The Effects of Mental Imagery and Embodied Action on L2 Word Learning

Laura M. Morett (morett@pitt.edu)
Department of Psychology, University of Pittsburgh
210 S. Bouquet Street, Pittsburgh, PA 15260 USA

Raymond W. Gibbs (gibbs@ucsc.edu)
Department of Psychology, University of California, Santa Cruz
1156 High Street, Santa Cruz, CA 95064 USA

Brian MacWhinney (macw@cmu.edu)
Department of Psychology, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213 USA

Abstract

Previous research has provided evidence that mental imagery and embodied action can facilitate word learning in a novel language. However, it is unclear *how* these factors interact—as well as *why* they play a role—in word learning. Through a set of four experiments, this research demonstrated that neither mental imagery nor embodied action directly promotes the acquisition of second language (L2) words by adult learners. Notably, both passive viewing of images and gestures, as well as active engagement in mental imagery and gesture enactment, were insufficient to enhance L2 word learning. These results suggest that adults are effective L2 word learners, and that, because of this, embodied action does not play an essential role in supporting L2 lexical acquisition.

Keywords: Second language acquisition, word learning, gesture, mental imagery, embodied cognition

Introduction

When setting out to learn a new language, it is common to start by learning the meanings of its conceptually simplest units: words. When learning a word, learners must associate the acoustic and/or orthographic form of the word with a representation of the word's meaning. These representations are based on real-world interactions between the learner and the object to which the word refers; thus, they are a composite of the object's properties, as perceived and acted upon by the learner. As a result, word learning likely depends heavily upon both mental imagery and embodied action, in conjunction with text/speech processing and memory. The present research sought to investigate—and dissociate—the contributions of mental imagery and embodied action to second language (L2) word learning. This research was based on two major goals: (1) elucidate the cognitive processes responsible for L2 word learning in adulthood; (2) determine how to engage these processes, thereby enhancing L2 word learning.

Several lines of research indicate that mental imagery plays a central role in L2 word learning. In the keyword method of L2 word learning (Atkinson, 1975), learners are instructed to choose a word from their native language that is phonologically similar to the target L2 word, and to formulate a mental image of the referents of these words interacting. For example, for the Spanish word *chico* (boy), a learner might choose the similar-sounding English word *chick,* and then could imagine a boy holding a chick. This method has been shown to be a more effective strategy for L2 word learning than verbal association of target words and translations (Atkinson & Raugh, 1975; Levin, McCormick, Miller, Berry, & Pressley, 1982; Raugh & Atkinson, 1975; Van Hell & Mahn, 1997). There is also evidence that, particularly in the early stages of L2 learning, the meanings of target words are learned more effectively via physical images, which cue mental imagery through isomorphism, than via verbal translation (Carpenter & Olson, 2011; Chun & Plass, 1996). These findings indicating that mental imagery promotes L2 word learning can be explained by dual coding theory (Paivio, 1990), which posits that long-term memory encoding can be enhanced via the simultaneous processing of visual and verbal information.

Aside from mental imagery, embodied action—action that learners perform using their bodies—also plays an important role in L2 word learning. Because learners' representations of word referents entail actions that learners have performed on these objects, the use of embodied action during word learning allows learners to create deeper, more meaningful word-referent associations. One method that has taken advantage of embodied action is the Total Physical Response Method (Asher, 1969), which has been successfully used to teach L2 words to novice learners (Asher, Kusudo, & Torre, 1974; Asher & Price, 1967). In this approach, the instructor teaches the target word by saying it while demonstrating its meaning using the body, and learners re-enact the instructor's actions using their own bodies in order to demonstrate comprehension of the words' meanings. Other evidence supporting embodied action comes from research showing that the enactment of gestures representative of word meanings (e.g., placing the hands together and opening them for *book*) during learning facilitates recall of target words to a greater degree than passively viewing pictures or representative gestures (Tellier, 2008). Taken together, these findings indicate that embodied action allows L2 learners to tap into their representations of word referents, creating robust,

multimodal associations between target words and the objects that they represent.

