Journal of Psycholinguistic Research (2019) 48:643–664
https://doi.org/10.1007/s10936-018-9623-2

The Power of an Image: Images, Not Glosses, Enhance Learning of Concrete L2 Words in Beginning Learners

Laura M. Morett[1]

Published online: 2 January 2019

Abstract
Prior research suggests that viewing still images and iconic gestures depicting concepts facilitates the learning of concrete words in the initial stages of second language (L2) acquisition. To date, however, the effect of viewing iconic gestures and images hasn't been systematically compared to the effect of glosses in the learning and retrieval of concrete words in early stage L2 acquisition. Therefore, it is unclear whether dual coding theory of embodied theories of cognition provide the most accurate account of these effects. This study demonstrates that concrete L2 words learned via viewing still images are recalled better than L2 words learned via viewing iconic gesture and that L1 glosses fail to facilitate L2 word learning in beginning learners. Together, these findings indicate that images facilitate the learning of concrete L2 words above and beyond glosses in learners unfamiliar with the target language, and that glosses are not always necessary for effective L2 word learning.

Keywords L2 word learning · Gesture · Images · Embodied cognition · Dual coding

Introduction

Given that word learning is typically the first step in second language (L2) acquisition, a long-standing question in the field concerns the most effective way to encode L2 words in memory. A related, and equally long-standing, question concerns the most effective way to retrieve newly-learned L2 words from memory. Both of these processes are essential to successful L2 word learning. Initially, encoding L2 words entails associating phonological forms in the target language with corresponding meanings, and retrieving L2 words entails drawing upon these associations. At both encoding and retrieval, meanings of L2 words can be conveyed verbally via L1 glosses (translations) or nonverbally via images or gestures, as well as in a variety of other ways. Thus, a central question is how the manner in which L2 word meanings are conveyed affects their encoding and retrieval. This question has important practical implications for how L2 words and their meanings should

✉ Laura M. Morett
lmorett@ua.edu

[1] Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, Box 870240, Tuscaloosa, AL 35401, USA

be presented when word-meaning associations are encoded, as well as at test when word-meaning associations are retrieved.

The impact of the type and form of word meaning depictions on L2 word learning also has important implications for theories relevant to L2 lexical representation. Dual coding theory (Paivio 1990) postulates that the combination of images and corresponding verbal information at encoding enriches memory traces. Thus, this theory predicts that, relative to L1 glosses alone, the presence of still images in conjunction with L1 glosses at encoding should facilitate L2 word retrieval, regardless of which of these depictions are used to elicit L2 word retrieval. Alternatively, embodied theories of language processing (Barsalou 2008; Gibbs 2006) postulate that representations underlying language are composed of sensations from the body (e.g., sight, sound, touch, etc.) and that perception and action are intertwined. Thus, these models predict that *iconic gestures*—hand motions conveying meaning via form—depicting concepts should enhance both encoding and retrieval of concrete L2 words to a greater extent than still images or L1 glosses because these gestures evoke sensations through multiple modalities, enriching lexical representations.

In light of theoretical issues concerning how newly-learned L2 words are represented and practical issues concerning how they are encoded in and retrieved from memory, the present study addresses three primary questions: (1) Do L1 glosses enhance L2 word encoding and/or retrieval?; (2) Does the combination of L1 glosses and nonverbal depictions of word meanings enhance L2 word encoding and retrieval beyond either of these depictions alone?; (3) Which type(s) of nonverbal word meaning depictions (still images or iconic gestures) result in the most effective L2 word encoding and retrieval? To address these questions, the presence of glosses, images, and iconic gestures conveying word meanings during the encoding and retrieval of novel concrete L2 words was manipulated. It was predicted that while L1 glosses would not facilitate L2 word encoding and that they would not be necessary for retrieval (i.e., recall), they would nevertheless facilitate retrieval. Moreover, based on dual coding theory, it was predicted that the presence of glosses in conjunction with nonverbal depictions of word meaning at encoding would enhance retrieval beyond the effects of gesture and still images alone. Finally, based on research demonstrating that viewing iconic gestures enhances beginning learners' retrieval of concrete L2 words (Kelly et al. 2009; Macedonia and Knösche 2011; Macedonia et al. 2010; Morett 2014; Tellier 2008), it was predicted that viewing iconic gestures would result in more effective L2 word encoding and retrieval than still images.

Aside from L1 glosses, nonverbal depictions of referents such as still images, videos, and gestures are often used to convey the meanings of L2 words to learners. These nonverbal depictions are particularly conducive to conveying the meanings of concrete (vs. abstract) words. Given that successful L2 word learning via nonverbal referent depictions requires association of these depictions with corresponding words via their meanings, theories explaining the role of imagery in L2 word learning must be able to account for the effects of both nonverbal depictions and glosses at encoding and retrieval. One such theory that accounts for these effects is dual coding theory (Paivio 1990). Dual coding theory posits that information can be simultaneously encoded via both the visual and verbal channels and that such encoding results in more effective retrieval. This is the case because visual codes are *analogue*, possessing perceptual features obtained via the senses, whereas verbal codes are *symbolic*, representing information conceptually without such features.

Much of the evidence supporting dual coding theory comes from research in memory and the learning sciences. This work has demonstrated that complex concepts are learned and remembered more effectively when they are presented to learners via both explanatory text and graphics than when they are presented via text alone (Mayer 2002; Mayer and

Moreno 2003). Dual coding theory has also been invoked to explain the positive effects of imagery on word learning, particularly in L2. One such effect involving mental imagery has been observed with use of the keyword method, in which learners envision the referent of an L2 word interacting with the referent of a phonologically similar L1 word (e.g., to learn the English–Spanish word pairing "boy"-"chico," the learner envisions a boy petting a chick) (Sagarra and Alba 2006; Van Hell and Mahn 1997). Evidence concerning the effect of passively viewing physical images on L2 word learning is mixed, however, with some studies showing a facilitatory effect (Carpenter and Olson 2011; Chun and Plass 1996), and others showing no effect or a negative effect (Chen 1990; Lotto and De Groot 1998), despite use of concrete words in all cases.

Another class of theories that takes the effects of imagery on L2 word learning into account is embodied theories of cognition (Barsalou 2008; Gibbs 2006). Unlike dual coding theory, embodied theories postulate that representations of both language and concepts are analogue and are composed of perceptual features obtained via the senses. Thus, a representation of the word *apple* consists of the sight, sound, taste, feel, and smell of an apple rather than a holistic symbolic concept of an apple. Because embodied theories of cognition postulate that perception and action are intertwined, these percepts can be activated directly by experiencing them or producing relevant actions with one's own body, as well as indirectly by viewing other people experiencing them or producing relevant actions with their bodies, although the effect of vicarious perception may be weaker. On a conceptual level, embodied theories of cognition postulate that language, imagery, and action are fully integrated rather than merely associated with one another. Thus, lexical representations are directly linked to the senses via these perceptual features, and this linkage is particularly strong during the initial stages of word learning when representations are novel and vivid and sensations of referents are often recent (Yu and Smith 2012). Moreover, embodied theories incorporate body movements in addition to visual imagery and language. Thus, these theories predict that viewing gesture should facilitate L2 word learning to a greater extent than viewing still images, whereas dual coding theory predicts that viewing gesture will not produce such an additional benefit. Finally, embodied theories propose that relevant body movements, such as iconic gestures depicting L2 word meanings, will produce episodic memories resistant to interference and conducive to retrieval over extended time periods (Wilson 2002). Thus, these theories predict that L2 words accompanied by such gestures will be retrieved more effectively than L2 words accompanied by images or glosses, particularly when considerable time elapses between encoding and retrieval.

