Articles

Learning Multiword Expressions with Flashcards: Deliberate Learning and L2 Implicit Knowledge Gains

Andrew Obermeier
Kyoto University of Education

This research investigated two aspects of second language learning: how implicit knowledge develops through explicit learning and how this is affected by multiword expression compositionality. More specifically, the experiment investigated how flashcard learning affected the implicit knowledge development of literal and figurative expressions. As these two types are composed differently, it was hypothesized that their implicit knowledge development would likewise differ. A lexical decision task was conducted in a masked repetition priming experiment to measure implicit knowledge gains, and response time data were analyzed in a linear mixed-effects model with participants and items set as random effects. Results showed that flashcard learning affected the implicit knowledge development of figurative and literal expressions differently.

Keywords: explicit learning; flashcards; implicit knowledge; interface; multiword expressions

本研究では、第二言語学習の2つの側面である、複単語表現の構成性と、明示的学習を通じて暗示的知識がどのように発達するかについて調査した。具体的には、フラッシュカードによる学習が、文字通りの表現と比喩表現の暗示的知識の発達にどのような影響を与えるかを調査した。この2つの表現は構成が異なるため、暗示的知識の発達も同様に異なるという仮説を立てた。暗示的知識の獲得を測定するために、マスク下の反復プライミング法を用いた実験で、語彙

https://doi.org/10.37546/JALTJJ45.1-1

JALT Journal, Vol. 45, No. 1, May 2023

性判断課題を実施し、応答時間データを、参加者と項目をランダム効果として設定した線形混合効果モデルで分析した。その結果、フラッシュカードによる学習は、比喩表現と文字通りの表現の暗示的知識の発達に異なる影響を与えることが示された。

キーワード: フラッシュカード、明示的学習、暗示的知識、複単語表現

Corpus linguists have found that *multiword expressions* (MWEs) make up about 59% of spoken and 52% of written English (Erman & Warren, 2000), so an essential issue for second language learners and teachers is understanding how they are acquired. Although various terms are used to refer to them (e.g., *formulaic sequences*, *chunks*, *collocations, idioms, conventional expressions*), this paper uses MWEs as an umbrella term covering all types of expressions (Siyanova-Chanturia, 2017). This research focused on two broad MWE categories: *figurative* and *literal expressions* (or *figuratives* and *literals*). The opaque meanings of figuratives (e.g., *kick the bucket, once in a blue moon*) make them more challenging to learn and process than literals (e.g., *all the time, get the idea*), which are transparent. As literals and figuratives are composed differently, the investigation focused on whether learning them is likewise different.

Another important issue is the intersection of explicit learning and implicit knowledge development because a high priority for language teachers is to foster these two processes for students. Explicit learning activities are conscious processes such as interpreting textbook explanations, doing worksheet exercises, practicing with drills, and rote memorizing. Explicit knowledge can be applied to monitoring language correctness or incorrectness and is often the focus of tests. Implicit knowledge develops unconsciously as the interlanguage system becomes fine-tuned through use, by which learned language can be accessed more fluently. Second language learners must learn explicitly and develop implicit knowledge to become proficient.

This study reports on a masked, repetition, priming experiment that compared the effects of learning literals and figuratives using flashcards. As this is an explicit learning method, and as masked repetition priming measures implicit knowledge development, the investigation addresses the interface regarding these two MWE types.

Multiword Expressions and Their Compositionality

A central issue to research on the processing of MWEs is that they vary widely regarding their *compositionality*, the degree to which the individual words that comprise them make up the meaning of the whole expression. The composition of MWEs varies along a continuum. Although some are transparent (i.e., *stay away*), others are less transparent but easy to process (i.e., *on the road*), and others are opaque (i.e., *once in a blue moon*). Grant and Bauer (2004) established major compositional categories showing how literal and figurative expressions generate meaning differently. The meanings of individual words in figuratives differ from those of the whole metaphorical expressions (e.g., *when pigs fly, walk on air*). Conversely, in literal expressions, each word directly contributes to the overall meaning (e.g., *get the idea, know better*).

Although great variation in the metaphorical makeup of figurative expressions exists (see Goatly, 2011), in this experiment, MWEs were allocated to two broad categories: either literal or figurative, depending on their opacity. In other words, although *get the idea* and *all the time* have figurative elements, they are nonetheless highly transparent, so they were classified as literals. Contrastingly, a few expressions such as *kick the bucket* are so opaque that Grant and Bauer (2004) classified them as *core idioms*, arguing that no discernable etymological metaphorical connection can be made. Nevertheless, such terms were classified as *figuratives* because learners could make metaphorical connections to remember them. Understanding literals involves naturally processing the words. This process is more straightforward than understanding figuratives, which involves deriving meaning from metaphors as well as rejecting the literal interpretation of each constituent word.

Collocation dictionaries (e.g., Kjellmer, 1994; Sinclair, 1995) contain thousands of entries and serve as valuable references for seeing examples of their use, but one reason these are not very practical guides for second language learners is that they do not address this issue of compositionality. Compositionality raises problems for language learners because even when they know the correct figurative meanings, they strongly favor literal word interpretations, (e.g., Cieślicka, 2006, 2012). To fill this gap, Martinez and Schmitt (2012) made the PHRASal Expressions List, composed of MWEs that are frequent, meaningful, and difficult for language learners to interpret. Martinez and Schmitt (2012) also provided frequency levels for the 505 MWEs on their list to facilitate prioritization for learning along with the first five thousand most frequent individual words on the British National Corpus. Some MWEs on the PHRASE List are difficult for learners due to their opacity (i.e., *end up*), and others cause problems because they are easily misinterpreted (i.e., although *at all* is very clear in its positive sense as in *at all times,* it is much less so in its negative sense, as in *Do you exercise at all?*).

Because highly opaque figurative expressions must be remembered as wholes, and transparent literal expressions can be understood when processed word by word, it may follow that MWE compositionality affects whether they are holistically processed. Research on MWEs shows that they may be retrieved holistically rather than being created from scratch by applying grammar (e.g., Sinclair, 1991; Tremblay et al., 2011; Wray, 2002), but the nature of this holistic processing is complex (Siyanova-Chanturia & Martinez, 2015). Holistic processing for figuratives entails both automatically interpreting the whole MWE's meaning to form link and processing the word sequence; whereas, the holistic processing of literals only entails recognizing the word sequence and processing it faster. By comparing the effects of deliberately learning literal and figurative MWEs, this research aims to shed light on whether holistic processing relates to compositionality.

