We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

A Software System for Inference-Based Context-Aware Vocabulary Learning

Yukiko Sasaki Alam

*Department of Digital Media, Hosei University
Tokyo, Japan*

1. Introduction

This chapter describes an educational software system for inference-based context-aware English vocabulary learning which has been put to use for three years at a university in Japan. It is a Web application so that the students can learn independently outside the classroom at their own individual paces anywhere anytime. Unlike a majority of commercial vocabulary learning software based on word-for-word translation, the current system is designed on effective learning principles from the point of view of vocabulary building as well as online learning, both of which would otherwise make learning tedious and boring. In addition, the system provides the instructor with a simple means of preparing and editing exercises according to his or her pedagogical needs.

A long history of research on the learning of vocabulary has demonstrated a high correlation between the learner's vocabulary knowledge and language proficiency (Golker & Yamini, 2007; Webb, 2005; Zareva et al., 2005; Qian, 2002, to name a few). Learning vocabulary would be probably the most important and time-consuming part of learning a foreign language (Atay & Ozbulagan, 2007). Educational software for learning second-language vocabulary, however, is not often designed on commendable learning principles. Many of the vocabulary exercises are based on word-for-word translation, which encourages the learner to learn individual words out of context and translate them into the equivalents in his or her native language. The result is that the learner has little knowledge about how the words are actually used in sentences.

The present system, on the other hand, is intended to help the students:

- (i) learn the usages of the words,
- (ii) infer the missing words from context,
- (iii) guess the words by reading the definitions in English
- (iv) think in English, and
- (v) develop the habit of using an English-to-English dictionary.

To achieve the above goals, each question is provided with a sentence with a word removed. The blank portion consists of as many blank spaces as the number of the letters of the word except for the initial letter already given. The question is also furnished with an English definition of the meaning of the word. The students should be able to infer the word from (a) the example sentence, (b) the definition of the meaning, (c) the initial letter and (d) the

number of the letters left out. At the end of the exercise session, they are provided with a feedback about mistakes so that they can review the mistaken words before challenging the exercise again.

In addition to pedagogical aspects, the developer of educational software should also realize the fact that online learning is indeed a monotonous and impersonal practice. The current system, therefore, made an effort to make online learning more stimulating by defining clear goals of achievement, giving graphical displays of the goals and progress of learning in real time, and making questions puzzle-like. Unlike usual commercial software, the instructor can choose words that deem appropriate for his or her students’ language levels, and therefore more meaningful to them.

This chapter will elaborate on the underlying educational background, the interfaces and designs of the system, together with findings and potential improvements that were suggested from three years of use as an online system for college second-language vocabulary learning.

2. Vocabulary learning

Vocabulary acquisition is said to occur intentionally, for instance, by doing exercises or incidentally, for example, through reading texts, and extensive research on the intentional vs. incidental acquisition issue has been conducted (Herman et al., 1987; Hulstijn et al. 1996; Webb 2008, to name a few). The current system is intended primarily for intentional vocabulary acquisition, because incidental acquisition would not occur often enough in foreign language learning material (Groot 2000).

Vocabulary knowledge could be divided between passive and active knowledge, as illustrated in Table 1. Even though the distinction between passive and active knowledge may not be as clear as it may deem (Read 2000), it is helpful in discussing vocabulary knowledge. Passive listening and reading vocabulary knowledge is useful in listening and reading comprehension respectively, but not necessarily so in such productive skills as speaking and writing. Active writing and speaking vocabulary knowledge is effective in writing and speaking respectively.

	Oral	Graphic
Passive	Passive listening	Passive reading
Active	Active speaking	Active writing

Table 1. Types of vocabulary knowledge

The current system aims at helping develop active writing vocabulary knowledge primarily, and as a by-product, passive reading vocabulary knowledge, because the students read many example sentences and definitions while doing the exercises. Therefore, many of the target words included in the current vocabulary exercise are those that the students are able to recognize in reading, but are unable to use properly in writing or speaking. The words are taken from the words at Levels 2 and 3 in the Hokkaido University Vocabulary List. The questions presented in the present system are not provided with multiple-choice answers, but ask students to fill in the blank by guessing the word from the definition of the meaning written in English, and the sentence in which it occurs. The other clues are the number of the letters of the word as well as the initial letter already given.

The main purpose of the use of the definitions of the words and the example sentences is to encourage students to obtain the meanings of the words without interference from their native language. Many vocabulary exercises are types of word-for-word translation, thus likely fostering habits of mental translation between words in target and native languages. This often results in an undesirable situation in which the learners know the equivalents of the words in their native language, but have little knowledge about the usages of the words.