In the domain of L2 word learning, considerable support for embodied theories of cognition comes from research examining the effect of viewing iconic gestures, which convey referent properties via handshape and motion, at encoding on word retrieval. In a seminal study investigating how iconic gesture impacts L2 word learning, Allen (1995) found that English-speaking college students who viewed iconic gestures when encoding the meanings of French expressions remembered their English meanings better than students who did not view gestures. More recent work has expanded upon this finding by showing that viewing iconic gestures depicting unrelated word referents at encoding hinders L2 word retrieval (Kelly et al. 2009), and that viewing meaningless gestures and gestures tracing the shapes of referents at encoding is not as effective at promoting L2 word retrieval as viewing iconic gesture conveying the meanings of L2 words at encoding (Macedonia and Knösche 2011; Macedonia et al. 2010; Mayer et al. 2015). Even when videos of iconic gestures depicting L2 word referents shown at encoding feature anthropomorphic virtual agents rather than humans, these gestures enhance L2 word retrieval (Bergmann and Macedonia 2013). In all of these more recent studies,

semantically congruent iconic gestures viewed at encoding enhanced recollection of L2 words over an extended period of time up to 6 months in length, providing evidence that these gestures enhance long-term memory for newly-learned L2 words. It is worth noting, however, that some research has failed to show evidence that viewing iconic gestures depicting L2 word referents at encoding enhances memory for L2 words (Krönke et al. 2013) or that the effect on memory of viewing iconic gestures at encoding depicting the meanings of L2 words is weaker than the effect of producing such gestures at encoding (Morett 2018). Given that the words tested in these studies vary in concreteness and part of speech, this finding suggests that the effect of iconic gesture on recollection may vary according to the characteristics of L2 words.

To date, only two published studies have directly compared the predictions of dual coding theory and embodied theories of cognition within the realm of L2 word learning, though neither was framed as such. In one of these studies (Tellier 2008), 4–5 year-old French-speaking children learned concrete English nouns and verbs by viewing either still images or iconic gestures depicting their referents in a between-participants design. Consistent with the predictions of embodied theories of cognition, the results showed that recall of words learned by viewing iconic gestures was superior to recall of words learned by viewing still images. However, no difference in recognition of word meanings by condition was observed. In another similar study with a larger sample size (Rowe et al. 2013), English-speaking preschool and kindergarten children (mean age: 4 yrs. 8 mos.) learned words for concrete objects from an artificial language in a within-subjects design via presentation of either translations alone, translations and still images depicting word meanings, or translations and iconic gestures depicting word meanings. In contrast with the findings of Tellier (2008), L2 word recall did not differ by condition, but L2 word recognition did. Specifically, consistent with the predictions of dual coding theory, monolingual children with lower language ability recognized L2 words that they had learned via viewing still images with greater accuracy than L2 words that they had learned via viewing iconic gestures, whereas recognition did not differ by condition for monolingual children with higher language ability. In both Tellier (2008) and Rowe et al. (2013), the respective effects of iconic gesture and still images on word recollection persisted for at least a week after learning, providing evidence that representative iconic gestures and still images enhance memory for newly-learned words over extended periods of time. However, the differing results of these two studies suggest that task and learner characteristics may influence whether embodied cognition or dual coding theory is best able to account for the impact of nonverbal referent depictions on L2 word learning.

The purpose of the current study was to test how well dual coding theory and embodied theories of cognition account for how L1 glosses, still images, and iconic gestures affect the learning of concrete words in early stage L2 acquisition. While most previous research on L2 word learning has tested the predictions of these theories separately, to date, the predictions of these theories have not been examined directly in relation to one another. Given that dual coding theory makes clear predictions regarding the effect of relevant imagery in combination with verbal information and that embodied theories of cognition make clear predictions regarding the effect of relevant iconic gesture, it is important to account for the effects of both types of depictions at encoding in evaluating how well these theories describe L2 word learning. In light of the ubiquity of glosses, still images, and iconic gestures, any comprehensive theory of L2 word learning—particularly one encompassing the early stages of L2 acquisition, when many words learned are concrete—must be able to account for the effects of each of these types of depictions on encoding and retrieval of L2 words.

To test the extent to which dual coding theory and embodied theories of cognition can account for early stage L2 word learning, the manner in which L2 words were *encoded* (i.e., how words and their meanings are entered into memory) and *retrieved* (i.e., how words and their meanings are extracted from memory) were manipulated. L2 word encoding was manipulated by presenting the meanings of words via L1 glosses, still images, or iconic gestures. L2 word retrieval was manipulated by using still images and iconic gestures or L1 glosses as cues both alone and in combination in a series of four tests progressively decreasing in difficulty. This study is the first examining the effects of these stimuli on L2 word learning to manipulate their presence at retrieval in addition to encoding. In the free recall test, which was administered first, all words were recalled in the absence of any cues; in the depiction test, which was administered second, words learned via still images or iconic gestures were recalled using only these depictions as cues; in the translation test, which was administered third, all words were recalled using only glosses; in the depiction-plus-translation test, which was administered fourth, words learned via gestures or still images were recalled using a combination of these depictions and glosses. The compounding of cues in retrieval tests allowed us to distinguish between whether L1 glosses were necessary for successful L2 word recall, whether they enhanced it, or whether they failed to do so. In all four tests, memory for L2 words was tested via recall to avoid possible ceiling effects for L2 word recognition with adult participants, which may mitigate the effects of verbal and nonverbal means of conveying L2 word meanings on the encoding and/or retrieval of these words. To gauge the impact of cues presented at encoding on both short- and long-term memory for newly-learned L2 words, all four recall tests were administered 5 min and 1 week following word learning.

In light of the preponderance of findings favoring embodied cognition and the mixed findings concerning dual coding theory, particularly for early stage L2 word learning, it was hypothesized that viewing iconic gestures depicting word meanings at encoding would enhance beginning learners' L2 retrieval of concrete nouns and verbs more effectively than viewing still images or L1 glosses alone at encoding, particularly following a 1-week delay. Moreover, it was hypothesized that L1 glosses would enhance beginning L2 learners' retrieval of word meanings but that they would not be strictly necessary for retrieval because word meanings can also be accessed via imagery and embodied action. Thus, it was predicted that more L2 words learned by viewing still images and iconic gestures would be recalled in a test including both nonverbal depictions of L2 word meanings and L1 glosses than would be recalled in a test including only nonverbal depictions of L2 word meanings. Furthermore, it was predicted that, in tests including only L1 glosses and both nonverbal depictions and L1 glosses, words learned via L1 glosses would not be recalled any better than words learned by viewing still images or iconic gestures. Aside from adjudicating between dual coding theory and embodied theories of cognition, the results hold important implications for how L2 words can be taught and tested in a manner that will enhance their recall by beginning L2 learners in instructional settings. As such, they speak to the issue of the most effective means of learning L2 words in the initial stage of L2 acquisition, helping to inform the practices of students and instructors alike.