Multiword Expression Flashcard Learning

Deliberate paired-associate vocabulary learning with flashcards involves repeatedly retrieving targets from meanings or meaning from targets. This systematic and repeated retrieval method is a well-established way for language learners to connect first language meanings with L2 vocabulary. Learners can remember vast numbers of paired associates in a short time. For example, Thorndike (1908) showed that 1,200 words studied for 30 hours showed remarkable persistence in memory. Digital flashcard applications now enable language teachers and learners to systematize a database of words to memorize conveniently. Nakata (2011) extensively reviewed free online flashcard applications, considering pedagogically essential features such as presentation mode variety, adaptive sequencing, and timing settings for spaced review. Retrieval using flashcards is more effective than word lists because learners can remove target items they have mastered. Also, cards can easily be shuffled, giving them another advantage over static lists of items in which the order is unchangeable. In static lists, the sequence of the list is also remembered, providing false memory support for the individual items, thus hindering proper lexical knowledge development.

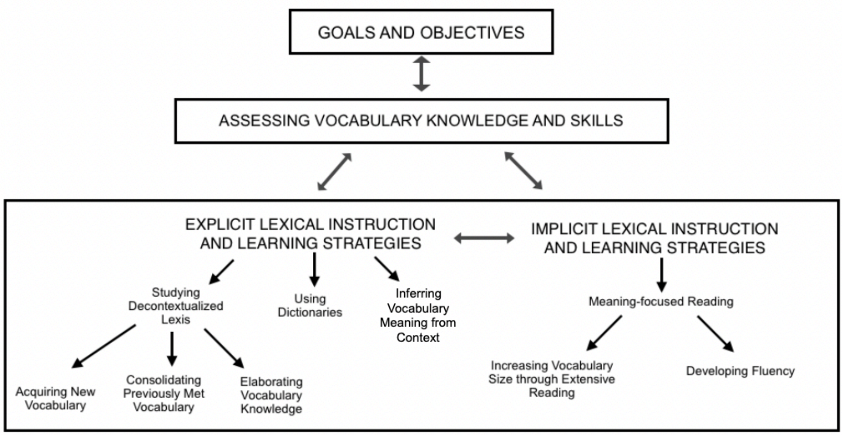
Once learners have a solid base of single-word knowledge, MWE learning is another important goal. Learning MWEs as wholes with flashcards may be an effective learning strategy. Learners can expand their collocation knowledge by practicing with MWE-to-meaning pairs. Given that literal expressions and figurative expressions generate meaning differently, practicing them with flashcards will facilitate learning in different ways. Each word matches its meaning for literal expressions, so flashcard practice will help with fluency development. For figurative expressions, each word must be relearned in its metaphorical context, so practicing with flashcards will both strengthen the meaning-to-form connection and foster processing fluency.

Explicit and Implicit Second Language Learning

Regarding the explicit/implicit interface, deliberate MWE flashcard learning is commonly classified as an explicit learning strategy that develops explicit knowledge. The current study is unique because it investigates whether deliberate MWE flashcard learning also develops implicit knowledge, which is more commonly associated with incidental learning. The interface has long been a central theme of second language acquisition research that reverberates strongly for language teachers, and Nick Ellis's (2005) review bridged connections to language learning with fields such as psycholinguistics, psychology, neurobiology, and cognitive science. He explained that explicit and implicit neurological processes are physiologically distinct but interact as learners develop their proficiency. Hulstijn (2005) defined and distinguished the interface parameters: implicit and explicit memory, implicit and explicit knowledge, implicit and explicit learning, inductive and deductive learning, and incidental and intentional learning. Rod Ellis (2005) operationalized the explicit/implicit distinction in terms of *awareness, accessibility*, and *use*. He explained that learners are not *aware* of implicit knowledge but are *aware* of explicit knowledge; they *access* implicit knowledge automatically, but *access* to explicit knowledge requires controlled processing; they *use* implicit knowledge in fluent performance, but explicit knowledge is *used* during introspective processing when learners encounter difficulties, plan to write, or make an utterance. Although children tend to learn implicitly, second language acquisition requires teenage and adult learners to develop explicit and implicit knowledge in tandem.

Frameworks for foreign language teaching, lesson planning, course design, and curriculum development often balance explicit and implicit learning. In the Four Strands framework (Nation, 2007), three of the strands develop implicit knowledge (meaning-focused input, meaning-focused output, and fluency development), and one strand develops explicit knowledge (language-focused learning). Textbooks are also designed to balance these two types of learning. Likewise, as Figure 1 shows, Hunt and Beglar (2005) explained how EFL reading program designers set goals, clarified objectives, and assessed knowledge gains in a curriculum structure built on a balance of explicit and implicit learning strategies.

Figure 1
Explicit and Implicit Learning Strategies in an EFL Reading Curriculum



Note: From "A framework for developing EFL reading vocabulary," by Hunt, A., & Beglar, D., 2005, *Reading in a Foreign Language*, *17*(1), p. 26.

However, although explicit and implicit teaching and learning methods can be balanced in course design, lesson planning, and teaching, implicit knowledge development is rarely the focus of formal assessment. Similarly, explicit learning gains are often investigated in second language acquisition research, but implicit knowledge gains are seldom the focus. This imbalance occurs because implicit knowledge gains are difficult to measure using traditional methods such as pen and paper tests.

Another concern with most interface research is that it has been chiefly focused on grammar acquisition (e.g., DeKeyser, 1997; Norris & Ortega, 2000; Rebuschat & Williams, 2012; Sorace, 2011; Suzuki & DeKeyser, 2017). In contrast, very little research on the implications of the explicit/implicit interface concerning lexical knowledge exists. Sonbul and Schmitt (2013) propose that this neglect of vocabulary interface research might be due to the traditional dictionary metaphor, which regards the mental lexicon as little more than a list of forms and meanings to associate through simple rote learning. Nation's (2013) framework of vocabulary knowledge has helped to overcome the mental dictionary metaphor by showing that know-ing words entails sophisticated knowledge aspects concerning form, meaning, and use. Nevertheless, Godfroid (2020) explained that this framework concerns explicit language knowledge that can be assessed offline rather than in real-time communicative situations. She transformed it to focus on automaticity, with criteria related to implicit knowledge development. Her framework explains ways to measure the automaticity of form, meaning, and use with real-time methods such as priming experiments, lexical decision tasks, self-paced reading, and eye-tracking. Godfroid (2020) shows how Nation's (2013) criteria may be adapted to consider implicit knowledge development by shifting the focus to real-time processing of form, meaning, and use. Table 1 shows a further adaptation of this framework that focuses on implicit MWE knowledge development criteria. It shows how this experiment measured response times for orthographical and lexical recognition, a narrow slice of the broader lexical knowledge spectrum.