3. The design of the system

The current software system has two major modules: the student module which is responsible for the students' vocabulary learning, and the instructor module for the instructor's creating and editing questions. The following section will first describe the student module, and then the instructor module.

3.1 The student module

The student module contains three panels, the *login*, *question* and *end* panels, the last two of which will be discussed in detail in this section.

3.1.1 The question panel

The question panel is comprised of three components: a set of three graphs, the question section and a number of buttons, as illustrated in Figure 1. The top bar graph indicates how many levels or units of the vocabulary exercise have been completed, and the lower two graphs, how many words were answered correctly one time and two times respectively. The student can go up to the next level or unit only after giving all the correct answers twice. These three graphs are intended to help the student see graphically his or her progress in real time, and how much more left to complete the exercise.

The question section on the panel provides an English sentence with an array of blank spaces. The first letter of the word is already in the first box as a clue, and the student is expected to fill in the remaining blank spaces by inferring the missing word from the meaning of the sentence as well as the definition of the meaning of the word situated above the example sentence. In this way, the student can guess the word from several clues.

As most words are used in more than one part of speech, and have several meanings even in each part of speech, it is important to specify which part of speech and the meaning are questioned by showing the word within a sentence. For instance, the word *steady*, which is the answer of the question on the panel in Figure 1, can be used in adjective, verb, adverb or noun. *Steady* in adjective has several meanings: not moving, continuous, not changing, reliable, and so on. The meaning of not moving is in question at present.

Besides the clues to the guessing of the word, the system allows the student to access a dictionary available online at another site. When finding an unfamiliar word that occurs in the definition as well as the question sentence, he or she just clicks on it to get the meaning.

At the bottom of the question panel is the button named *Send_My_Results_Now*, which allows the student to send the results intermittently to the database at the server so as not to lose the results so far during the exercise. On the other hand, the button *End-This-Session* finishes the session, and opens the end panel.

Fig. 1. The question panel

Fig. 2. The end panel

3.1.2 The end panel

The *End_This_Session* button on the question panel opens the end panel showing a list of all the words the student has failed to answer, as illustrated in Figure 2. As the Japanese glosses are provided, he or she can copy the feedback, and review the mistaken words before challenging the exercise again.

3.1.3 The student data file

The results of the student's performance is saved on the student data file in XML format, a portion of which is shown below:

```
<STUDENT>
  <ID>09k3030</ID>
  <FAMILY_NAME>sasaki</FAMILY_NAME>
  <GIVEN_NAME>yukiko</GIVEN_NAME>
  <GROUP>g</GROUP>
  <PASSWORD>secret</PASSWORD>
  <MINUTES>153</MINUTES>
  <LEVEL>13</LEVEL>
  <DONE>2</DONE>
  <STATUS>1 0 0 2 0 1 0 1 0 0 1 0 1 0 0 1 0 1 0 2 0 1 0 1 1 0 1 0 0 1 </STATUS>
</STUDENT>
```

Listing 1. A *STUDENT* element in the student data file

The content of the *MINUTES* element indicates how many minutes the student has so far spent on the exercise. The content of the *LEVEL* element denotes which level or unit he or she is currently working on, while that of *DONE*, how many words have been correctly answered two times, which is the number of times of repetition required at present for considering a word to have been learned. The values of *STATUS* shows how many times the words listed in the words file for that level have been answered correctly: the number 1 indicates one time, while the number 2, two times. Such a data structure on *STATUS* allows for an easy modification of the number of required repetition as well as the number of words assigned for one level or unit.

3.2 The instructor module

The instructor module has three panels, the main data input panel, the text editing panel and the last-word editing panel. The data input panel allows the instructor to add questions to the questions file in XML format, which is used for displaying the components on the question panel. The text editing panel allows the instructor to edit the questions file, in case he or she needs to modify or correct it. On the last-word editing panel, the instructor is able to change the last word which has been saved in a separate file storing the last word. This function is necessary when the instructor wants to see a list of words, and make sure that he or she has not skipped any important word.

3.2.1 The data input panel

When the data input panel opens, a word to work on and a number appear in the boxes located in the middle at the top of the panel, as illustrated in Figure 3. The number does not stand for the order of the word in the dictionary corpus, but the order in the list of words selected for making questions. However, the instructor can change his or her mind and decide not to use the word for some reason. The total number of words actually adopted for making questions is shown at the bottom of the panel next to the word *total*. Another number shown next to the phrase *words saved in this session* is the number of the words that have been stored during the session in question. The system, in reference to the last word which has been processed, retrieves the next word from the words list, and uploads to the panel the dictionary entry for the word.