Although extant research suggests that mental imagery and embodied action both contribute separately to L2 word learning, it is unclear from this work if—and how—they interact. One way by which their respective contributions could be clarified is by investigating whether gesture, which typically accompanies and complements speech (McNeill, 2005), facilitates L2 word learning. Iconic (i.e., representative) gestures are created via embodied action and evoke mental imagery via their iconicity, which conveys visuospatial properties of the referent. Several studies (Allen, 1995; Kelly, McDevitt, & Esch, 2009; Tellier, 2008) have shown that L2 words are recalled more accurately over longer intervals when they are learned via representative iconic gesture than when they are learned via speech. Additionally, when L2 words are accompanied by non-representative iconic gesture (e.g., placing the hands together and opening them as for *book* while the word *drink* is presented), they are learned less effectively (Kelly et al., 2009). All of these findings indicate that the combination of mental imagery and embodied action in iconic gesture is a powerful tool for enhancing L2 word learning when it is synchronous with the meanings of the target words.

The objective of the current research was to clarify the independent contributions of mental imagery and embodied action—as well as their interactions—to L2 word learning. To this end, target words were presented in four conditions in which words were accompanied by stimuli that crossed these factors in a 2-by-2 design (see Table 1). These stimuli were designed to elicit either active or passive processing of mental imagery and embodied action, depending on the task instructions. Based on the research discussed above, it was predicted that iconic gesture viewing and enactment would result in the highest number of L2 words recalled, followed by mental imagery, followed by meaningless embodied action and text/speech only. Furthermore, it was predicted that active learning conditions would allow participants to recall more words than passive conditions due to greater engagement of the sensorimotor system.

Table 1: Experimental design for Experiments 1-4.

	Mode	+ Mental Imagery	- Mental Imagery
+ Embodied Action	Passive	Iconic gesture	Beat gesture
	Active	Gesture enactment	Meaningless hand motion
- Embodied Action	Passive	Physical images	Text/Speech only
	Active	Mental imagery formation	Verbal repetition only

Experiment 1

Methods

Twenty-six undergraduate students (age: M = 20.25; SD = 1.5; sex: 11 males; 15 females) at a medium-sized public university in the US participated in return for partial course credit. All participants were fluent English speakers and had no knowledge of Hungarian.

Twelve Hungarian words and their English glosses were used in this research (see Table 2). Prior to this research, 15 English speakers who did not participate in this study were asked to rate the concreteness, imageability, and meaningfulness of the English glosses of 80 candidate words, and to gesture in a way that represented the meaning of each gloss. The 12 words with the most consistent responses from each of the categories were selected for the study. Videos of iconic gestures were created by recording a fluent Hungarian-English bilingual saying these words in each language while enacting the gestures produced by the most participants. Images were line drawings representing the English glosses of Hungarian target words from the International Picture Naming Project (Szekely et al., 2004). In order to control for possible vocal iconicity, audio of the pronunciation of Hungarian and English words was extracted from the iconic gesture videos and was played during presentation of referent images and text of words. Beat gestures—simple, non-iconic hand movements in time to speech prosody—represented embodied action without cuing mental imagery. Videos of these gestures were created by recording the Hungarian-English speaker saying the target words in each language while enacting beat gestures. These videos were presented with their own sound track in order to preserve prosodic speech-gesture synchrony.

The learning phase of this study consisted of three blocks comprising 12 trials apiece. In each learning trial, the

Table 2: Hungarian and English words from Expts. 1-4.

Hungarian	**English**
Betegség	Illness
Kalapács	Hammer
Kulcs	Key
Löni	To shoot
Mászni	To climb
Megütni	To hit
Orá	Watch
Öröm	Joy
Seprű	Broom
Tréfa	Joke
Unott	Bored
Varrni	To sew

following sequence of events was repeated twice: a randomly-selected Hungarian word was presented for 2000 ms., and after a 1000 ms. interstimulus interval, the corresponding English word was presented for 2000 ms. This study used a within-participants design; thus, for each experimental session, three Hungarian words and their English glosses were assigned at random to each of the following conditions: (1), iconic gesture, in which words were accompanied by video of a gesture representing their meaning; (2), beat gesture, in which words were accompanied by video of simple, non-iconic gestures made in time to speech; (3), image, in which words were accompanied by an image representing their meaning; (4), text, in which words were presented as text (see Materials section for a description of the stimuli used in each condition).

The test phase of this study consisted of a single block in which each Hungarian word that participants had learned was presented as text and speech. Participants responded by saying the corresponding English word or by saying "skip" if they could not remember it. In order to examine how learning conditions affected long-term L2 word recall, participants completed the test phase at three intervals following the learning phase: five minutes, one week, and one month.