Table 1

Real-Time Lexical Knowledge Aspects Learned with Multiword Expression Flashcards

Knowledge Aspect		**Receptive (R) and Productive (P) Criteria**	
Form	Spoken	R: Does the MWE have auditory representation in memory?	
		P: How rapidly can the MWE be spoken?	○
	Written	R: Does the MWE have an orthographic representation in memory?	○■
		P: How rapidly can the MWE be written or typed?	
	Word parts	R: What word parts are recognizable?	
		P: What word parts can be added or removed?	

Knowledge Aspect		Receptive (R) and Productive (P) Criteria	
Meaning	Form and meaning	R: How rapidly can the MWE's meaning be accessed?	○
		P: How rapidly can MWE be produced to express its meaning?	○
	Concept and referents	R: How rapidly can concepts and referents of the MWE be accessed?	○
		P: How rapidly can the MWE be produced to express a concept?	○
	Associations	R: Has the MWE been integrated into existing semantic networks?	○
		P: How rapidly can associates of the MWE be produced instead?	○
Use	Grammatical functions	R: Is the learner sensitive to the grammar involved with this MWE?	
		P: Can the learner use this MWE in actual conversation?	
	Collocations	R: Are the words of this MWE rapidly recognized?	○■
		P: Can this expression be rapidly produced?	○
	Constraints on use	R: Is the learner aware of constraints on how the MWE is used?	
		P: Can the learner use this MWE correctly?	

Note. ○ = aspects of implicit MWE knowledge developed by practicing with MWE flashcards; ■ = implicit knowledge aspects tested by the current experiment.

Priming to Test for Implicit Knowledge Development

Priming happens when exposure to one stimulus influences a response to a subsequent stimulus without conscious guidance or intention. For example, in semantic priming, the word *table* will be recognized more quickly when it follows *chair* than *dog* because *table* and *chair* often occur together, and thus neurons associated with these words will fire together. Other types of priming experiments focus on orthography, syntax, or perception. Reber (2013) explained that repetition priming is the most common method for investigating implicit knowledge, which he defined as a form of general plasticity and neural network adaptation. When the brain receives input and internally processes it, it stores the physical structure used. Such structures improve functionality and unconsciously facilitate future cognition.

Priming experiments in second language acquisition research aim to operationalize and measure this facilitation. Standard priming paradigms focus on form processing, grammatical sequencing, meaning interpretation, and lexical associations. When a word, MWE, or construction is learned so well that it primes a related target, it means the language learner has strong, well-integrated knowledge that can be accessed automatically. This automaticity signifies the quality of the knowledge, and evidence of priming illuminates how fluently the knowledge is processed.

Priming Research on Implicit MWE Knowledge Development

At the time of writing, research concerned with implicit MWE knowledge development resulting from flashcard learning was not found. However, Sonbul and Schmitt (2013) conducted a priming experiment to measure implicit knowledge development of technical medical MWEs (*cloud baby, iron lung*) resulting from three different learning conditions. In their enriched condition, participants encountered each MWE three times in a text they read. In the enhanced condition, the MWEs were in the same text but highlighted in red, which made the three encounters more explicit. In the decontextualized condition, learners were presented with the MWEs on PowerPoint slides and told to study them carefully. To test for implicit knowledge development, they conducted a lexical decision task experiment to see if the first words of the MWEs primed the processing of their final words. They did not find significant priming effects and proposed that their experimental learning treatment period was too brief and did not allow for recycling and review, which are needed to develop implicit knowledge. However, their explicit measures showed that all three learning conditions led to significant long-term recall and recognition. Their experiment demonstrates how readily explicit knowledge gains can be measured but how difficult it is to measure implicit knowledge gains.

In a replication and extension of Sonbul and Schmitt (2013), Toomer and Elgort (2019) tested the incidental reading conditions (reading only, bolding, and bolding plus glossing) with more participants and more time on task. The results of the primed lexical decision task only showed initial evidence of implicit knowledge development when the collocations were presented without enhancement. Their main finding was that repeated encounters with collocations in reading promoted the development of collocational knowledge. Bolding led to the development of explicit knowledge, and the absence of typographic enhancement promoted the development of implicit knowledge.

However, Toomer and Elgort (2019) did not replicate Sonbul and Schmitt's (2013) decontextualized explicit condition, which was most relevant to this current study. Furthermore, in this current experiment, the learners were each given their own sets of flashcards so that they could remove the MWEs they had learned and reshuffle them to enhance memorization. Elgort (2011) conducted encouraging research regarding implicit knowledge development from flashcard learning for single words (pseudowords). In Experiment 2, she conducted a masked repetition priming experiment displaying a mask (#######) for 522 ms, followed immediately by a pseudoword prime (e.g., "forfert") for 56 ms, and then a target ("FORFERT") for 522 ms. The participants made lexical decisions regarding the targets they had just seen while looking at the blank screen. They were instructed to treat the newly learned pseudowords as English words and answer YES for the lexical decision. This experiment showed that identity primes had a facilitation effect, 52 ms faster than the controls. These results indicated that learning the pseudowords with flashcards resulted in acquiring orthographic representations in implicit knowledge. That is, the quality of the knowledge of the newly learned pseudowords was strong enough to prime the targets that followed. In this current experiment, a masked repetition priming lexical decision task very similar to Elgort's (2011) Experiment 2 was employed to investigate changes in the quality of subconscious representations of the MWEs that participants learned with flashcards.