Fig. 3. The data input panel

The left column of the panel contains the meanings of the word as well as the example sentences, while the right column a blank template in XML format. The instructor is able to skip the word on the panel by pressing the *Next word* button, when he or she thinks it is not worth making a question about it. Looking at the left column, the instructor chooses one definiton from among many in the dictionary entry and one of the example sentences in order to fill in the XML template in the right column. Listing 2 below is an XML template already filled in with such information as the part of speech (*POS*), the definition, an example sentence, and a Japanese gloss:

```
<WORD>
  <ID>1318</ID>
  <LITERAL>willing</LITERAL>
  <POS>adj
  <MEANING> to be prepared to do something, or have no reason to not want to do
  it
    <EXAMPLE>How much are they (willing) to pay?
    </EXAMPLE>
    <JAPANESE>自発的にしようとする</JAPANESE>
  </MEANING>
</POS>
</WORD>
```

Listing 2. A *WORD* element in the questions file.

Except for the values of the *ID* and *LITERAL* elements, which are automatically entered, others must be entered by hand. The word which will be the missing word in the question must be placed in parentheses in the example sentence, so that the system leaves out the word and put blank spaces back according to the number of the letters of the word.

The design of the *WORD* element is based on the assumption that a word may belong in more than one part of speech, and that a word in different parts of speech may each possess more than one meaning.

3.2.2 The text editing panel

Fig. 4. The text editing panel

The instructor is able to edit directly the questions file which provides the question panel with information such as the definitions of the words and the question sentences. This editing is required when the instructor wants to make sure all the information has been placed properly or wants to check if important pieces of information or an important word have been skipped by referring to the words list displayed in the left column of the panel.

3.2.3 The last-word editing panel

The last-word editing panel shows up by pressing the *edit-last-word* button on the data input panel as well as the text editing panel. This panel allows the instructor to include a word he or she has skipped by replacing the last word, stored on a separate file, with the word preceding the skipped word, and going back to the data input panel to enter information on the word. Furthermore, it also allows the instructor to skip many words which otherwise must be skipped one by one on the data input panel. He or she just replaces the last word with a word much ahead in the words list shown in the left column.

4. Evaluation and future work

It would be certainly not so straightforward as deemed to measure effects of any educational software on learning, since one's learning depends on a variety of factors, and such factors can hardly be singled out. There was, however, an interesting result worth mentioning. The current system was incorporated as an assignment for two classes of the first-year English courses in the Faculty of Computer and Information Sciences at Hosei University, and not in the other two classes in the academic year of 2006. All the first-year students took TOEFL-ITP exams both at the beginning and end of the academic year, namely in April, 2006 and January, 2007. The students in the two classes who used the system showed significant improvement in their comprehensive as well as itemized scores, whereas the students in the other two did not. This seems to indicate a positive effect of the current system on the learning of a foreign language. However, it has not been possible to compare students' results since then, because the system has been made available to all the first-year students.

The system offers three parameters that can be changed by the instructor: (a) the number of words to be included in one level or unit, (b) how many times students should answer the questions correctly, and (c) the number of levels or units they should complete. At present thirty (30) words are assigned for each level or unit, and are contained in each different file. Each word is structured as an XML element, as illustrated in Listing 2 above. Earlier, the number of words for each level was fifty (50), but as the student has to repeat the questions twice, it has been reduced to thirty (30). The system is in use for two semesters of the first-year English classes: the students have to practice twelve (12) units in the first semester and fourteen (14) in the second semester.

According to the data collected from seventy three (73) and ninety seven (97) students respectively in the academic years of 2007 and 2008, the average time required for completing one level or unit in such settings mentioned above was about fifty (50) minutes overall in the first semesters. The time required, however, ranged from about thirty (30) to eighty (80) minutes. A wide gap has been observed in the time required for completion. For some students, the exercise was not so hard, but for others it was quite challenging initially. In the second semesters, the average time needed for completing one level or unit was about

thirty six (36) minutes, much shorter than that in the first semesters.

