

An investigation into the effectiveness of the keyword method for a group of Japanese EFL learners

著者	Burden Tyler
雑誌名	国際地域学研究
号	14
ページ	75-97
発行年	2011-03
URL	http://id.nii.ac.jp/1060/00003674/



An Investigation into the Effectiveness of the Keyword Method for a group of Japanese EFL Learners

Tyler Burden

Abstract

This study investigated the effectiveness of two variations of the keyword method compared to an 'own' strategy control condition for learning English vocabulary items with a group of Japanese EFL learners. The experiment employed a design mirroring self-study conditions.

The principal finding of this study was that, both one week and three weeks after treatment, the results on subsequent tests of vocabulary retention and recall showed a tendency favouring the participants in the control condition, whom it was subsequently discovered had all used rote learning techniques to memorise the treatment items. None of these tendencies, however, was found to be statistically significant.

Introduction

The learning of vocabulary is an essential part of achieving proficiency in a language and is justifiably seen by many learners as the single most important aspect of their studies. Learners frequently report that it accounts for their difficulties in both receptive and productive aspects of language use (Oxford and Scarcella, 1994: 232). However, the size of the lexicon is vast and for most learners the question of how best to commit items of vocabulary to memory is of crucial importance.

One technique for memorising L2 vocabulary items that has received a great deal of attention is the mnemonic keyword method which was first described by Atkinson (1975). The technique works by creating a link between the target L2 word and an L1 'equivalent' in the form of a mediating 'keyword'.

In order to remember a given L2 word, a keyword is chosen, usually from the learner's first language, which is acoustically similar to the target L2 word yet has a meaning of its own independent of the new word's meaning. A visual association through an image is then created between the keyword and the target word's meaning

and this provides a cue for subsequent retrieval. For example, to learn the Japanese word 'kaigi' (会議 = 'meeting'), English-speaking learners of Japanese might use the keyword 'kite' and imagine themselves flying a kite in a meeting. It is not necessary that the keyword sound identical to the foreign word, in the example above only the first syllable of the keyword is similar, but should be as similar as possible, and lend itself easily to a memorable image link connecting the keyword and the English translation (Atkinson and Raugh, 1975: 127).

Various theoretical rationales have been put forward which lend support to the effectiveness of mnemonic-based methods such as the keyword method. One such rationale is Paivio's "Dual Coding Theory" (1991) in which he stated that there is an additive effect on recall due to the keyword method employing two routes (verbal and imaginal) to retrieval of the L2 target word.

In addition, the "Depth of Processing Hypothesis" has been advanced by Craik and Lockhart (1972) and Craik and Tulving (1975) which supports the position that the persistence of a memory trace of a stimulus is a function of the level to which the stimulus is analysed and the degree to which it is enriched by associations or images (Craik and Tulving, 1975: 270).

Empirically, there is a wealth of research evidence that supports the efficacy of the method. However, as Cohen points out, the most successful studies have been performed under laboratory-like conditions and did not involve real language learners. The results of classroom studies, which are perhaps of greatest interest to practitioners in the field, have by contrast been mixed (Cohen, 1987: 47-50).

In Atkinson's seminal paper (1975: 824) the question was raised as to whether it is better for students to generate their own keywords or have the keyword supplied for them. Findings to date suggest that it may crucially depend on the age (and therefore cognitive maturity) of the learner with younger learners requiring a greater degree of support in terms of a mediator (and/or interactive image supplied for them) (Cohen, 1987: 49-50). This may also be related to cultural background as, in some cultures, learners may not be used to taking the initiative for their own learning and may struggle if not provided with teacher-generated solutions.

A further important issue centres on the long-term retention effect of treatment by keyword. In many studies retention has been measured immediately after treatment. Whereas such experiments have important theoretical bearing, it is the long-term effect that will be of most interest to practitioners in the field.

This study sets out to investigate these issues in the context of Japanese EFL

learners, a context which has so far received little attention. The group under investigation consists of university students assigned to three learning conditions: (a) keyword method using a self-generated keyword, (b) keyword method using a researcher-supplied keyword, and (c) students' own preferred strategies. They were required to learn a list of English words and were subsequently tested on two occasions.

Research Questions

The study aims to answer the following research questions:

- Is the keyword method a more effective facilitator of long-term retention than students' own strategies, and if so, which variation of the method is the most effective?
- Do any of the learning conditions have a significantly higher rate of forgetting than the others between the two test dates?