Results and Discussion

L2 word recall was quantified by scoring responses using a binary coding scheme (1 = correct, 0 = incorrect/skipped), and by converting scores into proportion correct for each participant and condition (in order to control for unscorable responses due to factors such as unintelligibility or technical errors in running the recall task). Proportional scores were submitted to repeated measures ANOVAs, in which participant and word were used as fixed factors.

This analysis revealed a main effect of recall interval, $F_{pp}(2, 51) = 12.16$, $p < .001$, $\eta_p^2 = .36$; $F_{word}(2, 18) = 6.42$, $p = .008$, $\eta_p^2 = .42$. Bonferroni-corrected post hoc analyses showed that participants recalled more L2 words after five minutes than they recalled after one week ($p_{pp} = .03$; $p_{word} = .02$) and one month ($p_{pp} = .001$; $p_{word} = .06$). There was also a main effect of learning condition, $F_{pp}(3, 77) = 7.70$, $p < .001$, $\eta_p^2 = .26$; $F_{word}(3, 27) = 7.04$, $p = .001$, $\eta_p^2 = .44$, see Figure 1. Post-hoc analyses showed that participants learned more words via text than they did via beat gesture ($p_{pp} = .002$; $p_{word} = .008$), iconic gesture (by participant; $p_{pp} = .05$; $p_{word} = .34$) and images ($p_{pp} = .08$; $p_{word} = .05$). However, the interval by condition interaction failed to reach significance. Together, these results suggest that orthographic representations of words may play a more integral role in L2 word learning than mental imagery or embodied action.

Experiment 2

One possible explanation why L2 words learned via text were recalled better than words learned via iconic gesture, beat gesture, or images is because Hungarian words were presented as text in test trials. This learning-test similarity may have resulted in transfer-appropriate processing, which occurs when similar cues are present at encoding and retrieval (Morris, Bransford, & Franks, 1977). In order to determine whether transfer-appropriate processing was responsible for the facilitatory effect of text observed in Experiment 1, recall trials were modified so that Hungarian words were presented as speech only, without any text. If, under these conditions, participants recalled similar numbers of L2 words presented as text, representational gesture, beat gesture, and speech, it could be concluded that the facilitatory effect of text observed in Study 1 was due to transfer-appropriate processing. Given a lack of conclusive reported evidence showing that L2 words learned as text are recalled better than words presented via other modalities (e.g., images), it was predicted that no significant advantage of text would be found in Study 2.

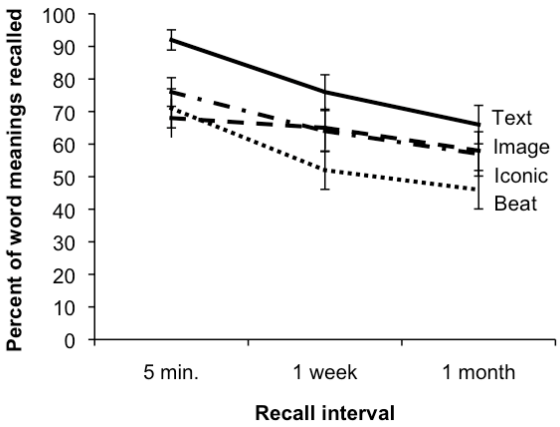

Figure 1: Percent of word meanings recalled by condition and recall interval for Expt. 1 (error bars represent *SEM*).

Methods

Twenty-six undergraduate students (age: $M = 20.64$; $SD = 1.56$; sex: 12 males; 14 females) at a medium-sized public university in the US participated in return for partial course credit. All participants were fluent English speakers and had no knowledge of Hungarian. Additionally, participants had not participated in Experiment 1.

Learning conditions and trials were identical to those used in Experiment 1. Recall trials were also identical to those of Experiment 1, except that Hungarian words were presented as speech only, while the task instructions were displayed on the screen as text.

Results and Discussion

As in Experiment 1, L2 word learning was quantified proportionally and was submitted to repeated measures ANOVAs with participant and word as fixed factors. This analysis revealed a main effect of recall interval, $F_{pp}(2, 51) = 7.11$, $p = .003$, $\eta_p^2 = .31$; $F_{word}(2, 10) = 47.14$, $p < .001$, $\eta_p^2 = .90$, see Figure 2. Bonferroni-corrected post-hoc analyses showed that participants recalled more words after five

minutes than they recalled after one week (p_{pp} = .04; p_{word} =.001) and one month (p_{pp} = .05; p_{word} = .004). However, learning condition failed to reach significance, as did the interval by condition interaction. These results indicate that the superior recall for words learned via text observed in Experiment 1 was due to the presence of text in both learning and test trials.

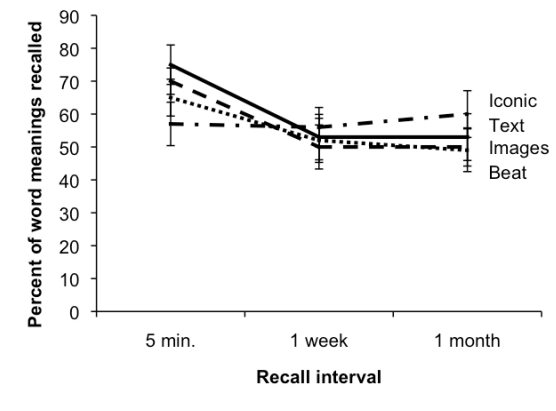

Figure 2: Percent of word meanings recalled by condition and recall interval for Expt. 2 (error bars represent *SEM*).

Experiment 3

Contrary to the predictions, Experiments 1 and 2 failed to show that iconic gesture, mental imagery, or beat gesture promote L2 word learning via mental imagery and embodied action. One possible reason why iconic gesture may not have produced the facilitatory effect observed in other similar studies of L2 word learning is that participants may have encountered difficulties mapping gestures onto the words that they were intended to represent. Despite the care that was taken in selecting gestures to represent target words based on gesture production data from the norming study, it is possible that the gestures chosen may logically map onto the meanings of more than one word (e.g., the gesture representing *to climb* may also map onto *ladder*). Unclear gesture-word mappings such as these may cause confusion, unnecessarily increasing cognitive load upon presentation of "correct" English glosses, which in turn may negatively impact L2 word learning and recall. Experiment 3 attempted to control for gesture-meaning mismatches by presenting English glosses prior to Hungarian words, in order to ensure that iconic gestures were immediately associated with their intended meanings.

In addition to the reversal of language presentation order, the text condition was replaced with a speech only condition. The rationale for this replacement was that speech with no visual representation may be a more appropriate baseline condition (- embodied action, - mental imagery) for this research than speech with text.

Methods

Twenty-seven undergraduate students (age: *M* = 21.05; *SD* = 1.83; sex: 13 males; 14 females) at a medium-sized public university in the US participated in return for partial course credit. All participants were fluent English speakers and had no knowledge of Hungarian. Additionally, participants had not participated in Experiments 1 or 2.

Learning trials were similar to those of Experiments 1 and 2, except that the order of language presentation was reversed (2000 ms. English gloss, 1000 ms. ISI, 2000 ms. Hungarian word). Learning conditions were also similar to those of Experiments 1 and 2, except that the text condition was replaced with a speech only condition, in which words were presented as speech concurrently with a blank screen. Recall trials were identical to those of Experiment 2.

Results and Discussion

As in Experiments 1 and 2, L2 word learning was quantified proportionally and was submitted to repeated measures ANOVAs with participant and word as fixed factors. This analysis revealed a main effect of recall interval, $F_{pp}(2, 53) = 24.81$, $p < .001$, $\eta_p^2 = .61$; $F_{word}(2, 14) = 36.59$, $p < .001$, $\eta_p^2 = .84$, see Figure 3. Bonferroni-corrected post-hoc analyses showed that participants recalled more words after five minutes than they recalled after one week (p_{pp} = .04; p_{word} = .002) and one month (p_{pp} = .04; p_{word} < .001). However, learning condition failed to reach significance, as did the interval by condition interaction. Taken together with the findings of Experiments 1 and 2, these results confirm that, in the passive form, mental imagery and embodied action do not significantly enhance the acquisition of L2 words.

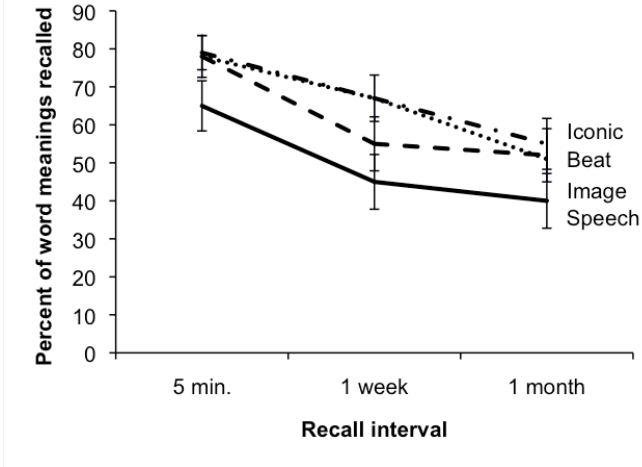

Figure 3: Percent of word meanings recalled by condition and recall interval for Expt. 3 (error bars represent *SEM*).

Experiment 4

The objectives of this experiment were to examine the effects of the active engagement in embodied motion and mental imagery on L2 word learning, and to compare them to the effects of their passive counterparts, as documented in Experiments 1-3. For young children, only enactment—not mere viewing—of iconic gestures enhances word learning (Tellier, 2008). Thus, it is possible that L2 word learning

may be enhanced via *enactment* of iconic gestures and *visualization* of mental imagery, even though word learning was not facilitated via passive *viewing* of iconic gestures and physical images. Experiment 4 tested this possibility by examining the effectiveness of L2 word learning via the active equivalents of the learning conditions of Studies 1-3.

Methods

Twenty-eight undergraduates (Age: M = 20.67; SD = 1.56; Sex: 12 males; 16 females) at a medium-sized public university in the US participated in return for partial course credit. All participants were fluent English speakers and had no knowledge of Hungarian. Additionally, participants had not participated in Experiments 1, 2, or 3.

As in Experiments 1-3, the learning phase of Experiment 4 consisted of three blocks of trials that varied by condition within participants. In learning trials, participants viewed video of a speaker saying an English gloss while producing a gesture representative of its meaning for 2000 ms., and after a 1000 ms. interstimulus interval, viewed video of the speaker saying a Hungarian target word while making the same gesture for 2000 ms. After one additional repetition of this sequence of events, participants repeated the English and Hungarian words aloud while performing the action corresponding to the condition to which the word had been assigned for that session. For words assigned to condition (1), gesture enactment, participants enacted the gesture that they had viewed in the video; for words assigned to condition (2), mental imagery formation, participants closed their eyes and visualized the words' meaning;[1] for words in condition (3), meaningless hand motion, participants made an X-shaped hand motion three times; for words in condition (4), repetition, participants repeated the words aloud while keeping their hands still. Recall trials were identical to those of Experiments 2-3.

Results and Discussion

As in Experiments 1-3, L2 word learning was quantified proportionally and was submitted to repeated measures ANOVAs with participant and word as fixed factors. This analysis revealed a main effect of recall interval, $F_{pp}(2, 56) = 12.90$, $p < .001$, $\eta_p^2 = .52$; $F_{word}(2, 20) = 10.37$, $p = .001$, $\eta_p^2 = .51$, see Figure 4. Bonferroni-corrected post hoc analyses revealed that participants recalled more words after five minutes than they recalled after one week ($p_{pp} = .05$; $p_{word} = .02$) or one month ($p_{pp} = .03$; $p_{word} = .01$). However, learning condition failed to reach significance, as did the interval by condition interaction. These results suggest that, although enactment facilitates L2 word learning by children, it does not seem to enhance L2 word learning by adults.

[1]The results of two participants who indicated in response to a question on the post-experimental survey that they did not follow the instructions regarding mental imagery were excluded.

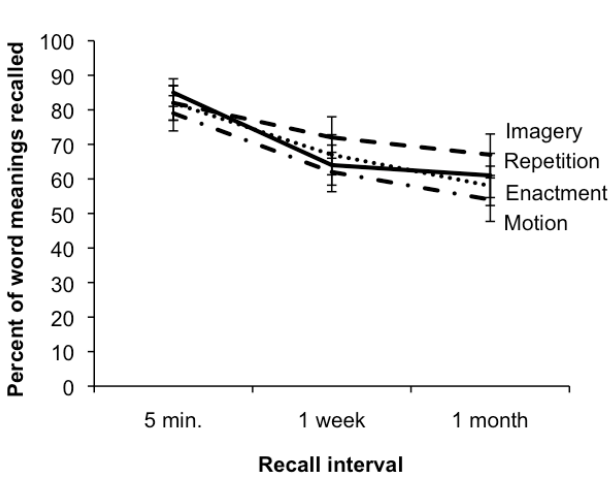

Figure 4: Percent of word meanings recalled by condition and recall interval for Expt. 4 (error bars represent *SEM*).

General Discussion

The present research examined the independent roles and interactions that passive and active engagement in mental imagery and embodied action play in L2 word learning. The lack of differences in word learning as a function of condition suggest that neither of these factors plays an integral role in L2 word learning by adults. Given that adults are particularly effective word learners (Snow & Hoefnagel-Höhle, 1978), it is likely that the participants of this study were using alternative methods to associate target words with their meanings, such as phonological association or generation. One notable difference between this study and others (Kelly et al., 2009) is that, in the interest of ecological validity, the stimuli accompanying the target words in the current study were either congruent or neutral to the word meanings. As a result, the variation in word learning between conditions in the current study was more modest than that observed in previous studies. This greater similarity between conditions may explain the lack of significant differences between conditions, as well as why the results of the current study failed to replicate those of previous work (e.g., Kelly et al., 2009; Tellier, 2008).

Surprisingly, the results failed to show that meaningful embodied action, in the form of representative iconic gesture, facilitated L2 word learning and recall more effectively than mental imagery. This finding is inconsistent with several previous studies showing superior recall of L2 words learned via viewing or enactment of iconic gesture (Kelly et al., 2009; Tellier, 2008) or enactment of meaningful motion using the body (Asher, 1969; Asher et al., 1974). There are several possibilities why iconic gesture may have failed to facilitate L2 word learning in the current research. One possibility is that the gestures chosen to represent target words may not have been sufficiently iconic, and thus, may not have been as imagistic and meaningful as gestures used in other studies. Another possibility is that the learning phase may have been too brief to allow for associations between gestures and target words

to be formed. Future research should test these possibilities by examining the acquisition of L2 words via imagistically-rich gesture during an extended learning period, and by comparing the results to those of the current research.

Overall, the results of the current research suggest that neither mental imagery nor embodied action plays a key role in L2 word learning by adults. More specifically, the results indicate that the viewing of iconic gesture, images, and text during L2 word learning result in comparable recall of target words across both long and short learning-test intervals. Finally, the results demonstrate that L2 words are recalled more accurately over short (5 min.) learning-test intervals than longer intervals (1 week, 1 month). Taken together, these findings fail to replicate the results of work showing that representative iconic gesture viewing and enactment enhance L2 word learning (Kelly et al., 2009; Tellier, 2008), suggesting that the word learning techniques of adult L2 learners are already so effective that mental imagery and embodied action have a negligible impact on them.

Acknowledgements

This research was supported by a National Defense Science and Engineering Graduate Fellowship and the Perlino Award to Laura M. Morett. The authors thank Jana Iverson for resources and helpful discussion.

References

Allen, L. Q. (1995). The effects of emblematic gestures on the development and access of mental representations of French expressions. *The Modern Language Journal*, *79*(4), 521–529.

Asher, J. J. (1969). The total physical response approach to second language learning. *The Modern Language Journal*, *53*(1), 3–17.

Asher, J. J., Kusudo, J. A., & Torre, R. de la. (1974). Learning a second language through commands: The second field test. *The Modern Language Journal*, *58*(1/2), 24–32.

Asher, J. J., & Price, B. S. (1967). The learning strategy of the Total Physical Response: Some age differences. *Child Development*, *38*, 1219–1227.

Atkinson, R. C. (1975). Mnemotechnics in second-language learning. *American Psychologist*, *30*(8), 821–828. doi:10.1037/h0077029

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 126–133. doi:10.1037/0278-7393.1.2.126

Carpenter, S. K., & Olson, K. M. (2011). Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0024828

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, *80*(2), 183–198.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*, 313–334. doi:10.1080/01690960802365567

Levin, J. R., McCormick, C. B., Miller, G. E., Berry, J. K., & Pressley, M. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies for children. *American Educational Research Journal*, *19*(1), 121 –136. doi:10.3102/00028312019001121

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. doi:16/S0022-5371(77)80016-9

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press US.

Raugh, M. R., & Atkinson, R. C. (1975). A mnemonic method for learning a second-language vocabulary. *Journal of Educational Psychology*, *67*(1), 1–16. doi:10.1037/h0078665

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

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C. C., et al. (2004). A new on-line resource for psycholinguistic studies. *Journal of Memory and Language*, *51*(2), 247–250. doi:10.1016/j.jml.2004.03.002

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, *8*, 219–235.

Van Hell, J. G., & Mahn, A. C. (1997). Keyword mnemonics versus rote rehearsal: Learning concrete and abstract foreign words by experienced and inexperienced learners. *Language Learning*, *47*(3), 507–546. doi:10.1111/0023-8333.00018