Obermeier (2022) measured semantic association gains in a self-paced reading experiment that likewise compared the effects of flashcard learning on literal and figurative MWEs. That experiment primarily focused on investigating the semantic components in Table 1. Like the current research, results in that experiment were analyzed in a repeated measures linear mixed-effects model with participants and items as crossed random effects. No statistically significant interaction for semantic association gains were found, as measured in an innovative priming paradigm wherein semantically related words that followed the MWEs in sentences were compared. Although the interaction was not significant, the semantic associates of literals were processed faster than those of figuratives in all three conditions: pre-test, learned post-test, and not-learned post-test. In a separate analysis of the data, a statistically significant interaction showed that deliberate learning resulted in substantial formulaic sequencing gains for literals but no such gains for figuratives. Obermeier (2022) concluded that the learning treatment was too brief to result in the strong semantic acquisition of the figuratives because of their high learning burden. This current research aims to complement findings from that self-paced reading experiment by investigating the effects of flashcard learning on orthographic MWE representations.

Methodology

This investigation focused on implicit knowledge development of literal and figurative MWEs, operationalized by response times in a masked repetition priming lexical decision task. The first research question was: *Does multiword expression flashcard learning develop implicit multiword expression knowledge?* As flashcard learning entails highly focused repetition and retrieval of meaning and form, it was hypothesized that implicit knowledge gains for both MWE types would be statistically significant. The second research question was: *Does implicit knowledge develop differently for the flashcard learning of literal and figurative expressions?* Because learning figuratives is more difficult than learning literals, it was expected that figuratives would be processed more slowly on the pretest. Flashcard learning should result in greater gains for figuratives when the meaning/form connection is established.

Participants

The study's participants (*N* = 43) were 21 male and 22 female students at a small national teacher training university in Japan. All had studied English for 4 to 8 hours a week for six years in junior high and high school in reading, writing, speaking, listening, and grammar courses. Their ages ranged from 19 to 22. They were enrolled in their first or second year of studies in the English Education Department, training to become elementary, junior high, or high school English teachers. Participants were in two intact classes, 26 in one class and 28 in the other (a convenience sample of 54). Teacher-training students often need to be absent from class for practicum training. For this reason, 11 participants missed one or more classes during the experiment and were excluded from the data analysis.

Soon after beginning their first year of studies, all students took the Global Test of English Communication (GTEC), designed for Japanese university and high school students. Their average total score was 623 (*SD* = 71.89), which, according to the GTEC instructional materials, classified them as Advanced Learners, the second-highest category of the test. Mean reading scores were 241 (*SD* = 29.32), earning them a level of assessment at which "reading a newspaper article with the occasional support of a dictionary is possible." The accompanying materials also state that the approximate TOEIC equivalent is 600, the approximate paper-based TOEFL equivalent is 480, and the approximate Internet-based TOEFL (iBT) is 60. Thirteen of the participants had studied English abroad for four weeks or more. The participants' motivation to learn English was high because they intended to eventually teach it professionally.

Before the experiment, the researcher explained the following three points verbally in English and then in writing in Japanese: (a) their participation in the study was optional; (b) their participation or lack of participation would have no effect on their grade; (c) no personal information would ever be shared. After they finished the experiment, they were debriefed on the purposes of the investigation and preliminary findings. Participants were also given a small gift as a token of appreciation and acknowledgment of their efforts.

Procedures

The experiment was conducted once weekly over five weeks. The primary experimental condition, flashcard learning, was counterbalanced across the two groups of participants. In Table 2, the schedule of the experiment is outlined.

Table 2

Schedule of the Experiment

Session	Minutes	Activity
Week 1	10	Introduction to the experiment
Week 2	30	Masked priming lexical decision pretest
Week 3	40	MWE flashcard learning
Week 4	30	Masked priming lexical decision posttest
Week 5	10	Debriefing

Learning Materials

The experimental materials and instruments were made using a list of 48 MWEs, 24 figuratives, and 24 literals. The target items were selected by two native speakers, who discussed each MWE and categorized it as literal or figurative according to how directly the constituent words matched the overall meaning. A third native speaker confirmed the literal/figurative categorizations. Next, the researcher matched the MWEs with Japanese meanings, and these paired associates were shown to four English learners who were not participants in the experiment to confirm whether the form-to-meaning connections made sense. For example, *next door* was matched to the Japanese meaning 隣の and confirmed. The literal and figurative expressions were counterbalanced across Study Lists A and B to create critical comparisons among the conditions. Therefore, each participant learned 24 MWEs: 12 literal and 12 figurative. All experimental contrasts were made on the items within participants. If participants in one group learned an MWE, the other group did not. Some examples of the paired-associates and the counterbalancing structure for the study lists are shown in Table 3.

Table 3
Examples from the List of Figuratives, Literals, and Japanese Meanings

MWE Composition	Study List	MWE	Japanese
Literal		above all	最も
	A	stay away	避ける
		take place	起こる
		deal with	扱う
	B	feel like	欲しい
		take it easy	のんびりする
Figurative		set out	始まる
	A	sinking ship	絶望
		play hardball	真剣
		high handed	攻撃的
	B	can of worms	複雑
		make waves	迷惑

Participants learned the MWE/Japanese pairs in the experimental treatment in which the English target was typed on one side of a piece of paper, and the Japanese meaning was typed on the other. They were also given a guidance sheet explaining the following instructions (written in Japanese) about the flashcard learning strategy. Before they started studying, the following were explained orally: (1) Practice with 8 MWEs at a time; (2) Recall the Japanese meanings from the MWEs; (3) When recalling the English MWE from the Japanese meaning, say it aloud; (4) When you feel you have learned an MWE well, remove the card; (5) When you remember the first 8 MWEs, add 8 more and study all 16 together; (6) After you remember these 16 MWEs, add the final 8 and study all 24 of them. Participants were given 20 minutes to study independently. Time announcements were made when 10, 5, and 1 minute(s) remained.