The students were tested for remembering the usages of the words at several mid-term and final exams. The students who completed the vocabulary exercise remembered about 55% on average, ranging from 20% to 100%. It seems that some students need to repeat the questions more than twice, while others do not. The retention rate is a little higher than those reported in Mondria & Wiersma, 2004, probably because many of the target words here are those the students are able to recognize in reading, but unable to put to use in writing. In spite of several clues including the definitions of the words, the example sentences, the numbers of the letters of the words and the initial letters, some students found it quite difficult to guess the missing words. There were several words even a native speaker failed to guess, and some modifications were required subsequently. It might be helpful to give more clues, for instance, by showing the constituent letters of the word in the wrong order or by introducing the words orally before testing students on them. It might also be beneficial to students at varying vocabulary levels to give clues step-by-step on demand, and not to give all the clues at once. In addition, students could be provided with an immediate feedback in the middle of the session when their answer is incorrect. At present, they receive a feedback on mistaken words at the end of the session so as to review them for the next attempt. Another possible future arrangement would be to add more example sentences.

5. Conclusion

Unlike a majority of commercial vocabulary learning software based on word-for-word translation, the present software system is designed so that students can learn the usages of words, not the equivalents in their native language, and think in English. For these purposes, most of the words included in the exercise are those students are able to recognize in reading, but are unable to put to use in writing or speaking. All the definitions of the missing words and the example sentences in the questions of the exercise are written entirely in English, with Japanese used only for the final feedbacks on mistakes. Students can infer the missing words from several clues including the example sentences as well as the definitions of the words, and become familiar with how they are used in sentences. This encourages them to understand English words and sentences directly in English, not through their native language.

In addition, the present system is intended to make online learning less tedious and boring by making use of visual aides such as graphs. The graphs indicate several types of students' performance results, and change in real time according to the results. Students receive feedbacks on mistaken words at the end of the session so that they are able to review them for their next attempt.

The system also saves the instructor much time and labor. It provides him or her with a simple means of preparing questions from a dictionary corpus in reference to a list of words the usages of which he or she thinks the students must learn. It also allows the instructor to edit the questions file with ease.

While some improvements seem desirable as mentioned above, the system has been implemented at a university in Japan for several years as a tool for the students' independent learning of vocabulary outside the classroom, and has proved to be a useful online application in teaching second-language vocabulary.

6. References

- Atay, D. & Ozbulagan, C. (2007). Memory strategy instruction, contextual learning and ESP vocabulary recall. *English for Specific Purposes* 26, pp. 39-51, Elsevier, Amsterdam, The Netherlands
- Golkar, M. & Yamini, M. (2007). Vocabulary, Proficiency and Reading Comprehension. *The Reading Matrix*, Vol. 7, No. 3, pp. 88-112
- Groot, Peter J. M. (2000). Computer assisted second language vocabulary acquisition. *Language Learning & Technology*, Vol. 4, No. 1, pp. 60-81
- Herman, P. A., Anderson, R. C., Nagy, W. E., & Pearson, P. D. (1987). Incidental acquisition of word meaning from expositions with varied text features. *Reading Research Quarterly*, 22, pp. 263-284
- Hulstijn, J., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, Vol. 80, No. 3, pp. 327-340
- Mondria, J-A & Wiersma, B. (2004). Receptive, productive, and receptive+productive L2 vocabulary learning: What difference does it make?, In: *Vocabulary in a Second Language*, Bogaards, P & Laufer, B, (Eds.), pp. 79-93, John Benjamins Publishing Company, Amsterdam
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52, pp. 513-536
- Read, J. (2000). *Assessing vocabulary*, Cambridge University Press, Cambridge, England
- Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27, pp. 33-52
- Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, Vol. 20, No. 2, pp. 232-245
- Zareva, A. Schwanenflugel, P. & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: variable sensitivity. *Studies in Second Language Acquisition*, 27, pp. 567-596

Advanced Learning

Edited by Raquel Hijn-Neira

ISBN 978-953-307-010-0

Hard cover, 444 pages

Publisher InTech

Published online 01, October, 2009

Published in print edition October, 2009

The education industry has obviously been influenced by the Internet revolution. Teaching and learning methods have changed significantly since the coming of the Web and it is very likely they will keep evolving many years to come thanks to it. A good example of this changing reality is the spectacular development of e-Learning. In a more particular way, the Web 2.0 has offered to the teaching industry a set of tools and practices that are modifying the learning systems and knowledge transmission methods. Teachers and students can use these tools in a variety of ways aimed to the general purpose of promoting collaborative work. The editor would like to thank the authors, who have committed so much effort to the publication of this work. She is sure that this volume will certainly be of great help for students, teachers and researchers. This was, at least, the main aim of the authors.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yukiko Sasaki Alam (2009). A Software System for Inference-Based Context-Aware Vocabulary Learning, *Advanced Learning*, Raquel Hijn-Neira (Ed.), ISBN: 978-953-307-010-0, InTech, Available from: <http://www.intechopen.com/books/advanced-learning/a-software-system-for-inference-based-context-aware-vocabulary-learning>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen