Research paper

Could a multimodal dictionary serve as a learning tool?
An examination of the impact of technologically enhanced visual glosses on L2 text comprehension

Takeshi Sato
Tokyo University of Agriculture & Technology, Japan

tsato@cc.tuat.ac.jp

Abstract
This study examines the efficacy of a multimodal online bilingual dictionary based on cognitive linguistics in order to explore the advantages and limitations of explicit multimodal L2 vocabulary learning. Previous studies have examined the efficacy of the verbal and visual representation of words while reading L2 texts, concluding that it facilitates incidental word retention. This study explores other potentials of multimodal L2 vocabulary learning: explicit learning with a multimodal dictionary could enhance not only word retention, but also text comprehension; the dictionary could serve not only as a reference tool, but also as a learning tool; and technology-enhanced visual glosses could facilitate deeper text comprehension. To verify these claims, this study investigates the effects of multimodal representations on Japanese students learning L2 locative prepositions by developing two online dictionaries, one with static pictures and one with animations. The findings show the advantage of such dictionaries in explicit learning; however, no significant differences are found between the two types of visual glosses, either in the vocabulary or in the listening tests. This study confirms the effectiveness of multimodal L2 materials, but also emphasizes the need for further research into making technologically enhanced materials more effective.

Keywords: Animated image, explicit learning, multimodal glosses, online dictionary, prepositions, L2 vocabulary acquisition.

1. Introduction
Many studies have examined and reported the positive effects of visual glosses (e.g., pictures or images) in second language (L2) vocabulary acquisition in multimodal environments (Chun & Plass, 1996; Lomicka, 1998; Al-Seghayer, 2001; Yoshii & Fraitz, 2002; Yeh & Wang, 2003; Sato & Suzuki, 2010). These results are underpinned by the Dual Coding Theory (Paivio, 1971) for multimedia learning (Mayer & Sim, 1994), which states that presenting information in both verbal and visual modes leads to longer retention of the target information than with only one code. Such representation is easily implemented under a multimodal environment, where several types of glosses can be displayed on a single screen. Owing to the nature of multimodal information presentation, recent language learning materials contain not only languages and still pictures, but also sounds or animations.

Despite the efficacy of visual glosses, however, this study emphasizes that overestimating multimodal capabilities may limit glosses’ effectiveness. Our study posits three limitations that previous studies did not discuss. The first challenge is the substantial focus on incidental learning, even though explicit instruction can have beneficial effects (Ellis, 1995; Groot, 2000; Boers, 2013). In that respect, the effects of visual glosses should be examined for explicit L2 vocabulary learning. Furthermore, the target vocabulary should be selected based on a theoretical criterion. Previous studies have chosen the target vocabulary using frequency, which indicates the amount of
words needed for successful L2 vocabulary learning. However, Littlemore (2009) stated that vocabulary-depth is more crucial than language-breadth in some cases. Finally, previous studies failed to examine the efficacy of different visual gloss configurations depicting the same image. Instead, they tended to focus on the appropriate combination of different glosses; Chun & Plass (1996) claim the combination of verbal and pictorial glosses is more effective in incidental L2 vocabulary learning than only presenting verbal or pictorial gloss. Yeh & Wang (2003) also stressed that the combination of verbal and pictorial glosses can increase target vocabulary retention more than integrating three gloss types: verbal, picture, and sound, and only one gloss type. As information presentation with multimodal functions has been developing, the impact of different visual glosses based on the same image should be examined.

2. From reference tool to learning tool

Taking these challenges into consideration, this study revalidates the efficacy of multimodal visual glosses in L2 vocabulary learning from the following perspectives.

This study would like to focus on explicit learning to increase students' language-depth and not language-breadth. Boers (2013) acknowledges incidental L2 vocabulary learning is ideal, but claims explicit learning should be utilized under the condition that time for learning is limited. Groot (2000) also states that explicit instruction is effective especially in a short period of learning time.