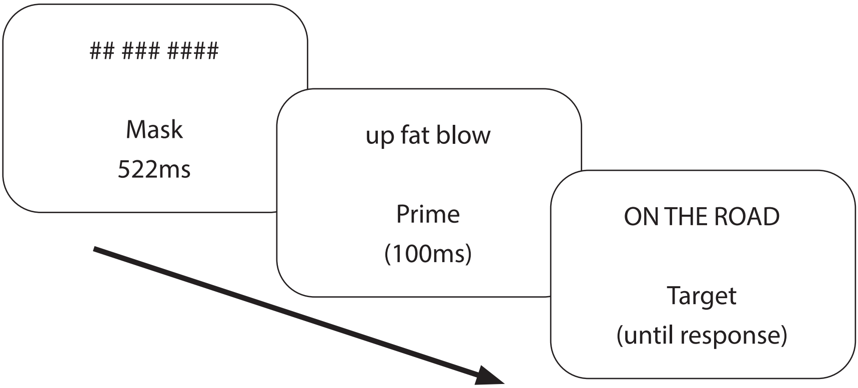
The Masked Repetition Priming Lexical Decision Task

The masked priming lexical decision task was conducted in a computer-assisted language learning classroom containing 48 Hewlett Packard Compaq® dc7700 desktop computers with 2.13 GHz Intel Core Duo® processors, displayed on 21.5-inch Iodata® liquid crystal display monitors. It was created using E-prime®, software for developing psychological experiments (Schneider et al., 2002). The pre-test and post-test each took approximately 30 minutes for participants to complete. Before beginning the actual task, participants did 20 practice trials to become familiar with the procedure.

The trial format is presented in Figure 2. Each mask, prime, and target word had the same number of characters as the corresponding word on the next slide. For example, in Figure 2, the mask's ##, the prime's *up,* and the target's *on* have two characters. Likewise, ###, *fat,* and *the* each have three characters, and so on. The mask was presented for 522 ms (slightly over half a second), followed by a prime that was presented for 100 ms (one-tenth of a second). This very short prime presentation time was crucial: it was brief enough to prevent conscious processing yet long enough to stimulate subconscious processing. After the experiment was finished, when asked what they saw, participants said the targets seemed to be slightly blurry at first but then came into focus, confirming the subconscious presentation paradigm.

The target was presented until a response was received. The hypothesis was that if the identity prime (in this example, "on the road", not shown in the figure) had been acquired in the flashcard learning treatment, it would facilitate the processing of the target ON THE ROAD faster than the unrelated prime “up fat blow”. Participants pressed different buttons to make lexical decisions on the targets, answering *YES* if all the words were English or *NO* if one or more words were not English. Figure 2 is an example of a trial in the unrelated priming condition wherein the correct response to the lexical decision was YES. In masked repetition priming, identity primes consistently facilitate the processing of the targets they precede. This procedure has had robust effects and has helped to understand subconscious lexical recognition processes (e.g., Adelman et al., 2014; Forster & Veres, 1998; Grainger, 1998). If an MWE is established in the mental lexicon, an identity prime will subconsciously pre-activate its lexical representation, and the target will be processed more fluently.

Figure 2
Example Trial for the Masked Repetition Priming Lexical Decision Task.



In each trial of the lexical decision task, participants decided whether all the words of the MWE target were English or not. The 144 targets were balanced half and half between 72 intact MWEs and 72 non-word MWEs. For each participant, 24 of the 72 intact targets had been learned, 24 had not, and another 24 were fillers (added to lower the percentage of critical trials and further prevent strategic processing). Although the nonword MWE trials were essential distractors for the lexical decision task, these were excluded from the analysis. In this way, 24 learned and 24 not-learned targets were the critical trials for the experiment and the focus of the analysis.

As a rule of thumb, Brysbaert and Stevens (2018) recommend that an adequately powered reaction time experiment has at least 1,600 observations per condition. In this experiment, 43 participants, 48 critical stimuli, and 2 test sessions yielded 4128 observations. Regarding research question 1, for the three conditions tested (Learning, Priming, and MWE Composition), each condition had 1376 observations, so the experiment had 86% of the observations needed to meet that criterion. Further power analyses by simulation were conducted using the *simR* package in R (Green & MacLeod, 2016). Based on 200 simulations, the *powerSim* function revealed that both Learning and Priming conditions had 100% of the statistical power necessary, but MWE Composition had only 50% of the power necessary. In the *powerSim* analysis for Research Question 2, in which 1972 literal observations and 1937 figurative observations were analyzed separately in a simpler model, Learning and Priming conditions both had 100% of the statistical power needed.

An example of each trial type (excluding the filler trials) is shown in Table 4. Every critical trial was tested under one level of all three two-leveled conditions: (a) Learning: Learned versus Not-Learned; (b) MWE Composition: Literal versus Figurative; (c) Priming: Identity vs. Unrelated. These contrasts were created in the trial list, in which each MWE was tested once per participant in one or the other level of each condition. That is, all participants experienced the same conditions on different MWEs. The experiment was a series of comparisons between conditions on items.

Table 4
Item Types for the Masked Repetition Lexical Decision Task

Lexical	Primes		
Decision	List A	List B	Targets
YES	dog eat dog (identity)	off the cat (unrelated)	DOG EAT DOG (intact figurative MWE)
	the fly door (unrelated)	all the time (identity)	ALL THE TIME (intact literal MWE)
NO	abobe all (identity)	sheep the (unrelated)	ABOBE ALL (nonword MWE)
	teh od nemes (unrelated)	han of wobes (identity)	HAN OF WOBES (nonword MWE)

Results

Data were collected for 4060 observations, but an initial phase of outlier trimming removed invalid trials. Baayen (2008) explains that extremely fast response times (RTs) signify non-engaged, automatic button-pushing, and extremely slow responses signify confusion or distraction. Accordingly, 80 observations (1.97% of the data) were removed with response times below 200 ms or above 4000 ms. Mean response times (RTs) for all conditions are shown in Table 5, and some comparisons of interest are as follows. For All Trials, mean RTs for the Learned trials (1303.57 ms) were 180.03 ms faster than the Not-learned trials (1483.60 ms). Regarding Priming, for Learned trials, MWE targets in the Identity Priming condition (1229.83 ms) were processed 146.88 ms faster than Learned MWEs in the Unrelated condition (1376.71 ms). Regarding MWE Composition, Literal expressions that were Learned (1264.20 ms) had 79.05 ms faster RTs than Figurative expressions that were Learned (1343.25 ms). Although these mean differences help to describe general trends in the data, more sophisticated modeling is required to interpret the effects of the Learning, Priming, and MWE Composition conditions and their interactions.

The analysis of the crossed linear mixed effects model was conducted using the lmer package in the R environment for open-source statistical software (Bates et al., 2015). The analysis followed the top-down model building strategy specified by West, Welch, and Gateki (2015). The first step was to confirm whether a random effects structure should be included. To test this hypothesis, a "loaded mean" structure containing both fixed and random effects was compared with a model containing only fixed effects (West et al. 2015, p. 66). The ANOVA comparison assessing whether the added random effect variances were zero was rejected with a *p*-value less than 0.0001, which indicated that the model including the random effects should be included for all subsequent stages of model building.

