Using Multimedia Vocabulary Annotations in L2 Reading and Listening Activities

JING XU
Iowa State University

ABSTRACT
This paper reviews the role of multimedia vocabulary annotation (MVA) in facilitating second language (L2) reading and listening activities. It examines the multimedia learning and multimedia language learning theories that underlie the MVA research, synthesizes the findings on MVA in the last decade, and identifies three underresearched areas on the subject.

KEYWORDS
CALL, Multimedia Vocabulary Annotation, Incidental Vocabulary Learning, Computer-mediated Reading and Listening

INTRODUCTION
Computers are of great benefit to education because they can present multimedia information for learning which, according to previous studies (e.g., Mayer & Gallini, 1990; Mayer & Anderson, 1992), makes learning more effective. In the past 10 years, multimedia technology has also been widely applied in computer-assisted language learning (CALL). Some CALL materials developed based on such technology, for example, are New Dynamic English (DynEd International Incorporation, 2007), Transparent Language, and Cyberbuch (Chun & Plass, 1997a, 1999). One prevalent application of the multimedia technology in these CALL materials is to provide learners with multimodal annotations for unknown vocabulary in reading and listening activities. Such annotations are comprised of a combination of text definitions, pictures, animations, sound, or even video clips and are intended to help learners gain vocabulary knowledge and better comprehend the listening or reading materials. To inform the development of multimedia CALL materials, this paper investigates the application of multimedia vocabulary annotations (MVAs) in second or foreign language (L2) reading and listening tasks. Specifically, it intends to answer the questions: Why and how effectively can MVA facilitate incidental vocabulary learning and at the same time enhance L2 reading and listening comprehension?

The investigation will begin by examining the theoretical rationales for multimedia learning and language learning, followed by a review of recent MVA studies in the L2 reading and listening contexts, respectively. Finally, the remaining research issues for MVA will be identified.

THEORETICAL BASIS FOR USING MVA IN CALL
The application of MVA in CALL has been motivated by multimedia learning and multimedia language learning theories. These theories serve as the basis for MVA research by explaining why vocabulary annotations presented in multiple modes are more beneficial for L2 learning than those delivered in a single mode.
**Dual-Coding Theory**

Dual-coding theory (Paivio, 1969, 1971, 1986) is the foundation for later multimedia learning and multimedia language learning theories. It hypothesizes that memory and cognition are served by two separate systems: one specialized in dealing with verbal information such as words and symbols and the other with nonverbal information such as pictures or objects. In the learning process, the human mind creates separate verbal and visual mental representations (encoding) for incoming information using each of the systems. Although the two systems work independently, they are interconnected: representations in one system can activate those in the other. For example, objects can be given names, and names can evoke the images of the objects in mind. Based on this theory, when learners use both systems to encode information, they will learn and retain the information better than when they use only one system.

The dual-coding theory has recently received supporting evidence from research in neural sciences. By examining the scanned brain images of learners who were studying German vocabulary items, Fliessbach, Weis, Klaver, Elger, and Weber (2006) found that an anterior region in precuneus—a bilateral region in the brain responsible for processing visual contents—was more strongly activated during the intentional encoding of concrete, more imageable words as compared to abstract, less imageable words. Such a finding implies that an additional image-based system in the brain might be involved in processing vocabulary, thus lending support to the dual-coding theory.

**The Cognitive Theory of Multimedia Learning**

Mayer’s (1997, 2001, 2002) cognitive (generative) theory of multimedia learning is probably the most influential theory for L2 learning via multimedia in the past 10 years because it has been referred to as a theoretical basis by many MVA studies (e.g., Al-Seghayer, 2001; Jones & Plass, 2002; Jones, 2003; Chun & Payne, 2004). The theory takes a step beyond dual-coding theory in that it models the detailed learning process in a multimedia environment. According to the theory, such a process contains three subcomponents: (a) selecting relevant verbal and visual information from the multimodal input, (b) organizing the selected information into the verbal and visual mental representations, and (c) integrating the resulting verbal and visual representations with each other (Mayer, 1997, 2001). Learning is therefore more likely to occur when learners can build meaningful connections between the verbal and visual mental representations. Based on this theory, many design principles for effective multimedia learning have been proposed, such as presenting relevant verbal and visual information simultaneously—a principle that is applied in the design of many multimedia CALL materials, reducing cognitive load (e.g., Mayer & Moreno, 2003; Moreno, 2004), and taking learner differences into consideration (e.g., Moreno & Durán, 2004).

**The Integrated Model of SLA with Multimedia**

Although Mayer’s cognitive theory of multimedia learning has served as the theoretical foundation for many MVA studies, similar to the dual-coding theory, it is intended to explain learning in general but not specifically for second language acquisition (SLA) in a multimedia environment. Therefore, a step forward is for SLA theorists to refine the multimedia learning theories in the SLA domain. Plass and Jones (2005) perceived a connection between Mayer’s cognitive theory and the interactionist perspective of SLA theories—the former emphasizes the means to enhance meaningful input through dual presentation of words (aural and/or written) and pictures (static and/or moving), while the latter addresses the importance of
comprehensible input (Krashen, 1982, 1985) to SLA. By virtue of the connection, they created an integrated model of SLA with multimedia (See Figure 1) which intertwines ideas from both sides. In such a model, L2 learners process multimodal input by first selecting useful verbal and visual information (apperception) and organizing this information into comprehensible verbal and visual mental representations (comprehension). They then develop the mental representations respectively into verbal and visual models (intake) which eventually become integrated in their linguistic systems (integration).

Figure 1
Integrated Model of SLA with Multimedia (Plass & Jones, 2005, p. 471)

Plass and Jones’ model theorizes the input-based L2 learning process in a multimedia environment and implies that the provision of additional visual information in parallel with L2 verbal information facilitates SLA. The model considers learner output as a reflection of the effectiveness of the modes of input in interaction with learners’ cognitive and metacognitive processing for comprehension (Jones, 2006a). In this respect, the model is in line with the information-processing theory which claims that, because comprehension and production draw on the same underlying knowledge source, input-based learning will facilitate both (VanPatten, 2007). However, the current version of the model offers little discussion over the roles that interaction (Long, 1985; Gass, 1997) and pushed output (Swain, 1985; Swain & Lapkin, 1995) play in L2 learning within a multimedia environment. Additionally, the model cannot be integrated with skill acquisition theory, which regards copious productive practice as the means to acquire automaticity of linguistic knowledge (DeKeyser, 2007).

A Hierarchical Model with Image
Plass and Jones’ model explains how combined verbal and nonverbal input makes L2 learning effective. However, it insufficiently explains the contribution of L1 verbal input to SLA in a multimedia environment where both L1 verbal input, L2 verbal input, and visuals are all available (See Yeh & Wang, 2003 for an example of such an environment). The hierarchical model with image that Yoshii (2006) adapted from Kroll and Stewart’s (1994) model illustrates the interaction between L1, L2, and image in the L2 learning process (See Figure 2).
In Yoshii’s model, the L2 can either be associated with the L1 via lexical links (the solid line) or with concepts via conceptual links (the broken line). That is to say, learners can mediate the L2 either through a L1 translation or through concepts without L1 assistance although, according to Yoshii, the former approach might be more effective than the latter in early stages of L2 learning. Image, on the other hand, is another source of learning in this model. As seen in Figure 2, it is conceptually linked with concepts and thereby provides additional cues for comprehension of the L2. This model indicates that learners may comprehend and acquire the L2 by selecting three different types of learning resources: L1 translation, modified L2 input, or visuals. It also hypothesizes that L1 translation + image is more helpful than modified L2 input + image for beginning-level learners because the lexical links between L1 and L2 are stronger than the conceptual links between L2 and concepts at this stage of development.

In summary, the theoretical basis for applying MVA in CALL is rooted in the dual-coding theory and the cognitive theory of multimedia learning from cognitive science. Recently, these theories have been moved to the SLA domain. Plass and Jones’ model is developed by integrating the cognitive theory of multimedia learning with the interactionist perspective of SLA theories. Yoshii’s model further differentiates the effects of L1 or L2 verbal input on L2 learning in a multimedia environment. These theories unanimously suggest that L2 learners would benefit from receiving input from a variety of sources that includes both verbal and visual information.

EMPIRICAL STUDIES OF MVA IN L2 READING AND LISTENING CONTEXTS

The multimedia learning theories, language learning theories, and resulting models above have explained, from a theoretical perspective, why MVA is beneficial for L2 learning in general. However, the effects of MVA specifically on facilitating incidental vocabulary learning and enhancing comprehension in L2 reading and listening activities need to be examined via empirical research. This section will synthesize the findings on these two issues.

*Incidental Vocabulary Learning in L2 Reading*

Incidental vocabulary learning refers to learners’ acquisition of vocabulary knowledge during a language learning activity that is not intended for vocabulary instruction (Read, 2004).
Previous research has confirmed the advantages of providing text annotations for enhancing incidental vocabulary learning in L2 reading (e.g., Hulstijn, Hollander, and Greidanus, 1996; Watanabe, 1997). Further, numerous studies have examined the effectiveness of different formats of text annotations, such as L1 versus L2 annotations (Jacobs, Dufon, & Hong, 1994; Chen, 2002; Miyasako, 2002), single versus multiple-choice annotations (Watanabe, 1997), and word-level versus sentence-level annotations (Grace, 1998, 2000; Gettys, Imhof, & Kautz, 2001). With the advancement and increased availability of computers, vocabulary annotations are no longer limited to text forms but have been expanded to embrace other modes of information, such as still pictures, sounds, animations, and even videos. Many researchers therefore have started examining the effectiveness of the nontext forms of vocabulary annotations for vocabulary learning.

The majority of MVA research in the last 10 years has compared MVA with single-mode vocabulary annotations in L2 reading (See Table 1).

Table 1
Summary of MVA Studies on Incidental Vocabulary Learning via L2 Reading

<table>
<thead>
<tr>
<th>Studies</th>
<th>Context</th>
<th>Independent Variable(s)</th>
<th>Dependent Variable(s)</th>
<th>Participants</th>
<th>Design</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chun &amp; Plass (1996a)</td>
<td>Reading</td>
<td>Annotation type: 1. L1 text 2. L1 text + picture 3. L1 text + video</td>
<td>Immediate/delayed vocabulary test scores (recognition/production)</td>
<td>160 adult German learners</td>
<td>within subject</td>
<td>L1 text + picture &gt; L1 text + video &gt; L1 text</td>
</tr>
<tr>
<td>Kost et al. (1999)</td>
<td>Reading</td>
<td>Annotation type: 1. L1 text 2. picture 3. L1 text + picture</td>
<td>Immediate/delayed vocabulary test scores (recognition/production)</td>
<td>56 German learners</td>
<td>within subject</td>
<td>L1 text + picture &gt; L1 text or picture</td>
</tr>
<tr>
<td>Al-Seghayer (2001)</td>
<td>Reading</td>
<td>Annotation type: 1. L2 text 2. L2 text + picture 3. L2 text + video</td>
<td>Immediate vocabulary test scores (recognition/production)</td>
<td>30 ESL learners</td>
<td>within subject</td>
<td>L2 text + video &gt; L2 text + picture or L2 text</td>
</tr>
<tr>
<td>Yoshii &amp; Flaitz (2002)</td>
<td>Reading</td>
<td>Annotation type: 1. L2 text 2. picture 3. L2 text + picture</td>
<td>Immediate/delayed vocabulary test scores (recognition/production)</td>
<td>151 ESL learners</td>
<td>between subject</td>
<td>L2 text + picture &gt; L2 text or picture</td>
</tr>
<tr>
<td>Yeh &amp; Wang (2003)</td>
<td>Reading</td>
<td>Annotation type: 1. L1/L2 text 2. L1/L2 text + picture 3. L1/L2 text + picture + sound</td>
<td>Immediate vocabulary test scores (recognition/production)</td>
<td>82 college EFL learners</td>
<td>between subject</td>
<td>L1/L2 text + picture &gt; L1/L2 text or L1/L2 text + picture + sound</td>
</tr>
<tr>
<td>Yoshii (2006)</td>
<td>Reading</td>
<td>Annotation type: 1. L1 text 2. L2 text 3. L1 text + picture 4. L2 text + picture</td>
<td>Immediate and delayed vocabulary test scores (recognition/production)</td>
<td>195 college EFL learners</td>
<td>A mixed design repeated measures</td>
<td>L1 group = L2 group Picture group &gt; no picture group</td>
</tr>
</tbody>
</table>
The results have consistently found that the former was more effective in facilitating vocabulary learning in terms of comprehension and retention (Chun & Plass, 1996a; Plass, Chun, Mayer, & Leutner, 1998; Kost, Foss, & Lenzini, 1999; Al-Seghayer, 2001, Yoshii & Flaitz, 2002; Yoshii, 2006). For instance, Yoshii and Flaitz (2002) examined the effects of three vocabulary annotation types—text only, picture only, and a combination of the two—on incidental vocabulary acquisition in a multimedia reading environment. By employing a between-subjects design, the researchers divided 151 ESL learners into three reading groups, treated each group with one annotation type, and measured learners’ vocabulary gains via immediate and delayed vocabulary posttests. The results of the study indicated that the group using combined text and picture annotations outperformed the other two groups in both vocabulary tests, revealing the advantage of MVA over traditional text glosses on enhancing vocabulary learning.

Besides the consistent findings favoring multimodal annotations, many researchers further explored which combination of modes was the most effective in facilitating vocabulary acquisition through reading. Chun and Plass (1996a) and Plass et al. (1998) discovered that the annotation type of text + still picture was more helpful than that of text + video for vocabulary learning, whereas Al-Seghayer (2001) found the opposite. Yeh and Wang (2003) investigated the effectiveness of a new annotation type—text + picture and sound. The additional aural part of the annotation provided learners with a native speaker’s voice pronouncing a word, spelling the word, and reading aloud the sentence in which the word was embedded. However, the study found the aural information lowered vocabulary gains due to participants’ learning styles and the difficulty of the aural information. Yeh and Wang’s study is also unique in that it offered learners both L1 and L2 texts, compared to other studies which used either L1 text mode (Chun & Plass, 1996a; Plass et al., 1998; Kost et al., 1999) or L2 text mode (Al-Seghayer, 2001; Yoshii & Flaitz, 2002). Inspired by Yeh and Wang’s study, Yoshii (2006) examined the effectiveness of annotations of L1 versus L2 text in combination with picture versus no picture on incidental vocabulary learning. His study found that both L1 and L2 types of annotations (regardless of the pictorial mode) were effective for vocabulary learning but they appeared to cause different patterns of vocabulary retention. Specifically, the L1 text-only group displayed lower vocabulary forgetting rates than the L2 text-only and L2 text + picture groups over the 2 weeks between the immediate and delayed vocabulary posttests. Yoshii thus suspected that the difference between L1 and L2 types of MVA would be revealed in longer term vocabulary learning.

**Incidental Vocabulary Learning in L2 Listening**

Following the path of MVA research in L2 reading, Jones conducted a series of studies to examine the effectiveness of dual-mode (L1 text + picture) versus single-mode (L1 text or picture alone) vocabulary annotations on incidental vocabulary learning in the L2 listening context (See Table 2).
Table 2
Summary of MVA Studies on Incidental Vocabulary Learning via L2 Listening

<table>
<thead>
<tr>
<th>Studies</th>
<th>Context</th>
<th>Independent Variable(s)</th>
<th>Dependent Variable(s)</th>
<th>Participants</th>
<th>Design</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones &amp; Plass</td>
<td>Listening</td>
<td>Annotation type: 1. None 2. L1 text 3. Picture 4. L1 text + picture</td>
<td>Immediate/delayed vocabulary test scores (recognition)</td>
<td>171 beginning French learners</td>
<td>between subject</td>
<td>L1 text + picture &gt; L1 text alone, picture alone, or none</td>
</tr>
<tr>
<td>(2002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jones</td>
<td>Listening</td>
<td>N/A</td>
<td>N/A</td>
<td>20 beginning French learners</td>
<td>qualitative</td>
<td>L1 text + picture &gt; L1 text alone, picture alone, or none</td>
</tr>
<tr>
<td>(2003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jones</td>
<td>Listening</td>
<td>Annotation type + collaborative strategy: 1. None + alone 2. None + in pairs 3. L1 text and picture + alone 4. L1 text and picture + in pairs</td>
<td>Immediate/delayed vocabulary test scores (recognition/production)</td>
<td>68 beginning French learners</td>
<td>between subject</td>
<td>L1 text + picture &gt; none regardless the presence or absence of collaboration</td>
</tr>
<tr>
<td>(2006b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the first study, Jones and Plass (2002) found that learners receiving dual-mode annotations outperformed learners receiving single-mode annotations or no annotation in vocabulary recognition posttests. Such a finding was confirmed in Jones’ (2003) follow-up qualitative study based on learners’ reflections of their experience using the annotations in the first study. A limitation of these two studies, however, was the use of recognition tests only to measure learners’ vocabulary gains. Jones (2004) expanded upon her previous research by refining the vocabulary knowledge measures. Specifically, learners were asked not only to recognize L1 translations and pictures for L2 vocabulary but to produce L1 translations as well. However, this study yielded a finding inconsistent with her previous research—learners using dual- and single-mode annotations performed equally well on vocabulary recognition. Also, learners best produced L1 translations when the mode of testing matched the annotation mode they experienced. Based on Vygotsky’s social learning theory (1978), Jones (2006b) investigated how peer-to-peer collaboration and dual-mode vocabulary annotations (L1 text + picture) together affected incidental vocabulary learning in L2 listening. The study revealed that annotations had a positive effect on vocabulary learning, while collaboration had no effect.
Interactive models of reading have suggested that readers comprehend text via both bottom-up processing at the microlevel and top-down processing at the macrolevel (e.g., Kintsch & Van Dijk, 1978; Rumelhart, 1977; Swaffar, Arens, & Byrnes, 1991). Readers’ bottom-up processing relies heavily on their knowledge of vocabulary. Studies by Laufer (1989) and Liu and Nation (1985) show that readers need to know over 90% of the words in a text to achieve adequate comprehension and to be able to guess the meanings of unknown words from the context. Thus, researchers wondered whether MVA intended for equipping learners with necessary L2 vocabulary knowledge would in turn facilitate overall reading comprehension.

Four studies examining the effects of MVA on reading comprehension have produced mixed results (See Table 3).

Table 3
Summary of MVA Studies on L2 Reading Comprehension

<table>
<thead>
<tr>
<th>Studies</th>
<th>Context</th>
<th>Independent Variable(s)</th>
<th>Dependent Variable(s)</th>
<th>Participants</th>
<th>Design</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chun &amp; Plass</td>
<td>Reading</td>
<td>Annotation type: 1. L1 text 2. L1 text + picture 3. L1 text + video</td>
<td>Scores of recall protocols</td>
<td>160 adult German learners</td>
<td>within subjects</td>
<td>L1 text + picture/video &gt; L1 text</td>
</tr>
<tr>
<td>Hong</td>
<td>Reading</td>
<td>Reading method: 1. A computer-based multimedia reading program with contextualized L1 glosses 2. Paper-pen-dictionary</td>
<td>Reading comprehension scores</td>
<td>20 beginning business Chinese learners</td>
<td>between subjects</td>
<td>Multimedia &gt; Paper-pen-dictionary</td>
</tr>
<tr>
<td>Lomicka</td>
<td>Reading</td>
<td>Types of images: 1. Still image 2. Dynamic image 3. Important image 4. One image per scene 5. No image</td>
<td>Reading comprehension scores, story reproduction scores</td>
<td>58 beginning Japanese learners</td>
<td>between subjects</td>
<td>Dynamic image &gt; = still image on comprehension; dynamic image &gt; still image on reproduction</td>
</tr>
<tr>
<td>Ariew &amp; Ercetin</td>
<td>Reading</td>
<td>Factors affecting comprehension: 1. prior knowledge 2. textual graphics 3. textual graphics audio 4. textual text 5. contextual graphics 6. contextual audio 7. contextual text 8. contextual video</td>
<td>Reading comprehension scores</td>
<td>84 intermediate and advanced ESL learners</td>
<td>within subjects</td>
<td>Visual annotations did not help reading comprehension; videos had negative effects on reading comprehension in the intermediate group.</td>
</tr>
</tbody>
</table>
Chun and Plass (1996b), Hong (1997), and Lomicka (1998) found that in-text vocabulary annotations enhanced by pictures, videos, or audio were effective in aiding overall reading comprehension; moreover, these annotations outperformed traditional text glosses (Chun & Plass, 1996b; Lomicka, 1998). For example, Lomicka (1998) found that learners looking up multimodal vocabulary annotations composed of L1 and L2 texts, images, references, questions, and word pronunciation produced more causal inferences—an indicator of high-level comprehension—in think aloud protocols than those who only used traditional L1 and L2 text glosses. The small sample size was, however, a limitation in the study. In contrast, Ariew and Ercetin (2004) found that learners’ access to MVA in the form of texts, graphics, or audio contributed little to their overall reading comprehension. One of the problems with this study, however, was that the participants were all intermediate and advanced L2 readers whose vocabulary knowledge was no longer an important predictor or a limiting factor for reading success (Chall, 1983; Juel, 1991; Stanovich, 1986). That is to say, readers at such proficiency levels were able to infer the meanings of unknown words based on the context and thus relied little on vocabulary annotations for help. Another problem of the study was the provision of both MVA for bottom-up processing and contextual annotations for top-down processing in the same reading passage. Due to learners’ combined use of the two types of annotations, the effects of the MVA on reading comprehension might have been obscured.

**L2 Listening Comprehension**

Participant use of the MVA was found to be helpful for listening comprehension as well. Jones and Plass (2002) and Jones (2003) found that learners understood an L2 aural passage better when they accessed dual-mode vocabulary annotations of L1 text + picture than when they used single-mode annotations or no annotation (See Table 4).

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Summary of MVA Studies on L2 Listening Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies</td>
<td>Context</td>
</tr>
<tr>
<td>Jones (2003)</td>
<td>Listening</td>
</tr>
<tr>
<td>Jones (2006b)</td>
<td>Listening</td>
</tr>
</tbody>
</table>
Jones (2006b) confirmed the benefits of dual-mode vocabulary annotations on L2 listening comprehension and found that such benefits could be further enhanced by peer-to-peer collaboration. In this study, learners who cooperated with a partner by taking and sharing notes while listening showed the highest level of aural comprehension. A limitation of these three studies, however, was that participants' prior knowledge of the aural passage was not pretested.

Given the previous studies investigating the effects of MVA on incidental vocabulary learning and comprehension in L2 reading and listening contexts, the following tentative conclusions can be stated:

In the context of L2 reading,
1. There is convincing evidence that MVA enhanced by static picture or video is beneficial and is better than traditional single-text-mode annotations for incidental vocabulary learning.
2. There is insufficient evidence to judge which type of verbal (L1 vs. L2) + nonverbal (static picture vs. video) MVA best facilitates incidental vocabulary learning.
3. All studies except Ariew and Ercetin (2004) found MVA to enhance overall comprehension.

In the context of L2 listening,
1. Research has produced mixed results regarding whether MVA outperformed single-mode annotations on facilitating vocabulary learning.
2. There is fairly convincing evidence that MVA with a combination of picture and text increase overall comprehension. Such positive effects of MVA can be further increased by learner collaboration.

UNRESOLVED PROBLEMS AND SUGGESTIONS
The aforementioned studies have contributed to the growing body of literature on MVA by finding evidence that including nonverbal information in annotations increases the chance for learners to acquire vocabulary (e.g., Chun & Plass, 1996a; Jones & Plass, 2002) and to comprehend written texts (Chun & Plass, 1996b; Hong, 1997; Lomicka, 1998) or aural texts (Jones & Plass 2002; Jones, 2003, 2006). However, research issues in this area remain. This section will discuss these issues and at the same time provide suggestions for future research.

Is MVA More Effective Than Single-Mode Vocabulary Annotations in Facilitating Incidental Vocabulary Learning in L2 Listening?
The advantages of MVA over single-mode annotations on enhancing vocabulary learning have been confirmed for L2 reading but not for L2 listening. The study by Jones (2004) found that using dual-mode (L1 text + picture) and single-mode (L1 text or picture) annotations during L2 listening resulted in equal performance on vocabulary recognition. Jones attributed this to a “factor of guessing” caused by the multiple-choice format of the recognition test (p. 133). However, such an explanation is problematic considering that in the same study, the
dual-mode group did not outperform one of the single-mode groups (L1 text only) in immediate and delayed vocabulary production tests in which the guessing factor no longer existed. Therefore, it is still unclear whether pictorial vocabulary annotations can enhance vocabulary learning in L2 listening activities.

To explore the issue, researchers need to refine vocabulary assessment to avoid learners’ guessing behavior as well as to minimize the impact of test mode on learner performance—learners performed the best when the test mode matched the annotation mode (Jones, 2004). Specifically, learners were not supposed to be tested on their ability to recall the exact information in annotations but on their understanding of the information. For example, if learners were only asked to recognize a picture for a vocabulary item (such a test appeared in Jones, 2004 and Kost et al., 1999), they would probably be able to do so even without understanding the meaning of the word—they could simply recall the picture they saw in annotations. In this sense, a test of production is a more valid way to assess learners’ vocabulary knowledge.

**What is the Most Effective Type of MVA?**

Current computer technology allows CALL developers to create MVA with various combinations of text (L1 or L2), visual (static picture, animation, or video), and even aural information (e.g., pronunciation). However, it is not clear yet which combination is the most effective.

**Static versus dynamic pictures**

A decision that CALL developers need to make is to select the type of visual—either static or dynamic (animation and video)—to enhance learning. Considering that creating dynamic pictures is much more expensive and time consuming than creating static pictures, it is meaningful to find out whether video is worth the extra money and effort.

The research comparing the effectiveness of dynamic versus static visuals on incidental vocabulary learning and overall comprehension is inconclusive in the context of L2 reading whereas this issue has not been explored in the context of L2 listening. In terms of the helpfulness of contextual visuals to L2 reading comprehension, Ikeda (1999) and Lin and Chen (2007) found dynamic pictures (animations) to be generally more beneficial than static pictures while Ariew and Ercretin (2004) found both types of visuals to be equally ineffective. Such contradictory findings on visuals also appeared in MVA research in L2 reading. Chun and Plass (1996a) and Plass et al. (1998) found that when combined with text glosses, static pictures were more helpful than videos for incidental vocabulary learning while Al-Seghayer (2001) found the opposite. Researchers also appeared to have different opinions regarding the usefulness of videos. Al-Seghayer (2001) thought of videos as intriguing visual cues in reading which increased readers’ concentration on the annotations whereas Ariew and Ercretin (2004) considered videos to be distractive—they interfered with reading when they were relied upon too much by readers for text comprehension. The controversy over the effectiveness of videos can be explained by Chun and Plass’ (1997b) views; video annotations are only helpful when learners’ interest directs their attention to the useful information for comprehension. In other words, when learners become interested in the irrelevant information contained in videos, they are distracted from the reading activity. Thus, the way learners select information from video annotations needs to be taken into consideration when the effectiveness of videos is evaluated.
It may be too early to judge which type of visuals in MVA is more effective when the potential of dynamic visuals has not been fully explored. Researchers therefore, need to determine the best video delivery method to efficiently direct learner focus to useful information for comprehension. There are several ways to manipulate the presentation of dynamic visuals. First, researchers may control the length of videos to reduce the amount of irrelevant information; Ariew and Ercetin (2004) noticed that lingering on videos for too long could possibly hurt reading comprehension. Second, by applying video-editing technology, researchers may highlight the useful information to make the video annotations more self-explanatory. For example, researchers may circle essential information or blur irrelevant information. Another way to increase the comprehensibility of videos is to provide written explanations for the videos (Hart, 2007). However, this approach may be controversial because it makes learning from MVA more demanding and complicated. In Hart’s view, the additional text increased the reading load, especially for learners with low proficiency in the language. Finally, if necessary, researchers may employ the eye-tracking technology to trace what learners are looking at in videos.