This study thus selects L2 prepositions as the target vocabulary for students to learn explicitly. As the target vocabulary is polysemic, it is considered difficult to acquire. Prepositions appear very frequently in discourse, but learners do not always understand them (Lindstromberg, 1996). They tend to learn prepositions as idioms or chunks, but they cannot use them according to the context, relying only on memorization (Lindstromberg, 2001). Despite the semantic complexity, inappropriate use of the senses might lead to a change of meaning (Nguyen & Rethinasamy, 2006). With regard to learning such vocabulary with a complicated semantic network, Ellis (1995, p.103) stresses that “acquisition of word meanings requires explicit learning processes with deep processing strategies like semantic elaboration and imagery mediation resulting in better acquisition.” Additionally, a linguistic theory also emphasizes that an image could motivate each sense of a polysemous word, and as a result, organize a semantic network where all the senses are conceptually motivated with respect to each other. The image is defined as an image schema, which is a key term of Cognitive Linguistics. Johnson (1987) defines image schemata as “abstract patterns in our experience and understanding that are not propositional” (p. 2). Figure 1 shows the image schema of the preposition “over” (Dewell, 1994). The image schema is an object, which conceptualizes a prototypical sense of over (e.g. “The plane is flying over the mountain.”) and then can be extended into other figurative senses (Langacker, 1987) such as “She got over her flu.” Such metaphorical extension mediated by the image schema results in a semantic network, in which all the senses of the word are cohesively embedded.

![Figure 1. Image schema of over (Dewell, 1994).](image)

Based on this advantage of the image schema, this study applied it to visual glosses for learning L2 prepositions. Boers (2004) suggests that metaphorical awareness facilitates L2 learning and information tends to be easily elicited once it is linked with a semantic network. As some studies recognize the positive impact of L2 vocabulary learning from the perspective of Cognitive Linguistics (Boers, 2013; Cho, 2010; Morimoto & Loewen, 2007; Yasuda, 2010), the advantage this study hypothesizes would be seen in L2 vocabulary learning within a multimodal environment.

Finally, this study examines the effectiveness of technologically advanced image schemata as visual glosses, so three-dimensional (3D) visual glosses are developed.
Thus, as the image schema in Figure 1 is shaped as a result of the embodiment of our bodily experience (Lakoff, 1987), the relationship between the elements in such a schematic image should be displayed not in a planar, but tactile way, to approximate our perception, which would serve as a more effective visual gloss for L2 learning. In fact, Littlemore (2009) also claims that 3D diagrams might be useful when displayed dynamically for L2 learning.

3. A web-based multimodal dictionary as a learning tool

To illustrate the efficacy of the image-schema glosses under a multimodal environment, two web-based bilingual dictionaries were developed. Each dictionary dealt with eight L2 prepositions (“above”, “across”, “along”, “below”, “in”, “into”, “on”, “over”), all of which depicted a spatial relationship between objects and held both literal and figurative senses. Figure 2 is the sample page of the dictionary. On the left page, indexes of the word are shown, whereas the right page shows the example sentences with L1 translation and the visual gloss based on the image schema developed by conceptualizing the schematic sense of the preposition “along.” Figure 3, on the other hand, illustrates the other dictionary with 3D visual glosses, which illustrates the same schematic image as Figure 2, but the image was developed with 3D animation to display the image as if the user perceives the situation (see Figure 4). This is based on the supposition that that image schema was the embodiment of our daily experience, so the schematic images should be built not from the objective viewpoint like Figure 2, but from a subjective one as if those who look at the image were at the spatial situation. Both dictionaries included the same verbal glosses, which were extracted from an English-Japanese dictionary (Eds. Tanaka, Takeda, & Kawade, 2003) with the permission of the chief editor.

This study postulates that a technologically enhanced dictionary can serve as both a reference tool and a learning tool for students acquiring the target vocabulary. The dictionary has traditionally been used to provide word meanings to help learners comprehend texts or to produce sentences, rather than as a resource to acquire vocabulary knowledge explicitly. A multimodal dictionary with several types of glosses, however, increases the saliency of target lexical items and their linguistic features. This creates the ideal environment for L2 vocabulary acquisition (Chapelle, 1998) and therefore leads to effective learning of the target vocabulary, although the literature has discussed the advantages of computer-assisted visual glosses mostly in terms of incidental learning.

Figure 2. Web-based dictionary with 2D visual gloss.

Thus, as the image schema in Figure 1 is shaped as a result of the embodiment of our bodily experience (Lakoff, 1987), the relationship between the elements in such a schematic image should be displayed not in a planar, but tactile way, to approximate our perception, which would serve as a more effective visual gloss for L2 learning. In fact, Littlemore (2009) also claims that 3D diagrams might be useful when displayed dynamically for L2 learning.

3. A web-based multimodal dictionary as a learning tool

To illustrate the efficacy of the image-schema glosses under a multimodal environment, two web-based bilingual dictionaries were developed. Each dictionary dealt with eight L2 prepositions (“above”, “across”, “along”, “below”, “in”, “into”, “on”, “over”), all of which depicted a spatial relationship between objects and held both literal and figurative senses. Figure 2 is the sample page of the dictionary. On the left page, indexes of the word are shown, whereas the right page shows the example sentences with L1 translation and the visual gloss based on the image schema developed by conceptualizing the schematic sense of the preposition “along.” Figure 3, on the other hand, illustrates the other dictionary with 3D visual glosses, which illustrates the same schematic image as Figure 2, but the image was developed with 3D animation to display the image as if the user perceives the situation (see Figure 4). This is based on the supposition that that image schema was the embodiment of our daily experience, so the schematic images should be built not from the objective viewpoint like Figure 2, but from a subjective one as if those who look at the image were at the spatial situation. Both dictionaries included the same verbal glosses, which were extracted from an English-Japanese dictionary (Eds. Tanaka, Takeda, & Kawade, 2003) with the permission of the chief editor.

This study postulates that a technologically enhanced dictionary can serve as both a reference tool and a learning tool for students acquiring the target vocabulary. The dictionary has traditionally been used to provide word meanings to help learners comprehend texts or to produce sentences, rather than as a resource to acquire vocabulary knowledge explicitly. A multimodal dictionary with several types of glosses, however, increases the saliency of target lexical items and their linguistic features. This creates the ideal environment for L2 vocabulary acquisition (Chapelle, 1998) and therefore leads to effective learning of the target vocabulary, although the literature has discussed the advantages of computer-assisted visual glosses mostly in terms of incidental learning.

Figure 2. Web-based dictionary with 2D visual gloss.
Furthermore, based on Ellis’ (1995) claim that word meaning acquisition requires semantic elaboration and imagery mediation, this study hypothesizes that explicit L2 preposition learning using this dictionary will increase learners’ awareness of the interrelationship between both words (Keane et al., 1997) and also the words’ organized semantic networks where their literal and figurative senses are reciprocally motivated. This will lead to effective vocabulary learning: selecting an appropriate word in various contexts L2 learners encounter. It is true that several studies have already concluded the advantages of animation for L2 learning (Sundberg, 1998; Al-Seghayer, 2001; Ling & Tseng, 2012). Therefore, our research questions for this research are as follows:

1. When Japanese L2 learners learn the locative prepositions with the multimodal bilingual dictionary, could their text comprehension be enhanced more than if they do not use it?
2. When they learn the words with the dictionary including three-dimensional animated aids, could their listening comprehension and sentence retention be enhanced more than when they use the dictionary with static visual glosses?

We will explain the detail of our experimental research in the next section.

4. Studies

Two experimental studies were conducted to examine the above research questions. Below are the descriptions of both sequentially.
4.1. Study 1

4.1.1. Participants

Fifty-two undergraduate students from a Japanese university participated in this research. As they major in either agriculture or technology, they do not specialize in English-related subjects. However, they are exposed to English during their studies; there is at least one compulsory English class for both freshman and sophomore students, while the junior and senior students have to read English journal articles related to their majors. In these respects, it would be estimated that their English language proficiency is at the lower to higher intermediate level. They were randomly divided into the control (n=26) and experimental (n=26) groups. Considering their constant exposure to English and homogenous English language proficiency, no test was conducted to divide them into these groups.

The experimental study could not be conducted simultaneously in the same location. Some sessions were held in the university’s computer room, while others were conducted in the author’s office with a maximum of five participants per session. All the studies were conducted under the author’s supervision.

4.1.2. Procedure

A personal computer with internet access was allocated to each participant (Windows OS). They were first asked to start the OS and to access the Moodle site developed only for this research.

The pre-test was then conducted on the Moodle site. The test consisted of forty-five fill-in-the-blank questions, where participants would provide the appropriate prepositions. Each question consisted of an English sentence with a blank and its Japanese translation. Within 20 minutes, the participants were asked to choose the most suitable preposition out of eight locative prepositions (i.e., above, across, along, below, in, into, on, over). This test was identical to the one used in my previous study (Sato & Suzuki, 2010). After the test, the correct answer was not given to the participants, and the next task was assigned.

The participants were asked to access the web-based bilingual dictionary for the eight target prepositions illustrated in Figure 2 and 3. They would then independently study the target prepositions with reference to the visual glosses. The images were different for each group: the control group referred to two-dimensional images (see Figure 2) derived from a paper dictionary (Tanaka, Takeda, & Kawade -Eds.-, 2003). The experimental group referred to three-dimensional animations of the images the author developed (Figures 3 and 4). Within 10 minutes, the participants were asked to understand the connections between the image and the words’ meanings.

The post-test was conducted immediately after the ten-minute study session. The test was the same as the pre-test, but the question order was randomized by the Moodle function. The method and duration of the post-test was identical to the pre-test. As this test comprised the final task, the participants were permitted to leave the PC room or the author’s office after they had completed the assessment.

4.1.3. Analysis

As all tests were conducted on Moodle, the test answers were automatically scored as the author set one correct answer as one point. The total scores were subsequently analyzed through a two-way (tests and treatments) repeated ANOVA measures.

4.1.4. Findings

In the control group, the average pre-test score was 19.23 (SD=4.30) and that of the post-test was 24.58 (SD=4.07). As for the experimental group, the average pre-test and post-test scores respectively were 19.23 (SD=4.18) and 24.31 (SD=3.47). The results from the analysis showed that both groups’ participants received higher scores in the post-test than those in the pre-test. As the average pre-test score of both groups was identical, the participants’ prior knowledge of the target prepositions would be almost the same; the score difference in the post-test could, therefore, be attributed to the treatments.
As seen in Table 1, the ANOVA result showed a significant difference in the within-subject factor ($F(1.50) = 112.49$, $p<.05$), whereas no significant difference was obtained in the between-subjects factor ($F(1.50) = 0.02$, $p>.05$).

<table>
<thead>
<tr>
<th>A = Image</th>
<th>B = Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>SV</td>
<td>SS</td>
</tr>
<tr>
<td>A subj</td>
<td>0.4712</td>
</tr>
<tr>
<td>B AxB</td>
<td>706.1635</td>
</tr>
<tr>
<td>SxB</td>
<td>313.8654</td>
</tr>
<tr>
<td>Total</td>
<td>2364.2212</td>
</tr>
</tbody>
</table>

Table 1: Results of the ANOVA analysis.

The significant difference in the within-subject factor supports the claim of previous CALL studies (e.g. Chun & Plass, 1996), stating that representing the target knowledge with both visual and verbal glosses could better facilitate vocabulary learning. On the other hand, no significant difference in the between-subject factor somewhat contradicts previous studies (e.g. Al-Seghayer, 2001), which claim that animation glosses could better facilitate L2 vocabulary acquisition. As a result, the first study’s results support the first research question but not the second. Therefore, a second study was conducted with the same treatment but a different research design.

4.2. Study 2

The second study examined how multimodal dictionaries facilitated listening to texts containing the target prepositions. As this study examined our second research question, our focus was only on the comparison between the treatments. We developed a fictional story that included the target prepositions as well as fifteen true-false questions about the story; participants needed to infer the answers to these questions by properly interpreting the prepositions’ spatial relationship. This reasoning task was based on how spatiality is crucial in constructing a text’s situation model referring to the deepest comprehension level (Zwaan & Radvansky, 1998).

4.2.1. Participants

Twenty college students joined this study. As all of them had participated in the first study, they simply remained in their original groups (9 in control, 11 in experimental).

4.2.2. Procedure

This study was conducted solely online through Moodle, but we could observe whether they had properly conducted the tasks because of Moodle’s management functions. The participants were asked to access the student Moodle site and to read the procedures displayed. The first task required participants to listen to the fictional story three times, where a woman provides directions to her friend on how to reach her flat from the nearest station (see Appendix 2). The story was read by a text-to-speech application (i.e., Speak it!) in American English, as this accent was familiar to the Japanese participants. After listening to the story, the participants were asked to understand the relationship behind the senses of the target prepositions with reference to the images in the dictionary in the same way as they did in the first study. They then answered fifteen true or false reasoning questions. These questions could not be answered correctly, even if they had memorized the text, which meant learners needed to have deeper textual comprehension in order to provide the correct responses.
4.2.3. Analysis
As in the first study, the participants’ answers were automatically collected and scored, and each correct answer was calculated as one point. The scores of each group were analyzed with a Mann-Whitney U test.

4.2.4. Findings
The average score of each group in the reasoning task is discussed below. The score of the control group was 13.00, whereas that of the experimental group was 8.46, which seems to be a large score difference. However, based on the U test’s results, no significant difference was obtained (p=0.08 >.05), although it was marginally significant at the 10% level. This result may indicate that two-dimensional images can better enhance deeper text comprehension than animated images despite the lack of significant difference between the groups. Therefore, our second research question was denied not only in the first study but also in the second study.

5. Discussion and conclusion
This study addresses the effectiveness of technology-enhanced visual glosses in explicit L2 preposition instruction. To test our hypotheses, two experimental studies were conducted on preposition acquisition using multimodal bilingual dictionaries: one of which displayed each word’s two-dimensional image schema while the other showed 3D animations of the schema. The findings showed that the visual glosses enhanced L2 vocabulary acquisition, regardless of the images’ configuration. On the other hand, no advantage was found in the technology-enhanced visual glosses, which showed the same result as obtained in our previous studies (i.e., Sato & Suzuki, 2010, 2011; Sato, Lai, & Burden, 2014).

The results could be interpreted in terms of the characteristics of the image schema and the influence of learner factors. In the field of Cognitive Linguistics, from which the image schema theory was derived, schematic images have flexibility and changeability in terms of their foregrounding, rotation, and focusing (Langacker, 1987). This implies that simple images are superior because they allow the learners to change the images in their minds to apply the images to each context, whereas the animated images may prevent learners from modifying the images due to their fixed configuration. Furthermore, individual factors may have affected the test results. Sato, Lai, & Burden (2014) suggest the influence of information processing styles, namely holistic or analytic cognitive style (Littlemore, 2001); this is based on Boers and Lindstromberg’s (2008) claim that L2 learning in the Cognitive Linguistics approach would be more suitable for those with holistic cognitive styles than those with analytic cognitive styles.

This study’s results show a pedagogical implication in the use of multimodal dictionaries in language classrooms. When L2 learners use online dictionaries accessible on their computers or mobile devices, positive learning effects are expected regardless of whether they use the dictionaries for their incidental or intentional tasks, even though their devices do not hold technologically advanced functions. As onscreen presentation with the multimodal functions can make target language and their linguistic features salient, a multimodal dictionary can be used as not only a reference tool but also a learning tool (Pachler, 2001). As an increasing number of institutions recommend that students bring their own devices into their classrooms, this study shows that technological functions of personal devices would not cause a big difference in students’ learning as long as web-based dictionaries are accessible and that more active use of the dictionaries for both incidental and intentional tasks are recommended.

There are some limitations to this study. In the first study, more analysis should have been conducted, such as a delayed test or a production task to write sentences using the prepositions. As for the second study, the number of the participants was not large enough to conduct a comparative analysis. Furthermore, the data’s validity could not be confirmed because all the tasks were conducted online without our observation. However, we believe that this study is not unreliable as our previous studies using different research designs received the same result (i.e., no difference between the treatments). To validate our findings and to optimize technological functions in CALL, future research is required: conducting an onsite study with a larger number of participants and taking into account individual learner factors.
The EUROCALL Review, Volume 24, No. 2, September 2016

References


**Appendix 1**

Links of the online bilingual dictionary for English locative prepositions

- Dictionary with two-dimensional static images: [http://goo.gl/seLOdk](http://goo.gl/seLOdk)
- Dictionary with animated images: [http://goo.gl/OCfI3A](http://goo.gl/OCfI3A)

* Click A, B, I, or O, and you will find the glosses of each target preposition as shown in Figure 2 and 3.

* Permission to use the sample sentences and translations was given by the chief editor of an English-Japanese dictionary (Tanaka, Takeda & Kawade -Eds.-, 2003) on condition that they were used only for research purposes.

**Appendix 2**

The script and questions for the listening task.

Dear Ken,

Thanks for your mail. I will tell you how to get to my flat.
When you come out of Hammersmith station, you’ll see a market across the street, which is Oxford Street. Turn right into Oxford Street and walk along the street towards St. Stephen’s Church. Pass the church on the right side and continue straight along the road. On the way, you’ll see a pub called Queen’s Pub. Just after the pub, on the left, is another pub called King’s pub. Turn right at the signpost “King’s Road” and walk along the path until you come to a bridge. Don’t cross over it but turn right and keep on walking along the river until you reach a restaurant called “Charles”.

Turn right into the narrow road in front of the restaurant. Follow the road and turn left just before you reach the park. At the end of this road is a row of houses. I live in the house in the middle. It’s number 3, and the number is on the door. The window of my room is on the second floor above the front door. Call my name when you get there, and I should hear you. If I’m not in, please find a spare key in the bucket which is upside down on the step beside your feet? You can find it when you turn the bucket over.

Best regards,

Lucy

True or False Questions:

1. When Ken goes back to the station on foot, he will turn right towards the Charles restaurant.
2. Lucy’s room is above house number 3.
3. The river is located on the left side of Lucy’s house.
4. St. Stephen’s Church is on the same side of Oxford Street as King’s pub.
5. The spare key is covered with the bucket.
6. King’s road is along the river.
7. The spare key is on the ground.
8. The number of the house is above the front door.
9. The park is along the river.
10. King’s pub is the farthest from the station out of King’s pub, Queen’s pub and Lucy’s house.
11. From the window of my room, the station can be seen.
12. Lucy’s house is in Oxford Street on the opposite side of King’s pub.
13. The Market is on the same side of Oxford Street as King’s pub.
14. There are houses on each side of Lucy’s house.
15. Ken will turn right into the road in front of Charles restaurant.