Haruko Miyakoda, Kei-ichi Kaneko, Masatoshi Ishikawa

Effective learning material for mobile devices: Visual data vs. Aural data vs. Text data

According to previous studies, visual data help to enhance vocabulary learning in a foreign language (e.g. Yeh and Wang 2003). However, it is not easy to create visual data for all lexical items. Particularly, we encounter problems when dealing with abstract words.

In this paper, we report on two experiments that tested the learning effects of four types of materials for mobile devices that were based on the following contents: 1. translation only, 2. aural data only, 3. visual data only, 4. aural and visual data. In Experiment 1, the 59 subjects were instructed to memorize the lexical items that were stored in 4 different types mentioned above, and vocabulary tests were administered in order to measure the memory retention rate of the meaning of each word.

Contrary to previous findings, the material employing aural data only drew out the best results. In addition, the material employing just translation scored better or did as well as either the translation + visual data or the translation + visual + aural data. This implies that contrary to claims found in the literature, visual data may not be so significant a factor in vocabulary learning. However, the difference was not statistically significant. In order to test the validity of this result, we conducted Experiment 2, following similar procedures. 40 subjects participated in the experiment. This time, again, visual data did not seem to provide much aid in facilitating vocabulary attainment, since the mean test scores were similar between translation only and translation + visual material.

The results from these two experiments suggest that contrary to claims in the literature, we may not need to rely too much on visual data in vocabulary attainment. Furthermore, material based on text data only may also prove to be an effective means of learning vocabulary in a foreign language.

1 Introduction

There has been a revival of interest in vocabulary teaching in recent years. This is partly due to the development of new approaches to language teaching, such as the lexical approach (Thornbury 2002). In addition to this, there are research findings showing how lexical problems can cause serious communication breakdowns, more severe is the nature of these problems than has been pointed out in the literature (Allen 1983).

Unlike the learning of grammar, vocabulary is largely a question of accumulating individual lexical items into long-term memory (Thornsbury, 2002). This means that one of the successful ways of achieving vocabulary attainment is to spend time on repetitive memorization activities (Schmitt and McCarthy, 2005). In this sense, ubiquitous autonomous learning can be seen as an ideal method of learning vocabulary, allowing learners to increase the time of exposure to the words to be learned, thus making good use of their time outside the classrooms.

With the advent of computers, new tools for studying vocabulary have been presented. Particularly, E-Learning based on mobile devices is getting more and more popular as a way of learning a foreign language (Amemiya et al., 2007). Employing mobile devices in vocabulary learning is an ideal way of studying because the mobility and portability of these devices provide the users with a ubiquitous environment, where they can study whenever and wherever they like. In addition to providing a ubiquitous environment, there is, of course, also the need to consider carefully the content of the learning material that is employed. This will be the topic to be taken up in the next section.

2 Learning material

Many papers dealing with learning material can be found in the literature, mostly supporting the effectiveness of visual data in facilitating the vocabulary learning process. For example, some studies investigating the difference between annotations by still images and those by movies, conclude that the learning effect by the annotations based on movies and texts are superior to those by still images and texts (cf. Al-Seyghayar 2001). On the other hand, other studies conclude that the annotations based on texts and still images are most effective (cf. Yeh and Wang 2003). Although these studies each have come up with different results as to whether movies or still images are effective, they all agree that visual data play an important role in vocabulary acquirement. However, as can easily be expected, although visual data may be effective, not all lexical items can be expressed visually. Furthermore, even if a word could be expressed using visual data, it does not necessarily mean that everyone will come up with the same visual image for the same lexical item.

In order to find out what role visual data play in vocabulary attainment, we conducted an experiment that compared the learning effects of different material, the details of which will be given in the following section. Before going into details of the experiment, we will briefly outline the vocabulary learning online system that we employed in the experiment.

3 Outline of the system

The online vocabulary system that we have developed consists mainly of three subsystems: 1. a system that supports or facilitates the creating process of the learning materials for mobile devices (Personal Super Imposer), 2. a system that supports its users in downloading the entities from the database and storing them for personal use (Personal Handy Instructor, 3. a system that allows users to share and evaluate the learning entities among themselves (SIGMA). Our emphasis in developing the system was to enhance the use of mobile devices in language learning and also to have the learners participate in creating the materials rather than just passively employ what has been prepared for them, because active involvement is expected to lead to effective learning.

3.1 Personal Super Imposer

As mentioned above, Personal Super Imposer (PSI) is a subsystem that supports its users in creating vocabulary learning material for mobile devices. When PSI is fed a five-second movie clip and the corresponding text data of the lexical item to be learned (namely the spelling of the word and its meaning), it automatically creates a multimedia learning entity. The process is outlined in Figure 1:

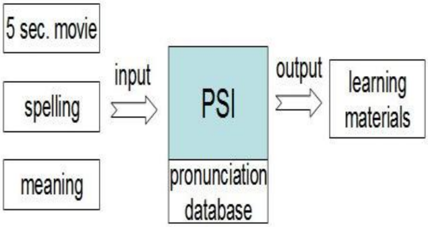

Figure 1: Personal Super Imposer

In Figure 2, a sample of the material made by PSI is given. The length of the movies for each entity is five seconds long. This length was determined based on the result of our pilot study that indicated that it was adequate in allowing most learners to process the visual/aural data repeatedly without feeling stress.

The pronunciation of each word is repeated twice in one learning entity. The spelling is displayed from the beginning of the entity but the corresponding meaning in the native language appears two seconds later. This time gap is necessary in order to give the learner a chance to concentrate first on the spelling before trying to memorize the meaning.

Figure 2: Sample of learning material (English-Japanese)

One of the advantages of creating material with PSI is that the same material can be reused or recycled so that it can be applied to virtually any language or dialect. For example, the item given in Figure 2 was originally created for Japanese learners of English. That is, the English word (foreign lexical item) appears on the first line and the corresponding meaning in Japanese appears on the second. If we change the typed-in information from Japanese to Chinese, the system automatically transforms the entity into an English-Chinese item (Figure 3):

Figure 3: Sample of learning material (English-Chinese)

3.2 Personal Handy Instructor

Personal Handy Instructor (PHI) is a vocabulary learning system for mobile devices such as iPods and mobile phones. PHI employs the five-second movie created by PSI mentioned above as its learning material. Figure 4 shows the process of learning vocabulary using PHI.

First, the learner selects the learning material that he/she wants to use from the learning-material list managed by PHI. The chosen material is copied into a folder called a 'vocabulary book.' Then, the users import the learning materials to their mobile devices such as iPods by dragging and dropping the folder onto the iTunes window. Finally, the users can download the learning material from iTunes to iPod. An outline of this process is summarized in Figure 4.

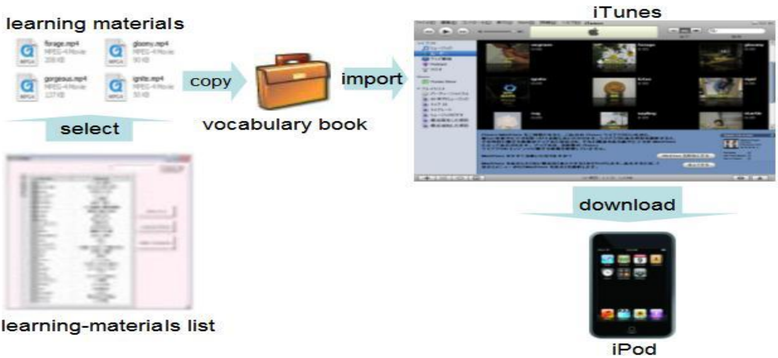

Figure 4: Personal Handy Instructor

3.3 SIGMA

The SIGMA system is a Web application that uses Apache, PHP and MySQL. It was designed to let learners register their own learning material and download the material created by other users.

In addition, it allows the users to evaluate the learning material and also give comments to each one. Figure 5 shows the main frame of the SIGMA system

If a user just wants to browse through the evaluation scores or comments of the learning materials, no login operations are required. However, if users want to evaluate the learning material or give comments on them, they need to become authorized users. Only authorized users can register and manage their own material and give evaluation scores and comments back on all material after login operation.

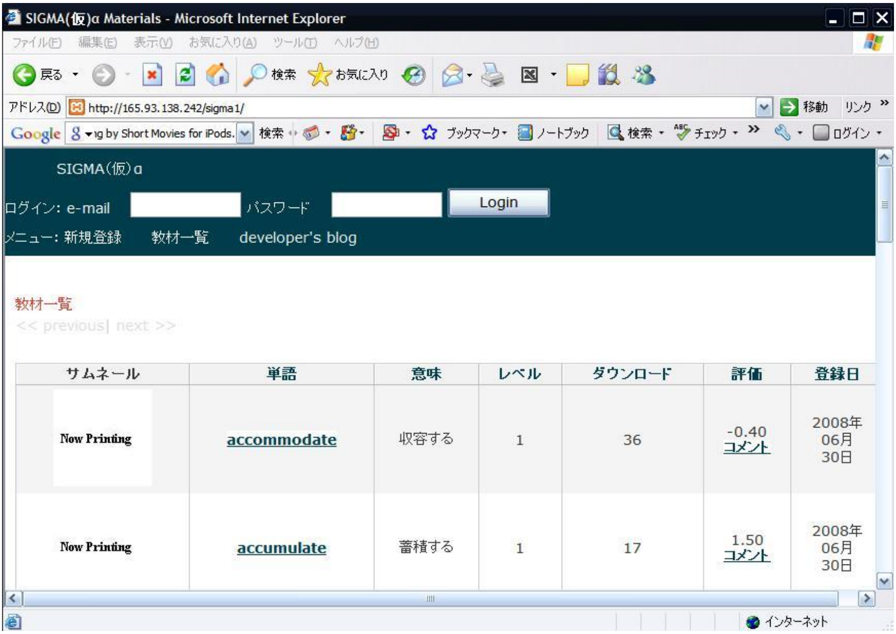

Figure 5: Main frame of the SIGMA

4 The Experiment No. 1

In order to see which factor is most effective in vocabulary learning, we conducted an experiment employing the online vocabulary system mentioned in the previous section.

4.1 The procedures

The main purpose of the experiment is to test the efficacy of visual data in vocabulary learning. Particularly, we wanted to compare the following four different types of material: 1. translation only (hereafter tr), 2. translation + aural data (hereafter aur), 3. translation + visual (hereafter vis), 4. translation + visual + aural (hereafter vis+aur). 59 undergraduate and graduate students attending a university in Tokyo, Japan participated in the experiment. The procedures prior to the experiment are as follows:

1. A vocabulary test was conducted by all participants in order to distinguish the lexical items that they were familiar with from the ones that were least familiar. The items that were least familiar among the participants were considered to be candidates for the experiment.

2. Based on the result of the vocabulary test in 1, we selected the following 15 items for use in the experiment: *ajar, beckon, bib, bicep, detour, disheveled, diverge, faucet, gargle, glimpse, hibernate, lament, perspire, pollen, stroll*.
3. The learning material for the 15 words above was created using the PSI system. For the (tr) and (aur) entities, no visual data is provided, and the screen would look something like Figure 6, where only the English word and the corresponding Japanese translation are provided as subtitles on a blank screen. The only difference between the (tr) and (aur) entities is that for the (aur) material, the pronunciation of the English word is additionally provided by sound data.