L1 versus L2

Besides the type of visuals, another issue that CALL developers need to consider is the type of text for MVA (i.e., L1, L2, or both). However, research so far has not given a clue which text type best facilitates vocabulary learning. Research comparing L1 versus L2 text glosses on incidental vocabulary learning in reading has produced mixed results. Two studies found that there was no difference between L1 and L2 glosses (Chen, 2002; Jacobs et al., 1994) while a third one revealed the advantage of L2 glosses on immediate vocabulary comprehension (Miyasako, 2002). Studies on MVA in L2 reading have shown that both L1 annotations (Chun & Plass, 1996a; Plass et al., 1998; Kost et al, 1999) and L2 annotations (Al-Seghayer, 2001; Yoshii & Flaitz, 2002) are helpful for vocabulary learning when they are combined with visuals. In the L2 listening context, only the MVA with L1 text were found to be effective (Jones & Plass, 2002; Jones 2003, 2004, 2006). Yoshii (2006) is one of the few researchers who compared MVA with L1 versus L2 annotations on vocabulary learning. However, he did not find that one text type was superior to the other.

To tackle the issue of text type, researchers may consider refining the vocabulary assessment in MVA studies. Yoshii’s (2006) study revealed that learners’ long-term retention of vocabulary may differ when they refer to L1 versus L2 type MVA. Following this line of reasoning, researchers may, on the one hand, prolong the interval between the immediate- and delayed-vocabulary posttests (2-3 weeks in the majority of previous studies) and, on the other hand, increase the number of delayed posttests to track learners’ vocabulary forgetting rate. Besides, vocabulary assessment should not be limited to definition recognition or picture identification tasks in a multiple-choice format, which only reflect learners’ receptive vocabulary knowledge (Nation, 2001). Learners’ gains of productive vocabulary knowledge, such as the ability to use a word in a sentence, should also be assessed. With better designed measures of vocabulary knowledge, the impact of MVA text type on vocabulary learning is more likely to be disclosed.

Does Word Concreteness Make a Difference for the Effectiveness of MVA?

All studies on MVA have investigated how the manipulation of annotation type would improve incidental vocabulary learning and comprehension in L2 reading and listening activities. However, none of these studies has explored how the helpfulness of annotations would vary
across different types of words, such as concrete versus abstract words. As Paivio (1969, 1986) pointed out, concrete words that refer to objects, materials, or persons are more easily imaged than abstract words that refer to concepts. Fliessbach et al.’s (2006) study also confirmed that learners’ imaging system was more active when processing concrete words than abstract words. Therefore, learners looking up MVA may benefit from visuals to a different extent depending on word concreteness. Specifically, visuals for words of high concreteness may be more comprehensible and thus more beneficial for learning than those for words of low concreteness. For example, the picture for a concrete word *pipe* (see Figure 3)—a picture of a real pipe should be better—is more self-explanatory than the picture for a relatively abstract word *reunion* (see Figure 4) which can be reasonably misinterpreted by learners as *celebration, feast, or holiday*.

The scene of reunion is not as easy to visualize as a pipe because various learners, especially those of different cultural backgrounds, likely have different prior experience of attending family gatherings on holidays and meeting old friends in various situations. So, future research needs to compare the effectiveness of MVA for concrete versus abstract vocabulary and to look for the best way to visualize abstract vocabulary.

**CONCLUSION**

Using MVA in L2 reading and listening has obtained support from both multimedia (language) learning theories and related empirical research. From a theoretical perspective, MVA provides learners with rich input in multiple modalities and thus creates an optimal environment for SLA. From a research perspective, many studies have revealed the positive effects of MVA on facilitating vocabulary learning and enhancing comprehension in L2 reading and listening processes. Even so, three issues related to this field remain as yet unanswered. To better inform developers of CALL materials, researchers are encouraged to further investigate the effects of pictorial MVA on vocabulary learning in the context of L2 listening, to explore the best combination of modes for MVA to facilitate SLA, and to look into the influence of word concreteness on the effectiveness of MVA.

**NOTE**

1 For example, Borrás & Lafayette (1994) investigated L2 French learners’ speaking performance with assistance of a multimedia language learning courseware.
REFERENCES


**AUTHOR’S BIODATA**

Jing Xu is a Ph.D. student in the program of Applied Linguistics and Technology at Iowa State University. His research interests include multimedia language learning, hypermedia reading, and validity issues in second language assessment.

**AUTHOR’S ADDRESS**

Jing Xu  
319 Ross Hall  
Iowa State University  
Ames, IA 50011  
Phone: 515 294 7460  
Email: jxu2@iastate.edu