The random-effects specification was improved through the inclusion and exclusion of different structures. Subjective model comparisons included comparison of Akaike's Information Criterion and the Bayesian Information Criterion. Objective comparisons were assessed using likelihood ratio tests of models using results in the lmer output until the best model was identified (Baayen, 2008; Baayen et al., 2008). The inclusion of random intercepts (Participant and Item) and random slopes (Trial Order and Learning) were judged to best capture the overall random effects structure.

In the initial model, the Learning, Priming, and MWE Composition conditions were all statistically significant fixed-effect predictors. The constant

Table 5
Response Times for the Masked Priming Repetition Lexical Decision Task (Milliseconds)

	Learning Condition		
	Pretest	Not-learned	Learned
M	1611.86	1483.60	1303.57
SEM	16.91	21.64	17.93
n	1947	978	984
SD	746.57	676.93	562.48
95% CI Lower	1578.68	1441.12	1268.38
95% CI Upper	1645.05	1526.06	1338.76

	Priming Condition					
	Identity			Unrelated		
	Pretest	Not-learned	Learned	Pretest	Not-learned	Learned
M	1571.02	1433.63	1229.83	1651.68	1532.97	1376.71
SEM	22.10	31.08	30.95	21.82	30.89	30.83
n	961	486	490	986	492	494
SD	763.69	655.05	537.39	727.68	695.03	577.61
95% CI Lower	1527.68	1372.69	1169.14	1608.89	1472.40	1316.27
95% CI Upper	1614.35	1494.56	1290.51	1694.46	1593.53	1437.15

	MWE Composition					
	Literal			Figurative		
	Pretest	Not-learned	Learned	Pretest	Not-learned	Learned
M	1500.47	1382.88	1264.20	1725.69	1586.41	1343.25
SEM	21.35	27.83	26.32	24.71	35.91	26.75
n	984	494	494	963	484	490
SD	698.26	690.81	562.62	776.91	741.40	560.12
95% CI Lower	1461.13	1328.07	1212.64	1678.63	1526.51	1288.53
95% CI Upper	1546.08	1441.28	1314.59	1776.21	1656.22	1394.07

variance, linearity, independence, and normality assumptions were assessed using the *mcp* function in R's *LMERConvenienceFunctions* package. In this initial model, the distribution of the residuals had a severe negative skew and a very long positive tail, so 71 positive and negative outliers (1.78% of the data) were trimmed, resulting in a bell-shaped distribution of the residuals that resembled the normal distribution.

After confirming differences between the levels of the independent variables, the next step was to investigate the interactions of interest, as specified per the experimental hypotheses. The interaction between Priming (Identity versus Unrelated) and Learning (Pretest, Learned, or Not-learned) tested how flashcard learning affected the RTs of the different prime types. The interaction between MWE Composition (Literal versus Figurative) and Learning tested how flashcard learning affected their RTs differently. The notation for the specification of the final model was as follows:

Fixed Effects:

RT ~ Priming*Learning + MWEComposition*Learning

Random Effects:

(1 + TrialOrder + Learning | Participant) + (1 | Target)

The results of the random and fixed effects for the model are shown in Table 6. The intercept represents the reference levels of the independent variables: Unrelated Priming, Pre-test Learning Condition, and Figurative MWE Composition. The other estimates are in comparison to the intercept level. For simple effects, pairwise effect size calculations were made following Brysbaert and Stevens (2014).

The focus of the investigation for Research Question 1 was the statistically significant interaction explained in Table 6 between the Learned condition and Priming (β = -0.04; t = -2.42; $p < .05$) and how it contrasts with the non-significant result for interaction between the Not-learned condition and Priming. This difference indicates that the MWEs were learned well enough to produce priming effects through flashcard learning. For the simple effects, the estimate associated with the Learned (Posttest) condition (β = -0.14; t = -5.11; d = -0.34) was larger than that for the Not-learned condition (β = -0.07; t = -2.93; d = -0.17). These results, as well as the small but statistically significant Learned x Priming interaction (β =0.038; t = -2.42, p > .05) were evidence of priming effects associated with flashcard learning. As the reference level was Pre-test, the small effect for the Not-learned condition indicates testing effects.

Table 6

Linear Mixed-Effects Model for the Masked Repetition Priming Lexical Decision Task

Random Effects		
	Variance	*SD*
Target (Intercept)	0.01	0.14
Participant (Intercept)	0.05	0.23
Trial order (Slope)	0.01	0.96
Learned (Slope)	0.03	0.18
Not-learned (Slope)	0.02	0.14
Residual	0.05	0.22

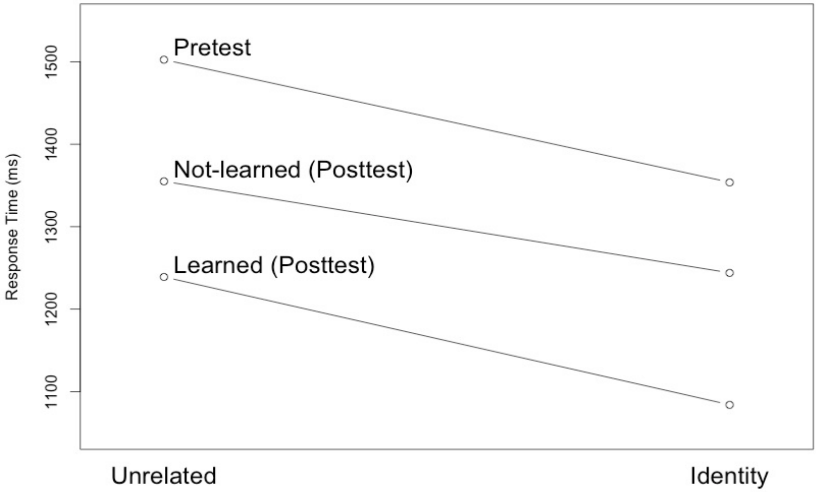
Fixed Effects				
	β	*SE*	*t* value	*d*
Intercept	-0.67	0.44	-15.26*	
Priming (identity)	-0.73	0.10	-7.23*	-0.18
Learned (Posttest)	-0.14	0.28	-5.11*	-0.34
Not-learned (Posttest)	-0.07	0.25	-2.93*	-0.17
MWE Composition (Literal)	-0.11	0.41	-2.61*	-0.31
Learned x Priming	-0.04	0.02	-2.42*	
Not-learned x Priming	-0.01	0.02	0.42	
Learned x MWE Composition	0.04	0.02	2.17*	
Not-learned x MWE Composition	0.03	0.02	1.59	

* $p < .05$

Figure 3 shows faster RTs for Identity priming for all three learning conditions. It also shows incremental facilitation overall as exposure increases because the Learned trials are faster than the Not-learned trials, and both are faster than the Pretest baseline.