Figure 6: Example of iPod screen for (tr) and (aur) materials

4. For the (vis) and the (vis + aur) entities, in addition to the original English word and its translation in Japanese, visual data corresponding to the meaning of the word is provided, as the following example in Figure 7 indicates.

Figure 7: Example of iPod screen for (vis) and (vis+aur) materials

The procedures for the experiment itself are as follows:

1. The subjects were randomly divided into four learning groups (Group A-D), according to the type of material they were assigned to employ (i.e. Group A, (tr); Group B, (aur); Group C, (vis); and Group D, (vis+aur)).
2. The subjects were administered a test (Test 1) which included the 15 words mentioned above. In the test, the subjects were asked to write down the translation of the 15 lexical items in Japanese.
3. The subjects were given 5 minutes learning time. They were instructed to memorize the meaning of the 15 words using mobile devices.
4. Test 2 was conducted, which was based on English to Japanese translation tasks.
5. Test 3 was conducted a week later. The task involved was the same as Test 2.

4.2 Result and Discussion

The result of the experiment is summarized below in Table 1.

Test 1	A	B	C	D	Total
aver.	0.73	0.47	1.00	0.13	**0.58**
SD	1.87	1.13	2.32	0.52	**1.59**
Var.	3.50	1.27	5.38	0.27	**2.52**
Test 2					
aver.	21.20	22.47	20.00	20.20	**20.98**
SD	5.88	6.30	4.62	6.71	**5.88**
Var.	34.60	39.70	21.38	45.03	**34.57**
Test 3					
aver.	11.33	13.40	11.73	11.14	**11.96**
SD	7.45	6.63	7.94	5.60	**6.73**
Var.	55.52	43.97	63.02	31.36	**45.29**

Table 1: Result of Experiment 1

The full score for each test was 30 points (2 points x 15 lexical items). The participants received 2 points if they were able to answer the meaning of the lexical item correctly, 1 point if the meaning was partially right (e.g. “to be hot” for “perspire”), and 0 point if they were not able to come up with the correct answer.

When we focus on Test 1, we find that the results among the four groups vary slightly, with Group C having the highest average score of 1.00 out of 30.00, and Group D having the lowest score of 0.13. The average for all 59 subjects came out as 0.58. Since the 15 words employed in the experiment were all supposed to be least familiar to the subjects, the low average score for Test 1 was what we had expected. When we shift our attention to Test 2, which was administered to the participants right after the 5 minute learning process, we find that Group B, which employed the (aur) material, had the highest average score of 22.47 out of 30.00, followed by 21.20 for Group A (tr), 20.20 for Group D (vis+aur), and 20.00 for Group C (vis). For Test 3, which was administered to the participants one week after the learning process, here again, Group B had the highest average score of 13.40 out of 30.00, followed by Group C (11.73), Group A (11.33) and then Group D (11.14). In both Tests 2 and 3, the group that came out with the highest average score was Group B, which had used the translation and the sound data to memorize the meaning of the words.

In addition to these findings, we calculated the retention rate of the words' meanings between Tests 2 and 3. The retention rate was obtained by comparing each participant's test score for Test 2 with that of Test 3.

The overall results are summarized below:

Group A (tr) 52.70%
Group B (aur) 57.90%
Group C (vis) 52.53%
Group D (aur + vis) 57.25%

Here again, Group B scored the highest, followed by Group D, Group A, and then Group C.

In all cases, the material employing aural data only drew out the best results. Interestingly, in addition, the data employing translation only scored better or did as well as the visual data. This implies that contrary to claims found in the literature, visual data may not be playing as crucial a role as one might expect. As mentioned above, previous studies related to vocabulary acquirement generally emphasize the importance of employing visual data for effective learning. However, this is easier said than done. For one, it is very time-consuming to create visual data for use in the classroom. Finding the right visual data that corresponds to the word is a bit of a burden, and even with the aid of technology, it still requires much time and effort in the creation process. A further problem arises in the case of abstract words; visual images are hard to create for these words to begin with. You cannot easily find images for words like "lament", as you would for words like "lollypop" or "tiger." Therefore, even though material employing movies and visual images may seem useful for vocabulary learning, we must not forget that it has its limitations. It is hardly practical for advanced learners, who, especially must cope with abstract terminology most of the time. The result of our experiment indicates that there may not be the need to rely on visual data, and that employing either aural data only or translation only may be effective ways in vocabulary attainment. However, the results were not statistically significant. In order to verify whether visual data do play a role in vocabulary acquisition, we did a follow-up experiment, whose details will be given in the next section.

5 The Experiment No.2

40 university students participated in this experiment. They were randomly divided into four groups. Just as for Experiment 1, the four groups were divided according to the type of material employed in the learning process. That is, translation only (Group A), translation + aural (Group B), translation + visual (Group C), and translation + visual + aural (Group D). This time, however, in order to test the effectiveness of repeated learning, we conducted two sets of pre-test and post-test. In other words, after conducting the first set of pre-test and post-test on day one, a second round of the pre-test and post-test was conducted on day two, which was one week after day one.

The following 15 items employed in this experiment: *crumble, dilute, dormant, evalcuate, evaporate, glare, immerse, meditate, nudge, provoke, reconcile, sacrifice, seize, tempt, wither*.

The results of the mean scores for these two sets are given below:

pre-test 1	A	B	C	D	Total
	1.1	2.1	1.1	1.7	**1.5**
post-test 1					
	24.2	18.7	24.3	19.8	**21.7**
pre-test 2					
	13.7	8.6	13.8	9.9	**11.5**
post-test 2					
	29.3	26.8	29.7	28.4	**28.5**

Table 2: Mean Scores of Experiment 2

Just as for Experiment 1, the low average score for Test 1 was again what we had expected, since the 15 words employed in the experiment were all supposed to be least familiar to the subjects.

The pre-test 1 results for Group A and C were exactly the same (1.1), indicating that the starting point was exactly the same for the group that was provided translation only and the one that had both translation and visual data. The mean scores for post-test 1 conducted right after the learning session were also very similar for these two groups. The mean score for Group A in post-test 1 was 24.2, and 24.3 for Group C. The difference observed between the two is only 0.1.

If we shift our attention to the remaining two groups, we find it interesting to see that the mean scores for Group B and D in pre-test 1 were slightly higher than those of either Group A or C, yet, the mean scores for post-test 1 in the former were lower than the latter. This tendency for Group A and C to outscore the remaining two groups persists throughout. In pre-test 2, which was conducted a week later from the learning session, Group C scored the highest (13.8), followed by Group A (13.7), but the difference again was merely 0.1. In the end, all four groups were able to attain fairly good scores, but here again, Group C had the highest mean score followed by Group A, then D, then B, but the difference between Group A and C was slight (0.4). If we exclude the pre-test 1 scores, the result summarized in Table 2 depicts the fact that the learning effect varied greatly as to whether the material used aural data or not. Unlike the result obtained for Experiment 1, however, aural data did not work in favor of enhancing the learning effect. On the contrary, the mean scores for Group B (aur) and Group D (vis + aur), the two groups that employed aural data in the learning material, were constantly lower than the remaining two groups that did not employ aural data.

In addition to the mean scores for each group, we decided to measure the memory retention rates. Table 3 shows the experimental results of the memory retention rates of the four types of learning materials. The memory retention rate here refers to the difference observed between the results for post-test 1 and pre-test 2, conducted one week apart.

We conducted ANOVA based on the result of the memory retention rates, and found the F number for Factor 2 (aural) to be 7.81, as shown in Table 4

Factor 1 (visual)	Without		With	
Factor 2 (aural)	Without (= Group A)	With (= Group B)	Without (= Group C)	With (= Group D)
Data 1	0.333	0.267	0.567	0.133
Data 2	0.233	0.308	0.833	0.357
Data 3	0.233	0.333	0.267	0.607
Data 4	0.545	0.267	0.429	0.517
Data 5	0.367	0.067	0.367	0.375
Data 6	0.444	0.071	0.267	0.300
Data 7	0.367	0.133	0.130	0.300
Data 8	0.533	0.133	0.321	0.214
Data 9	0.733	0.364	0.233	0.000
Data 10	0.600	0.462	0.633	0.100
Number of Data	10	10	10	10
Average	0.439	0.240	0.405	0.290
SD	0.155	0.127	0.203	0.177

Table 3: Learning material and memory retention rates

Factors	Square sums	DOF	Mean squares	F numbers
Factor 1	6.17E-04	1	6.17E-04	1.97E-02
Factor 2	2.45E-01	1	2.45E-01	7.81E+00
Interaction	1.78E-02	1	1.78E-02	5.67E-01
Residual	1.13E+00	36	3.13E-02	
Total	1.39E+00	39		

Table 4: Result of ANOVA

If we put forward the null hypothesis that “there is no difference between the memory retention rates of learning material with sound and without,” the result obtained makes it possible to refute this with a significance level of 0.01. That is, from the average memory retention rates, we can conclude that the learning material without sound is superior to the one with sound. On the other hand, no significant difference could be observed for Factor 1 (visual). Furthermore, there was no significant difference in the interaction of these two factors either.

Since Experiment 2 was a follow-up to Experiment 1, both experiments were conducted under similar conditions. Experiment 2 parallels Experiment 1 in that visual data did not have a particularly positive effect on the participants' learning. The findings obtained from both experiments counter the general claim in the literature that visual data enhance vocabulary learning. However, as we have already seen above, the results of the two experiments came out as completely different in terms of the role that aural data play in vocabulary attainment. In Experiment 1, the material employing aural data only brought about the best result. The result for Experiment 2, however, showed that the two groups using aural data (whether with or without visual data) did not do as well as the groups that did not use aural data. The statistics based on the memory retention rates clearly indicated that aural data actually had a bad effect on vocabulary learning. Furthermore, the group that employed the translation only material actually did quite as well as the group that employed visual data in addition to the translation.

Visual data may be useful, but there are limitations to the type of words that can be expressed visually. Our results indicate that vocabulary attainment may be achieved without relying too much on visual data. If so, then this may shed new light on ways of designing learning material for online use. Instead of trying to force oneself to link text data with visual data or aural data for all lexical items, incorporating flexibility into the vocabulary learning system may be the key for providing a better learning tool. It is generally the case that most of the words that we need to learn in a foreign language are abstract and difficult to express as visual data. If the text only data (translation) can do as well a job as a text (translation) + visual data, then we may not have to worry about finding suitable visual data for these types of abstract words. We, however, do not intend to claim that visual data should be totally denied in vocabulary learning.

It is true that the effectiveness of visual data was not proven in terms of the scores obtained in the experiments. However, the results of the questionnaire that we had conducted on participants of Experiment 2 (right after post-test 2) indicate that there were many who felt that using visual data actually facilitated the learning task. Some participants in Group C (vis) and D (vis + aur) mentioned that the learning task was, in fact, actually fun and entertaining. Another point worth mentioning is that when we were observing the participants during the test session, we found that many participants of Group A (tr) were trying to answer the questions in alphabetical order, that is, they started answering from *crumble* (the first alphabetical word on the list) and ended with *wither* (the last alphabetical word). Since we had presented the vocabulary list in alphabetical order in the pre-tests and the learning process, these participants may merely have been memorizing the words by rote. Since the lexical items on the post-tests were all randomly ordered, they may have been rearranging the words in the order that were presented in the learning process. This tendency to prefer alphabetical order was not observed for the participants in Group C. In other words, almost all of the participants in this group were writing down the answers to the test questions regardless of the order. They may have been able to answer in any order because the way they memorized the words was not by rote (as was the case for the participants in Group A).

If we take this point into consideration, although Group A and C both brought about similar results throughout the two sets of pre-test and post-test that we had conducted, the quality of their understanding of each lexical item may be different. If participants in Group C could answer the test questions regardless of word order, this may mean that they may have achieved a higher level of understanding compared to those in Group A. If we had conducted the post-tests on a longer time span, then we may have been able to measure the difference of memory retention rate between the two groups more clearly. Furthermore, by increasing the number of lexical items, or testing the meaning retention from a different perspective other than translation (e.g. reading comprehension test, definition test, etc.), this difference may further easily be teased out. This is left for future research.

Before ending, we need to consider some of the possible reasons as to why aural data had such a bad effect on the learning effects in Experiment 2. One possible factor may have been the environment under which the experiment was conducted. Although we made best effort to conduct the experiments in quiet a place as possible, there were several occasions during the learning/testing sessions where we could not prepare as suitable an environment as Experiment 1. Both experiments were conducted at different universities, and this may have affected the results. It may be necessary to prepare better controlled environments so that the participants will not be influenced by the surrounding sounds such as passing cars and so on.

6 Concluding remarks

In this paper, we presented the results obtained from the two experiments that tested the learning effects of different types of vocabulary material. By employing the vocabulary creating system that we had developed for mobile devices, we created the following 4 different types of material: 1. translation only, 2. translation + aural data, 3. translation + visual data, 4. translation + aural + visual data. In Experiment 1, the 59 subjects were instructed to memorize the lexical items that were stored in 4 different types just mentioned above, and vocabulary tests were administered in order to measure the memory retention rate. Contrary to claims found in the literature, visual data did not provide much aid in facilitating vocabulary attainment, since the mean test scores were similar between translation only and translation + visual material. A similar tendency could be read off from the results from Experiment 2. This finding has some important implications for language teaching, especially for lexical items that are difficult to relate to visual images, such as abstract words. It is interesting to note that in the second experiment, the learners who employed the translation only material did as well as those who used both translation and visual data. Furthermore, the translation only group outscored the group that employed translation + aural + visual data. Based on this fact, we conclude that it is most likely that vocabulary attainment may be achieved without relying too much on visual data.

Acknowledgments

The authors are grateful to Takeshi Goto and Marie Matsumoto for their contribution to the experiments presented. This study is partly supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT) Fund for Promoting Research on Symbiotic Information Technology. It is also partly supported by the MEXT Special Coordination Funds for Promoting Science and Technology.

Bibliography

ALLEN, V. F. (1983). *Techniques in teaching vocabulary.* Oxford and New York, Oxford University Press.

AL-SEYGHAYAR, K. (2001). "The effect of multimedia annotation modes in L2 acquisition: a comparative." *Language learning and technology, 5*, 202-232.

AMEMIYA, S., HASEGAWA, K., KANEKO, K., MIYAKODA, H, AND TSUKAHARA, W. (2007). "Long-term memory of foreign-word learning by short movies for iPods." *Proc. of the 7th IEEE International conference on advanced learning technologies*, 561-563.

HASEGAWA, K., AMEMIYA, S., ISHIKAWA, M., KANEKO, K., MIYAKODA, H, AND TSUKAHARA, W. (2007). "Promoting autonomous learning: a multilinguistic word learning system based on iPod." *Proc. of the 2007 International conference on ESL/ EFL*, 70-83.

SCHMITT, N., AND MCCARTHY, M. (2005) *Vocabulary: description, acquisition, and pedagogy*. Cambridge, Cambridge University Press.

THORNBURY, S. (2002). *How to teach vocabulary*. London, Longman.

YEH, Y., AND WANG, C. (2003). "Effects of multimedia vocabulary annotations and learning styles on vocabulary learning." *CALICO Journal*, 21 (1), 131-144.