Research Paper

Vocabulary Learning: an Ideal Candidate for the Application of CALL

ALLUM Paul

Key words: 語彙、修得、CALL、ソフトウェア、デザイン
vocabulary, learning, CALL, software, design

Abstract

Vocabulary learning is well researched (e.g. Nation, 2001), clear pedagogic guidelines are available (e.g. Nakata, 2011), and numerous papers have been published on the use of CALL to enhance such learning (e.g. Mohsen & Balakumar, 2011; Esit, 2011; Fehr et al., 2012). However, as Nakata (2011) points out, there is considerable room for improvement in the design of software created to support vocabulary learning. In this paper, I hope to outline a potentially 'ideal' generic design for vocabulary-learning software, one that takes into account the processes required to successfully acquire new items. While pedagogic principles are the starting point of the design, an attempt is made to consider HCI processes (e.g. Redmond-Pyle, 1995) in the analysis of the learning task. Results produced in actual use help refine the theoretically derived model and indicate more precisely how data needs to be collected and used to automate the process, as well as what kinds of database are required. The resultant design reported here would require impractical levels of human intervention if attempted in an educational environment lacking in computers, but the final aim would be to produce an automatic system, one that fully realizes the capacities of programming and that can generate an appropriate and pedagogically sound learning path for new vocabulary items, whether operated by students themselves, working directly from the text, or a teacher, with minimum effort.

1. Introduction

Vocabulary learning has been the focus of considerable research. The 'classic' text on the topic, Nation's (2001) *Learning Vocabulary in Another Language* indicates broadly the steps that provide the greatest chance of successful learning outcomes. Some of these steps have been refined (e.g. Laufer, Elder, Hill, & Congton, 2004; Laufer & Goldstein, 2004). Additional factors that make for more effective task design have been adduced by others (Bjork, 1999; Pyc & Rawson, 2007; Laufer & Hulstijn, 2001). Finally, work on memory has indicated yet other considerations that should benefit the design of vocabulary learning tasks (Baddeley, 1997; Nation, 2001; Pimsleur, 1967). Bringing together the fruits of these strands of research can provide a designer of vocabulary learning tasks with a clear path to follow. However, when implementing such a path in a CALL application, there are considerations as to what kind of interface is needed and how the tasks that are often performed using traditional classroom facilities or pencil and paper can be realized within both the interface and the functionality of the computer. For this aspect of CALL design, it is beneficial to refer to Human Computer Interface (HCI) design principles, which have been well elucidated (e.g. Redmond-Pyle, 1995). Finally, beyond these considerations it is also useful to refer to reports of practical application of these ideas in order to make the design into one that works in reality. In sum, reference is made to the theory and application of both language learning and human-computer interface design, as well as to the results of efforts to implement these ideas in earlier software. Each of these is considered in more detail below.

2. Learning Theory

In an earlier paper (Allum, 2012) general second language principles were referred to, these being based on the work of Chapelle (1998a, 2001) or Jamieson, Chapelle & Preiss (2005). These should be taken into account, however, in this paper I will focus more closely on the principles that apply specifically to vocabulary learning. These were also summarized in earlier papers (Allum, 2004, 2012) so here I will briefly resume the basic ideas and add further points. Nation (ibid.) indicates that the three basic processes for vocabulary learning are noticing, retrieval and generative use. Noticing involves decontextualisation, taking the word out of its background context, and finding its meaning. Effective ways to do this are, for example, giving a definition or an L1 translation. Retrieval is the process that strengthens the memory of a word. Retrieval can be further divided into receptive and productive retrieval. Receptive retrieval involves retrieving the meaning when seeing the word, while productive retrieval means being

able to retrieve the word for a particular meaning. This idea has been further refined by sub-dividing retrieval into two types, recall and recognition (Laufer, Elder, Hill, & Congton, 2004; Laufer & Goldstein, 2004). Recognition means that when we are presented with a word or meaning and a choice of several options for the appropriate meaning or word respectively, we can recognize which is the correct one out of the selection of choices. The typical example of a recognition exercise would be a multiple choice type question where the learner is presented with a word or meaning and has to choose which of the options matches it. This process can be either receptive or productive. For example, if a word is provided and a choice of several meanings is also provided, when the learner manages to choose the correct meaning (for example in a multiple choice format), this is receptive *recognition*. Alternatively, if the learner is provided with the meaning (typically in the case of vocabulary the L1 word) and has to choose the matching L2 word from several options, this is productive *recognition*. Receptive *recall* on the other hand, as stated above, involves having the word and recalling the meaning, while productive recall involves having the meaning and recalling the correct word, without having any options presented in either case. A further refinement when considering designing exercises is the idea that retrieval, of whatever type, only strengthens memory when it is successful (Modigliani, 1976, Barcroft, 2007). The last process, generative use, involves production or perception of the word in a different grammatical form, in a new context, or with a new meaning.