Figure 3

Learning and Priming Conditions for all Multiword Expressions



Research Question 2 concerned whether flashcard learning affects Literal and Figurative expressions differently. Table 6 above shows the statistically significant interaction between MWE Composition and Learning ($\beta = 0.04$; $t = 2.17$, $p < .05$), indicating that flashcard learning had different effects on the Figurative and Literal targets. To better understand the effects of learning conditions, priming conditions, and MWE Composition, separate investigations were conducted on the literals and figuratives by specifying the following model for each:

Fixed Effects:

RT ~ Priming*Learning

Random Effects:

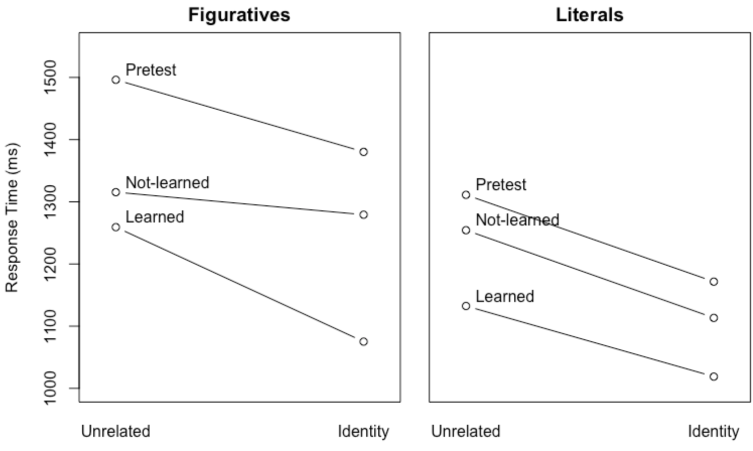
(1 + Trial Order + Learning | Participant) + (1 | Target)

As in the previous analysis, the model structure was confirmed step by step, and the interaction between Priming and Learning was tested for significance.

The interaction between Priming and Learning was statistically significant for Figuratives but not for Literals. Figure 4 below shows the different effects of learning on the two MWE types. For Literals, the RTs for targets in Identity and Unrelated priming conditions decrease in equal progression for both the Learned and Not-learned conditions, showing that the priming effects between Unrelated and Identity primes were the same at Pre-test and Post-test in both the Learned and Not-learned conditions. Furthermore, the overall pre-test to post-test response time changes for the Literals (from around 1350 ms at the Unrelated Pretest to 1000 ms at Identity Learned Postest) shows that these were processed more consistently and faster than Figuratives (which changed from around 1500 ms to 1100 ms over the same conditions). Most importantly, comparing the Figurative Not-learned and Learned line pairs with the corresponding Literal line pairs shows the dramatic difference in effects that deliberate learning had on these two different types. Learned Figuratives had greater gains in Identity Priming effects than Learned Literals.

Figure 4

Interactions between Priming and Learning Conditions for the Literal and Figurative Expressions



The statistical results confirmed the effects shown in Figure 4. For Figurative expressions, the interaction between Learning and Priming was statistically significant. For the Learned Figuratives, Identity priming was significantly faster than Unrelated priming (β = -3.14; t = -3.037; p = .002). Thus, the line is steeply sloped. For the Not-learned Figuratives, this difference was not significant (as shown by the nearly horizontal line). Identity and Unrelated priming effects were constantly incremental for the Learned and Not-learned Literals, as shown by the nearly parallel three lines on the right side of Figure 4.

Discussion and Conclusion

For the full data set with all the MWEs, the interaction between Priming and Learning conditions was statistically significant for the Learned targets but not for those in the Not-learned condition. These different effects confirmed Research Question 1, showing that flashcard learning resulted in strong Identity priming effects, evidence of facilitated subconscious orthographic processing for MWEs overall. Separate analyses of Literals and Figuratives were conducted to investigate Research Question 2. For Figurative expressions, the statistically significant interaction between Learning and Priming conditions showed priming effects for Learned targets but not for Not-learned targets. However, no such priming effects were found for the Literal expressions. Together, these results showed that learning with flashcards through repetition and retrieval facilitated the development of implicit orthographic knowledge for all MWEs, but the effects of learning were more substantial for Figurative expressions, which have a heavier learning burden.

Although this research had some valuable findings, its limitations must also be mentioned. First, although prior learning was accounted for with a pretest, this is not standard in priming research because of the strong tendency to produce testing effects. A better way to control for prior MWE knowledge would be to use highly specialized unknown targets like the medical MWEs that Sonbul and Schmitt (2013) used. The second limitation was the convenience sample. Severe participant attrition resulted in insufficient statistical power to test the full model for Research Question 2, so a second separate analysis was required with a simpler model without the MWE Composition predictor. Another constraint resulting from the convenience sample (taken during scheduled class time) was limited time on task. A third limitation is also concerned with time on task. As the participants were guided to remove flashcards once they remembered them, they may have removed literal flashcards sooner than figurative ones, and this imbalance of study time may have influenced the results.

Keeping these limitations in mind, it nevertheless seems fair to argue that the pedagogical implication of these findings is that flashcard learning benefits the learning of figuratives but not literals. For both MWE types, automatic orthography and word sequence recognition must be developed. However, literal and figurative semantic compositions entail different learning processes. For literals, the direct meaning-to-form connection for each word is also automatized each time it is encountered, meaning their integration into the mental lexicon is straightforward. Contrastingly, when learning a figurative expression incidentally, the learner must reject the direct semantic interpretation of each word, and this is not possible until the whole figurative expression is recognized (Cieślicka, 2006; 2012). Thus, processing figurative expressions entails the extra steps of rejecting individual word meanings and then learning the metaphor of the whole expression. These two additional steps seem to hinder the development of automaticity for figuratives.

Researchers have explored explicit learning methods for deeply processing figurative expressions such as focusing on etymology (Boers, Eyckmans, & Stengers, 2007), cognitive semantics (Boers, 2011), and pictorial elucidation (Boers et al., 2009). Such methods may entail rich and thoughtful processing that fosters durable associations, but they may also require learning superfluous explicit knowledge that cannot be applied in real-time communicative situations. In deliberate flashcard learning, such deep processing is not the aim. Instead, the strategy aims to automate the association of the metaphorical meaning to the whole expression through repetition and retrieval.

Strategic flashcard learning of the PHRASE List (Martinez & Schmitt, 2012) is undoubtedly an effort wisely spent, as these frequent MWEs will likely be encountered in natural English use. Abundant, thematic flashcard learning focused on specialized MWEs found in accompanying texts would balance explicit and implicit learning strategies as supported by research cited herein (Hunt & Beglar, 2005; Sonbul & Schmitt, 2013; Toomer & Elgort, 2019). Flashcard learning of figuratives entails bypassing misleading (yet normal) individual word processing and automatizing the connection between the whole expression and its metaphorical meaning. Whether learned incidentally or learned deliberately as a whole, each encounter with a literal expression entails processing facilitation. In sum, findings from this study call for strategic flashcard learning of frequent MWEs with opaque meanings accompanied by massive exposure that will provide incidental learning opportunities.

Andrew Obermeier is an associate professor at Kyoto University of Education. His research interests include vocabulary acquisition and positive psychology in language learning.

References

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., Perry, J. R., Lupker, S. J., Forster, K. I., Cortese, M. J., Scaltritti, M., Aschenbrenner, A. J., Coane, J. H., White, L., Yap, M. J., Davis, C., Kim, J., & Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods, 46*(4), 1052–1067. https://doi.org/10.3758/s13428-013-0442-y

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

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

Bates, D., Mächler, 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

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(9), 1–20. https://doi.org/10.5334/joc.10

Boers, F. (2011). Cognitive semantic ways of teaching figurative phrases: An assessment. *Review of Cognitive Linguistics, 9*(1), 227–261.

Boers, F., Piquer Píriz, A. M., Stengers, H., & Eyckmans, J. (2009). Does pictorial elucidation foster recollection of idioms? *Language Teaching Research, 13*(4), 367–382. https://doi.org/10.1177/1362168809341505

Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics? *Language Teaching Research, 11*(1), 43–62. https://doi.org/10.1177/1362168806072460

Cieślicka, A. (2006). Literal salience in online processing of idiomatic expressions by second language learners. *Second Language Research, 22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa

Cieślicka, A. (2012). Do nonnative language speakers chew the fat and spill the beans with different brain hemispheres? Investigating idiom decomposability with the divided visual field paradigm. *Journal of Psycholinguistic Research, 42*, 475–503. https://doi.org/10.1007/s10936-012-9232-4

DeKeyser, R. M. (1997). Beyond explicit rule learning. *Studies in Second Language Acquisition, 19*(2), 195–221.

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning, 61*(2), 367–413. https://doi.org/10.1111/lang.2011.61.issue-2

Ellis, N. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27*(2), 305–352.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141–172.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk, 20*(1), 29–62.

Forster, K. I., & Veres, C. (1998). The prime lexicality effect: Form-priming as a function of prime awareness, lexical status, and discrimination difficulty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 498.

Goatly, A. (2011). *The language of metaphors* (2nd ed.). Routledge.

Godfroid, A. (2019). Sensitive measures of vocabulary knowledge and processing: Expanding Nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (1st ed., pp. 433–453). Routledge.

Grainger, J. (1998). Masked priming by translation equivalents in proficient bilinguals. *Language and Cognitive Processes, 13*(6), 601–623. https://doi.org/10.1080/016909698386393

Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics, 25*(1), 38–61.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498. https://doi.org/10.1111/2041-210x.12504

Hulstijn, J. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition, 27*(2), 129–140.

Hunt, A., & Beglar, D. (2005). A framework for developing EFL reading vocabulary. *Reading in a Foreign Language, 17*(1), 23–59.

Kjellmer, G. (1994). *A dictionary of English collocations: Based on the Brown corpus*. Clarendon Press.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 33*(3), 299–320. https://doi.org/10.1093/applin/ams010

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning, 24*(1), 17–38.

Nation, P. (2007). The four strands. *Innovation in Language Learning & Teaching, 1*(1), 2–13. https://doi.org/10.2167/illt039.0

Nation, P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning, 50*(3), 417–528.

Obermeier, A. (2022). The effects of learning multiword expressions deliberately: Implications regarding implicit knowledge development. *JACET Kansai Journal, 24*, 96–115.

Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia, 51*(10), 2026–2042. https://doi.org/10.1016/j.neuropsychologia.2013.06.019

Rebuschat, P., & Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(4), 829–856. https://doi.org/10.1017/S0142716411000580

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Psychology Software Tools.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. M. (1995). *COBUILD English collocations*. Harper Collins.

Siyanova-Chanturia, A. (2017). Researching the teaching and learning of multi-word expressions. *Language Teaching Research, 21*(3), 289–297. http://doi.org/10.1177/1362168817706842

Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics, 36*(5), 549–569. https://doi.org/10.1093/applin/amt054

Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*(1), 121–159. https://doi.org/10.1111/j.1467-9922.2012.00730.x

Sorace, A. (2011). Pinning down the concept of "interface" in bilingualism. *Linguistic Approaches to Bilingualism, 1*(1), 1–33. https://doi.org/10.1075/lab.1.1.01sor

Suzuki, Y., & DeKeyser, R. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning, 67*(4), 747–790. https://doi.org/10.1111/lang.12241

Thorndike, E. L. (1908). Memory for paired associates. *Psychological Review, 15*(2), 122.

Toomer, M., & Elgort, I. (2019). The development of implicit and explicit knowledge of collocations: A conceptual replication and extension of Sonbul and Schmitt (2013). *Language Learning, 69*(2), 405–439. https://doi.org/10.1111/lang.12335

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569–613. https://doi.org/10.1111/lang.2011.61.issue-2

West, B., Welch, K., & Gateki, A. (2015). *Linear mixed models: A practical guide using statistical software* (2nd ed.). CRC Press.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.