Background

The learners

An important difference between this study and other keyword studies lies in the cultural background of its participants. All the participants in the present study are Japanese native speakers. In a study by Schmitt, which examined Japanese learners of English of various age ranges, it was discovered that by far the most used strategies, and those perceived as being the most helpful, for remembering a word once its meaning had been established were written repetition, followed closely by verbal repetition (1997: 222). In short, rote learning. By contrast, he found that the keyword method ranked as one of the least used strategies and, in the perceptions of the respondents, ranked the second least helpful strategy (Schmitt, 1997: 208). These findings were supported in a study by Burden (2002) which looked at, specifically, the strategies of university students.

There is evidence to suggest that learners in Japan are already experienced users of mnemonic techniques in general. Throughout the curriculum a range of published study aids exists which are devoted entirely to mnemonics and provide students with alternative methods of committing academic information to memory. For example, books such as Shiraishi (1999) and Miura and Kaneko (1997) are specifically devoted

to English vocabulary and typically employ a mnemonic technique which involves linking the sound similarities of the parts of the English word to Japanese words and providing a cue to retrieval of the meaning through a story.

In addition, a sub-skill involved in the learning of 'kanji' characters is imagery. Learners are required to mentally link a pictorial/visual image to the shape of the strokes of a given kanji character. This provides a cue to subsequent retrieval of the character's meaning, a process similar to the imagery stage required for the keyword method. It is surprising, then, that the Japanese learners' surveyed in the Schmitt study had a negative attitude towards the method.

A possible explanation for this may lie in the creativity requirement that is placed upon the learner in generating either a keyword or an imagery link or both. There is evidence to suggest that Japanese learners, although adept at tasks that involve copying and/or repetition, find tasks that require imagination and creativity rather difficult (Reischauer, 1990: 200). Also, the education system favours an approach whereby learners expect information and answers to be provided for them (LoCastro, 1996: 50). Thus, it was decided that in the present study two variations of the keyword method be investigated, one involving the researcher supplying the keywords and the other involving the learner generating his or her own keywords.

Research into Keyword

The keyword method was first described by Atkinson (1975) and subsequently investigated in relation to (American) university students learning Russian vocabulary items (Atkinson and Raugh, 1975) and Spanish vocabulary items (Raugh and Atkinson, 1975). In both studies a significant effect for treatment by keyword was discovered. This prompted a variety of further studies, the results of which supported the initial findings. Not only did the method appear to be effective for learning vocabulary from a variety of languages, but it also appeared to be more effective than other learning strategies (Pressley et al., 1982: 67-69).

However, a problem with many earlier keyword studies was that they were performed in laboratory-like conditions with subjects who were not real language learners (Cohen, 1987: 47). Hence, the results of such studies may be of little practical use to language educators. A study by Merry (1980), however, demonstrated that such results could be obtained in classroom conditions. In his experiments a significant effect for treatment by keyword was found in both short and long-term retention for a group of 11 year-olds learning French in secondary education in Britain.

This classroom result was supported by the findings of Cohen and Aphek (1980) who demonstrated that the keyword method could facilitate both immediate and long-term retention of Hebrew vocabulary items for a group of adult English-speaking learners. However, as Cohen observes, the results of classroom studies have been rather mixed. He cites student attention, motivation, and prior patterns of vocabulary learning behaviour as confounding problems in this area of research (1987: 48).

Another area where results have been mixed concerns the effectiveness of the keyword method in long-term retention. The studies reviewed above suggest it is facilitative. However, a study by McDaniel *et al.* (1987) found the keyword condition to be no better than a contextual condition in a test of recall one week after treatment. Similarly, a classroom study performed by Brown and Perry (1991) with Arabic-speaking ESL adult learners found no statistically significant result in long-term memory.

Ellis and Beaton (1993) compared the effect of three learning strategies (keyword, repetition and 'own') on native English-speaking university students' ability to learn German foreign words. They found keyword to be efficient for receptive vocabulary learning but repetition to be more effective for productive learning in both immediate and long-term retention.

A classroom study by Elhelou (1994) examined the effect of keyword instruction on a group of Arab 7-8 year old children studying English at elementary school. The keyword group out-performed the control group on immediate recall. The finding is significant as it demonstrates that the keyword method may facilitate learning of English vocabulary for learners educated in a different orthographic script (such as the Japanese students in the present study).

Avila and Sadoski (1996) and Rodriguez and Sadoski (2000) have also investigated the effects of the keyword method in a classroom context. In the Avila and Sadoski (1996) study, a significant effect was discovered for the keyword method group one week after treatment, whereas in the Rodriguez and Sadoski (2000) study, the results of the keyword treatment condition were found to decline rapidly between immediate and delayed recall to a position where the keyword participants scored worst of all the four treatment conditions in a delayed test of recall. This leaves open not only the question of the type of learner who benefits from use of the method, but also, the question of how significant is the rate of forgetting with treatment by keyword. In the McDaniel *et al.* (1987) study mentioned above the keyword condition participants' initially superior results similarly declined over time.

Design

A two-retention interval, three-treatment condition design was used for this study. Intact second-year university classes were randomly assigned to three learning conditions: keyword method using a self-generated keyword (condition A), keyword method using a researcher-supplied keyword (condition B), and students' own preferred strategies (condition C). Testing was administered one week and three weeks after treatment.

A feature of this design, similar to that employed in various other classroom studies (e.g. Elhelou, 1994; Rodriguez and Sadoski, 2000) was that the subjects were not randomly assigned to treatment conditions; i.e. treatment was administered to intact classes. This was done, principally, because it was not a practical possibility within the host institution to facilitate the random assignment of students (with the consequent breaking up of classes). Such a design has its weaknesses in so far as the results may be less generalisable from the viewpoint of randomising individual differences. However, with fully randomised experimental designs there is often a lack of 'ecological validity' due to the inauthentic situations in which such studies take place (Brown and Perry, 1991: 660). The present design, then, represents an authentic learning environment.

Participants

The participants were selected from an initial pool of around 70 Japanese second-year non-English major students from a university in Tokyo. The students were around 20 years of age. To control for variations in the participants' English ability, only students who had attained level 3 on the Japanese-administered STEP (Society for the Testing of English Proficiency) test were selected from the initial larger pool. This puts them at around the lower intermediate level. In total 27 participants were selected (9 participants in each learning condition).

Materials

Vocabulary

An initial list of twenty-four words was drawn up which, in the main, comprised colloquial terms that students were unlikely to have met in previous English courses. No words with a cognate relationship to a Japanese word and no words that had a

transparent meaning (i.e. whose meaning may be inferred from its constituent word parts) were selected. The principal criterion for selection, however, was that the participants should not know the words prior to treatment. To satisfy this condition a multiple-choice pre-test was administered. From the pre-test results, the final list of fifteen words was arrived at.

Keyword Selection

The keywords presented in learning condition B were chosen in collaboration with a native Japanese speaker. In accordance with Raugh and Atkinson's (1975: 2) advice, the keywords were selected to satisfy three principal criteria: (i) Share as much as possible an acoustic similarity with the target word, (ii) lend themselves easily to an imagery link with the target word, and (iii) be unique from the other keywords presented to the participants. Nouns were preferred (where possible) which, as Ellis and Beaton (1993: 604) observe, are generally more successful as mediating links because of their greater imageability.

Instruction Materials

In the majority of the keyword studies reviewed in the sections above, the participants have been provided with the target word and an L1 translation. However, in practical terms, learners rarely encounter new vocabulary items out of context. In most cases there will be some contextual information available, such as through the surrounding sentences of a written text. Thus, in an effort to make this investigation a more realistic reflection of what may occur in an actual study situation for this group of students, the materials presented to the participants (appendix A) incorporated a Japanese translation, grammatical word class information and a short example sentence, such as may be found in a typical learners dictionary. In addition, a brief English definition was included.

Testing Instrument

In order to assess learning across the three groups two forms of testing instrument were devised, one a cued-recall test, the other a multiple-choice test (appendix A). The rationale behind having two testing formats to assess learning of the same set of vocabulary items is that there may be an incremental difference between retention and retrieval of those items (Brown and Perry, 1991: 662). A single testing format would not be sensitive to such differences.

The cued-recall test was designed to assess retrieval. A list of the English words was presented with information about their grammatical word class and a space was provided for participants to translate the target word into an L1 (Japanese) equivalent.

Questionnaire

An important question to emerge from the mixed results of earlier group-administered keyword studies was the need to find out what students actually did in practice (Pressley *et al.*, 1987: 114-116). In an effort to address this issue, two forms of questionnaire were devised, one dealing specifically with the strategies used for remembering the words which was administered directly after the first testing session, the other a more general questionnaire dealing with, in addition to the strategies used, the participants' reactions to the study in general. The first questionnaire simply listed the target words and provided a space for the students to write down an explanation of how they remembered them. For the keyword groups, this included a space for them to write the keyword they had used.

The second questionnaire, given to the keyword treatment groups, sought to discover participants' reactions to using the keyword method. It was also used to identify whether the participants had had any contact with the target words outside the context of the study (between treatment and either of the testing sessions).

For the control group, a similar questionnaire was devised, but with the keyword references omitted. To try and ascertain what these students did to memorise the words a table of options was presented for which the students were invited to simply tick 'yes' or 'no'.

Procedure

Piloting

With the exception of the pre-test, all of the materials described above were piloted with individual Japanese colleagues. This led to a number of revisions both in format and in error eradication. For the multiple-choice test, more distracter options were included and the rubric was simplified.

Training

A week after the pre-test the classes met once again. The keyword groups (conditions A and B) received a period of training in the technique lasting

approximately 60 minutes. The technique was introduced along with some explanation as to its origins and some of the claims that have been made in its favour. An explanation was provided in which the classroom teacher gave two personal examples of how English keywords had been used to learn Japanese (foreign language) words and the imagery links were explained through illustrations. Then, an example English word was written on the board with its definition and word class information in a deliberately identical format to the layout adopted on the instruction materials. Students were subsequently re-assigned to their groups and asked to think of a mediating keyword. The words were elicited and, through discussion with the students, the best two or three were retained and approximations of the corresponding interactive images were drawn on the board. A brief survey was performed asking students to put their hands up if they felt the method would be useful for them.

In addition, in eliminating some of the students' suggested keywords, a discussion was performed of why certain mediating words were more useful than others. This method of presentation, involving a pooling of ideas and a subsequent attempt to involve the students in an evaluation of the technique's usefulness, was based on the recommendations of Hulstijn (1997: 218). The procedure was repeated for three further words and although the period of training was brief (totaling around 60 minutes) it was felt to be consistent with the amount of time allocated to training in various other classroom studies (e.g. Elhelou, 1994; Rodriguez and Sadoski, 2000).

Implementation

In all three learning conditions, the students were given their instruction materials containing the target words and, in line with many other keyword studies (e.g. Merry, 1980; Cohen and Apeh, 1980), the words were read aloud to ensure familiarity with the words' phonological features. The students in all conditions were given around 30 minutes to learn the 23 items presented on their respective instruction materials. This time-allowance was set because, in a number of previous studies, lack of 'processing time' has been identified as a possible problem (Pressley *et al.*, 1987: 117).

Students in condition A were told to write down the keyword they had devised to help them remember the word. In condition B, students were told to use the keyword provided to help them remember the words. In these two conditions the students were told that the reason they were being asked to learn the words, was to see if the technique could be helpful to them, and for the teacher to find out what techniques the students liked. In condition C, students were told to try and learn the words as best

they could. None of the students (in any of the groups) were told that they were going to be tested.

The classes met again a week later and the subjects were set the cued-recall test followed by the multiple-choice test. They were also subsequently given the first of the questionnaires to complete. The students were surprised to receive a test, and were consequently assured that this had no bearing on their course grades. It was observed after administration of the tests that a number of the students were keen to find out the answers and were talking with their peers. This was discouraged and at this point it was revealed that there would be a further test for which they were requested not to study.

The two tests were administered again two weeks later. On this occasion the second questionnaire, about the participants reactions to using the keyword method, was administered.

The method of implementation adopted here differs qualitatively from that adopted in many recent classroom studies. In the Elhelou (1994) study, and a number of the studies involving Pressley and his collaborators, items have been presented individually (one by one) through the researcher, and subjects have been given explicit instructions about what to visualise, or even in some cases (e.g. Merry, 1980) have been given a picture to memorise (exceptions to this approach can be found in the studies of Brown and Perry (1991) and Cohen and Aphek (1980)).

The merits of such an approach are that the time devoted to each item is strictly controlled, and the researcher can exert greater influence over the participants' attention. However, these approaches with their high degree of teacher input (and control), ultimately measure whether the technique is effective as a classroom resource for the teacher. It was the intention of this research to ascertain whether the method is effective as a self-study resource for the students. Thus, whole-class item presentation was rejected in favour of a list of the items.

Analysis of Results

Experimental results

An initial observation to be made from the test results is that the scores were low. The group means ranged from a low of 1.56 (group A on the first administration of the cued-recall test), to a high of 4.89 (group C on the second administration of the multiple-choice test) out of a possible 15. In other words, the mean group scores

were between approximately 10 and 33 per cent. Thus, even in the most successful condition, the participants did not learn a high proportion of the items.

The group means and standard deviations for both the test-types and retention intervals are shown in Table 1 (mean scores expressed as a percentage in brackets).

Table 1,
Group means and standard deviations for the four tests

Group		1 week after treatment		3 weeks after treatment	
		Cued-recall	Multiple-choice	Cued-recall	Multiple-choice
A	Mean	1.56 (10.37)	3.56 (23.70)	1.78 (11.85)	3.00 (20.00)
	SD	1.51	1.24	1.48	1.80
	N	9	9	9	9
B	Mean	1.78 (11.85)	3.78 (25.19)	2.11 (14.07)	3.44 (22.96)
	SD	1.39	1.30	1.05	1.33
	N	9	9	9	9
C	Mean	3.22 (21.48)	4.56 (30.37)	3.00 (20.00)	4.89 (32.59)
	SD	2.68	3.00	2.65	3.02
	N	9	9	9	9
Total	Mean	2.19 (14.57)	3.96 (26.42)	2.30 (15.31)	3.78 (25.19)
	SD	2.02	1.99	1.86	2.24
	N	27	27	27	27

(Note: SD = Standard Deviation, N = number of values Group mean and standard deviations rounded to 2 decimal places)

From the table above three further observations can be made:

- (i) There is a clear trend in favour of group C (the control group) who score better on all four tests than the other two (keyword) groups. Group B (researcher-supplied keyword) score slightly better than group A (self-generated keyword), again on all four tests. The difference between the test results of group C and those of either of the keyword groups (A or B) are substantial. The difference between the results of the two keyword groups A and B are minimal. However, the variability in the results, as indicated by the standard deviation, increases from group B to A to C.
- (ii) There appears to be no clear trend to either gain or lose scores between the two retention intervals for either version of the tests. Some scores increase slightly between retention intervals, some scores decrease.
- (iii) There is a clear tendency for the multiple-choice test scores to be higher than the cued-recall test scores.

To verify whether any of the differences between the group scores observed above are statistically significant, two null hypotheses were formulated:

- a) There are no significant differences between the group means for any of the four tests, either separately, or for the combined scores on the four tests.
- b) There are no significant differences between the mean scores for the four tests for each group, or for the four groups combined.

Four separate one-way ANOVA's were performed to test the hypothesis in (a) (Appendix B, Table a) and it was found that there were no significant group effects (at $p < 0.05$).

To test the hypothesis in (b), a two-way ANOVA was performed (Appendix B, Table b). (This operation differs from the one-way operation performed above in that it investigates the possibility of an overall group mean effect, rather than separate group mean effects) Again, no significant group effect was found ($F[2,24]=1.70$, $p=0.204$) although, a highly significant effect was found for test ($F[3,72]=19.30$, $p<0.001$).

Thus, in response to the first research question that this study sought to investigate, the results indicate that neither of the two keyword conditions are more effective facilitators of long-term retention than the students' own strategy use. In fact, there is an apparent tendency for the group using their own strategies to perform better in both the tests of long-term retention. However, neither this tendency nor the tendency for group B to outperform group A is statistically significant.

The second research question was concerned with the rate of forgetting between the two test dates. Re-examination of Table 1 (above) indicates that some group mean scores increase from one administration of the test to another, and some decrease. Looking at the total means, for the cued-recall test there appears to have been a slight improvement in performance (from 2.19 to 2.30), whereas for the multiple-choice total means, there is a slight decline in performance (from 3.96 to 3.78). There appears, then, to be no clear pattern in these results.

Questionnaire Results

The main questionnaire investigated the participants' reactions to the study and they were found to be mixed. Some students reported they found the technique to be helpful whilst others did not. Criticisms were directed towards the difficulty and time-

consuming nature of generating images together with its limited application. Some participants reported that it was easy to make mistakes using the technique. However, a slight majority of the participants in the keyword conditions reported that they would use the technique again in the future.