Laufer & Hulstijn (2001) have suggested another factor that affects vocabulary learning, 'task involvement'. There are three elements that determine the strength of this: need, search and evaluation. For example, need for a vocabulary item is moderate if the task demands it but strong if the learner feels the need for it. Search is moderate if learners have to search for the meaning of the item and strong if learners have to search for the form to express the meaning; in other words, productive retrieval requires more task involvement than receptive retrieval. Evaluation is the process of deciding if a word choice is appropriate or not. Evaluation is moderate if the context is provided and is strong if the learner has to create a context. Overall, the stronger the task involvement is, the greater the strengthening of the memory of that word.

One further very important idea derived from memory theory is the concept of spacing (Baddeley, 1997; Nation, 2001; Pimsleur, 1967). Spacing refers to the period of time between each retrieval of a word. This theory suggests that successively longer spacing (length of time) between *successful* retrieval attempts strengthens memory more than frequent presentation or repetition of an item. Learning based on application of this principle is called 'expanded learning' (see Nakata, 2008, 2011 for more detailed explanation).

2. 1. Application of Learning Theory

How can these principles be applied in reality, and what would be the resulting learning tasks? In this paper, I will focus on one typical case of students encountering new vocabulary, a high intermediate reading class focusing on current news English. However, the same principles and approach could be applied to almost any learning situation in which students encounter new vocabulary.

The first step is simple. Students need to notice the word and get its meaning. The simplest way to do this is to get a translation into L1. While not the only way, Nation (ibid.) indicates that the key point is to have form and meaning matched accurately, unambiguously and quickly.

Once a word has been matched with its meaning, a series of retrievals of that word should take place. Given the importance of success in the retrieval process (Modigliani, 1976; Barcroft, 2007), it would seem reasonable to grade the retrieval steps in order of difficulty or effort, taking into account the concepts elucidated by Laufer et al. (2004) and Laufer & Hulstijn (2001) (cf. Nakata, 2011). Thus recognition tasks should probably come before recall tasks and in both cases receptive tasks should come before productive. This would result in four task types done in this order: receptive recognition, productive recognition, receptive recall, productive recall. While this would seem to maximize the learning potential by increasing the chance of successful retrieval by means of sequencing the tasks in order of difficulty from easier to harder, it is also necessary to increase the spacing between tasks which itself adds a dimension of difficulty. Two recommendations emerge from the literature. Increasing the absolute spacing is very important (Karpicke & Bauernschmidt, 2011) and such spaced retrieval should progress from relatively short periods, a few minutes, to much longer periods of a month or more and continue over years (Dempster, 1987). Mizuno (2000, 2003b) has suggested an algorithm to obtain the optimal spacing between retrieval attempts, and Nakata (2006a, 2006b) has further investigated its application to vocabulary learning. However, these recommendations on spacing are based mainly on experiments with productive recall only (possibly some receptive recall), using flash cards or similar, that present the Japanese word for which the learner then has to produce the English word. Thus the application of this algorithm to the other varieties of memory-strengthening task suggested above is not quite as sure and, does not take *task difficulty* into account as a variable. It is based on performing the same type of task at increasing intervals. Given the importance of success in retrieval on the effectiveness of strengthening long-term memory of a word, there is a need for an ideal algorithm that will consider both the difficulty arising from retrieval task type and the difficulty arising from spacing of retrieval. Such an algorithm does not seem to be

available yet in the literature. It would be possible, for example, to have students perform one task, receptive recognition, at increasingly spaced intervals, then move onto productive recognition, with shorter intervals at first, followed by increased intervals, then follow the same procedure with the two more difficult task types, recall. Alternatively, it would be possible to keep a short interval between retrieval tasks but increase the difficulty of the task each time, then gradually increase the spacing between occurrences of the most difficult of these types, productive recall, or between these four types of task in parallel. There are obviously quite a few permutations of task difficulty and increased spacing. In brief, learning theory is not precise enough to provide clear guidelines for the best method of combining these two important aspects of memorization, retrieval difficulty and retrieval spacing. The compromise reached in the design suggested below and its rationale are, therefore, tentative.

