

# UC Berkeley

## UC Berkeley Previously Published Works

### Title

Context Synthesis Accelerates Vocabulary Learning Through Reading: The Implication of Distributional Semantic Theory on Second Language Vocabulary Research

### Permalink

<https://escholarship.org/uc/item/36m6x4jf>

### Authors

Wang-Kildegaard, Bowen  
Ji, Feng

### Publication Date

2023-05-03

### DOI

10.1093/applin/amad014

### Data Availability

The data associated with this publication are in the supplemental files.

Peer reviewed

**Context Synthesis Accelerates Vocabulary Learning Through Reading: The Implication of Distributional Semantic Theory on Second Language Vocabulary Research**

Bowen Wang-Kildegaard<sup>1</sup>, Feng Ji<sup>1 2 3</sup>

<sup>1</sup> Berkeley School of Education, University of California, Berkeley

<sup>2</sup> Department of Applied Psychology and Human Development, University of Toronto

<sup>3</sup> Division of Biostatistics, University of California, Berkeley

This is a pre-copyedited, author-produced version of an article accepted for publication in *Applied Linguistics* following peer review. The version of record, Wang-Kildegaard, B.-W., Ji, F. (2023). Context synthesis accelerates vocabulary learning through reading: The implication of distributional semantic theory on second language vocabulary research. *Applied Linguistics.*, is available online at:

[https://academic.oup.com/applij/advance-article-abstract/doi/10.1093/applin/amad014/7150752?utm\\_source=advanceaccess&utm\\_campaign=applij&utm\\_medium=email](https://academic.oup.com/applij/advance-article-abstract/doi/10.1093/applin/amad014/7150752?utm_source=advanceaccess&utm_campaign=applij&utm_medium=email)

University of California's Open Access Policy grants us the right to share this author accepted manuscript on University of California's eScholarship immediately upon publication in the journal.

### Author Bio

Bowen Wang-Kildegaard <https://orcid.org/0000-0001-6943-7073>

Bowen Wang-Kildegaard is a PhD candidate in Learning Science and Human Development in the Berkeley School of Education at UC Berkeley. His research interests include the interface between vocabulary and reading, as well as the measurement of word knowledge. He has presented his works in the annual meetings of American Educational Research Association and Society for the Scientific Study of Reading. His address is 2121 Berkeley Way, Berkeley, CA 94704, USA. His email is [bowenwang6266@berkeley.edu](mailto:bowenwang6266@berkeley.edu)

Feng Ji <https://orcid.org/0000-0002-2051-5453>

Dr. Feng Ji is an assistant professor at the University of Toronto. His research interests include developing, evaluating, and applying innovative quantitative methods in social and behavioral sciences research. He recently received his Ph.D. in Biostatistics from UC Berkeley. His recent work appeared in *Psychometrika*, *Psychological Methods*, *Multivariate Behavioral Research*, and *Child Development*. His address is 252 Bloor Street West, Toronto, Ontario, M5S 1V6, Canada. His email is [f.ji@utoronto.ca](mailto:f.ji@utoronto.ca)

The authors would like to thank the children and parents who participated in this study. We also thank Dr. Jessica Briggs Baffoe-Djan for her mentorship on experimental design, and Dr. Richard Kern for his feedback. We are grateful to Dr. Sophia Rabe-Hesketh for her invaluable guidance on data analysis using hierarchical modeling. We also appreciate Dr. Anne Cunningham and her research group for their feedback on early drafts and revisions. Last but not least, we thank Brittney Cooper and Christopher Wang-Kildegaard for their suggestions.

This study received no external funding.

We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Bowen Wang-Kildegaard.

## ABSTRACT

Besides explicit inference of word meanings, associating words with diverse contexts may be a key mechanism underlying vocabulary learning through reading. Drawing from distributional semantic theory, we developed a text modification method called *reflash* to facilitate both word-context association and explicit inference. Using a set of left and right arrows, learners can jump to a target word's previous or subsequent occurrences in digital books to synthesize clues across contexts. Participants read stories with target words modified by *reflash*-only, gloss-only, gloss + *reflash*, or unmodified. Learning outcomes were measured via Vocabulary Knowledge Scale and a researcher-developed interview to probe word-context association. We modeled the learning trajectories of words across five weeks among three adolescent L2 English learners (113 word-learner pairings) using Bayesian multilevel models. We found that *reflash*-only words made more gains than words in other conditions on both outcomes, controlling for key covariates such as types of existing knowledge. Our analysis also revealed that context synthesis may be particularly useful for learning specific types of words like homonyms, which has significant pedagogical implications.

*Keywords:* Vocabulary learning; Reading; Distributional semantics; Second language acquisition; Psycholinguistics; Computer-assisted language learning

## INTRODUCTION

Researchers agree on the importance of vocabulary knowledge in second language learning (Schmitt 2010; Cobb and Horst 2019). However, *how* vocabulary should be learned or taught is debated by researchers. One of the debates concerns incidental and intentional vocabulary learning. Incidental vocabulary learning is generally conceptualized as a byproduct of reading or listening to language materials. In contrast, intentional vocabulary learning involves explicit memorization of word-meaning pairs, such as using flashcards. Decades of research indicates that word learning can occur through reading in both first language (L1; Nagy *et al.* 1985; Sternberg 1987; Cunningham 2005) and second language (L2; Day and Swan 1998; Webb 2007; Pellicer-Sánchez 2016). However, experimental studies have found that intentional learning is more efficient (Hulstijn 1992; Laufer 2005; Lin and Hirsh 2012). In these debates, neither side has fully addressed *how* word learning occurs through reading. Most studies assume that incidental vocabulary learning is mainly a result of learners explicitly inferring what a word means by applying inferential skills and metalinguistic knowledge to contextual, intralinguistic (e.g. morphology), or interlinguistic clues such as L1-L2 cognates (P. Nation 2001; Wesche and Paribakht 2009; Ender 2016). However, explicit inference may not be the whole story.

According to the distributional semantic theory, semantic learning is a result of associating words with various contexts and implicitly extracting semantic representations from global distributional patterns and contingencies across contexts (Landauer and Dumais 1997; Mander *et al.* 2017; Günther *et al.* 2019). Under this theory, the semantic knowledge of a word is not a dictionary-style formal definition, but patterns of activation in a distributional neural network. Computational simulations based on this theory can learn words at the same rate as L1 children

engaging in natural reading (Landauer and Dumais 1997). Distributional semantics may be applicable to L2 acquisition but has been under-investigated.

A goal of this paper is to provide initial evidence on the relevance of distributional semantics to L2 research. Drawing from this theory, we investigate whether synthesizing multiple contexts accelerates L2 word learning through reading. To do so, we developed a text modification method called *reflash*, enabling learners to jump to previous or subsequent occurrences of target words while reading digital books. We hypothesize that learners can synthesize information from different contexts by reviewing or previewing multiple occurrences of a word while reading. This process of context synthesis will conjecturally facilitate implicit learning via word-context association and explicit learning via inferencing.

To test our hypothesis, we completed a longitudinal experiment where students read digital books embedded with *reflash* and other text modification methods over several weeks. This study has the following methodological merits: 1. Besides explicit word knowledge, we also measured word-context association to capture the associative learning suggested by distributional semantic theory. 2. The analysis was conducted at word level, enabling granular investigation into what words were better learned and why. 3. The target words included partially learned words, permitting a nuanced examination of how existing knowledge affects learning. 4. We analyzed data using Bayesian multilevel models, which possess better finite-sample properties than frequentist models (see **Methods**).

## BACKGROUND

### **Distributional semantics: Mechanism of word learning through reading**

Connectionist theories, such as distributional semantics, have proposed a plausible mechanism of how word learning through reading occurs. According to the *distributional hypothesis*, words with similar meanings tend to occur in similar contexts (Harris 1954; Firth 1957) and thus a word's meaning corresponds with its distributional patterns over contexts (Landauer 2007; Mander *et al.* 2017). In psychology research, Latent Semantic Analysis (LSA; Landauer and Dumais 1997), one of the best-known computational implementations of this theory, explains how L1 children in higher grades add thousands of new words to their vocabulary every year. Landauer and Dumais (1997) attributed most of this achievement to reading because these children only learn a few hundred words from direct instruction per year and they already know almost all the words they hear in speech. They estimated that 5th to 8th graders acquire 10-15 new words every day, which equates to one word per five paragraphs read. However, explicit inference, the presumed mechanism behind word learning through reading, is insufficient to achieve this rate because the information in most contexts is too scant to decipher a word's exact definition. Experimental studies with L1 children in higher grades (e.g. Nagy *et al.*, 1985) have shown that the learning rate for words embedded in passages is only one word in twenty paragraphs, three times slower than the estimated rate of learning from natural reading (Landauer & Dumais, 1997). Landauer and Dumais (1997) argued that these experiments failed to measure much of the word knowledge children gained from reading.

Drawing on distributional semantic theory, models like Landauer and Dumais' (1997) LSA provide an alternative explanation of word learning through reading. These models all adopt the *distributional hypothesis* as the theoretical foundation of word meaning and learning. Traditionally, they quantify the semantic representations of words by constructing a word-by-context matrix and counting word frequency in each context. Context is operationalized either as documents where

the word occurs (e.g. Landauer and Dumais 1997) or other words within a fixed-size window around that word (e.g. Lund and Burgess 1996). If two words both occur frequently in similar contexts, these models can infer that these two words are semantically similar. For instance, ‘doctor’ and ‘physician’ both occur frequently in contexts that contain words like ‘hospital’, ‘medicine’, and ‘patient’, even though these two words seldom co-occur with *each other* (Günther *et al.* 2019). These models then derive *latent* semantic dimensions from such global distributional patterns, similar to factor analysis. The resulting semantic representation of a word is a vector of that word’s factor scores in each latent dimension. For instance, words like ‘hospital’ and ‘patient’ may form a semantic dimension about medical settings; both ‘doctor’ and ‘physician’ will have a high factor score on this dimension but low factor scores on dimensions unrelated to medicine. Landauer and Dumais (1997) found that LSA performed as well as human TOEFL test takers on synonym multiple-choice questions by choosing the option whose semantic vector is most similar to the target word (and thus has the most similar meaning). They also found that LSA simulated the same learning rate as L1 children in natural reading (Landauer and Dumais 1997), providing initial evidence for the *distributional hypothesis* and shedding light on how word learning through reading occurs.

Published in the *Journal of Memory and Language*, Mandera *et al.* (2017) introduced a more recent distributional semantic model called The Continuous Bag of Words model (CBOW) to psycholinguistics and found that CBOW performed extremely well on predicting human performances in semantics-related tasks. CBOW was originally developed in neural network research (Mikolov *et al.* 2013a; 2013b). Each word in a corpus is represented as an input node and an output node in a neural network with hidden layers. A word’s semantic representation is operationalized as an activation pattern of hidden nodes and is represented as a numeric vector of



weights. This type of model is trained to predict the occurrence of each word using the context, which is operationalized as a window of a specific number of words (e.g. six words in Mander *et al.* 2017) before and after the target word. Each time a word is encountered, the model uses its *existing knowledge* (the current vector of weights) of the surrounding words to predict the target word; if the prediction is incorrect, the surrounding words' vectors will be updated to improve future predictions. Each encounter incrementally refines one's knowledge of all words encountered.

Echoing Günther *et al.* (2019), we emphasize that these models are not merely language-engineering tools; they are implementations of the *distributional hypothesis*, a cognitive *theory* on the nature of human semantic representations and semantic learning through exposure. Indeed, the underlying mechanism of implicitly predicting one event (the presence of a target word, in this case) from associated events using global distributional patterns and contingencies is psychologically plausible. Research shows that models like CBOW are mathematically equivalent to established psychological models of associative learning, such as Rescorla and Wagner's (1972) Reinforcement Learning (Baayen *et al.* 2011; Mander *et al.* 2017). The model-derived semantic vectors encode interpretable information such as concreteness, valence, and functionality-related properties (e.g. 'tiger' has the property of 'being dangerous', see Sommerauer and Fokkens 2018). Additionally, distributional semantic models perform well in semantic tasks, such as word-pair analogy (e.g. king-man is analogous to queen-woman, see Mikolov *et al.* 2013c). Semantic similarity measures generated by these models are robust predictors of human performances on semantics-related tasks, such as response time in primed lexical decision (Baroni *et al.* 2014; Günther *et al.* 2016). These models also show human-like learning behaviors; for instance, they can acquire the same implicit bias as humans, such as gender bias (Caliskan *et al.* 2017). Therefore,

robust empirical evidence supports the *distributional hypothesis* as a psychologically plausible theory of semantic representation and learning. Next, we will discuss the potential implications of distributional semantic theory on L2 vocabulary research.

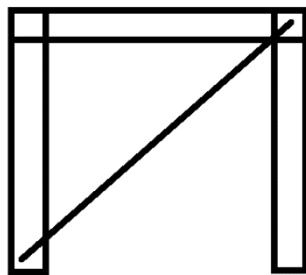
### **The implications of distributional semantics on L2 vocabulary research**

The associative learning mechanism described by the distributional semantic theory is domain general (Mandera *et al.* 2017); we can reasonably hypothesize that this mechanism may apply to L2 learning. We acknowledge that L1 and L2 learners differ in many ways. Cobb (2007) estimated that English L2 learners read, on average, 175,000 words per year in English, much less than the reading volume of English L1 fifth graders (an average of a million words, estimated by Anderson and Nagy 1992). Additionally, successful word meaning inference from reading requires knowing 98% of words in the text (P. Nation 2001; Laufer and Ravenhorst-Kalovski 2010), which may be hard for L2 learners. However, as argued before, explicit inferences may not be the whole picture; even successful experimental L1 studies do not measure up to children's actual rate of learning because these studies did not capture growth in the distributional semantic representations of words. Similarly, most L2 studies only measure explicit knowledge of word definition, which may be the tip of the iceberg. Therefore, whether this powerful associative learning occurs for L2 learners is an empirical question yet to be investigated.

Another implication of this theory is that reading not only develops knowledge of new words but also partially learned words. Many L2 studies only examined entirely unknown words; however, according to distributional semantics, learners receive new data with each encounter, thus updating their semantic representations of known words. Therefore, studying partially known words is also informative for understanding how L2 learners develop semantic knowledge.

The distributional semantic theory also has important implications on the role of context informativeness in vocabulary learning. Existing L2 research found that more informative contexts led to better word meaning inference when learners encountered words in isolated sentences (Webb 2008; Teng 2019). However, most contexts are not informative in connected texts, thus, explicit inferences are mostly unsuccessful. Yet, after being exposed to words in various contexts (and more importantly, *not* being exposed to them in other contexts, more on this later), most of which are uninformative, learners acquire a large number of words (Landauer and Dumais, 1997). The reason is that weak individual contexts, when combined, produce strong *inductive* effects. Landauer and Dumais (1997) used the analogy of how a diagonal brace increases the rigidity of a three-beam structure (see Figure 1). The diagonal brace connects all three beams, which is analogous to connecting three different contexts. Before the brace is installed, the angles formed between the beams have no constraints; the brace then forms a triangle that completely constrains the angles. Analogously, the data from each context alone provide weak constraints on the semantic representation of the target word; however, after connecting various contexts, one can establish a more specified representation of that word.

*Figure 1: Diagonal brace analogy for context synthesis.*



Another important proposition of the *distributional hypothesis* is that learning occurs not only through contexts where a word occurs, but also (more importantly) through contexts where that word does *not* occur (Landauer and Dumais 1997). According to the *distributional hypothesis*, one's semantic representation of a word is derived from global co-occurrence patterns between that word and all other words (Mandera *et al.* 2017; Günther *et al.* 2019); thus, '... data solely about other words, for example a text sample containing words Y and Z, but not X, can change the representation of X because it changes the representations of Y and Z, and all three must be accommodated in the same overall structure' (Landauer and Dumais 1997: 223). Landauer and Dumais used their LSA model to simulate the word learning of 7th-grade children and found that reading led to *indirect* learning of words that did *not* occur in the text and the learning rate was 0.15 words per paragraph, three times the rate of *direct* learning of words that did occur (0.05 words per paragraph). The combination of *direct* and *indirect* learning is equivalent to the estimated natural rate of word learning (1 word every 5 paragraphs) among higher-grade L1 children (Landauer and Dumais 1997: 222).

These findings suggest that all words in all contexts, regardless of those words' association strength with the target word, may contribute to the learning of the target word. Through reading one learns that some words co-occur frequently (i.e. higher association strength), some words could co-occur albeit less frequently (i.e. lower association strength), and some never co-occur (i.e. no association), etc. Each context provides data about which co-occurrence pattern is more or less frequent; the relative co-occurrence frequencies between the target word and all other words form the basis of incrementally deriving semantic representation of that word (Landauer and Dumais 1997; Mandera *et al.* 2017; Günther *et al.* 2019). Uninformative contexts contain many words that have lower association strength with the target word and provide data that these words co-occur

less frequently; these data are an indispensable part of the global co-occurrence patterns. The following example illustrates this point. One can learn that ‘surgeon’ and ‘psychiatrist’ are semantically *similar* after encountering both words in informative contexts that contain words like ‘hospital’ and ‘patient’. However, one also needs to learn that ‘surgeon’ and ‘psychiatrist’ still *differ*; this knowledge comes from the differences in their global co-occurrence patterns. For instance, ‘surgeon’ also frequently co-occurs with ‘scalpel’ and ‘suture’, but ‘psychiatrist’ co-occurs with them much less frequently. ‘He told the psychiatrist that he thought the suture would leave a scar’ is an uninformative context for ‘psychiatrist’; but it provides data that the co-occurrence of ‘psychiatrist’ and ‘suture’ is much less frequent than the co-occurrence of ‘surgeon’ and ‘suture’, which contributes to learning the difference between ‘psychiatrist’ and ‘surgeon’.

To summarize, contexts with varying degrees of informativeness could play a role in developing in-depth word knowledge by establishing a network of words with various degrees of association strengths. Synthesizing different contexts may facilitate word learning.

## PRESENT STUDY

### ***Context synthesis via reflash***

Inspired by the *distributional hypothesis*, we developed a text modification method called *reflash* to facilitate word learning through reading. In an e-text modified by *reflash*, each occurrence of a target word is followed by a left arrow ← and a right arrow →. By clicking the arrows, the learner can review or preview the previous or subsequent occurrences of words ‘in a flash’ (hence the name *reflash*). A word’s derivational forms are connected to their root word via *reflash* (e.g. ‘crash’ and ‘crashed’ share the same *reflash*).

We hypothesize that *reflash* may facilitate *word-context association*. By reviewing previous occurrences or previewing future occurrences of a word, learners can strengthen the associations between that word and multiple contexts, including its lexical contexts (i.e. co-occurring words) or sentence-, paragraph-, or text-level contexts. According to distributional semantic theory, the acquisition of semantic representations results from abstracting and generalizing from word-context co-occurrences, which reflects ‘the transition from episodic memory, capturing concrete instances of co-occurrence of entities, to semantic memory, capturing more fundamental, conceptual relations between them’ (Günther *et al.* 2019: 5). Therefore, *reflash* may facilitate learning by enhancing the episodic memory of local co-occurrence patterns, the first step of extracting the global distributional semantic representations.

We hypothesize that reflash may also facilitate explicit inference of word meaning via *context synthesis*. By synthesizing clues from multiple contexts, learners notice and actively process the co-occurrence, grammatical, and semantic information. Table 1 shows some *reflash* examples for ‘crash’ in the reading material used in this study and summarizes these three types of information about ‘crash’ that can be derived from each context.

Table 1: Context synthesis of ‘crash’

Reflash examples	Co-occurrence	Grammar	Meaning
his journey home, an <b>aeroplane</b> (飞机) from Sydney <b>crashed</b> → into the sea just south of Hong Kong.	aeroplane + crash	verb (past tense: crashed)	An action that can be done by an aeroplane
	crash + into + sea	non- transitive: + into	An action related to movement from one location to another

I heard about the plane <b>crash</b> $\Leftarrow \Rightarrow$ on television.	plane + crash	noun	An event related to plane
			An event that is big enough to be on television
'Seven years,' answered my mother. 'My husband died in a plane <b>crash</b> $\Leftarrow \Rightarrow$ last year, so we've	die + in a crash	countable	An event that can cause death

---

Table 1 demonstrates that the semantic properties, grammatical features, and co-occurrence patterns of a word are intertwined in contexts. For example, one can learn that ‘crash’ is ‘a movement from one location to another’ from the co-occurrences of ‘crashed’ + ‘into the sea’, which also reflects the grammatical feature of ‘crash’ as a non-transitive verb. By attending to the co-occurrence patterns and grammatical features, learners can glean information on the semantic properties. Therefore, whether any specific context is informative enough for learners to infer an exact word definition may not be crucial. The synergy of different contexts incrementally leads to “a rich and nuanced database about a word, its connections to other words and its lexical history within an individual’s experience” (K. Nation 2017: 2).

*Reflash* is analogous to the diagonal brace that stabilizes the three-beam structure in Figure 1. Using *reflash* to synthesize multiple contexts may bring out the power of contexts with varying degrees of informativeness to facilitate learning. *Reflash* may be especially valuable for learning over long periods. In natural reading, the occurrences of lower frequency words usually spread far apart; thus, the learners may have forgotten the previous encounter before they encounter the word again. Seeing the left *reflash* arrow reminds them that they have encountered the word before; using *reflash* to review the previous occurrences can reconsolidate the previously laid memory trace. By noticing, deeply processing, and encountering a word repetitively in diverse contexts via

*reflash*, learners may acquire in-depth knowledge of its co-occurrence, grammatical, and semantic features. Extant research also highlights the importance of noticing (Schmidt 2001), depth of processing (Laufer and Hulstijn 2001), and repetition (Webb 2007).

## Research questions

This study investigates the effect of *reflash* on vocabulary learning through reading among middle- and high-school English-as-a-Foreign-Language (EFL) learners in China. Participants read digital books embedded with target words and were tested on these words after reading. Target words were assigned to four conditions: *reflash*-only, gloss-only, gloss + *reflash*, and unmodified. *Reflash*-only provided the *reflash* arrows only. Gloss-only provided a parenthesized Mandarin translation (e.g. lightning (闪电)) only. Gloss + *reflash* provided both the Mandarin translation and *reflash* (e.g. lightning (闪电) <- ->). Unmodified means the word was not highlighted or modified.

Learning outcomes were examined by two measures:

- a. Vocabulary Knowledge Scale (VKS; Wesche and Paribakht 1996) measured knowledge of recognizing the word form and providing a Chinese translation.
- b. *Word-context association* questions asked students to recall at least one scene where the target word occurred in the story **or** provide at least one content word that co-occurred with that word in the story (see **Instruments**). According to the *distributional hypothesis*, the episodic memory of word-context co-occurrence is the basis for deriving semantic representations (Günther *et al.* 2019). Thus, *reflash may facilitate word learning by facilitating word-context association*.

The specific research questions are as follows:



1. Are *reflash*-only words scored higher on VKS than words in other conditions, controlling for the words' pretest VKS score, number of occurrences in the story, and type of existing knowledge?

2. Is the probability of establishing *word-context association* higher for *reflash*-only words than words in other conditions, controlling for the same covariates?

We controlled for the type of existing knowledge because the 'learning burden' or the effort required to learn a word depends on what is predictable from existing knowledge (P. Nation 2001: 36). For instance, if a new word contains a familiar morpheme, the learning burden decreases; but breaking the word down does not always lead to the correct meaning and sometimes may inhibit learning. Based on existing literature (see **Discussion**) and stimulated recall data (see **Method**), we summarized the types of existing knowledge in Table 2.

*Table 2: Types of existing word knowledge*

Type of existing knowledge	Definition
<i>Homonym</i>	knowing one meaning of a <i>homonym</i> (i.e., a word with multiple unrelated meanings) but not the correct meaning in the text
<i>Morphology</i>	knowing a word that is <i>morphologically</i> related to the target word (e.g., knowing 'dress' but not 'dressed')
<i>Etymology</i>	knowing a word that is <i>etymologically</i> related to the target word (e.g., knowing 'secret' but not 'secretary')
<i>Compound</i>	knowing part of a <i>compound</i> word but not the whole word (e.g., knowing 'dressing' but not 'dressing-gown')

*Uncrystallized**uncrystallized* memory of word meaning; can recall the meaning when encountering the word during reading but not in a decontextualized setting (e.g., word lists) or can only retrieve the meaning in some contexts but not others.

## METHOD

### Participants

Three Mandarin-speaking EFL learners from a city in China participated in this study, including two middle-school (Grades 7-9) students and one high-school (Grades 10-12) student. Table 3 summarizes their backgrounds.

*Table 3: Background of participants*

Participant	Gender	Age	Grade	# of years since 1 <sup>st</sup> English lesson*
B	Female	15	9	9
T	Female	15	9	6
X	Male	16	10	10

\* Formal English lessons in their schools started in Grade 3, but Participants B and X started taking private English lessons in Grade 1.

Note that because we used multilevel modeling (see **Data analysis**) with words as the unit of analysis (Baayen *et al.* 2008; Barr *et al.* 2013) and entered each individual as a fixed effect in our model, the sample size was the number of repeated measures for words (see **Supplementary materials** for data). Therefore, the statistical power of our models was not determined by the number of participants. Nevertheless, we acknowledge that the small number of participants

limited the external validity of our study; future larger-scale studies are necessary to examine the generalizability of our findings.

## Reading materials

The reading materials were from the Oxford Bookworms Graded Readers. To select the appropriate book level, we administered the Oxford Bookworms Level Tests (OBLTs; Oxford University Press 2017) and the Vocabulary Levels Test (VLT; P. Nation 1983) two weeks before the experiment. The OBLTs are multiple-choice cloze questions at different levels designed by the Bookworms publisher. Table 4 shows the participants' performance details.

*Table 4: VLT and OBLTs results*

Participant	VLT (1k level) *	VLT (2k level)	OBLTs (Starter)	OBLTs (Level 1)	OBLTs (Level 2)
B	90%	61%	29/30	29/30	28/30
T	70%	33%	27/30	25/30	18/30
X	85%	50%	28/30	27/30	23/30

\* The results were the average of the two versions of the 1k-level test

Participant B scored the highest on all measures, followed by X, and then T. Hence, the participants' OBLTs results corresponded with their vocabulary level. According to the official OBLTs guideline (Oxford University Press 2017), scoring 29-30 means that book level is 'too easy'; 24-28 means 'just right'; 0-23 means 'too difficult'. The results indicate that Level 2 was 'just right' for B, on the edge of 'just right' and 'too difficult' for X, and 'too difficult' for T.

Level-3 tests were not administered as per OBLTs guideline to administer tests from the starter level up to the ‘just right’ level.

Two Level-2 books were chosen as the reading material: *Dead Man’s Island* (Escott 2012) and *Dracula* (Stoker 2014). Even though Level 2 was below ‘just right’ for Participants T and X, the participants ended up having almost equal text coverage (i.e. the percentage of known words in the text) because we provided glosses as part of the experimental design. Counting glossed words and proper nouns as known words, the text coverage was 99%, 98.5%, and 98.75% for Participants B, T, and X respectively. Hence, glosses presumably made the text equally comprehensible across participants.

## Target words

One week before the experiment, participants were given a list of all the 701 unique words (excluding function words) in the reading material, and they rated each word against the Vocabulary Knowledge Scale (VKS; Wesche and Paribakht 1996):

- 1 = I don’t remember having seen this word before.
- 2 = I have seen this word before, but I don’t know what it means.
- 3 = I have seen this word before, and I think it means \_\_\_.
- 4 = I know this word. It means \_\_\_.
- 5 = I can use this word in a sentence, e.g. \_\_\_.

We focused on receptive knowledge and thus only used scales 1-4. Students who rated a word as 4 but provided an incorrect Chinese translation were scored as 3. We assigned seventy-two words rated as 1-3 by at least one participant to four conditions (*reflash*-only, gloss-only, gloss + *reflash*, unmodified), balancing the proportions of words with low and high frequency in the story

across conditions. Words that occurred only once were randomly assigned to gloss-only or unmodified, because *reflash* requires at least two occurrences.

## Instruments

Participants' learning outcomes were measured by VKS and a researcher-developed interview to probe *word-context association*. After reading the materials in each session, the participants rated the target words using VKS, and then answered whether they remembered seeing the word in the current or previous sessions. When they claimed to have seen the word, they were asked to recall that word's meaning in the story (by providing a Chinese translation), scenes where that word occurred, and other words that co-occurred. Figure 2 shows an example recording sheet.

If the participant recalled at least one scene (i.e. sentence-, paragraph-, or text-level context) or content word (i.e. lexical context), their *word-context association* score was 1, and 0 if they recalled neither.

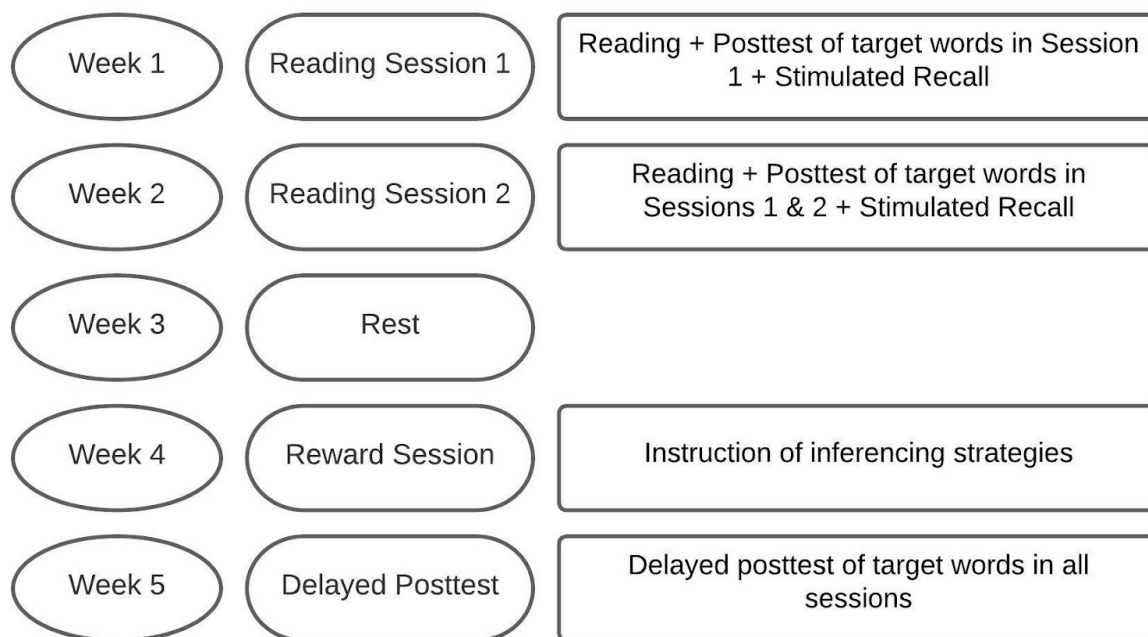
*Figure 2: Example posttest recording spreadsheet.*

Word	VKS	Chinese Translation	Have you seen it during this/previous reading session?	In what scenes of the story did the word occur?	What words appeared together with this word in the story?
coach	3	教练 [athlete coach]	√	教练带我去找大老板 [The coach took me to see the big boss]	
moustache	3	大金牙/坏嘴/口腔溃疡 [big golden teeth/bad mouth/ mouth ulcer]	√	Ross有; 揭穿秘密前, 不仅有还带眼镜 [Ross had it; before the secret was revealed; (he) not only had it but also wore glasses]	have/get moustache
poster	3	邮递员? [postman?]	√	有印象但想不起来在哪里 [I remember seeing it but cannot remember exactly where (in which scene)]	
dressing-gown	4	睡衣 [sleepwear]	√	穿上/脱下睡衣 [put on/take off]	turn off/put up

## Procedure

The experiment included two reading sessions, a reward session, and a delayed posttest session (see Figure 3).

Figure 3: Procedure flowchart.



In Reading Session 1, the participants read the first three chapters of *Deadman's Island* (3,332 words) on a laptop. A screen-recording software was running while they read the text. Their reading was followed by a posttest (detailed earlier) of the target words that had occurred. No feedback was provided. Afterwards, the screen recording video was shown to the participants for stimulated recall, which was recorded using the same software. We paused the video whenever a *reflash* arrow was clicked or skipped, and asked the participants the following questions: a. Why did/didn't you click this *reflash* arrow? b. (If they clicked the arrow) What were you thinking after

clicking the arrow and jumping to this part of the text? If you were trying to infer the word's meaning, how were you doing so? We also paused the video whenever a glossed word was encountered and asked if they noticed that word and its gloss. The stimulated recall video was transcribed, and a taxonomy of existing knowledge (described in Table 2) emerged through transcript coding. The same word could be of different types of existing knowledge for different participants.

In Reading Session 2 (Week 2), they read the last two chapters of *Deadman's Island* and the first chapter of *Dracula* (3,443 words). After that, they were given a posttest on all target words that had occurred (including those that only occurred in Session 1), followed by the same stimulated recall protocol as Session 1. Two weeks later, as a reward, we taught the participants how to better use contextual clues to infer word meaning when reading the last five chapters of *Dracula*; no tests were administered. Participant T dropped out after that due to schoolwork. In Week 5, Participants B and X completed a surprise delayed posttest on all target words. If a target word was given instruction in the reward session, we removed the observation of that word in the delayed posttest from the analysis (13 words for Participant X; 8 words for B).

## **Data analysis**

Each target word's VKS and *word-context association* scores at three posttest time points were the outcome variables. If a participant rated a word as VKS=4 (i.e., knowing the correct meaning) in the pretest, the observations of that word for that participant were excluded from the analysis. We ended up analyzing 25 words for Participant B, 50 for T, and 38 for X (i.e. 113 word-by-person combinations). We subtracted 1 from the original VKS ratings so that 1 was coded as 0, 2 was coded as 1, etc. We analyzed the data using Bayesian multilevel models (Gelman *et al.* 2013; Lee



*et al.* 2018) where observations at each time point are nested within each word. We used Bayesian multilevel modeling over its frequentist counterpart for its improved numerical stability in small samples and intuitive explanation for statistical inferences (e.g. Gelman and Hill 2006). The models were fitted using the *rstanarm* package (Goodrich *et al.* 2020) in R computing environment. Bayesian models report 95% credible intervals instead of p values. Significance is indicated by credible intervals not containing 0 (linear models) or 1 (logistic models).

The models examined the effect of different text modifications on the growth trajectory of VKS (Model 1) and *word-context association* (Model 2). The modification conditions were dummy coded with Unmodified as the reference group. Covariates include the word's frequency in each session, pretest VKS, and type of existing knowledge (dummy variables for *Homonym*, *Morphology*, *Etymology*, *Compound*, and *Uncrystallized*, with None (no existing knowledge) as the reference). We used fixed effects for the time points (dummy variables for Session 2 and Delayed posttest, with Session 1 as the reference) and the dummy-coded student identifier to account for possible unmeasured individual-level heterogeneity. For both models, random intercepts were included for each word; for Model 1, random by-word slopes were assumed for the fixed effects of time points, with correlation between random intercepts and random slopes (Baayen *et al.* 2008; Barr *et al.* 2013). Model convergence was ensured using Rhat (Gelman *et al.* 2015). A satisfied level of the goodness of model fit was evidenced by applying multiple model checking techniques (see **Supplementary materials**).

## RESULTS

### **Question 1: The effect of different conditions on VKS scores**

To answer Question 1, we used a Bayesian multilevel linear model, as shown in the left columns of Table 5.

Table 5: Posterior summaries of parameters in Bayesian multilevel models

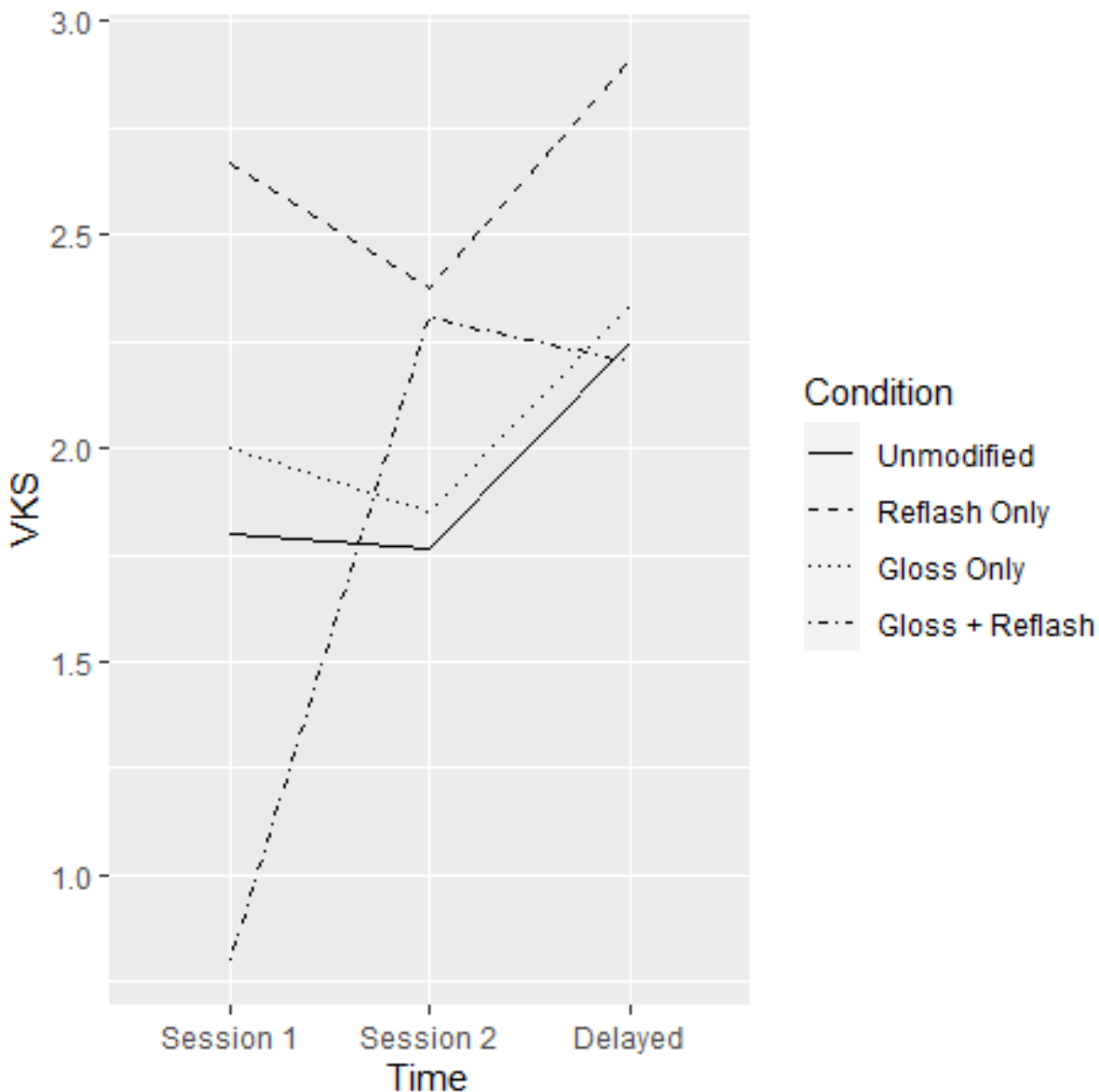
Predictors	Model 1 (VKS)		Model 2 (Association)	
	Coefficient	Credible Interval	Odds Ratio	Credible Interval
(Intercept)	0.37	-0.07 – 0.82	0.01	0.00 – 0.11
Pretest VKS	0.04	-0.19 – 0.28	1.03	0.37 – 2.91
<i>Reflash</i> -only	0.46	0.01 – 0.91	11.17	2.08 – 71.19
Gloss-only	0.21	-0.15 – 0.56	1.00	0.24 – 4.44
Gloss + <i>reflash</i>	0.24	-0.21 – 0.69	1.60	0.25 – 9.51
<i>Homonym</i> words	0.46	-0.00 – 0.89	7.59	1.51 – 50.17
<i>Morphology</i> words	0.67	0.13 – 1.25	1.81	0.24 – 13.92
<i>Etymology</i> words	0.43	0.05 – 0.82	2.65	0.56 – 12.76
<i>Compound</i> words	1.01	0.52 – 1.48	9.67	1.30 – 87.96
<i>Uncrystallized</i> words	0.98	0.57 – 1.40	0.73	0.12 – 4.40
Session frequency	0.08	-0.00 – 0.16	1.05	0.78 – 1.45
Participant B	0.52	0.19 – 0.87	15.75	3.63 – 80.79
Participant X	0.81	0.51 – 1.10	9.34	2.48 – 42.69
Session 2	0.48	0.22 – 0.75	4.86	1.45 – 16.97
Delayed posttest	0.81	0.37 – 1.24	3.60	0.83 – 16.94

Note. See **Supplementary Table S1** for random effects

This model shows significant effects of *reflash*-only on VKS scores. The mean VKS is estimated to be 0.46 points higher for *reflash*-only words than for unmodified words, controlling for pretest VKS and all other covariates. The probability is 95% that the true effect is within the

credible interval [0.01, 0.91], which does not contain 0. In contrast, controlling for all other covariates, the mean VKS is estimated to be 0.21 points higher for gloss-only than for unmodified, but the 95% credible interval contains 0. Similarly, the mean VKS is estimated to be 0.24 points higher for gloss + *reflash* than for unmodified, but the effect is also not significant. To summarize, *reflash*-only led to higher VKS than unmodified, but gloss-only and gloss + *reflash* may not have. Figure 4 shows the average VKS for words in different conditions across three posttests.

*Figure 4: Graphs of average VKS for words in different conditions across three posttests (at the end of Reading Sessions 1 & 2, and delayed posttest).*



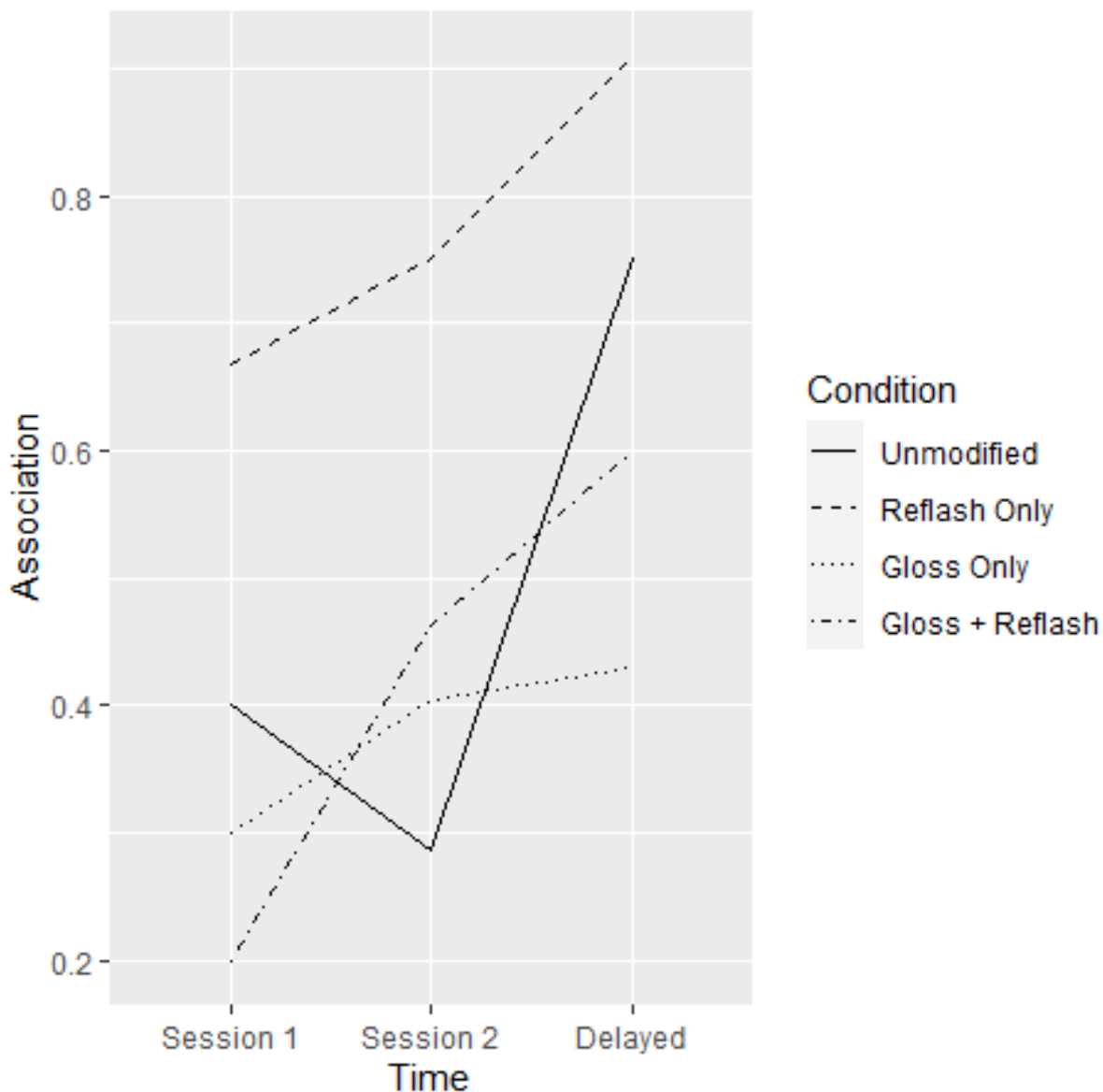
Note. We subtracted 1 from the original VKS ratings in data analysis. Thus, 3 on the graph means 4 in the original VKS scale.

### **Question 2: The effect of different conditions on *word-context association***

To answer Question 2, we used a Bayesian multilevel logistic model, as shown in the right columns of Table 5. The odds ratio of *reflash-only* shows that the estimated odds of successfully recalling at least one scene or co-occurring word were 10.17 times higher for *reflash-only* than for

unmodified, controlling for all other covariates. The 95% credible interval [2.08, 71.19] does not contain 1. In contrast, the credible intervals for gloss-only and gloss + *reflash* contain 1. To summarize, *reflash*-only facilitated *word-context association* whereas gloss-only and gloss + *reflash* may not have. Figure 5 shows the proportion of correct *word-context association* for words in different conditions across three posttests.

*Figure 5: Graphs of the proportion of correct word-context association for words in different conditions across three posttests (at the end of Reading Sessions 1 & 2, and delayed posttest).*



## DISCUSSION

### **Efficiency of *context synthesis* via *reflash***

We found that words modified with *reflash*-only were better learned than other conditions, controlling for all covariates in the model. Students improved more on VKS and had higher probability of recalling at least one context for words modified with *reflash*-only.

We hypothesized that *reflash* may strengthen *word-context association*, which may, in turn, explain why *reflash* words made the most progress on VKS. The stimulated recall reveals that *reflash* brought learners' attention to the target word and reminded them to think about the previous contexts even without jumping to those contexts. Therefore, *reflash* may have led to deeper engagement with the word and contexts, resulting in better episodic memory and retention of word meaning. However, whether *word-context association* is a causal mediator between *reflash* and VKS is a question for future research.

We also hypothesized that *reflash* may facilitate word learning via *context synthesis*. The stimulated recall shows that the participants tried to synthesize clues from different contexts via *reflash*. For instance, participant T inferred that “(‘secretary’) is a type of job in companies” when she encountered the word ‘secretary’ in ‘My father was a businessman there and my mother worked as a secretary’. She then used *reflash* to read each occurrence of ‘secretary’ and tried to think of as many types of company jobs as possible, checking whether they fit each context (e.g. ‘And I needed help with my work. I needed a good secretary’). She made several close guesses, including ‘assistant’.

*Reflash* may also be particularly valuable in the following situation: The learner inferred the correct meaning for a word in an informative context; after a period, they re-encountered the word in a less informative context. If the learner uses *reflash* to reread the previous context, they may retrieve their previous inference and synthesize both contexts for better learning. For instance, after encountering ‘crash’ in four informative contexts (see Table 1) in Session 1, Participant T improved from VKS-1 to VKS-4. However, her VKS dropped to 2 in Session 2, after another five exposures. During the stimulated recall, she reported that she found the word’s contexts in Session 2 unhelpful (e.g. ‘I faked the car crash’), and she did not realize that she could use *reflash* to jump

to the contexts in Session 1. She then proceeded to use *reflash* and successfully recovered her previous inference.

Our results echo studies on corpus-based learning methods such as concordance lines. Learners can also use the various contexts in concordance lines to infer word meaning and compare word usage across contexts. Several meta-analyses showed positive effects of corpus-based methods on L2 vocabulary learning (Cobb and Boulton 2015; Boulton and Cobb 2017; Lee *et al.* 2019). Despite the similarity, *reflash* has the following potential advantages. Firstly, in concordance activities, learners only see incomplete sentences. In contrast, learners use *reflash* while reading connected texts, so they are aware of the larger context beyond the sentence and can use it to make more informed inferences. Additionally, learners may have an easier time recalling the contexts from stories they read than random contexts from a corpus; being able to recall contexts (i.e. retrieve episodic memory) may facilitate learning because learners sometimes retrieve word knowledge by recalling the contexts where they saw the word (see next section). Lastly, students usually study concordance lines intensively in one setting (i.e. massed repetition). In comparison, learning spreads across a long period when one uses *reflash* during reading so that spaced repetition (Baddeley 1990) occurs naturally, leading to better learning than massed repetition.

### **How words of different types of existing knowledge may benefit from *reflash***

As in Table 2, we derived five types of existing knowledge from the stimulated recall, including *homonym*, *etymology*, *morphology*, *compound*, and *uncrystallized*.



*Homonym* A *homonym* is a word with multiple *unrelated* meanings (e.g. ‘pen’ and ‘bat’). We distinguish *homonyms* from *polysemous* words, which have multiple *related* senses (e.g. ‘chicken’ as both an animal and its meat).

We hypothesize that learners may benefit from *reflash* when learning a homonym’s new unrelated meaning. For instance, for the unmodified word ‘coach’, all participants knew its ‘athlete trainer’ meaning in the pretest but failed to infer the ‘carriage’ meaning in the story. According to the stimulated recall, they found the ‘athlete trainer’ meaning fitted in the first sentence where the word occurred (‘I had six hours to wait before the coach came to take me there’). However, if *reflash* had been provided and the learners had used *reflash* to synthesize clues from other contexts (e.g. ‘When the coach arrived and I got into it’), they could have realized that ‘athlete trainer’ no longer fits and tried to infer a new meaning.

Our observation echoes previous studies where both L1 and L2 learners experience interference from the known meaning of homonyms when learning an unrelated second meaning. L1 children tended to disregard the context and cling to the known meaning of homonyms when they initially encountered unrelated novel meanings (Mazzocco, 1997; Casenhiser 2005). An event-related-potentials (ERPs) study on L2 learners found that learning the second meaning of homonyms induced higher negative N400 than learning the first meaning, indicating that the activation of the first meaning interferes with the acquisition of the second meaning (Zhang et al. 2020).

Because the types of existing knowledge only emerged after we analyzed the stimulated recall, no homonyms were in the *reflash* condition. Future studies that directly manipulate types of existing knowledge as an independent variable in experimental design are needed to confirm whether *reflash* indeed facilitates homonym learning.

*Etymology* Extant research indicates that a novel word is easier to learn if part of that word is known and the whole word's meaning is *related* to the known word part's meaning (P. Nation 2001; Vidal 2011). For example, knowing 'cup' and 'board' is conducive to learning 'cupboard'. In contrast, when the novel word is composed of word parts whose meanings are *unrelated* to the whole word, the existing knowledge of the word parts may inhibit learning the new word (Bensoussan and Laufer 1984). For instance, learners may assume that 'nevertheless' means 'never less'. Even when learners successfully infer the meaning of 'nevertheless' *during* reading, the activation of 'never' and 'less' may still interfere with recalling the correct inference *later* in other contexts or vocabulary tests, because their knowledge is not fully crystallized yet.

When a novel word is *etymologically* related to a known word, these two words are usually no longer *semantically* related. For instance, 'secretary' meant 'person entrusted with secrets' when first coined; however, as 'secretary' gained its modern meaning, the transparent relationship between 'secret' and 'secretary' was lost. When learners see 'secretary', the activation of 'secret' could be counterproductive to learning 'secretary'. Using *reflash*, Participant T made roughly correct inferences of 'secretary' during reading (e.g. 'company job' and 'assistant'). However, when asked what 'secretary' meant in the immediate posttest, she answered 'secret', indicating that the activation of 'secret' interfered with retrieving her inferences. However, when asked to recall what words co-occurred with 'secretary', she recalled the phrase 'work as a secretary' and then recalled the 'company job' meaning, indicating that *reflash* may be valuable for this type of words by helping learners notice and process co-occurrence patterns; the episodic memory of contexts can be used for retrieving the newly acquired semantic knowledge. Thus, *reflash* may

help learners overcome the interference from a known word when learning a novel word that is etymologically related but not semantically related to that known word.

Another illuminating case is ‘moustache’, which is etymologically related to ‘upper lip’ and ‘mouth’ (Doric words ‘mystax’ and ‘mastax’). In this case, the novel word is weakly related to the known word part. According to the stimulated recall, all participants inferred that ‘moustache’ should be related to ‘mouth’, based on the first four letters ‘mous’ and the contexts (e.g. ‘You've got short hair, you've got a moustache now, and you wear glasses’). Using *reflash*, Participant T made several inferences, such as mouth ulcer, decayed teeth, and beard. Participant B inferred that it probably meant something bad on one's face, such as scars. Their inferences are all negative things related to the mouth, indicating that they considered the larger story context that the character is an anti-social recluse. Both participants could recall their inferences and at least one context in the posttest.

*Morphology* We coded the type of existing knowledge as *morphology* if the learner knew a word that is morphologically related to the target word (e.g. knowing ‘dress’ but not ‘dressed’). *Reflash* could potentially be valuable when the target word is an irregular variation of a known word. For instance, Participant T knew ‘shake’ but not ‘shook’ in the pretest; when encountering the unmodified ‘shook’ in the story, she did not know that ‘shook’ was related to ‘shake’. If *reflash* had been provided to connect ‘shook’ with occurrences of ‘shake’, she might have realized that ‘shook’ was the past tense form of ‘shake’. However, because no *morphology* words were in the *reflash* condition, this hypothesis needs to be tested by future research.

*Compound* Compound words may be easier to learn when at least one word part is known. Using *reflash*, all participants inferred the correct meaning of ‘dressing-gown’ by combining the contextual clues (e.g. ‘I took off my dressing gown and went back to bed’) and their knowledge of ‘dress’, even though they did not know what ‘gown’ meant.

*Uncrystallized* We categorize a word as *uncrystallized* if the learner had acquired partial knowledge of the word before the experiment, including the following two scenarios.

The first scenario where reflash may be helpful is when learners can recall a word’s meaning in some contexts but not others. For example, during the stimulated recall in Session 1, Participant T recalled the meaning of ‘corner’ in ‘I walked along the passage and turned a corner. Then I saw the door at the end of the passage,’ and reported that she had learned this word at school. However, when the word re-occurred in Session 2 in ‘I walked along the passage and turned the corner. There it was, the locked room,’ she failed to recall its meaning. After returning to the Session-1 context using *reflash*, she recalled what ‘corner’ meant.

The second scenario is when learners can recall the meaning of a word when encountering that word during reading but not in a decontextualized setting (e.g. word lists). For example, in the pretest, Participant T reported that she had difficulty distinguishing ‘pleased’ from ‘honored’. However, she immediately remembered that ‘pleased’ meant ‘happy’ when reading the story and used *reflash* to double check. She retained this knowledge in both the immediate and delayed posttests.

### **Inefficiency of gloss-only and gloss + *reflash***

Although prior studies found that gloss facilitated vocabulary learning (e.g. Hulstijn *et al.* 1996; Ko 2012; Yanagisawa *et al.* 2020), we did not find statistically significant evidence that gloss-only or gloss+reflash led to more improvement on VKS score and *word-context association* than the unmodified condition, controlling for pretest VKS and the other covariates.

A possible reason is that the participants did not process words deeply when the word meaning was already provided. The presence of gloss even caused learners to ignore words entirely sometimes, which may be particularly problematic for homonyms. For instance, all participants knew that ‘passage’ meant ‘article’ in the pretest. However, ‘passage’ meant ‘corridor’ in the story. In the stimulated recall of Session 1 where ‘passage’ occurred twice, all participants reported that they did notice the gloss ‘(走廊)’ but ignored the word ‘passage’ itself. Therefore, in the posttest, they did not remember seeing the word and cannot provide its correct meaning in the story. ‘Passage’ occurred another six times in Session 2; Participants T and X succeeded in recalling the correct meaning in the posttest, whereas participant B was still stuck at the ‘article’ meaning and could not recall any context. Model 2 also shows that participants were, indeed, no more likely to recall a scene or co-occurring word for glossed words than for unmodified words. An early study (Holley and King 1971) also reported that learners sometimes ignored glossed words because the meanings were provided directly, so they assumed that they were not expected to learn these words.

Surprisingly, gloss+reflash did not improve word learning. Analyzing the quantitative data and stimulated recall, we found that similar to gloss-only, the gloss in this condition caused learners to ignore the words in many cases; consequently, they also did not use *reflash*. The experiment would have been more informative if we had included another group who were forced to use *reflash* and compared the voluntary vs. forced group. Our finding does not suggest that facilitating learning via a gloss+reflash combination is impossible. A better method may be

providing gloss *only if* learners have used *reflash* and *only after* the learners have finished reading the passage. This method may encourage students to use *reflash* and then use gloss to check their inferences and reinforce their learning.

Our within-subject design may also partly explain why we found different results than previous studies that adopted between-subject design. In between-subject studies, glossed words are the most salient words, so learners are more likely to notice them. In contrast, our participants may have assumed that glossed words were less important than *reflash* words because of the novelty of *reflash*, so they ignored most glossed words. Therefore, future research should investigate in what circumstances learners may ignore glossed words and how to increase learners' attention to the word itself besides noticing the gloss.

## IMPLICATIONS AND LIMITATIONS

We introduced the theory of distributional semantics to the field of second language acquisition. Drawing on the *distributional hypothesis*, we critically discussed the role of context informativeness on word learning and the importance of encountering words in contexts of varying degrees of informativeness. We also provided initial evidence that the powerful associative learning postulated by this theory may be applicable to L2 learners, and that *reflash*, a tool inspired by this theory, can facilitate vocabulary learning. Using *reflash*, students may engage with contexts more deeply via *context synthesis* to better infer word meaning and establish stronger *word-context association*, which is fundamental to the incremental associative learning of distributed semantic representations. *Reflash* may be useful for strengthening uncrystallized word knowledge, learning additional meanings of homonyms, and retrieving knowledge of newly learned words by recalling

contexts. In contrast, in-text gloss may cause students to only pay attention to the gloss and ignore the words under some circumstances, which may be especially problematic for learning homonyms.

Because we did not force the learners to use *reflash* or use it in any specific way, our findings suggest that *reflash* may motivate learners to *voluntarily* process words more deeply, leading to better learning. Therefore, *reflash* has potentially important practical implications. For instance, eBook devices like Kindle currently have a Word Wise function that provides in-text gloss for difficult words; embedding *reflash* for these words may enhance learning outcomes.

Despite the promising findings, the present study has its limitations. The three low-intermediate proficiency participants are not representative of all EFL learners, so the external validity of this study is limited. Our findings only provide initial evidence that *reflash* may have merits for some learners; the generalizability of these findings should be examined by larger-scale studies. Furthermore, the participants had different proficiency levels; even though the gloss almost equalized the proportion of known words across the participants, they may have still differed on other skills related to reading and word learning, such as syntactic knowledge and depth of word knowledge. Even though we held unobserved individual-level characteristics constant statistically, future research may recruit more participants to explore how specific individual-level characteristics affect learning and interact with word-level characteristics.

The target words from two graded readers are not representative of all words that learners need to learn. The types of existing word knowledge were analyzed post-hoc when they emerged from transcript coding. Thus, our discussion of how words of different types of existing knowledge may have benefitted from *reflash* is exploratory. Future studies should manipulate this variable directly in the experimental design and include more words to examine statistically whether the effect of *reflash* differs by types of existing knowledge (e.g. whether *reflash* is more effective for

homonyms than compound words). A larger sample of words would also enable the analysis of more word-level factors (e.g. imageability and valence).

Another limitation is that we did not control for context informativeness. However, operationalizing context informativeness is difficult in studies that investigate word learning through reading long passages. While reading long passages, learners may rely on different kinds of contexts to infer word meaning. Some may only rely on the adjacent words within the sentence, others may seek for clues in the paragraph or page. The same person may sometimes only rely on adjacent words but consider larger contexts at other times. In other words, there are both interindividual and intraindividual differences in how learners use contexts. Therefore, similar studies typically did not control for context informativeness (e.g. Nagy et al. 1985, 1987; Horst et al. 1998); some studies excluded words whose contexts were judged to be uninformative by the researchers (Pellicer-Sánchez and Schmitt 2010). More importantly, as we discussed, contexts of varying informativeness levels may contribute to word learning. Future research is needed to overcome these methodological challenges and account for the nuanced effect of context informativeness in both experimental design and statistical analysis.

Even though we emphasized the potential importance of implicit associative learning, implicit semantic knowledge was not measured directly. Future studies can use psycholinguistic measures such as primed lexical decision tasks and semantic relatedness judgment tasks. For instance, if the target word has a priming effect on the lexical decision of a semantically related word, then the learner may have acquired at least some semantic representation of the target word. Future research can also explore whether *word-context association* mediates the effect of *reflash* on both implicit and explicit word knowledge.



## REFERENCES

- Anderson, R. C.** and **W. E. Nagy.** 1992. 'The vocabulary conundrum,' *American Educator* 16: 14–8, 44–6.
- Baayen, R. H., D. Davidson,** and **D. Bates.** 2008. 'Mixed-effects modeling with crossed random effects for subjects and items,' *Journal of Memory and Language* 59/4: 390-412.
- Baayen, R. H., P. Milin, D. F. Đurđević, P. Hendrix,** and **M. Marelli.** 2011. 'An amorphous model for morphological processing in visual comprehension based on naive discriminative learning,' *Psychological Review* 118/3: 438-81.
- Baddeley, A.** 1990. *Human Memory*. Lawrence Erlbaum Associates.
- Baroni, M., G. Dinu,** and **G. Kruszewski.** 2014. 'Don't count, predict! A systematic comparison of context-counting vs context-predicting semantic vectors'. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics Vol. 1*, pp. 238-47.
- Barr, D. J., R. Levy, C. Scheepers,** and **H. J. Tily.** 2013. 'Random effects structure for confirmatory hypothesis testing: Keep it maximal,' *Journal of Memory and Language* 68/3: 255–78.
- Bensoussan, M.** and **B. Laufer.** 1984. 'Lexical guessing in context in EFL reading comprehension,' *Journal of Research in Reading* 7/1: 15-32.
- Boulton, A.** and **T. Cobb.** 2017. 'Corpus use in language learning: A meta-analysis,' *Language Learning* 67/2: 348-93.
- Caliskan, A., J. J. Bryson,** and **A. Narayanan.** 2017. 'Semantics derived automatically from language corpora contain human-like biases,' *Science* 356/6334: 183-86.
- Casenhiser, D. M.** 2005. 'Children's resistance to homonymy: An experimental study of pseudohomonyms,' *Journal of Child Language* 32/2: 319-43.

- Cobb, T.** 2007. 'Computing the vocabulary demands of L2 reading,' *Language Learning and Technology* 11/3: 38–63.
- Cobb, T.** and **A. Boulton.** 2015. 'Classroom applications of corpus analysis' in R. Biber and R. Reppen. (eds): *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press, pp. 478-97.
- Cobb, T.** and **M. Horst.** 2019. 'Bringing home the word,' *Canadian Modern Language Review* 75/4: 285-98.
- Cunningham, A. E.** 2005. 'Vocabulary growth through independent reading and reading aloud to children' in E. H. Hiebert and M. L. Kamil (eds): *Teaching and Learning Vocabulary: Bringing Research to Practice*. Routledge, pp. 45-68.
- Day, R. R.** and **J. Swan.** 1998. 'Incidental learning of foreign language spelling through targeted reading,' *TESL Reporter* 31/1: 1-9.
- Ender, A.** 2016. 'Implicit and explicit cognitive processes in incidental vocabulary acquisition,' *Applied Linguistics* 37/4: 536-60.
- Escott, J.** 2012. *Dead Man's Island (Level 2 Oxford Bookworms Library)*. Oxford University Press.
- Firth, J. R.** 1957. *Papers in Linguistics, 1934–1951*. Oxford University Press.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin.** 2013. *Bayesian Data Analysis*. CRC Press.
- Gelman, A.** and **J. Hill.** 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.
- Gelman, A., D. Lee, and J. Guo.** 2015. 'Stan: A probabilistic programming language for Bayesian inference and optimization,' *Journal of Educational and Behavioral Statistics* 40/5: 530-43.

- Goodrich, B., J. Gabry, I. Ali, and S. Brilleman.** 2020. *rstanarm: Bayesian Applied Regression Modeling via Stan*. R package version, 2(1).
- Günther, F., C. Dudschig, and B. Kaup.** 2016. ‘Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies,’ *The Quarterly Journal of Experimental Psychology* 69/4: 626–53.
- Günther, F., L. Rinaldi, and M. Marelli.** 2019. ‘Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions,’ *Perspectives on Psychological Science* 14/6: 1006-33.
- Harris, Z. S.** 1954. ‘Distributional structure,’ *Word* 10/2-3: 146-62.
- Holley, F. M. and J. K. King.** 1971. ‘Vocabulary glosses in foreign language reading materials,’ *Language Learning* 21/2: 213-9.
- Horst, M., T. Cobb, and P. Meara.** 1998. ‘Beyond a *Clockwork Orange*: Acquiring second language vocabulary through reading,’ *Reading in a Foreign Language* 11/2: 207-23.
- Hulstijn, J. H.** 1992. ‘Retention of inferred and given word meanings: Experiments in incidental vocabulary learning’ in P. J. L. Arnaud and H. Bejoint (eds): *Vocabulary and Applied Linguistics*. Macmillan, pp. 113-25.
- Hulstijn, J. H., M. Hollander, and T. Greidanus.** 1996. ‘Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words,’ *The Modern Language Journal* 80/3: 327-39.
- Ko, M. H.** 2012. ‘Glossing and second language vocabulary learning,’ *TESOL Quarterly* 46/1: 56-79.
- Landauer, T. K.** 2007. ‘LSA as a theory of meaning’ in T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch (eds): *Handbook of Latent Semantic Analysis*. Erlbaum, pp. 3–34.

- Landauer, T. K.** and **S. T. Dumais.** 1997. 'A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge,' *Psychological Review* 104/2: 211-40.
- Laufer, B.** 2005. 'Focus on form in second language vocabulary learning,' *EUROSLA Yearbook* 5/1: 223–50.
- Laufer, B.** and **J. Hulstijn.** 2001. 'Incidental vocabulary acquisition in a second language: The construct of task-induced involvement,' *Applied Linguistics* 22/1: 1-26.
- Laufer, B.** and **G. C. Ravenhorst-Kalovski.** 2010. 'Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension,' *Reading in a Foreign Language* 22/1: 15-30.
- Lee, H., M. Warschauer,** and **J. H. Lee.** 2019. 'The effects of corpus use on second language vocabulary learning: A multilevel meta-analysis,' *Applied Linguistics* 40/5: 721-53.
- Lee, J., Sim, N., Ji, F.,** and **Rabe-Hesketh, S.** 2018. 'Introduction to multilevel modeling using Rstanarm: A tutorial for education researchers,' *Stan Case Studies* 5/1: 1-27.
- Lin, C.** and **D. Hirsh.** 2012. 'Manipulating instructional method: The effect on productive vocabulary use' in D. Hirsh (ed): *Current Perspectives in Second Language Vocabulary Research*. Peter Lang, pp. 117–48.
- Lund, K.** and **C. Burgess.** 1996. 'Producing high-dimensional semantic spaces from lexical co-occurrence,' *Behavior Research Methods, Instruments, & Computers* 28/2: 203–8.
- Mandera, P., E. Keuleers,** and **M. Brysbaert.** 2017. 'Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation,' *Journal of Memory and Language* 92: 57-78.
- Mazzocco, M. M.** 1997. 'Children's interpretations of homonyms: A developmental study,'

*Journal of Child Language* 24/2: 441-67.

- Mikolov, T., K. Chen, G. Corrado, and J. Dean.** 2013. 'Efficient estimation of word representations in vector space,' arXiv doi: 1301.3781 [cs], 7 Sep 2013, preprint: not peer reviewed.
- Mikolov, T., I. Sutskever, K. Chen, G. Corrado, and J. Dean.** 2013. 'Distributed representations of words and phrases and their compositionality'. *Advances in Neural Information Processing Systems Vol. 26*, pp. 3111–9.
- Mikolov, T., W.-T. Yih, and G. Zweig.** 2013. 'Linguistic regularities in continuous space word representations'. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746–51.
- Nagy, W. E., R. C. Anderson, and P. A. Herman.** 1987. 'Learning word meanings from context during normal reading,' *American Educational Research Journal* 24/2: 237-70.
- Nagy, W. E., P. A. Herman, and R. C. Anderson.** 1985. 'Learning words from context,' *Reading Research Quarterly* 20/2: 233-53.
- Nation, K.** 2017. 'Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill,' *npj Science of Learning* 2/1: 1-4.
- Nation, P.** 1983. 'Testing and teaching vocabulary,' *Guidelines* 5: 12-25.
- Nation, P.** 2001. *Learning Vocabulary in Another Language*. Cambridge University Press.
- Oxford University Press.** 2017. 'Oxford Bookworms and Dominoes Level Tests', retrieved April 2017, from <https://elt.oup.com/student/readersleveltest/>.
- 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.
- Pellicer-Sánchez, A. and N. Schmitt.** 2010. 'Incidental vocabulary acquisition from an authentic

novel: Do things fall apart?,' *Reading in a Foreign Language* 22/1: 31-55.

- Rescorla, R. A.** and **A. R. Wagner.** 1972. 'A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement' in A. H. Black and W. F. Prokasy (eds): *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, pp. 64–99.
- Schmidt, R.** 2001. 'Attention' in P. Robinson (ed): *Cognition and Second Language Instruction*. Cambridge University Press, pp. 3–32.
- Schmitt, N.** 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Springer.
- Sommerauer, P.** and **A. Fokkens.** 2018. 'Firearms and tigers are dangerous, kitchen knives and zebras are not: Testing whether word embeddings can tell'. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pp. 276–86.
- Sternberg, R. J.** 1987. 'Most vocabulary is learned from context' in M. G. McKeown and M. E. Curtis (eds): *The Nature of Vocabulary Acquisition*. Lawrence Erlbaum Associates, pp. 89-105.
- Stoker, B.** 2014. *Dracula (Level 2 Oxford Bookworms Library)*. Oxford University Press.
- Teng, F.** 2019. 'The effects of context and word exposure frequency on incidental vocabulary acquisition and retention through reading,' *The Language Learning Journal* 47/2: 145-58.
- Vidal, K.** 2011. 'A comparison of the effects of reading and listening on incidental vocabulary acquisition,' *Language Learning* 61/1: 219-58.
- Webb, S.** 2007. 'The effects of repetition on vocabulary knowledge,' *Applied Linguistics* 28/1: 46-65.
- Webb, S.** 2008. 'The effects of context on incidental vocabulary learning,' *Reading in a Foreign Language* 20/2: 232-45.

- Wesche, M. B.** and **T. S. Paribakht.** 1996. 'Assessing second language vocabulary knowledge: Depth versus breadth,' *Canadian Modern Language Review* 53/1: 13-40.
- Wesche, M. B.** and **T. S. Paribakht.** 2009. *Lexical Inferencing in a First and Second Language: Cross-Linguistic Dimensions.* Multilingual Matters.
- Yanagisawa, A., S. Webb,** and **T. Uchihara.** 2020. 'How do different forms of glossing contribute to L2 vocabulary learning from reading?: A meta-regression analysis,' *Studies in Second Language Acquisition* 42/2: 411-38.
- Zhang, Y., Y. Lu, L. Liang,** and **B. Chen.** 2020. 'The effect of semantic similarity on learning ambiguous words in a second language: An event-related potential study,' *Frontiers in Psychology* 11/1633: 1-11.