Method

Participants

Twenty-eight undergraduate students (14 males; 14 females; mean age = 18.70; SD = 2.02) at a medium-sized public university in the US participated in return for partial course credit. All participants had normal or corrected-to-normal vision. All participants were fluent English speakers with no knowledge of Hungarian, the language of the L2 words used in this study.

Materials

Twenty Hungarian words and their English glosses were selected for use in this study (see Table 1). These words consisted of ten concrete words whose English glosses have been rated as highly imageable and concrete by native English speakers (Brysbaert et al. 2014). These words were selected to ensure that their meanings could be depicted clearly via still images and iconic gestures and expressed succinctly via glosses, that they were typical of words learned in the earliest stages of L2 learning, and that they were as similar as possible to words used in other studies examining the effects of imagery on L2 word learning.

Three depictions of each word's meaning were created: a gloss, an image, and an iconic gesture. Glosses, which were displayed as text, were the L1 translation of each L2 word. Images were photos of each L2 word's meaning. These photos were obtained by conducting a Google Image search for free-use images using L1 glosses as search terms and by selecting a photo that clearly depicted the meaning of each gloss. Iconic gestures were videos of gestures depicting each L2 word's meaning via handshape and motion. These videos were produced by recording a Hungarian-English bilingual producing an iconic gesture conveying an action associated with the meaning of each word (see Table 1 for descriptions). In all conditions, Hungarian words consisted of audio clips of them being pronounced by the same Hungarian-English bilingual recorded in iconic gesture videos. Moreover, in all conditions, English translations of Hungarian words consisted of audio clips of translations being pronounced via synthesized speech to ensure acoustic uniformity and avoid any affective or accentual cues.

Procedure

To learn the Hungarian words, participants first completed two blocks comprising 15 trials apiece (one trial per word for a randomly-selected subset of 15 of the 20 possible words). In each trial, a Hungarian word was presented aurally, a 1000 ms interstimulus interval occurred, the English translation of the Hungarian word was presented aurally, and this sequence was repeated. Depictions of word meanings were presented visually concurrent with auditory presentation of both Hungarian words and English translations.

To obtain maximum possible statistical power, this study used a within-participants design; thus, for each participant, five Hungarian words and their English translations were assigned at random to each of the following learning conditions: (1) iconic

Table 1 Hungarian words and iconic gestures presented in learning test with English glosses

Hungarian word	English gloss	Iconic gesture description
madár	Bird	Pantomimes flapping by moving fingers back and forth with one hand in front of the other and fingers oriented horizontally
ablak	Window	Pantomimes closing window by moving hands down simultaneously while closing fingers into fists
kés	Knife	Pantomimes chopping by moving right hand downward quickly three times over left hand; left hand held flat with palm up
harang	Bell	Pantomimes ringing bell by wagging right fist back and forth horizontally
könyv	Book	Pantomimes opening book by placing hands together and opening them outward horizontally
seprű	Broom	Pantomimes sweeping by moving fists held one above the other back and forth horizontally simultaneously three times
zászló	Flag	Pantomimes waving flag by holding right fist up and waving it back and forth horizontally in wide motions
kalapács	Hammer	Pantomimes hammering by swinging right fist downward three times above left hand held flat with palm oriented up
kulcs	Key	Pantomimes turning key in lock with right fist
kesztyű	Glove	Pantomimes putting glove on; left hand oriented vertically with palm inward and fingers spread, right hand moves downward parallel to left hand and arm while closing fingers
fogott	To grab	Pantomimes grabbing by extending right hand and then moving it inward towards body abruptly while closing fingers
csókolt	To kiss	Pantomimes kissing
emett	To lift	Pantomimes lifting by moving both hands upward simultaneously with palms upward
evett	To eat	Pantomimes spooning food while opening and closing mouth with right hand over left hand held flat with palm up
festett	To paint	Pantomimes painting by moving closed right hand up and down with broad strokes
mosott	To wash	Pantomimes scrubbing by placing hands with palms together and moving in quick circular motions against one another

Table 1 (continued)

Hungarian word	English gloss	Iconic gesture description
kötött	To tie	Pantomimes tying a bow by moving hands in small circular motions near one another simultaneously and then moving them outward horizontally simultaneously
dobott	To throw	Pantomimes throwing by pulling right hand back and then pushing it forward quickly while opening the fingers
ivott	To drink	Pantomimes drinking by cupping hand and moving in upward arc towards mouth
nyitott	To open	Pantomimes opening door by moving right hand in arc from left to right horizontally with fingers cupped

Table 2 Interrater reliability by test

Test	Cohen's κ	*N*
Free recall	0.70	540
Depiction	0.82	402
Translation	0.84	535
Depiction-plus-translation	0.87	379

gesture, in which each word was accompanied by a video of an iconic gesture depicting its meaning; (2) still image, in which each word was accompanied by a photograph depicting its meaning; (3) gloss, in which each word was accompanied by its L1 gloss presented as text.

After learning the words, participants completed four recall tests: free, depiction, translation, and depiction-plus-translation. These tests were always presented in this order, given that it represents the number of recall cues available in increasing magnitude (and, thus, difficulty in decreasing magnitude).

In the free recall test, participants recalled as many Hungarian words and corresponding English translations as they could by saying them aloud in any order with no time limit. In the depiction test, nonverbal depictions presented with words at learning were re-presented as recall cues. This test evaluated recall of words learned in the still image and iconic gesture conditions but not the gloss condition.[1] In the translation test, L1 glosses of words learned in all conditions were presented as text. In the depiction-plus-translation recall test, the nonverbal depictions presented with words at learning and text L1 glosses of words were presented simultaneously. Similar to the depiction test, the depiction-plus-translation test evaluated recall of words learned in the still image and iconic gesture conditions but not the gloss condition.

In the depiction, translation, and depiction-plus-translation tests, participants responded by saying the Hungarian word corresponding to the nonverbal depiction and/or text gloss presented or by saying "skip" if they could not remember the word. Participants completed all tests 5 min and, to examine how learning conditions affected extended recall, 1 week following the learning phase.

Scoring

For all tests, L2 word recall was quantified by scoring responses using a binary coding scheme (1 = correct, 0 = incorrect/skipped). Raters were blind to the conditions in which words were learned. Because participants were unfamiliar with Hungarian, words were scored as correct as long as their pronunciation resembled that of the target word and did not conflict with that of another word.

All responses were scored by a single individual (the author) as well as a second individual unaware of the purpose and design of the experiment. Using Landis and Koch's interpretation, agreement between the first and second coders using Cohen's Kappa was substantial for the free recall test, excellent for the depiction test, excellent for the depiction

[1] Because still images and iconic gestures were used to test L2 word recall in the depiction and depiction-plus-translation tests, recall of words presented in the translation condition during the learning phase could not be evaluated in these tests.

test, and excellent for the depiction-plus-translation test (see Table 2). 0.02% of the data was unscorable due to factors such as unintelligibility or technical errors.

Results

Given that all responses examined in this study were binary (correct vs. incorrect/skipped), multi-level logit models were used to model the log odds of correct responses. The multi-level logit models described below were fit using the *R* statistical programming language (R Core Team 2015) with Laplace estimation using the glmer() function of the lme4 package (Bates et al. 2015). These models included random effects of participants and items as well as fixed effects for independent variables of interest and their interactions as specified below for each analysis. Orthogonal planned contrasts corresponding to the hypotheses were used to determine the direction of observed effects because, similar to level-by-level comparisons, they partition the total variance attributed to each condition. Thus, these models maximized the statistical power of the limited number of participants and items tested in this experiment.

Comparison of Recall Tests

It was first examined whether memory for L2 words differed by recall test. To do so, a model was fit that used recall test (free recall, depiction, translation, depiction-plus-translation), study-test interval (5 min, 1 week), and the interaction of recall test and study-test interval as fixed factors. All fixed effects were coded using mean centered (Helmert) contrast coding to obtain main effects analogous to those that would be obtained from ANOVA.

One contrast comparing recall at 5 min to recall at 1 week was used to describe study-test interval. Three contrasts were used to describe recall test: The first compared recall in the depiction test to free recall; the second compared recall in the translation test to the free recall and recall in the depiction tests; and the third compared recall in the depiction-plus-translation test to recall in all of the other tests. The first of these comparisons was made to determine whether nonverbal depictions alone (present in the depiction test) facilitated word recall beyond the lack of any depictions (present in the free recall test). The second of these comparisons was made to determine whether text glosses alone (present in the translation test) facilitated word recall beyond nonverbal depictions alone (present in the depiction test) and the lack of any depictions (present in the free recall test). The third of these comparisons was made to determine whether text glosses in combination with nonverbal depictions (present in the depiction-plus-translation test) facilitated word recall beyond text glosses alone (present in the translation test), nonverbal depictions alone (present in the depiction test), and the lack of any depictions (present in the free recall test).

Three parallel contrasts were used to describe whether L2 word recall differed by recall test as a function of study-test interval. The first compared the difference between recall in the depiction and free recall tests at 5 min and 1 week; the second compared the difference between recall in the translation and the free recall and depiction tests at 5 min and 1 week; and the third compared the difference between recall in the depiction-plus-translation test and all other tests at 5 min and 1 week. These comparisons were made for a similar purpose as the comparisons for learning conditions described above while also examining differences by study-test interval. This model included random slopes for test

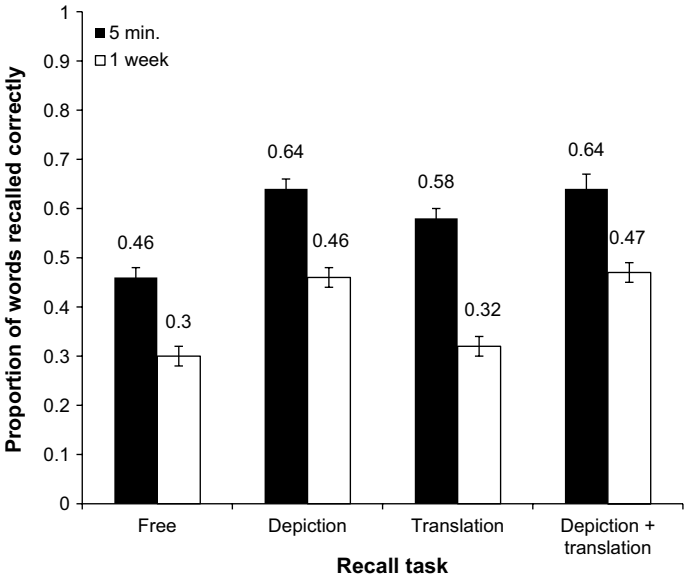

Fig. 1 Mean recall of Hungarian words as a function of study-test interval and recall test (error bars represent SE)

Table 3 Fixed effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall across tests by study-test interval (observations = 2671, log-likelihood: − 1489.4)

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.11	0.27	0.43	.67
5 min interval (vs. 1 week)	0.61	0.08	7.80	< .001
Depiction (vs. free recall)	0.76	0.14	5.61	< .001
Translation (vs. free recall and depiction)	0.64	0.11	5.95	< .001
Depiction-plus-translation (vs. others)	0.44	0.12	3.76	< .001
Depiction test (vs. free recall) * 5 min (vs. 1 week)	0.59	0.13	4.39	< .001
Translation (vs. free recall and depiction test) * 5 min (vs. 1 week)	− 0.12	0.11	− 1.15	.25
Depiction-plus- translation (vs. others) * 5 min (vs. 1 week)	− 0.19	0.12	− 1.66	.096
Random effect				s^2
Participant				1.08
Participant * test				0.31
Word				0.72

by participant;[2] all other fixed effects were modeled using random intercepts. Mean recall is displayed in Fig. 1 as a function of recall test and study-test interval.

[2] These random slopes were justified by a linear regression test. Given that they contributed significantly to the model, excluding them results in an anticonservative test that inflates the rate of Type I error (Barr et al. 2013). In models excluding random slopes, all contrasts reaching or trending towards significance in this and other models remained so.

As can be seen from Table 3,[3] the tests varied in difficulty: Across study-test intervals, the odds of correct L2 word meaning recall were 2.13 times (95% CI: [1.64; 2.78]) greater in the depiction test than in the free recall test, 1.89 times (95% CI: [1.53; 2.33]) greater in the translation test than in the free recall and depiction tests, and 1.54 times (95% CI: [1.23; 1.94]) greater in the depiction-plus-translation recall test than in the free recall, depiction, and translation tests. In addition, performance declined over time: across recall tests, the odds of correct L2 word meaning recall were 1.84 times (95% CI: [1.58; 2.15]) greater after 5 min than they were after 1 week. There was also one significant and one marginal interaction for recall test by study-test interval. The significant interaction indicated that the difference in recall between the depiction test and the free recall test was 1.80 times (95% CI: [1.39; 2.35]) greater after 5 min than it was after 1 week. The marginal interaction indicated that, after 5 min, the difference in recall between the depiction-plus-translation test and all other tests was 0.83 times (95% CI: [0.66; 1.04]) what it was after 1 week. Consistent with dual coding theory, these results indicate that, relative to no cues (present in the free recall test), both images (present in the depiction and depiction-plus-translation tests) and lexical cues (present in the translation and depiction-plus-translation tests) enhance recall of concrete L2 words. Moreover, they indicate that these differences in recall of concrete L2 words resulting from the presence of imagery and lexical cues at encoding increase with longer study-test intervals.

To examine the relationship between cues presented at encoding and retrieval, it was examined whether memory for L2 words varied by learning condition depending on recall test. To do so, a model was fit that used recall test (free recall, depiction, translation, depiction-plus-translation), learning condition (still image, iconic gesture[4]), and the recall test by learning condition interaction as fixed factors. Again, all fixed effects were coded using mean centered (Helmert) contrast coding.

One contrast comparing recall of words learned by viewing still images to recall of words learned by viewing iconic gesture was used to describe learning condition. The purpose of this contrast was to determine whether viewing iconic gestures depicting the meanings of L2 words enhanced their learning beyond viewing still images depicting their meanings, as postulated by embodied theories of cognition (Barsalou 2008; Gibbs 2006). The same three contrasts used in the previous model were used to describe recall test: the first compared recall in the depiction test to the free recall test; the second compared recall in the translation test to the free recall and depiction tests; and the third compared recall in the depiction-plus-translation test to recall in all of the other tests. Again, the purpose of these contrasts was to determine whether glosses were necessary for L2 word recall, as well as whether they enhanced it. Three parallel contrasts were used to describe how recall varied by learning condition. The first compared the difference between recall in the depiction test and the free recall test under the still image condition to the difference between recall in the depiction test and the free recall test under the iconic gesture condition; the second compared the difference between recall in the translation test and both the free recall and the depiction tests under the still image condition to the difference between recall in the

[3] All model statistics reported in tables reflect log odds, upon which logit mixed effects models are based. For ease of interpretation, these are converted in the text to odds.

[4] Recall of words learned in the translation only condition couldn't be assessed in the depiction and depiction-plus-translation tests because words assigned to this condition were learned without nonverbal depictions. Thus, the translation only condition was dropped from this comparison due to its absence from these tests.

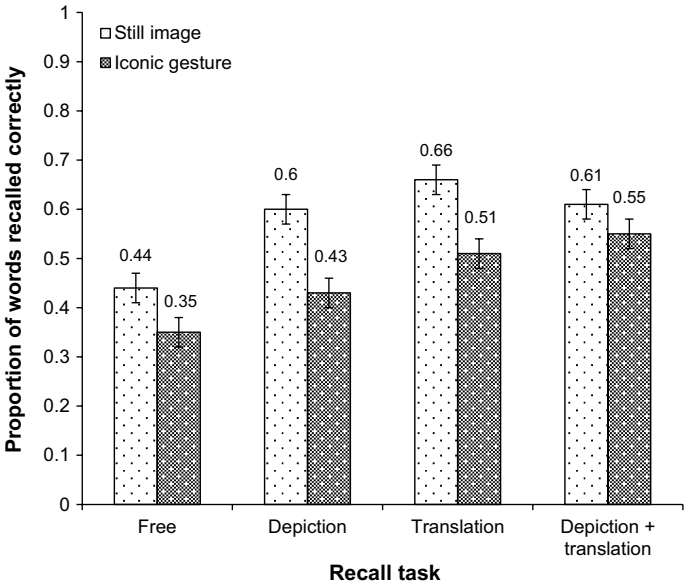

Fig. 2 Mean recall of Hungarian words as a function of learning condition and recall test (error bars represent SE)

translation test and both the free recall and depiction tests under the iconic gesture condition; the third compared the difference between recall in the depiction-plus-translation test and recall in all other tests under the still image condition to recall in the depiction-plus-translation test and recall in all other tests under the iconic gesture condition. This model included random slopes for condition by item and depiction test and depiction-plus-translation test by participant; all other fixed effects were modeled using random intercepts. Mean recall is displayed in Fig. 2 as a function of learning condition and recall test.

As can be seen in Table 4, the results of this analysis confirmed that the tests varied in difficulty: across learning conditions, the odds of word recall were 1.95 times (95% CI: [1.44; 2.63]) greater in the depiction test than in the free recall test, 1.95 times (95% CI: [1.44; 2.63]) greater in the translation test than in the free recall and depiction tests, and 1.95 times (95% CI: [1.54; 2.47]) greater in the depiction-plus-translation test than in all other recall tests. Word recall also differed by learning condition: the odds of word recall were 1.76 times (95% CI: [1.32; 2.35]) greater in the still image viewing condition than in the iconic gesture viewing condition. None of the interactions between recall test and learning condition reached significance. In contrast to the predictions of embodied theories of cognition, these results indicate that concrete L2 words learned by viewing still images depicting their meanings are recalled better than words learned by viewing iconic gestures depicting their meanings, regardless of the cues present at recall.

Free Recall Test

Having examined differences in memory for L2 words and their meanings across all recall tests, it was next examined whether word recall differed by study-test interval and learning

Table 4 Fixed Effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall across tests by learning condition (observations = 2134, log-likelihood: − 1264.7)

Fixed effect	Coefficient	SE	Wald *z*	*p*
Intercept	0.08	0.24	0.33	.74
Still image (vs. iconic gesture)	0.57	0.15	3.84	< .001
Depiction (vs. free recall)	0.67	0.15	4.36	< .001
Translation (vs. free recall/depiction)	0.67	0.12	5.49	< .001
Depiction-plus-translation (vs. others)	0.40	0.13	3.15	.002
Depiction (vs. free recall) * still image (vs. iconic gesture)	0.43	0.28	1.53	.13
Translation (vs. free recall/depiction) * still image (vs. iconic gesture)	0.11	0.24	0.44	.66
Depiction-plus-translation tests (vs. others) * still image (vs. iconic gesture)	− 0.25	0.23	− 1.08	.28
Random effect				s^2
Participant				0.94
Participant * depiction				0.15
Participant * depiction-plus-translation				0.06
Word				0.63
Word * still image				0.46

condition in each recall test. To examine whether L2 word recall differed by study-test interval and learning condition in the free recall test, in which participants freely recalled Hungarian words and their English meanings in the absence of any cues, a model was fit that used learning condition (iconic gesture, still image, gloss), study-test interval (5 min, 1 week), and learning condition by study-test interval interaction as fixed factors. Additionally, part of speech was included as an additional fixed factor to determine whether free recall for nouns and verbs differed.

One contrast comparing recall at 5 min to recall at 1 week was used to describe study-test interval. Two contrasts were used to describe learning condition: the first compared recall in the still image condition to recall in the iconic gesture condition; and the second compared recall in the gloss condition to recall in both the iconic gesture and still image conditions. The first contrast was made to determine whether, consistent with embodied theories of cognition (Barsalou 2008; Gibbs 2006), relevant iconic gesture enhanced word learning to a greater degree than relevant still images. The second contrast was made to determine whether glosses are necessary for word recall. Likewise, two contrasts were used to describe how word recall varied by learning condition as a function of study-test interval: the first compared the difference between recall in the still image condition at 5 min and 1 week to recall in the iconic gesture condition at 5 min and 1 week; and the second compared the difference between recall in the gloss condition at 5 min and 1 week to recall in the iconic gesture and still image conditions at 5 min and 1 week. This model included random slopes for study-test interval by participant; all other random effects were modeled using random intercepts. Mean recall by learning condition and study-test interval is displayed in Fig. 3a.

As can be seen from Table 5, across study-test intervals, recall odds were 1.23 times (95% CI: [0.99; 1.52]) greater in the still image condition than in the iconic gesture condition. Moreover, across learning conditions, the odds of correct L2 word meaning recall were 1.52 times (95% CI: [1.16; 1.98]) greater after 5 min than 1 week, indicating that

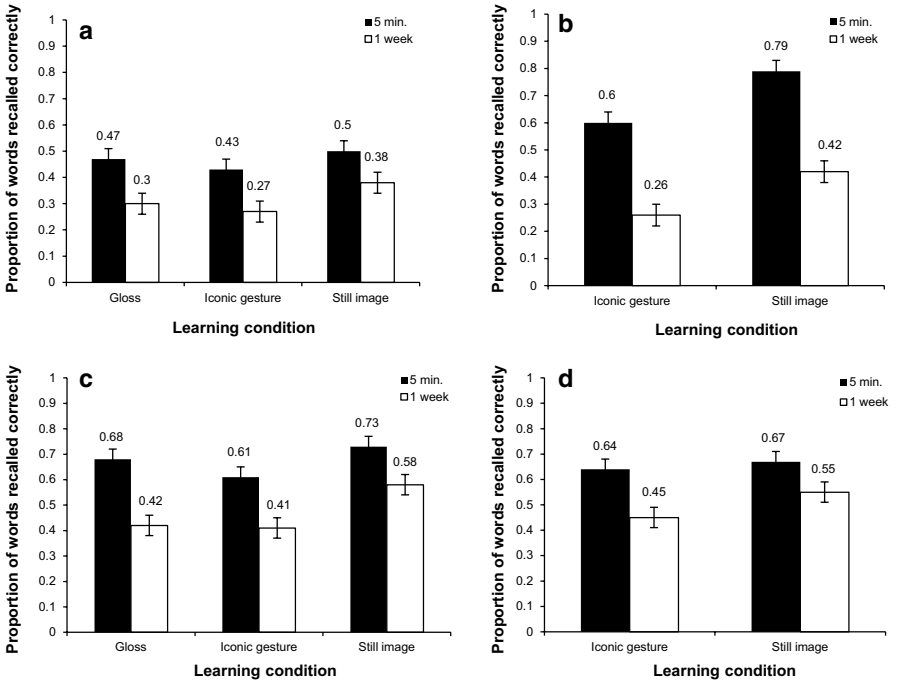

Fig. 3 Mean recall of Hungarian words in **a** the free recall test; **b** the depiction test; **c** the translation test; **d** the depiction plus translation test as a function of learning condition and study-test interval (error bars represent SE)

Table 5 Fixed effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall in the free recall test (observations = 810, log-likelihood: −453.7)

Fixed effect	Coefficient	SE	Wald *z*	*p*
Intercept	−0.63	0.32	−1.98	.047
5 min interval (vs. 1 week)	0.42	0.14	3.08	.002
Still image (vs. iconic gesture)	0.20	0.11	1.89	.06
Translation (vs. others)	−0.01	0.06	−0.24	.81
Still image (vs. iconic gesture) * 5 min (vs. 1 week)	−0.04	0.11	−0.41	.68
Translation (vs. others) * 5 min (vs. 1 week)	0.02	0.06	0.34	.73
Random effect				s^2
Participant				1.21
Participant * 5 min interval				0.52
Word				0.86

some forgetting occurred between sessions. There were no significant interactions between study-test interval and any of the learning conditions. In contrast to embodied theories of cognition, these results suggest that concrete L2 words and their meanings are freely recalled more accurately when they are learned by viewing still images than by viewing

Table 6 Fixed effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall in the depiction test (observations = 539, log-likelihood: − 317.4)

Fixed effect	Coefficient	SE	Wald *z*	*p*
Intercept	0.10	0.19	0.54	.59
5 min interval (vs. 1 week)	0.88	0.11	8.33	< .001
Still image (vs. iconic gesture)	0.45	0.10	4.29	< .001
5 min interval (vs. 1 week) * still image (vs. iconic gesture)	0.05	0.10	0.45	.65
Random effect				s^2
Participant				0.76
Word				0.33

iconic gestures. Furthermore, these results indicate that concrete L2 words learned via glosses were not recalled any more accurately than words learned by viewing still images or iconic gestures.

Depiction Test

To examine whether L2 word recall differed by study-test interval and learning condition in the depiction test, in which participants saw nonverbal depictions of word meanings presented at learning and produced corresponding Hungarian words, a model was fit that used learning condition (iconic gesture, still image), study-test interval (5 min, 1 week), and the learning condition by study-test interval interaction as fixed factors. One contrast comparing recall at 5 min to recall at 1 week was used to describe study-test interval. One contrast comparing recall in the still image condition to recall in the iconic gesture condition was used to describe learning condition. One contrast comparing the difference between recall in the still image condition at 5 min and 1 week and the iconic gesture condition at 5 min and 1 week was used to describe how word recall in different learning conditions varied by study-test interval. No random slopes reached significance by participant or item; thus, all random effects were modeled using random intercepts. Mean recall by learning condition and study-test interval is displayed in Fig. 3b.

As can be seen from Table 6, across study-test intervals, the odds of correct L2 word meaning recall were 1.56 times (95% CI: [1.27; 1.92]) greater in the still image condition than the iconic gesture condition. Across learning conditions, the odds of correct L2 word meaning recall were 2.41 times (95% CI: [1.96; 2.97]) greater after 5 min than 1 week. There was no significant interaction between learning condition and study-test interval. In contrast to embodied theories of cognition, these results again indicate that concrete L2 words learned and recalled by viewing still images depicting their meanings are remembered more accurately across study-test intervals than words learned by viewing iconic gestures depicting their meanings.

Table 7 Fixed effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall in the translation test (observations = 806, log-likelihood: −430.0)

Fixed effect	Coefficient	SE	Wald z	p
Intercept	0.46	0.37	1.25	.21
5 min interval (vs. 1 week)	0.63	0.09	6.80	< .001
Still image (vs. iconic gesture)	0.38	0.11	3.27	.001
Translation (vs. others)	−0.11	0.07	−1.72	.08
Still image (vs. iconic gesture) * 5 min (vs. 1 week)	−0.06	0.11	−0.53	.59
Translation (vs. others) * 5 min (vs. 1 week)	0.08	0.06	1.29	.20
Random effect				s^2
Participant				1.34
Word				1.09

Translation Test

To examine whether L2 word recall differed by study-test interval and learning condition in the translation test, in which participants heard English translations and produced corresponding Hungarian words, a model was fit that used learning condition (iconic gesture, still image, gloss only), study-test interval (5 min, 1 week), and the learning condition by study-test interval interaction as fixed factors. One contrast comparing recall at 5 min to recall at 1 week was used to describe study-test interval. Two contrasts were used to describe learning condition: the first compared recall in the still image condition to recall in the iconic gesture condition; and the second compared recall in the gloss condition to recall in the still image and iconic gesture conditions. Likewise, two contrasts were used to describe the interaction between study-test interval and recall test: The first compared the difference between recall in the still image condition at 5 min and 1 week to recall in the iconic gesture conditions at 5 min and 1 week; and the second compared the difference between recall in the gloss condition at 5 min and 1 week to recall in the still image and iconic gesture conditions. No random slopes reached significance by participant or item; thus, all random effects were modeled using random intercepts. Mean recall by learning condition and study-test interval is displayed in Fig. 3c.

As can be seen from Table 7, across study-test intervals, recall odds were 1.46 times (95% CI: [1.16; 1.83]) higher for words learned with still images than for words learned with iconic gestures. Additionally, for words learned with glosses, recall odds were 0.89 times (95% CI: [0.79; 1.02]) what they were for words learned with still images and iconic gestures. Across learning conditions, the odds of correct L2 word meaning recall were 1.88 times (95% CI: [1.56; 2.25]) greater after 5 min than 1 week. There was no interaction between either of the learning conditions and study-test interval. In contrast to embodied theories of cognition, these results provide additional evidence that concrete L2 words learned by viewing accompanying relevant still images were recalled more accurately than words learned by viewing iconic gestures regardless of study-test interval. Furthermore, these results suggest that concrete L2 words learned with accompanying L1 glosses are recalled less accurately than words learned by viewing still images or iconic gestures, even when glosses are the only cue present at recall.

Table 8 Fixed effect estimates (top) and variance estimates (bottom) for multi-level logit model of word recall in the depiction-plus-translation test (observations = 516, log-likelihood: − 318.8)

Fixed effect	Coefficient	SE	Wald *z*	*p*
Intercept	0.39	0.25	1.60	.11
5 min interval (vs. 1 week)	0.41	0.10	4.01	< .001
Still image (vs. iconic gesture)	0.20	0.11	1.85	.06
5 min (vs. 1 week) * still image (vs. iconic gesture)	− 0.08	0.10	− 0.76	.45
Random effect				s^2
Participant				0.96
Word				0.55

Depiction-Plus-Translation Test

In the depiction-plus-translation test, participants both saw nonverbal depictions of word meanings presented at learning and heard English translations before producing corresponding Hungarian words. To examine whether L2 word recall differed in this test by study-test interval and learning condition, a model was fit that used learning condition (iconic gesture, still image), study-test interval (5 min, 1 week), and their interaction as fixed factors. One contrast comparing recall at 5 min to recall at 1 week was used to describe study-test interval. One contrast comparing recall in the still image condition to recall in the iconic gesture condition was used to describe learning condition. Finally, one contrast comparing the difference between recall in the still image condition at 5 min and 1 week and recall in the iconic gesture condition at 5 min and 1 week at was used to describe how word recall under different learning conditions varied by study-test interval. No random slopes reached significance by participant or item; thus, all random effects were modeled using random intercepts. Mean recall by learning condition and study-test interval is displayed in Fig. 3d.

As can be seen from Table 8, across study-test intervals, recall odds were 1.22 times greater (95% CI: [0.99; 1.50]) in the still image condition than in the iconic gesture condition. Moreover, across learning conditions, the odds of correct L2 word meaning recall were 1.51 times (95% CI: [1.23; 1.84]) greater after 5 min than 1 week. There was no significant interaction between learning condition and study-test interval. In contrast to embodied theories of cognition, these results provide further confirmatory evidence that concrete L2 words learned by viewing still images depicting their meanings were recalled more accurately than words learned by viewing iconic gestures depicting their meanings regardless of study-test interval.

Discussion

The purpose of the current study was to examine how L1 glosses, still images, and iconic gestures affect the encoding and retrieval of concrete words in early stage L2 learning. In particular, this study addressed the following three questions: (1) Do L1 glosses enhance

L2 word encoding and/or retrieval; (2) Does the combination of L1 glosses and nonverbal depictions of word meanings enhance L2 word encoding and retrieval beyond either of these depictions alone; (3) Which type(s) of nonverbal word meaning depictions (still images or iconic gestures) result in the most effective L2 word encoding and retrieval? The results showed that L1 glosses are unnecessary for effective recall of concrete L2 words by beginning learners and that they do not always enhance it. Furthermore, the results show that still images depicting the meanings of concrete L2 words enhance their encoding and recall by beginning learners more effectively than representative iconic gestures. This contrasts with the results of other similar studies examining the effect of viewing iconic gesture on the learning of concrete L2 words by individuals unfamiliar with the target language (Kelly et al. 2009; Macedonia et al. 2010; Tellier 2008); see below for explanation. This finding supports the predictions of dual coding theory, which predicts that viewing iconic gesture at encoding does not promote L2 word retrieval any more effectively than viewing still images, rather than embodied theories of cognition, which posit that viewing iconic gesture at encoding promotes L2 word retrieval more effectively than viewing still images. Thus, it suggests that the impact of viewing iconic gesture on the learning of concrete words in early stage L2 acquisition is limited.

Regarding the role of glosses in L2 word learning, the results of the free recall and translation tests indicated that words learned with L1 glosses were recalled no more accurately by individuals unfamiliar with the target language than words learned by viewing still images or iconic gestures. L1 glosses were also less helpful as cues to recall for L2 words: in a comparison of the recall tests, recall accuracy was higher when still images or iconic gestures were viewed at retrieval than when glosses were viewed at retrieval. However, greater recall accuracy in the translation and depiction-plus-translation tasks than the free recall task provide evidence that L1 glosses enhanced L2 word retrieval. Together, these results demonstrated that glosses were unnecessary for L2 recall and that they enhanced L2 word recall in some, but not all, cases. In general, these results indicate that although L1 glosses may enhance retrieval of newly-learned L2 words in some cases, they do not always do so, providing evidence that L1 glosses have a limited impact on L2 word acquisition in beginning learners. These results are consistent with evidence indicating that L2 words can sometimes activate concepts of their referents directly in the mind, rather than indirectly by activating L1 translations concurrently with these concepts (Altarriba and Mathis 1997; Finkbeiner and Nicol 2003; Poarch et al. 2015). Thus, these results suggest that, for individuals unfamiliar with the target language, viewing nonverbal depictions of concepts such as representative still images or iconic gesture is a more effective means of learning the meanings of concrete L2 words than viewing L1 glosses.

With regard to which types of word meaning depictions result in the most effective learning of concrete L2 words for beginning learners, the results provide evidence that viewing still images enhances learning to a greater extent than viewing iconic gestures. More specifically, words learned with accompanying still images were remembered better in all recall tests across learning–test intervals, regardless of which types of depictions were used at retrieval. The robustness of this beneficial effect of viewing still images on word learning is consistent with previous findings showing a similar effect in individuals unfamiliar with the target language (Carpenter and Olson 2011; Chun and Plass 1996), as well as research showing that viewing images illustrating passages of text enhances learning of complex concepts (Mayer 2002; Mayer and Moreno 2003; Paivio 1990). This finding provides additional support for dual coding theory by demonstrating that concrete L2 words are learned and recalled best by beginning learners in conjunction with images. Thus, the results of this study suggest that a combination of analogue (imagery)

and symbolic (language) information is most effective in facilitating encoding and recall of concrete L2 nouns and verbs in individuals unfamiliar with the target language.

The finding that viewing still images facilitated beginning learners' acquisition of concrete L2 words above and beyond viewing iconic gestures was surprising in light of the findings of work suggesting that novel concrete words learned via viewing relevant iconic gestures are recalled better by this population than novel concrete words learned via viewing relevant still images over extended time periods (Allen 1995; Kelly et al. 2009; Macedonia and Knösche 2011; Tellier 2008). There are several possible reasons why the benefit of viewing iconic gesture on beginning learners' concrete L2 word learning may have failed to exceed that of still images in this study. First, the current study examined L2 word learning in adults rather than in young children as in Tellier (2008), and there is evidence that the effect of viewing relevant body motions, including iconic gestures, may be particularly profound for children (Gogate et al. 2000; Iverson et al. 1999). Second, all of the tests used in the present study required L2 words to be recalled and produced in response to L1 glosses, which is more demanding than the recognition and L1 recall tests used in most other studies of L2 word learning in individuals unfamiliar with the target language. Third, some work suggests that the effect of iconic gesture production is stronger than the effect of iconic gesture viewing on L2 word learning in individuals unfamiliar with the target language (Morett 2018). Further study is needed to distinguish between these possibilities.

The finding that, for individuals unfamiliar with the target language, the beneficial effect of learning the meanings of concrete L2 words by viewing accompanying still images relative to accompanying iconic gestures was stronger in the depiction test than in the free recall test suggests that the presence of similar cues at encoding and retrieval facilitates L2 word learning. This interpretation is reinforced by the trend towards better recall in the translation test for concrete L2 words learned via glosses than words learned by viewing still images and iconic gestures. On the other hand, it is qualified by the strength of the benefit for words learned by viewing still images versus iconic gestures in the translation test, as well as the weakness of this benefit in the depiction-plus-translation test. However, in the depiction-plus-translation test, the impact of viewing still images on concrete word learning may have been attenuated by the presence of glosses in addition to still images at retrieval. Unfortunately, a direct comparison of the effects of viewing still images and viewing iconic gestures to the effect of glosses on word recall wasn't possible in the depiction-plus-translation test because recall of words learned in the gloss condition could not be tested. Nevertheless, all of these findings indicate that it is important to evaluate the influence of the way in which L2 word meanings are depicted at both encoding and retrieval, as well as to account for similarity and differences in depictions.

In conclusion, the findings of this study demonstrate that L1 glosses often fail to enhance effective recall of concrete L2 words in early stage L2 acquisition. Moreover, the findings show that viewing still images enhances concrete L2 word learning to a greater extent than viewing iconic gesture, as well as glosses, in beginning learners. These findings are inconsistent with embodied theories of language processing (Barsalou 1999; Gibbs 2006) but are consistent with dual coding theory (Paivio 1990). Thus, they indicate that models of L2 lexical development acknowledging the importance of imagery are most accurate, at least for early stage L2 acquisition. From an applied perspective, these findings show that, when individuals with limited familiarity with the target language learn concrete L2 words, the presence of images depicting the meanings of these words at encoding is particularly effective, and L1 glosses are not necessary—or always helpful—as recall cues. Given that only concrete L2 words were learned in the current study, it is important to note that the findings may be limited to these types of words; additional research is needed to

determine the effectiveness of viewing still images and viewing iconic gestures for learning abstract words. Furthermore, given that this study focused on L2 word learning in individuals unfamiliar with the target language, the findings may apply primarily to learners in the initial stages of L2 acquisition; again, additional research is needed to determine whether the findings also apply to more advanced L2 learners. Nevertheless, the findings provide insight into the cognitive processes involved in L2 word learning, contributing to the understanding of how concrete L2 words can be learned most effectively, as well as how instruction and assessment can be improved, in early stage L2 acquisition.

Acknowledgements Portions of this research were presented at the 22nd annual Second Language Research Forum (October 2013). I thank Brian MacWhinney for inspiring the idea for this project and the faculty of the Language and Reading Group at the Learning Research and Development Center at the University of Pittsburgh for providing the space and resources necessary to complete it.

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,* 521–529. https://doi.org/10.1111/j.1540-4781.1995.tb05454.x.

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36,* 550–568. https://doi.org/10.1006/jmla.1997.2493.

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660. https://doi.org/10.1017/S0140525X99002149.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59,* 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

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

Bergmann, K., & Macedonia, M. (2013). A virtual agent as vocabulary trainer: Iconic gestures help to improve learners' memory performance. In *International Workshop on Intelligent Virtual Agents* (Vol. 8108, pp. 139–148). Berlin: Springer.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46,* 904–911. https://doi.org/10.3758/s13428-013-0403-5.

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, 38,* 92–101. https://doi.org/10.1037/a0024828.

Chen, H.-C. (1990). Lexical processing in a non-native language: Effects of language proficiency and learning strategy. *Memory & Cognition, 18,* 279–288. https://doi.org/10.3758/BF03213881.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal, 80,* 183–198. https://doi.org/10.1111/j.1540-4781.1996.tb01159.x.

Finkbeiner, M., & Nicol, J. (2003). Semantic category effects in second language word learning. *Applied Psycholinguistics, 24,* 369–383. https://doi.org/10.1017/S0142716403000195.

Gibbs, R. W. (2006). *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development, 71,* 878–894. https://doi.org/10.1111/1467-8624.00197.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother-child interactions. *Cognitive Development, 14,* 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5.

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. https://doi.org/10.1080/01690960802365567.

Krönke, K.-M., Mueller, K., Friederici, A. D., & Obrig, H. (2013). Learning by doing? The effect of gestures on implicit retrieval of newly acquired words. *Cortex, 49,* 2553–2568. https://doi.org/10.1016/j.cortex.2012.11.016.

Lotto, L., & De Groot, A. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning, 48,* 31–69. https://doi.org/10.1111/1467-9922.00032.

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education, 5,* 196–211. https://doi.org/10.1111/j.1751-228X.2011.01129.x.

Macedonia, M., Müller, K., & Friederici, A. D. (2010). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping, 32,* 982–998. https://doi.org/10.1002/hbm.21084.

Mayer, K. M., Yildiz, I. B., Macedonia, M., & von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current Biology, 25,* 530–535. https://doi.org/10.1016/j.cub.2014.11.068.

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 41, pp. 85–139). New York: Academic Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38,* 43–52. https://doi.org/10.1207/S15326985EP3801_6.

Morett, L. M. (2014). When hands speak louder than words: The role of gesture in the communication, encoding, and recall of words in a novel second language. *The Modern Language Journal, 98,* 834–853. https://doi.org/10.1111/modl.12125.

Morett, L. M. (2018). In hand and in mind: Effects of gesture production and viewing on second language word learning. *Applied Psycholinguistics, 39*, 355–381. https://doi.org/10.1017/S0142716417000388.

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

Poarch, G. J., Van Hell, J. G., & Kroll, J. F. (2015). Accessing word meaning in beginning second language learners: Lexical or conceptual mediation? *Bilingualism Language and Cognition, 18,* 357–371. https://doi.org/10.1017/S1366728914000558.

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved July 24, 2016 from http://www.R-project.org/.

Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology, 38,* 109–117. https://doi.org/10.1016/j.cedpsych.2012.12.001.

Sagarra, N., & Alba, M. (2006). The key is in the keyword: L2 vocabulary learning methods with beginning learners of Spanish. *The Modern Language Journal, 90,* 228–243. https://doi.org/10.1111/j.1540-4781.2006.00394.x.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture, 8,* 219–235. https://doi.org/10.1075/gest.8.2.06tel.

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,* 507–546. https://doi.org/10.1111/0023-8333.00018.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9,* 625–636. https://doi.org/10.3758/BF03196322.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition, 125,* 244–262. https://doi.org/10.1016/j.cognition.2012.06.016.