all results reported here are based on the remaining 40 participants (30 female; 10 male). All participants began learning English in school at age 5 or later. See Table 1 for complete biographical information.

Stimuli

Participants were randomly assigned to either List 1 or List 2. Each list contained 12 base words (e.g., *daffodil*), 12 matching novel competitor words (e.g., *daffodat*), and 12 foil words (e.g., *daffodan*) from the items used in Henderson et al. (2015) (The complete study materials are available on the Open Science Framework (OSF) at https://doi.org/10.17605/OSF.IO/E34WB). The novel words and novel foils differed from the base words from the onset of the last syllable, such that the initial fragment (e.g., *daffo__*) had no other potential completions. Base words between the two lists were matched on the number of syllables, phonemes, length, and frequency. We initially gave five highly-proficient L1 Polish-L2 English speakers, who were knowledgeable about the English proficiency of the target participant population (i.e., university-level L1 Polish-L2 English speakers), the list of base words from Henderson et al. (2015). Based on their feedback, we replaced two base-novel words from Henderson et al. (2015; *bramble-brambooce* and *lantern-lantobe*) with two new base-novel words pairs (*cathedral-cathedruke* and *napkin-napkess*), because these informants indicated that university-level L1 Polish-L2 English speakers might not know the original two base words. This modification aimed to ensure familiarity with the base words among participants. Thirty-seven L1 English speakers rated the naturalness of the content and the words of the original story versions and the revised story versions with these new novel words. Mann–Whitney tests revealed no difference in ratings between versions (all $ps > .4$; see Supplementary Materials on OSF for more details). All base words were picturable nouns (see Henderson et al., 2015 for further details about stimuli and design). Each

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

participant learned the novel competitor words from one list while the original base words from the other list served as control words in the pause detection task.

Materials

Fictitious story

In Session 1, the experimenter told participants that they would read and listen to a story through headphones about an "alien family who lives in space and is having a day trip to an alien zoo," and that "there may be some words that you have not heard before but please continue to listen and avoid asking questions so that you do not miss any part of the story." This was the same instruction and story used by Henderson et al. (2015). Two versions of the story were recorded by a native English speaker with a North Midland English dialect. She was an intermediate-level L2 speaker of German, but spoke primarily English in her daily life. One version contained the 12 novel words from List 1, and another version contained the 12 novel words from List 2. In both versions, the 12 novel words were presented 5 times. The length, intonation, and pitch of the story for both versions were matched as closely as possible. The story was presented using E-Prime 3.0 (Psychology Software Tools, Inc, 2016).

Story comprehension task

Following the story, participants answered six questions on paper about the people, actions, and settings in the story. The story comprehension task was administered to ensure participants had understood the story.

Testing measures

The pause detection task served as the primary measure of lexical competition. The word recognition task, the cued recall task, and the definitions task were primary measures of explicit memory of the novel words. The remaining tasks served as control measures to assess knowledge of the target base words (L1-to-L2 translation task) and general English proficiency (Cambridge English Assessment; LexTALE).

Pause detection task

In Sessions 1 and 2, the pause detection task served as an on-line measure of lexical competition. In the pause detection task, 200 ms pauses were inserted at the onset of the last syllable of the 24 base words (e.g., *daffo_dil*), both for the 12 base words the participants had heard the novel word competitor during the story (competitor condition) and for the 12 base words they had not heard any novel word competitor during the story (control condition). The words were recorded by the same native English speaker who recorded the fictitious story. The participants heard the 24 base words and 24 filler words through headphones and were asked to decide whether a pause was present in each word by pressing a button on a button box. Half of the words in the competitor and control conditions contained a 200 ms pause. Each list also contained 24 filler words. Half of the fillers contained a 200 ms pause. The item order was randomized for each participant and stimuli were presented using

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

E-Prime 3.0 (Psychology Software Tools, Inc, 2016). Reaction times (RTs) were measured from pause onset.

Word recognition

In Sessions 1 and 2, the two-alternative forced-choice (2AFC) word recognition task measured participants' ability to recognize the novel words from the story. Participants heard the 12 novel words they had encountered in the story (e.g., *daffodat*) paired with an accompanying novel foil word (e.g., *daffodan*). They had to choose which word they had actually heard in the story by pressing the appropriate key on a computer keyboard. For half of the items, the novel word was presented first and for the other half of the items, the novel foil was presented first. The item order was randomized for each participant and stimuli were presented using E-Prime 3.0 (Psychology Software Tools, Inc, 2016). Participants received 1 point for each accurate response in the task.

Cued recall task

In Sessions 1 and 2, the cued recall task measured participants' ability to recall the 12 novel words from the fictitious story. Participants heard the first syllable (e.g., *daff*) of the 12 novel words from their respective list and were prompted to complete the word based on what they had heard in the story (e.g., *daffodat*). The item order was randomized for each participant and stimuli were presented using E-Prime 3.0 (Psychology Software Tools, Inc, 2016). Participants' answers were recorded and transcribed by the first author. Responses were then scored as correct (1 point) if participants correctly reproduced each phoneme of the final syllable and incorrect (0 points) if they did not accurately reproduce each phoneme of the final syllable. A second person transcribed and coded 25% of the data, with an interrater agreement of 99.2% ($\kappa = 97.5$).

Definitions task

In Session 2, the definitions task measured whether participants had learned information about the meaning of the target words. With the experimenter present in the testing booth, the participants heard a novel word through headphones and were asked to define each of the 12 novel words from the story. They were prompted to describe the words as if they were speaking to someone who had never heard the word before and to describe the word with as much detail as possible. If the participants could no longer recall any details about the words, they were prompted by the experimenter with "Can you tell me anything else about a/an ___ (e.g., *daffodat*)?" until they could provide no more details. Participants were given up to 3 points per word (max score = 36), with one point for each defining feature (Henderson et al., 2015). For example, the *daffodat* was a flying vehicle (1 feature, 1 point) with seatbelts (1 feature, 1 point) that runs on stardust (1 feature, 1 point). Participants' recorded responses were transcribed and scored by the first author. A second coder scored 25% of the data, with an interrater agreement of 94.2% ($\kappa = 88.9$).

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

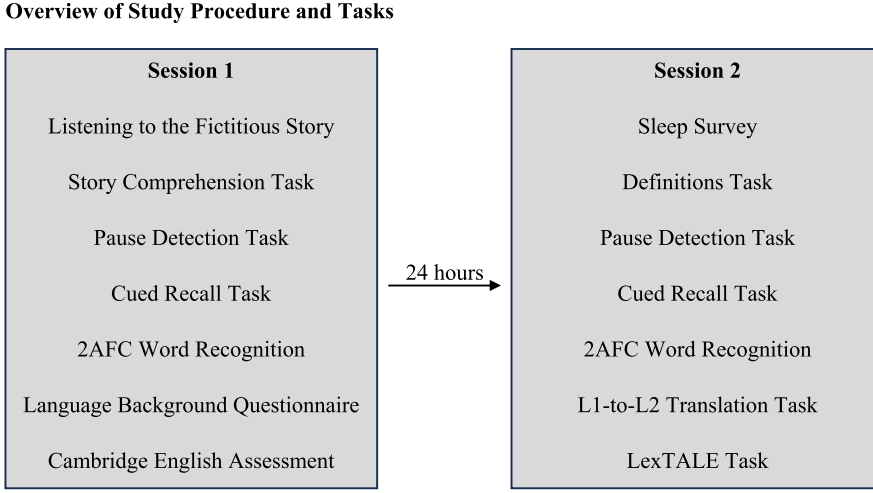


Figure 1. Overview of study procedure and tasks.

L1-to-L2 translation task

In Session 2, participants completed a written L1-L2 translation task that assessed if they knew the form of the L2 English base words for the novel word competitors from the story. Participants saw the L1 Polish translation (e.g., *narcyz* "daffodil") of the L2 base words from their assigned list of novel competitor words and were prompted to provide an L2 English translation for this word on paper. Participants received one point for every correct translation.

English proficiency

Participants completed the 25-question Cambridge English Assessment ("Cambridge University Press and Assessment," n.d.) as a measure of overall L2 English proficiency. A score of 18 or higher indicated at least B2-level English proficiency on the CERF scale. Participants also completed the LexTALE task (Lemhöfer & Broersma, 2012) as a measure of L2 English vocabulary knowledge.

Procedure

The participants were tested individually in a soundproof booth. They participated in two different sessions, one day apart. As depicted in Figure 1, in Session 1, the participants listened to and read a fictitious story and then completed the story comprehension task, a pause detection task, a cued recall task, a 2AFC word recognition task, a language background questionnaire, and the Cambridge English Assessment ("Cambridge University Press and Assessment," n.d.). Session 2 occurred approximately 24 hours later, providing participants an opportunity to sleep between Sessions 1 and 2. In Session 2, the participants completed a sleep survey, the pause detection task, the cued recall task, the 2AFC task, a definitions

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

task, an L1-L2 written translation task, and the LexTALE task (Lemhöfer & Broersma, 2012).

Data analysis

The data files and analysis code are available on OSF at https://doi.org/10.17605/OSF.IO/E34WB. As the primary measure of lexical integration, we relied on the RT and accuracy data from the pause detection task. For these analyses, both RT and error data were collapsed across the pause and no-pause conditions (Henderson et al., 2015). RTs for both conditions were calculated from pause onset in the pause condition. For the RT analysis, we excluded inaccurate responses, RTs greater than 4000ms, and any RTs greater than 2 SDs away from the individual participant mean for that condition (competitor vs. control), leading to the exclusion of 6.4% of the RT data. To address RQ1, regarding the process of lexical integration and word learning after a story-based learning task, results for the primary measure of lexical integration (Pause Detection task) and the explicit knowledge of the novel words (2AFC Task, cued recall, Definition tasks) were analyzed using mixed-effects logistic regression analyses for accuracy data and linear mixed-effect models for both the RT data and the results from the definitions task with the lme4 package version 1.1-34 (Bates et al., 2015) and *lmerTest* package version 3.1-3 (Kuznetsova et al., 2017) in R version 4.5.2 (R Core Team, 2025). For the pause detection task, Session (Session 1 vs. Session 2) and Condition (Control vs. Competitor) were entered as fixed effects, contrast coded −0.5 and 0.5. The 2AFC Task and cued recall included session as a fixed effect, also contrast coded −0.5 and 0.5.

To address RQ2, regarding the role of individual factors and task-based performance measures on word learning and lexical integration, the following individual difference variables were also entered as continuous variables, centered at the sample mean, for all tasks: (1) L2 Proficiency, based on participants' score on the Cambridge English Assessment, (2) Comprehension Score, based on participants' score on the 6-point comprehension task they completed after listening to the fictitious story, (3) Base Word Knowledge, based on participants' score on the L1-to-L2 translation task, and (4) 2AFC Accuracy, based on participants' score on the 2AFC task from Session 1, as a measure of word recognition accuracy immediately after learning.

For each analysis reported here, the random effect structure was determined by starting with the maximum structure justified by the experimental design (Barr et al., 2013), including the fixed effects of Session and Condition (for the Pause Detection task) and their interactions, removing slopes due to non-convergence (removing which variable contributed the least amount of unique variance to the random effect structure in a stepwise fashion; see R scripts on OSF for more details). This led to different random-effect structures for each model, but the most parsimonious model for each analysis. The specific random effect structure for each analysis is provided in the R code at the end of each results table (Tables 3–7). For all analyses, VIF values were < 2.0 for all fixed effects and their interactions, indicating acceptable levels of collinearity for all factors (see Supplementary Materials on OSF for full results).

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Table 2. Descriptive measures for all tasks

	M(SD)	*95% CI*	*Range*
Pause detection errors (out of 12)			
Session 1			
Competitor	0.3 (0.6)	0.9, 0.51	0–3
Control	0.5 (1.1)	0.2, 0.9	0–5
Session 2			
Competitor	0.2 (0.5)	0.1, 0.4	0–2
Control	0.3 (0.5)	0.1, 0.5	0–2
Pause detection RT (ms)			
Session 1			
Competitor	914 (183)	856, 973	
Control	921 (216)	838, 976	
Session 2			
Competitor	943 (260)	860, 1026	
Control	915 (220)	844, 985	
L1-to-L2 translation (out of 12)	9.9 (1.4)	9.5, 10.4	7–12
2AFC word recognition (% correct)			
Session 1	66.0 (20.1)	59.6, 72.5	25–100
Session 2	67.3 (17.6)	61.7, 72.9	33–100
Cued recall (out of 12)			
Session 1	2.1 (2.4)	1.3, 2.9	0–10
Session 2	4.0 (2.8)	3.1, 4.8	0–12
Definitions (max = 36)	6.4 (4.9)	4.8, 7.9	0–21
Story comprehension (out of 6)	3.6 (1.2)	3.2, 3.9	1–5

Results

Descriptive results for all tasks are presented in Table 2, and the results of all statistical analyses are presented in Tables 3–7. Results from the sleep survey conducted in Session 2 confirmed that participants slept 6.7 hours on average (SD = 1.3; range: 3–9 hours) between sessions.

Pause detection task

As seen in Table 3, there was a significant 2-way interaction between 2AFC Accuracy x Session.[2] As seen in Figure 2, while there was a decrease in the number of pause detection errors as a function of 2AFC accuracy in Session 1, this pattern

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Table 3. Results from the mixed-effects model for the pause detection task: errors

Fixed effects	Estimate	*SE*	*z*-Value	*Pr(>|z|)*
Intercept	−4.74	0.41	−11.48	<.001
Session	−0.80	0.51	−1.57	.116
Condition	−0.89	0.48	−1.85	.065
Proficiency	0.09	0.13	0.71	.476
Comprehension score	0.33	0.22	1.47	.143
Base Word knowledge	0.02	0.19	0.10	.924
2AFC accuracy	0.01	0.01	1.21	.227
Session × Condition	−0.66	0.88	−0.75	.455
Session × Prof.	0.24	0.22	1.09	.275
Session × Comp. Score	0.00	0.39	−0.01	.992
Session × Base Word K.	0.37	0.32	1.16	.245
Session × 2AFC Acc.	0.05	0.02	2.39	**.017**
Condition × Prof.	0.19	0.22	0.87	.385
Condition × Comp. Score	−0.16	0.38	−0.42	.677
Condition × Base Word K.	−0.18	0.32	−0.57	.570
Condition × 2AFC Acc.	0.03	0.02	1.55	.121
Session × Condition × Prof.	0.38	0.41	0.95	.345
Session × Condition × Comp. Score	0.23	0.70	0.33	.742
Session × Condition × Base Word K.	0.48	0.58	0.82	.412
Session × Condition × 2AFC Acc.	0.04	0.04	1.06	.290
ACC~Session*Condition*Prof+Session*Condition*Comp+Session*Condition*Base + Session*Condition*AFC+(1+Session2||item)+(1+Session2+Type2||participant)				

reverses in Session 2—although this is within the context of few errors overall, as the maximum number of errors for any participant in Session 2 was 2 out of 24 items. No other effects were statistically significant (all *p*s > .06).

As seen in Table 4, in the RT data, there was a significant 2-way interaction between Condition x 2AFC Accuracy and a significant 3-way interaction between Session x Condition x 2AFC Accuracy. As seen in Figure 3, while 2AFC Accuracy did not predict the degree of lexical competition—defined as participants' average RT for competitor words minus their average RT for control words—in Session 1, increased 2AFC Accuracy in Session 1 led to increased lexical competition effects in Session 2. There were no other significant effects (all *p*s ≥ .05). Overall, participants also exhibited high accuracy on the L1-to-L2 translations of the base words from their respective lists (see Table 2).[3]

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Table 4. Results from the mixed-effects model for the pause detection task: reaction times

Fixed effects	Estimate	*SE*	*t*-Value	*Pr(>|t|)*
Intercept	923.37	29.69	31.10	<.001
Session	17.96	33.81	0.53	.599
Condition	17.31	14.11	1.23	.220
Proficiency	−25.31	12.98	−1.95	.059
Comprehension score	−47.22	23.29	−2.03	.050
Base word knowledge	4.48	20.44	0.22	.828
2AFC accuracy	2.29	1.34	1.71	.096
Session × Condition	21.80	28.09	0.78	.438
Session × Prof.	−28.71	16.73	−1.72	.095
Session × Comp. Score	−3.59	30.02	−0.12	.905
Session × Base Word K.	26.25	26.34	0.10	.326
Session × 2AFC Acc.	0.46	1.73	0.26	.793
Condition × Prof.	−2.31	6.98	−0.33	.741
Condition × Comp. Score	−2.04	12.77	−0.16	.873
Condition × Base Word K.	−0.83	10.96	−0.08	.940
Condition × 2AFC Acc.	2.22	0.87	2.54	**.011**
Session × Condition × Prof.	9.39	13.87	0.68	.500
Session × Condition × Comp. Score	−12.41	24.83	−0.50	.617
Session × Condition × Base Word K.	−16.16	21.78	−0.74	.458
Session × Condition × 2AFC Acc.	3.24	1.44	2.25	**.025**
RT∼Session*Condition*Prof+Session*Condition*Comp+Session*Condition*Base + Session*Condition*AFC+(1|item)+(1+Session||participant)				

Table 5. Results from the mixed-effects model for the 2AFC task

Fixed effects	Estimate	*SE*	*z*-Value	*Pr(>|z|)*
Intercept	0.93	0.33	2.82	.005
Session	0.08	0.16	0.47	.636
Proficiency	0.03	0.06	0.54	.588
Comprehension score	0.16	0.11	1.43	.154
Base Word knowledge	0.01	0.10	0.10	.919
Session × Prof.	0.03	0.08	0.40	.691
Session × Comp. Score	−0.17	0.14	−1.20	.232
Session × Base Word K.	−0.02	0.13	−0.14	.890
ACC∼ Session*Prof+Session*Comp+Session*Base+(1|item)+(1|participant)				

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Table 6. Results from the mixed-effects model for the cued recall task

Fixed effects	Estimate	*SE*	*z*-Value	*Pr(>|z|)*
Intercept	−1.71	0.30	−5.60	<.001
Session	1.28	0.26	4.95	**<.001**
Proficiency	0.26	0.10	2.54	**.012**
Comprehension score	0.32	0.18	1.75	.082
Base Word knowledge	0.33	0.16	2.06	**.039**
2AFC accuracy	0.00	0.01	−0.41	.680
Session × Prof.	−0.12	0.10	−1.23	.220
Session × Comp. Score	0.15	0.17	0.90	.368
Session × Base Word K.	0.02	0.14	0.16	.871
Session × 2AFC Acc.	0.02	0.01	1.76	.078
ACC∼Session*Prof+Session*Comp+Session*Base + Session*AFC+(1 + Session|item)+(1|participant)				

Table 7. Results from the mixed-effects model for the definitions task

Fixed effects	Estimate	*SE*	*t*-Value	*Pr(>|t|)*
Intercept	0.53	0.07	7.12	<.001
Proficiency	0.05	0.03	2.09	**.044**
Comprehension score	0.19	0.05	4.03	**<.001**
Base Word Knowledge	−0.07	0.04	−1.62	.114
2AFC accuracy	0.00	0.00	0.50	.617
Score∼Prof+Comp+Base + AFC+(1|item)+(1|participant)				

2AFC task

As seen in Table 5, there was no effect of Session, nor were there any effects or interactions with any individual difference variables (all *p*s > .15). As seen in Table 2, the changes in novel word recognition accuracy were minimal from Session 1 to Session 2.

Cued recall task

As seen in Table 6, there was a significant effect of Session. Although performance on the cued recall task was low overall, participants' scores increased from Session 1 to Session 2 (see Table 2). There were also significant effects of Proficiency and Base Word Knowledge. The positive estimates in Table 6 indicate that overall accuracy on the cued recall task increased with increased proficiency, increased knowledge of the base words, and increased accuracy on the story comprehension task. No other effects were significant (all *p*s > .07).

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

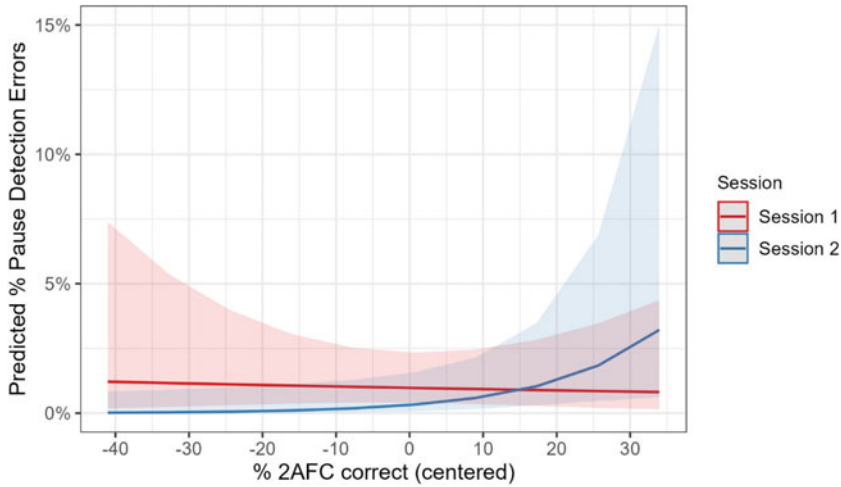


Figure 2. Predicted percentage of pause detection errors as a function of 2AFC accuracy in Session 1.

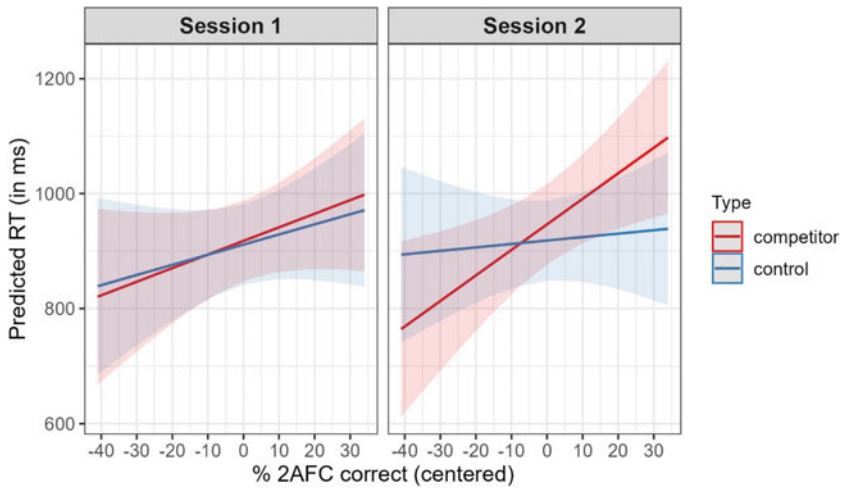


Figure 3. Predicted reaction times as a function of condition and 2AFC accuracy in Session 1 (left) and Session 2 (right).

Definitions task

As seen in Table 2, performance on the definition task, which was only included in Session 2 (as in Henderson et al., 2015), was low.[4] As seen in Table 7, there were significant effects of Proficiency and Comprehension Score. The positive estimates in Table 7 indicate that participants' scores on the definition task increased with increased proficiency and increased accuracy on the story comprehension task. No other effects were significant ($ps > .1$).

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Discussion

This study investigated the impact of a story-based word learning task on the consolidation and integration of newly learned L2 lexical-semantic knowledge in the mental lexicon. Testing adult L1 Polish-L2 English learners, our first research question (RQ1) asked whether lexical integration and word learning occur after a 24-hour consolidation period for L2 words learned in an incidental learning task. In the pause detection task, no RT differences emerged between competitor and control words in either Session 1 or 2, implying no evidence of lexical integration for this group of participants overall. Explicit knowledge of the novel target words increased, with the highest scores occurring with target word recognition, as measured by the 2AFC task. Scores on the definitions and cued recall tasks were low, although cued recall of the target words improved from Session 1 to Session 2. Our second research question (RQ2) asked whether individual differences and task-based performance factors have an impact on word learning and lexical integration. Participants with higher L2 English proficiency and participants with higher accuracy on the story comprehension questions from the learning task performed better on the definitions task. Participants with higher L2 English proficiency and participants with higher knowledge of the base words—as measured by the L1-to-L2 translation task—performed better on the cued recall task. Notably, participants with higher word recognition accuracy—as measured by the 2AFC task—showed greater lexical competition effects in the pause detection task in Session 2, after the 24-hour consolidation period.

Contrary to our prediction for RQ1, we did not replicate the results from Henderson et al. (2015), who found form-based lexical integration effects after a 24-hour consolidation period among L1 children and adults using the same story-based learning task as the present study, even though the number of exposures to each novel word was identical across studies. We also did not replicate previous L2 vocabulary consolidation studies that found form-based (Tartaro et al., 2021) and semantic-based (Bakker-Marshall et al., 2021) lexical integration effects after a 24-hour consolidation period, even though our participants reported overnight sleeping between Sessions 1 and 2, thereby providing optimal conditions for consolidation to occur. The key difference between our study and these previous L2 word learning studies was the learning method. In our study, the training phase consisted of listening to and reading a fictitious story that contained novel words, followed by answering comprehension questions about the story. This contextualized, incidental approach to word learning aligns with vocabulary learning practices in adult L2 learning (Reynolds, 2025; Webb & Nation, 2017), but is in stark contrast to the intentional word learning methods in Bakker-Marshall et al. (2021) and Tartaro et al. (2021). Both studies used paired-associate learning in their training, namely L1-L2 translation pairs in Bakker-Marshall et al. (2021) and picture-novel word pairs in Tartaro et al. (2021). Furthermore, their training phases contained extensive practice activities with the novel words. These differences could explain why Tartaro et al. (2021) found a general lexical integration effect in the pause detection task after an offline consolidation period while we did not. Their intentional instruction method trained participants more intensively on their formal knowledge of the target items, and, indeed, their participants' word recognition

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

accuracy for words learned 48 hours before testing (93%) and shortly before testing (92%) was very high. Therefore, group-level form-based lexical competition may have been more likely to emerge after a consolidation period compared to our study. As such, differences between these previous findings and the present results suggest that consolidation and lexical integration, as outlined in the CLS model (Davis & Gaskell, 2009), may not proceed uniformly among L2 learners when learning new L2 words, but are qualified by instruction method. Specifically, incidental learning tasks with only five exposures to new L2 words, rather than intentional learning activities that involve more extensive and focused practice with new L2 words, are less effective in integrating newly learned L2 words into the mental lexicon and forming connections between these new L2 words and existing word knowledge.

In Tartaro et al. (2021) and Bakker-Marshall et al. (2021), the novel L2 forms also referred to existing semantic concepts, indicated either by pictures of objects (Tartaro et al., 2021) or L1 translation equivalents (Bakker-Marshall et al., 2021). In the present study, the novel L2 forms (e.g., *daffodat*) were related to novel semantic concepts (e.g., *a flying space vehicle with seatbelts that runs on stardust*) (see also Bakker et al., 2015; Liu & Van Hell, 2020, for a similar approach in L1 word learning). These variances in training stimuli create different L2 word learning situations in terms of the lexical-conceptual links assumed by bilingual lexical-semantic memory models (e.g., the Revised Hierarchical Model (RHM; Kroll & Stewart, 1994; Kroll et al., 2014) and the developmental Bilingual Interactive Activation model (d-BIA; Grainger et al., 2010)). For example, the RHM assumes that access to conceptual information of L2 words is mediated via L1 translation equivalents, particularly at initial stages of word learning. Bakker-Marshall et al.'s (2021) paired-associate training with translation pairs reinforced L1-L2 links so that accessing conceptual information connected to L2 words would rely on the L1 word form, while Tartaro et al.'s (2021) training connected L2 words directly to their corresponding concepts (via pictures). In the present study, the combination of novel L2 forms with novel semantic concepts meant there was no L1-based concept for the novel L2 form to attach to (see Reynolds, 2025, for further discussion). Paralleling Henderson et al.'s (2015) study on L1 word learning, our L2 word learning study did not include a measure of meaning-based lexical integration, only of form-based lexical integration (the pause detection task). It is possible that the necessity of establishing new conceptual information alongside new L2 lexical information had an overall negative impact on the consolidation process for novel words.

A recent model on L2 lexical representation, the Ontogenesis Model of the L2 Lexical Representation (OM), describes in more detail how lexical-semantic knowledge develops in L2 learners' mental lexicon (Bordag et al., 2022). This model postulates that each component of an L2 lexical representation (orthographic, phonological, and semantic knowledge, mappings between domains, and networks around each domain) follows its own ontogenetic curve, or developmental trajectory. Furthermore, L2 lexical representations are characterized by fuzziness, meaning any component of an L2 lexical representation may be imprecisely encoded rather than fully specified. Therefore, L2 lexical representations that contain some already established—and thus less fuzzy—components, such as an L2 word that draws on L1 semantics, may engage with the rest of the L2 mental lexicon

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

differently than new—and thus highly fuzzy—representations. Such differences could also contribute to the observed differences in the present results versus those in Tartaro et al. (2021) and Bakker-Marshall et al. (2021).

For RQ2, we confirmed our prediction that higher L2 proficiency and vocabulary knowledge would lead to better explicit memory of the novel word forms, paralleling research on L2 incidental vocabulary learning (e.g., Feng & Webb, 2020; Webb & Chang, 2015; Webb et al., 2023; see also Tartaro et al., 2021). However, higher L2 proficiency and vocabulary knowledge did not correlate with the emergence of lexical integration effects after a consolidation period. This suggests that while individual differences related to language knowledge may impact L2 learners' performance on measures of word recall, such factors are less likely to predict L2 learners' ability to integrate newly learned words into their existing mental lexicon (see also Tartaro et al., 2021). This contrasts with the L1 children in Henderson et al. (2015), where expressive vocabulary size positively correlated with both explicit memory of the novel word forms and the size of their lexical integration effect, as measured by the pause detection task. Instead, in the present study, it was a task-based factor, namely performance on the 2AFC recognition task in Session 1, that significantly correlated with the emergence of lexical integration effects at Session 2. Only those participants with higher accuracy on the 2AFC recognition task exhibited evidence of lexical competition on the pause detection task in Session 2. We hypothesize that in L2 learning, initial learning of the novel lexical form plays a crucial role in consolidation, especially when learners are exposed to new words more incidentally through a meaning-focused task like reading and listening to a story. This initial learning of novel lexical forms is crucial to initiate the integration process and is more difficult in an L2, particularly in an incidental learning context. At least a partial representation for the new word must be established in the mental lexicon before the new word can be consolidated and integrated with other items in the mental lexicon. However, models of L2 lexical development propose that establishing well-specified lexical representations in the L2 is not necessarily easy or quick. According to the OM (Bordag et al., 2022), each component of the lexical representation follows its own ontogenetic curve from imprecise (fuzzy) to optimum encoding, and some components of the representation may never reach their optimum state (see Jiang, 2000, 2004, for a three-stage model that also emphasizes the development of a formal representation early on in L2 word learning). Furthermore, fuzzy encoding in any linguistic domain (orthographic, phonological, or semantic) has a negative impact on the quality of the mappings between domains and the network connections within domains, as all dimensions of the L2 lexical representation are interconnected.

When looking at the scores on the 2AFC word recognition task—as an index of whether some representation of the L2 form has been established—a stark difference between our results and Henderson et al.'s (2015) emerges. Henderson et al.'s (2015) L1 adult participants had a mean score of 90.9% on the 2AFC task at Session 1, with a range from 75% to 100%, meaning all paticiapnts scored above chance, and 24 out of 33 were essentially at ceiling, with scores of 92% or higher. In contrast, our L2 participants' mean score on the 2AFC task at Session 1 was 66.0%,

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

with a range from 25% to 100%. Further, the L1 children in Henderson et al. exhibited group-level lexical integration effects after 24 hours, even though their performance on the 2AFC task at Session 1 was also more variable, with a mean score of 71.0% and a range of 41.7–100%. This suggests that establishing an initial formal representation, especially in an incidental learning context, and then harnessing these initial representations to integrate the novel forms into the mental lexicon, may be more difficult in the L2 than the L1—and learners' success in doing so is more variable. That said, in our study, those participants with higher 2AFC scores at Session 1—indicating the ability to recognize the novel forms after five exposures to the novel words—showed evidence of form-based lexical integration at Session 2. If the initial formal representation is too weak or "fuzzy" (Bordag et al., 2022), or not established at all, the item cannot interact or compete with other items in the network, and we would not expect to see evidence of lexical integration. In this sense, the OM provides a way to understand the impact that initial word recognition performance has on the consolidation process. This interpretation of our results also aligns with Bakker-Marshal et al.'s (2021) proposition that lexical integration proceeds (or can proceed) more slowly and gradually in the L2 than the L1.

A final contribution of the present study is that it leverages findings from the burgeoning literature on L2 lexical consolidation to provide broader insights into how vocabulary learning proceeds in the L2. Specifically, it can help explain why incidental vocabulary learning leads to relatively small gains in L2 word knowledge, by highlighting the importance of initial stages of learning. While learning new words incidentally through context rather than targeted training is an important part of L2 vocabulary development (Teng, 2018; Webb & Nation, 2017), research has generally shown that it leads to smaller word learning gains than intentional learning (e.g., Laufer, 2003). A recent meta-analysis of incidental L2 vocabulary learning studies showed average gains (in terms of proportion of target words learned) of 9–18% at immediate posttest and 6–17% at delayed posttest (Webb et al., 2023). By comparison, a meta-analysis of intentional L2 vocabulary learning studies showed learning gains of 18–77% at immediate posttest and 24–73% at delayed posttest (Webb et al., 2020), with the large variation attributed to which specific word learning tasks were used. Our participants' form recognition of the novel words at immediate posttest (following the training in Session 1), as measured by the 2AFC task, was actually much higher than the average form recognition tasks found in Webb et al.'s (2023) meta-analysis of incidental L2 vocabulary learning, 66.0% compared to 18.1%, though that could be due to the nature of the 2AFC task.

The initial learning account that we propose to explain the high variability in the emergence of form-based lexical integration effects could also explain the overall small learning gains for incidental L2 vocabulary learning, as reported in Webb et al.'s (2023) meta-analysis. Assuming that establishing an initial representation of the L2 phonological and orthographical form is a key component for further lexical development, such as attaching semantic, syntactic, and morphological meaning to the form and developing associations with the broader lexical network (Bordag et al., 2022; Jiang 2000, 2004), it follows that learning tasks need to provide the opportunity to acquire this L2 form knowledge. In contrast to intentional vocabulary learning

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

paradigms, which often involve repetition of and focused practice with the target words, incidental learning methods typically provide fewer exposures to the target words and direct the learners' attention away from the words themselves and towards some broader meaning-bearing context. Therefore, the opportunities for initial learning are more limited. If this initial learning does not occur, performance on subsequent tasks of form recognition, form recall, or meaning recognition will be low. A limitation of the current study is that we only tested one type of incidental vocabulary learning condition, reading while listening. Given the wide variety of ways to learn vocabulary from context (e.g., reading, listening, and watching videos) and the differing cognitive demands they place on the learner (see, e.g., Reynolds, 2025), future research should investigate and compare the impact of different modalities of incidental learning on initial encoding and the trajectory of L2 lexical development and consolidation.

Conclusion

The current study presents new evidence to suggest that form-based L2 lexical integration can occur as the result of an incidental learning task, but that the integration of the novel words into the mental lexicon heavily depends on successful initial learning of the L2 form. Our results suggest that the processes that underlie the consolidation of L2 words in a story-based incidental learning task proceed largely in accordance with the CLS model, originally developed to account for intentional word learning and consolidation in L1 (Davis & Gaskell, 2009), but with greater variability than for L1 consolidation. Especially under more incidental learning conditions, consolidation and integration may not occur for all learners and seems constrained by their initial learning of the novel L2 form. More specifically, the current study offers insight into the role of initial L2 form learning in L2 vocabulary development, pointing to how initial learning creates the foundation for further lexical development and integration into the mental lexicon, and helps to explain the variable success of such initial L2 form learning during incidental vocabulary learning tasks. Future studies should explore the connections between consolidation and the broader arc of L2 vocabulary learning more systematically. In particular, researchers should build on research from the L2 vocabulary acquisition literature showing that factors like the number of exposures or longer periods of practice lead to greater learning gains and test whether such factors similarly lead to greater consolidation effects across a broader range of learners (e.g., Brown et al., 2008; Rott, 1999). Similarly, one could directly compare different learning methods, including more intentional teaching methods—perhaps in combination with an initial incidental learning task—to investigate the relationship between initial learning, subsequent memory of the form and meaning of newly learned words, and patterns of consolidation and lexicalization among L2 learners (e.g., Bisson et al., 2014). Additionally, future research should test for both meaning-based and form-based lexical competition. Together, such research would contribute to our understanding of both L2 learning processes and the structure of the L2 mental lexicon.

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Replication package. Replication data and materials for this article can be found at https://osf.io/e34wb/.

Acknowledgements. This research was supported by NSF OISE grant 1545900 to Paola Dussias, Judith Kroll, John Lipski, and Janet van Hell.

Competing interests. The authors declare none.

Notes

1 Qiao and Forster (2017) did not observe a prime lexicality effect in Chinese-English bilinguals who learned new L2 English words. The absence of the L2 lexicality effect in their masked priming task may be driven by differences between L1 Chinese and L2 English orthographies, as discussed by Ekerdt et al. (2023). This suggests the observed differences in consolidation patterns between L1 and L2 word learning may be influenced by language-specific orthographic characteristics.

2 A preliminary analysis included Pause Type (No Pause vs. Pause) as an additional fixed effect in the analysis for the error data. This analysis revealed no main effect ($p = .152$) or interactions between Pause Type, Condition and Session (all $ps > .4$) so this variable was not included in the primary analyses reported here.

3 Participants only completed the L1-to-L2 translation task for the 12 base words that formed the novel words from their experimental list, rather than for all 24 words (both the 12 base word competitors and the 12 control words). Therefore, it was not possible to exclude items from the pause detection task that participants did not demonstrate prior knowledge of, as in previous studies (e.g., Henderson et al., 2015). However, descriptive results from the six participants who knew the L2 English translations of all 12 base words they received on the L1-to-L2 translation task—thus demonstrating high enough vocabulary knowledge that they likely knew all 24 base words—paralleled the overall results reported here. Only two of these six participants exhibited the predicted lexical competition effect, and four of the six participants exhibited negative lexical competition effects in which the control words had longer RTs than the competitor words. Therefore, it is unlikely that the absence of overall lexical competition effects is due to the inclusion of unknown L2 English base words in the analyses.

4 In general, participants scored 0 or 1 for the 12 novel words, meaning they remembered at most 1 feature about the novel words. Participants with higher definition scores tended to answer 2 or 3 features on a few words and 0 or 1 for the remaining novel words.

References

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**, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Bakker-Marshall, I., Takashima, A., Fernandez, C. B., Janzen, G., McQueen, J. M., & Van Hell, J. G. (2021). Overlapping and distinct neural networks supporting novel word learning in bilinguals and monolinguals. *Bilingualism: Language and Cognition*, **24**(3), 524–536. https://doi.org/10.1017/S1366728920000589

Bakker-Marshall, I., Takashima, A., Schoffelen, J. M., Van Hell, J. G., Janzen, G., & McQueen, J. M. (2018). Theta-band oscillations in the middle temporal gyrus reflect novel word consolidation. *Journal of Cognitive Neuroscience*, **30**(5), 621–633.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Batterink, L. J., Westerberg, C. E., & Paller, K. A. (2017). Vocabulary learning benefits from REM after slow-wave sleep. *Neurobiology of Learning and Memory*, **144**, 102–113. https://doi.org/10.1016/j.nlm.2017.07.001

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Bisson, M-J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of repeated exposure to multimodal input in incidental acquisition of foreign language vocabulary. *Language Learning*, **64** (4), 855–877. https://doi.org/10.1111/lang.12085

Bordag, D., Gor, K., & Opitz, A. (2022). Ontogenesis model of the L2 lexical representation. *Bilingualism: Language and Cognition*, **25**(2), 185–201. https://doi.org/10.1017/S1366728921000250

Brodt, S., Inostroza, M., Niethard, S., & Born, J. (2023). Sleep: A brain state serving systems memory consolidation. *Neuron*, **111**(7), 1050–1075. https://doi.org/10.1016/j.neuron.2023.03.005

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language*, **20**(2), 136–163.

Cambridge University Press and Assessment. (n.d.). English Language Assessment – General English. Cambridge English. Retrieved July 21, 2023, from https://www.cambridgeenglish.org/test-your-english/general-english/.

Chang, L., & Ma, J. (2018). Comparing the effects of listening input and reading input on EFL learners' incidental vocabulary acquisition. *Chinese Journal of Applied Linguistics*, **41**(2), 169–181. https://doi.org/10.1515/cjal-2018-0010

Chen, Y. (2021). Comparing incidental vocabulary learning from reading-only and reading-while listening. *System*, **97**, 102442. https://doi.org/10.1016/j.system.2020.102442

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

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

Ekerdt, C., Takashima, A., & McQueen, J. (2023). Memory Consolidation in L2 Neurocognition. In K. Morgan-Short & J. G. van Hell (Eds.), *The routledge handbook of second language acquisition and neurolinguistics* (pp. 353–367). Routledge. https://doi.org/10.4324/9781003190912-33

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, **42**(3), 499–523. https://doi.org/10.1017/S0272263119000494

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, **89**(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Godfroid, A., Ahn, J., Choi, I. N. A., Ballard, L., Cui, Y., Johnston, S.L., Sarkar, A., & Yoon, H. J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, **21**(3), 563–584. https://doi.org/10.1017/S1366728917000219

Gore, K. R., Woollams, A. M., Bruehl, S., Halai, A. D., & Lambon Ralph, M. A. (2022). Direct neural evidence for the contrastive roles of the complementary learning systems in adult acquisition of native vocabulary. *Cerebral Cortex*, **32**(16), 3392–3405. https://doi.org/10.1093/cercor/bhab422

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-Thinking the Bilingual Interactive-Activation Model from a Developmental Perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 267–284). John Benjamins. https://doi.org/10.1075/lald.52.18gra

Henderson, L., Devine, K., Weighall, A., & Gaskell, M. G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology*, **51**(3), 406. https://doi.org/10.1037/a0038786

Henderson, L. M., van Rijn, E., James, E., Walker, S., Knowland, V. C., & Gaskell, M. G. (2021). Timing storytime to maximize children's ability to retain new vocabulary. *Journal of Experimental Child Psychology*, **210**, 105207. https://doi.org/10.1016/j.jecp.2021.105207

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, **21**(1), 47–77. https://doi.org/10.1093/applin/21.1.47

Jiang, N. (2004). Semantic Transfer and Development in Adult L2 Vocabulary Acquisition. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 101–26). John Benjamins. https://doi.org/10.1075/lllt.10.09jia

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

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, **33**(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science*, **23**(3), 159–163. https://doi.org/10.1177/0963721414528511

Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The revised hierarchical model: A critical review and assessment. *Bilingualism: Language and Cognition*, **13**(3), 373–381. https://doi.org/10.1017/S136672891000009X

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, **82**(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, **59**(4), 567–587. https://doi.org/10.3138/cmlr.59.4.567

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, **44**, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lindsay, S., & Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, **60**, 45–63. https://doi.org/10.1111/j.1467-9922.2010.00600.x

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **39**(2), 608. https://doi.org/10.1037/a0029243

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

Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. John Benjamins Publishing. https://doi.org/10.1075/lllt.24

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language*, **47**, 343–359. https://doi.org/10.1016/S0749-596X(02)00037-2

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, **102**(3), 419–457. https://doi.org/10.1037/0033-295X.102.3.419

Na, L., & Nation, I. S. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, **16**(1), 33–42. https://doi.org/10.1177/003368828501600103

Palma, P., & Titone, D. (2021). Something old, something new: A review of the literature on sleep-related lexicalization of novel words in adults. *Psychonomic Bulletin & Review*, **28**, 96–121. https://doi.org/10.3758/s13423-020-01809-5

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition from and while reading: An eye-tracking study. *Studies in Second Language Acquisition*, **38**(1), 97–130. https://doi.org/10.1017/S0272263115000224

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

Qiao, X., & Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition*, **158**, 147–152. https://doi.org/10.1016/j.cognition.2016.10.026

R Core Team (2025). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/.

Reynolds, B. L. (2025). *Researching second language incidental vocabulary acquisition through reading*. Springer Nature.

Rott, S. (1999). The effect of exposure frequent on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, **21**(4), 589–619. https://doi.org/10.1017/S0272263199004039

Sunderman, G., & Kroll, J. F. (2006). First language activation during second language lexical processing: An investigation of lexical form, meaning, and grammatical class. *Studies in Second Language Acquisition*, **28**(3), 387–422. https://doi.org/10.1017/S0272263106060177

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *Journal of Neuroscience*, **30**(43), 14356–14360. https://doi.org/10.1523/JNEUROSCI.3028-10.2010

Tartaro, G., Takashima, A., & McQueen, J. M. (2021). Consolidation as a mechanism for word learning in sequential bilinguals. *Bilingualism: Language and Cognition*, **24**(5), 864–878. https://doi.org/10.1017/S1366728921000286

Teng, F. (2018). Incidental vocabulary acquisition from reading-only and reading-while-listening: A multi-dimensional approach. *Innovation in Language Learning and Teaching*, **12**(3), 274–288. https://doi.org/10.1080/17501229.2016.1203328

Webb, S., & Chang, A. C. (2012). Vocabulary learning through assisted and unassisted repeated reading. *Canadian Modern Language Review*, **68**(3), 267–290. https://doi.org/10.3138/cmlr.1204.1

Webb, S., & Chang, A. C. (2015). Second language vocabulary learning through extensive reading with audio support: How do frequency and distribution of occurrence affect learning? *Language Teaching Research*, **19**(6), 667–686. https://doi.org/10.1177/1362168814559800

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press. https://doi.org/10.25170/ijelt.v12i1.1458

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, **56**(2), 161–180. https://doi.org/10.1017/S0261444822000507

Webb, S., Yanagisawa, A., & Uchihara, T. (2020). How effective are intentional vocabulary-learning activities? A meta-analysis. *The Modern Language Journal*, **104**(4), 715–738. https://doi.org/10.1111/modl.12671

Cite this article: Padt, R. T., Kerschen, K., Wodniecka, Z., & Jackson, C. N. (2026). Second language vocabulary consolidation through incidental learning. *Applied Psycholinguistics*. https://doi.org/10.1017/S0142716426100927

https://doi.org/10.1017/S0142716426100927 Published online by Cambridge University Press