Finally, generative use of the word is important. This means we have to recognize or produce the word in a new context, in a new form (e.g. adjective rather than adverb) or with a new meaning. There are no guidelines from learning theory about whether there is some ideal order or timing for this kind of use, whether, for example, it is better to initially keep the meaning and form the same and vary the context, or keep the context and meaning the same but vary the form etc. Nor is it clear whether, for example, context should be varied during early retrieval exercises or wait till the succession of basic retrieval processes has been completed. For example, much research has been done on flash cards. With this method of learning, the only variable is spacing and in almost no cases is generative use practiced. The only likely guideline we can derive from theory is that receptive processes are likely easier than productive and recognition processes easier than recall when we encounter a word in a new context, or with a new form or meaning.

3. HCI Approaches

The aspect of HCI knowledge that is most relevant here is task analysis. This term may cover two tasks – the task that is performed by the user (student) when doing the task without a computer and secondly when doing the task with the computer. Here we will look at how students typically handle vocabulary learning in a particular setting, a Japanese university and a high intermediate class studying reading of the news in English without a computer. However, one initial point should be made. The motive of task analysis of the first type is to clarify in detail what *expert* users do when they perform a particular task in a non-computer environment, and use this knowledge to make the design easy to use and with all necessary functions. What is apparent from a rather coarse level analysis of how students learn vocabulary and even how it is presented in textbooks

is that neither student nor the average textbook approach to vocabulary learning shows expertise. Thus they do not provide a model of how the task is effectively done. Nevertheless, this approach does reveal shortcomings which can help in the design of the final software.

Typically, in a reading class, students annotate the text with Japanese translations of words they do not know. Thus they notice, decontextualize and match form to meaning – the first process of Nation's three processes. They may review the text several times over a short period before the class and before the test, but this will typically not involve retrieval as the text is annotated. In the class period focusing on a particular news story, they may use the word in the same or a slightly different context or form as part of a comprehension or discussion exercise. Thus there may be some productive retrieval at this stage, but it is neither certain nor structured. This approach is likely not all that effective for learning the vocabulary, though the first stage, matching form to meaning is obviously the essential first step.

More keen students make vocabulary lists. Some of these students will attempt to learn vocabulary by covering the Japanese or English entries and attempting either receptive or productive recall. Again, this is probably not the most effective process as they are trying a more difficult task, recall, which will likely lead to errors and, as shown in arguments above, this is not a good way to strengthen memory. In addition, it is highly unlikely that students in these cases think about spacing and it is also unlikely that they skip over words or meanings they have retrieved successfully. Rather they may tend to go down the list repetitively retrieving items just successfully retrieved while trying to be more successful with those they failed a few moments before.

We can get some more useful data from looking at somewhat more expert teaching of vocabulary. For example, the teacher may, if following the sequence of tasks suggested for strengthening retrieval above, do as follows. Preselect vocabulary items that are considered unlikely to be known by students, and put them in a list with their translation, while possibly instructing students not to annotate the text. Students may then be asked to read the text referring to the list and trying to answer comprehension questions as homework. The beginning of the class may then consist of vocabulary learning exercises. For example, following the steps for strengthening retrieval outlined above, students might do a receptive recall matching exercise, English to Japanese. This might be followed by a cloze type exercise where the blanks indicate the meaning and students have to choose from a list of words to fill that blank, productive recall. This could then be followed by an exercise in which the English word is provided but students have to provide the Japanese. Finally there would be the reverse case, providing the Japanese or a context that supplies the meaning and having students recall the English word. Such

exercises would normally be done with pencil and paper. Typically, either the teacher will check students' results in class time, or alternatively, take in the papers and mark them. In the case that most students get the answer right, it is likely that such a vocabulary item will not be systematically recycled at increasing intervals. If most students get it wrong, it may be that the teacher will revise the item in the following class. It is unlikely that it will be systematically processed after that. Students are typically left with a vocabulary list, and a few exercises manipulating those items. While in pedagogic terms it is a relatively impoverished cycle of activity, on the plus side, it does allow students to revise, albeit typically in a manner that is not considered the most effective, in a way that is convenient: papers can be easily carried and referred to anywhere anytime.

What becomes apparent even when considering how vocabulary is handled in a slightly sounder way in the classroom is that there are many aspects of classroom teaching of vocabulary that are entirely insufficient if we refer to the theoretically driven approach outlined in the previous section.

However, there is one more method of learning vocabulary that the most keen students employ, flash cards. Typically this involves writing the Japanese for the word on one side and the English on the other, or possibly an English definition on one side and the word on the other of a card designed for the purpose. Students typically use them for both receptive and productive recall. However, recognition exercises require the input of distractors and such cards are typically not created by students. Furthermore, although such cards could allow spaced learning if students keep a record of their successes and failures and then re-arrange the card set to include only those items that they failed on the first attempt, it is more likely that students go through the set skipping those they think they know and checking on those they feel hesitant about. Again, most flash cards made by hand are unlikely to include recognition exercises. And they do not include much if any generative use of a new word.

Analysis of the steps taken by students or the teaching processes of 'good' teachers indicate only partially what the software needs to provide. Simply put, vocabulary learning may not be being implemented as well as it could be, and so analysis of these processes does not yield a good model to implement. However, the analysis is not wasted as it brings more clearly to light the problems in current non-computer settings. In the next section, an attempt is made to derive a process that would be in line with theory.

4. A Sound Model of Vocabulary Learning

4. 1. Functionality

The model described here is realized in the context of a current English reading class, and assumes that a student-driven process is more likely to lead to best results than a teacher-led one. This will be explained in more detail later, but it means that the selection of vocabulary to learn is decided by the student while the materials from which the vocabulary is picked may or may not be. The part of the model focused on here is mainly the functionality. Reference will be made to a possible interface design later.

If we take the start of the process as that described as 'noticing', within a written text on screen what would be the ideal way for students to handle an unfamiliar word? The word or phrase needs to be identified to the computer, probably by highlighting it or clicking on it. As mentioned above, annotating text may not be conducive to encouraging effective retrieval steps, so it would seem the next step is to automatically save such words to a database. It then needs to be matched to its meaning. Given the prevalence of electronic dictionaries, it is feasible already to have possible translations shown to the student by, for example clicking on or highlighting the word in the database (which for the moment can be envisaged as an on-screen list next to the text). The student should select from the definitions the one that is appropriate for the context and this would then be entered into the appropriate field of the record for that word in the database. Given the fact that the database will be used as the tool to time delivery of retrieval exercises, the time and date of this first encounter with a new word would also need to be kept in the record. This process would continue as the student reads through the text and would result in a database with a record of the word, its L1 translation, and the date and time when it was entered into the database. From the students point of view it would be a vocabulary list with each new English word matched with its appropriate translation.

Once the vocabulary has been noticed, decontextualized and matched to its meaning, we need to strengthen the memory of it by the various types of retrieval. Remembering the order suggested in Section 2, it would be reasonable to make the first retrieval exercise one where students have to engage in receptive recognition, the least burdensome of retrieval types. If we take the example of a 10-word list, one easy way to do this is to have a matching exercise. In such an exercise a list of words on the left would be the English words selected from the text and a list on the right the Japanese translations but not matched with their English counterparts. The students' task would be to match the words in the right list with those in the left. Once done, a check should reveal which were matched correctly, which not. The creation of the exercise can be automated and does not necessarily require student or teacher action. The result, in terms

of a successful match or not, should be added to the record for that word in the database along with the date and time. For the purposes of understanding the system, we will imagine that 6 words were correctly matched and four not. Those words that were correctly matched should now be compiled into an exercise that is one step harder, productive recognition. That could be, for example, a multiple choice exercise in which the Japanese translation is provided and students have to choose between four alternative English words, those choices either being four words selected from the group just recalled or from the new word and three other unrelated words. Again, the exercise can be generated automatically. And again, results should be stored in the record for that word along with the date and time. As for the four words that were not successfully recalled, they should be presented to the student briefly in list form and then again in a receptive recognition task. The principle applied in each step up in difficulty is that words that are not successfully retrieved at that level of difficulty should be re-presented at a task one step easier. For the sake of argument, let's say that five out of the six words are correctly matched in productive recognition, these five should now be retrieved in receptive recall. This could be done by having the word displayed in English and students filling in a blank with the correct Japanese. As with the previous exercises, the result should be added to the record for that word with the time and date. If four of these five are correctly recalled, then they should be retrieved in a productive recall exercise. That could be a simple reverse of the previous exercise, with students provided with the Japanese and having to type in the English, or more interestingly in the form of a sentence with a blank which requires the meaning of one of the newly learnt words.

The core functionality depends on the database. The creation of exercises only requires a program that can pick up relevant words from the database and insert them into the appropriate variable in exercise creation software, the same function that is carried out by an author using exercise-creation software. For example, all the necessary exercise types are available in a software suite called Hot Potatoes that runs inside course management software called Moodle. The type of exercise which is created would depend on the record of success or failure for that particular word, which only means that the program has to access the fields in that word's record which track correct or incorrect answers and to have an instruction about the order in which exercises are to be presented. Though this is a simplification, it is clearly a not particularly complex design. In effect, the word, once inserted into the database, is put onto a learning track, a series of exercises that follow a pedagogically sound path.

In addition to putting the word onto a track that presents it to the student in an increasingly difficult series of retrieval exercises but with provision for repeated review of more difficult items, the keeping of times and dates with both the original entry and each

subsequent retrieval effort allows spaced learning. In brief, if a student learns a word quickly and is successful in the four types of exercises explained above, then it is not difficult to create an algorithm that would, for example, in such a case, present the word again, perhaps a day later, at either the same level of difficulty or perhaps one lower. If this was successful, then the interval could again be increased etc. Given the computer's capacity for storage, this kind of algorithm could present words over periods of a year or more.

While the above design is based on initial strengthening of retrieval, there is also the potential to include certain types of generative use. For example, if we take the simplest progression, and have generative use occur subsequent to consistently successful retrievals, then we would need to have access to a data base of sentences which use the word in the original meaning, maybe, but in a different form, for example past tense versus present tense or in a different collocation etc. How would this work? Various corpora are already available. However, from the point of view of providing sample sentences that themselves do not contain more new vocabulary, it is probable that either the a corpus used in creating a learner dictionary or a specially modified corpus would be needed. Once a bank of suitable sentences was available, the algorithm keeping track of the learner's success in retrieving the word could access the database of sentences, extract one, blank out the word to be tested and either let the learner produce it by means of a variety of exercises, for example productive recognition or productive recall. Such a database of sentences, if tagged appropriately, could allow generative use in many ways – new collocations, new forms and new meanings.

In this section I have tried to indicate in outline how software designed to enhance vocabulary learning would be designed, looking closely at the kinds of function that would be needed and, in broad terms, how that might be implemented. In the next section I will briefly consider what kind of interface would be appropriate.

4. 2. Interface

The interface is the means by which the human and computer communicate. The details of the design depend on the intention with which the vocabulary learning software is created, in particular where the locus of control lies.

The first requirement, however, is not concerned with control but is just a means for the program to understand in what mode the student is operating. For example, if we take the case of a student reading a text and creating a list of new words, then it is not appropriate for the software to generate a matching exercise until the student has finished reading and adding new words. There should, therefore, be some means, perhaps a button with 'Start' and another with 'Finish' that indicate to the computer that the

process of adding to the word database is over.

However, at this stage, if the software is seen as having 'teacherly' functions, being in a more controlling role, it may be that it will decide what the student *should* be doing. This could result in a more or less forceful insistence on the student going through a series of retrieval exercises within a certain time. In other words, the software will insist the student at least go through early retrieval exercises keeping a short time interval between them in order to maximize the chances of building an initially strong memory. In that case a 'Start Learning' button alone might appear with, perhaps, an estimate of the time needed to complete the minimum number of initial retrieval steps as designated by the program.

Alternatively, the software could provide the option for the student to proceed onto retrieval exercises in his or her own time but keep a track of this timing and adjust the exercises accordingly in order to maintain the high success rate of retrieval that theory advocates. Thus if a student creates a vocabulary list but then decides to take a break before coming back to the text later, the software could first re-present the vocabulary list before moving on to presenting a matching exercise. Alternatively, if a student goes through the first two levels of retrieval and then takes a break, depending on the length of time between this and the next attempt at learning the items, the software could take the learner back to the first step. In terms of interface, if the software is seen as controlling the process, it may be that a warning would be given to the student that he or she needs to have 'x' minutes available in order to proceed, while if the student is seen as being in control, buttons to allow starting and stopping of different exercises as the student wishes would feature more prominently.

This issue becomes even more complex as we consider the spacing requirement. If, for example, once a word has been successfully retrieved, it is decided that the next retrieval effort should ideally occur twenty-four hours later, how can this be handled? The student may or may not be logged on. Several possible solutions are available. One is that the spaced learning is made a little flexible and the difficulty of the retrieval task is manipulated to account for variation in length of spacing that results from students not being online at the time they are due for the next round of retrieval. Another way to tackle this problem is to make the software mobile compatible thus increasing the chance that students will perform the retrieval task close to the optimal spacing, though as Stockwell (2010) points out, mobile platforms are not necessarily the preferred option for students.

Given the long-term nature of spaced learning, with retrievals recommended over periods of up to a year, it is likely that a certain level of student control will be necessary as study should continue irrespective of semesters, and even during the semester, it will

be done outside of the classroom. Thus an interface that allows student control over timing but limits control over sequencing would seem to be the best compromise. And while students may control when they do retrieval exercises, it would seem best that the program contains an advisory function to let them know the ideal schedule on the one hand, and on the other adapts the type and sequencing of exercises in response to the timing of students' use. Thus the interface would not allow control over the sequence.

Designing the interface requires more detailed decisions on the locus of control, since the controls that appear on screen depend on these. This is an area where research and experimentation are required to discover what the software, in the role of teacher, should specify and what the leaner should be allowed to manipulate.

5. Considerations from Practice

The early stages of vocabulary learning, with repeated retrieval of increasing difficulty were implemented by the writer with a small class (23 students) in a blended learning context with a reading class. Students did vocabulary exercises online outside class and their knowledge was tested in class. In effect, students were given the vocabulary list with the English words only, told to check the Japanese, and then did a series of exercises following the four steps suggested for initial learning online before class. Typically they did them in quick succession a day or two before the class. At the beginning of the class a test of productive recall was administered. The average score was 70%. Items that most students succeeded in retrieving were then put into two types of exercise, productive recognition and productive recall with access online only allowed 3 days after the class in which the test was administered, to create spacing, but which had to be completed 2 days before the next class, classes being weekly, to create a second spacing. Items that most students had failed were put into online exercises that started at an easier level, starting with receptive recognition, and working back up to productive recall, without any access time limit. These items would then be tested along with the items successfully retrieved in the first class and items from the next text for that week at the beginning of the second class. These processes, with some additional retrievals over longer intervals, were repeated over an 8-week period focusing on ten vocabulary items a week that were new to all or almost all students. This far cruder model of the ideal process described above produced very effective learning. On an unannounced test in the 9th week of a sample of 30 of the 80 items learnt, the average score was 93%. An unannounced test two weeks later with no intervening online exercises on the same sample returned an average score of 87%.

From this small trial it seems that planned, spaced retrieval does lead to effective

initial learning of vocabulary. It is reasonable to assume that with more refinement it would lead to even better learning over a longer time frame. What was equally clear is that the amount of work involved makes it impossible to do this manually. As described above, however, this process is an ideal candidate for computerization.

6. Conclusion and Afterword

I have tried to make a case for the use of CALL software in the case of vocabulary learning and presented a rough draft of a suitable design, referring mainly to vocabulary learning theory, and to a lesser extent HCI ideas and practice. It should be possible to create software that can allow an ideal learning path, that is strongly supported by theory, in a way that is not possible in the conventional classroom or through conventional means. Such a program would also be able to provide support in an area in which textbooks and classroom practice are typically rather weak.

Given the power of databases, the improvements in language learning and course management software, the maturing of CALL as a discipline, and the increasing uptake of Internet learning by publishers and other commercial interests, it seems strange that such an eminently suitable candidate for the use of CALL has not been developed to the standards suggested here. Presentation of this concept was made at a major CALL conference, EuroCALL 2013, to an audience that included two academic computer scientists researching ways to improve Gaelic teaching online, and the manager of a CALL support center at a major university. On the one hand they expressed a strong interest and on the other, in the case of the computer scientists, were very positive about the practicality of such a system.

It is time for forward-thinking universities keen to promote language learning to take the leadership in developing CALL applications such as this that have a sound design and great potential to enhance the learning process.

Bibliography

Allum, P.H. (2004). Evaluation of CALL: Initial vocabulary learning. *ReCALL 16* (2) 488-501.

Allum, P.H. (2013). Principles and Practices Applicable to the Design of Successful Blended Language Learning. *Language, Culture and Communication*, *5*, 1-14.

Baddeley, A.D. (1997). *Human memory: Theory and practice*. East Sussex: Psychology Press.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52* (2), 323–363.

Bjork, R.A. (1999). Assessing our own competence: Heuristics and illusions. In D. Gopher &A. Koriat (Eds.), *Attention and performance XVII: Cognitive regulation of performance:*

Interaction of theory and application. Cambridge, MA: MIT Press.

Chapelle, C.A. (1998). Multimedia CALL: lessons to be learned from research on instructed SLA. *Language Learning & Technology*, *2* (1), 22-34.

Chapelle, C.A. (2001). *Computer Applications in Second Language Acquisition*. Cambridge: CUP

Dempster, F. N. (1987). Effects of variable encoding and spaced presentations on vocabulary learning. *Journal of Educational Psychology, 79*, 162-170.

Esit, O. (2011). Your verbal zone: an intelligent computer-assisted language learning program in support of Turkish learners' vocabulary learning. *Computer Assisted Language Learning*, *24* (3), 211–232.

Fehra, C.N., Davison M.L., Graves M.F., Sales, G.C., Seipel, B., Sekhran-Sharma, S. (2012). The effects of individualized, online vocabulary instruction on picture vocabulary scores: an efficacy study. *Computer Assisted Language Learning, 25* (1), 87–102.

Jamieson, J., Chapelle, C.A., Preiss, S. (2005). CALL Evaluation by Developers, a Teacher, and Students. *CALICO Journal, 23* (1).

Karpicke,J.D. & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 2011, *37* (5), 1250-7.

Laufer, B., Elder, C., Hill, K., & Congton, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21* (2), 202–226.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54* (3), 399–436.

Laufer, B. & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics 22*, 1–26.

Mizuno, R. (2000). A test of the effectiveness of the Low-First Spaced Learning Method applied to CAI. *Educational Technology Research, 23* (1/2): 35–44.

Mizuno, R. (2003b). The application of cognitive scientific findings related to spacing effects to web-based CAI systems. In: Hesse, F. W. and Tamura, Y. (eds.) *Proceedings of the Joint Workshop of Cognition and Learning through Media-Communication for Advanced e-Learning*. Berlin: Japanese-German Center Berlin, 128–133.

Modigliani, V. (1976). Effects on a later recall by delaying initial recall. *Journal of Experimental Psychology: Human Learning and Memory*, *2* (5), 609–622.

Mohsen, M.A. & Balakumar, M. (2011). A review of multimedia glosses and their effects on L2 vocabulary acquisition in CALL literature. ReCALL *23* (2), 135–159.

Nakata, T. (2006a). English vocabulary learning based on the reactivation theory of spacing effects. *Language and Information Sciences*, *4*: 229–242.

Nakata, T. (2006b). Implementing optimal spaced learning for English vocabulary learning: towards improvement of the Low-First Method derived from the reactivation theory. *The JALT CALL Journal, 2* (2): 19–36.

Nakata, T. (2008). English vocabulary learning with word lists, word cards and computers: implications from cognitive psychology research for optimal spaced learning. *ReCALL*

20(1): 3-20.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software. *Computer Assisted Language Learning, 24* (1), 17–38.

Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: CUP

Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, *51* (2), 73–75.

Redmond-Pyle, D. (1995). *Graphical User Interface Design and Evaluation*. Prentice Hall.

Pyc, M.A., & Rawson, K.A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, *35* (8), 1917–1927.

Shneiderman, B. (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. (3rd ed.). Menlo Park, CA: Addison Wesley

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of the platform. *Language Learning & Technology, 14* (2), 95-110.