For the control group, the main questionnaire included an item asking the students to explain the strategies they had used. In all cases, the participants reported using either written or verbal repetition or a combination of the two. Four participants reported supplementing this with self-testing.

For groups A and B the first questionnaire revealed that the participants had been able to remember the meaning of some of the words without the need to use the keyword mediating link in the retrieval process. (In group A, the proportion was 9 from 48, in group B, 15 from 37).

Discussion

The principal result of this study was that an 'own' strategy use control group showed a tendency to out-perform two groups of students using the keyword method in delayed tests of retention and recall. How, then, can this result be accounted for?

Rote learning

A feature of the questionnaire feedback was that, without exception, all the members of the control group used a repetition strategy (either verbal or written) as their principal method of memorising the target items. One possible way of accounting for the present results may be that, for these learners, rote learning is a more effective strategy for learning a list of words.

The education system in Japan places heavy emphasis on rote learning, particularly in acquiring the characters of its (complex) writing system (Locastro, 1996: 49). Since learners are exposed to, and encouraged to use this type of strategy from an early age they may have developed a skill in its application. Such a possibility is highlighted in a study by Tinkham who compared Japanese and American high school students in their ability to remember FL vocabulary lists through rote learning. He found that Japanese students scored significantly higher in both recall and recognition (1989: 697).

The possibility that an apparently shallow repetition strategy may be used to greater effect than the more elaborate keyword method does not sit easily with the

Depth of Processing Hypothesis noted above. Although the keyword method shares aspects consistent with the depth of processing hypothesis, there remains a question mark as to how deep the semantic processing involved in linking the target word to its L1 equivalent is given the arbitrary nature of the selection process involved. In any case, the exact nature of the processing undertaken by the control group participants is not clear. It may be that these participants were supplementing repetition with more elaborate forms of semantic processing but were not able to articulate this in their feedback responses.

Keyword Training

An assumption, upon which this study is based, is that a period of training may result in learners being able to use a particular learning strategy effectively. It was noted in the Schmitt (1997) study above that the Japanese learners surveyed believed the keyword method to be of little value in helping them to learn vocabulary. Instead, they placed far greater value in the effectiveness of other strategies, in particular, rote learning. The participants' questionnaire results, noted above, were somewhat mixed. This suggests the possibility that, despite the measures taken to positively affect the participants' attitudes, many of the participants in the study may not have been as committed as the participants in the control group. Such an effect was noted in a study involving a group of Asian students by O'Malley (1987). He concluded that learners' culturally determined attitudes may be of critical importance in determining the effectiveness of strategy training.

In addition, there is the possibility that the students were not comfortable enough with the mechanics of the technique to use it effectively. In the questionnaire feedback, many of the participants reported problems in regard to such things as generating a keyword, or visualising images. Perhaps a longer period of time was necessary for training. This, however, was felt unnecessary given the learners' familiarity with visualisation/imagery techniques as well as mnemonic techniques in general. A number of successful classroom keyword studies exist where the period of training has been comparable. Two notable examples, described above, are the Merry (1980) and the Elhelou (1994) studies. Both these studies differ, however, in that more support was given by the researchers in providing visual images.

Number of Words/Applicability

It has been argued that there is a limit on the applicability of the keyword

method. In particular, Hulstijn (1997: 218) argues that there are plenty of times when words naturally find a strong trace in the mental lexicon and thus techniques such as the keyword method are only necessary as an aid for the troublesome words that do not, although the reported successes of the Linkword course (Gruneberg and Jacobs, 1991) suggest that the keyword method can be used for a large number of words.

In contrast to the Linkword course materials (Gruneberg and Jacobs, 1991) and many keyword studies, the present study employed self-pacing, in that the participants were not given a per-word time limit (rather, they were given an overall time limit). The rationale was that it would make the experiment more comparable to self-study conditions, where learners generally choose themselves when to switch their attention from one item to another. Thus the onus was placed on the participants to methodically work through the items. The possibility exists that some of the participants became demotivated and did not give their full attention to all of the items. This would help to account for the disparity in test performances and the high variability (as indicated by the standard deviations) in the test results. However, there were only a total of 23 items to memorise and if demotivation were a factor then it would not account for the highest variability being detected in the control group.

Rate of Forgetting

The final research question of this study was concerned with comparing the rate of forgetting of vocabulary items across the three groups, and as was noted above no clear pattern was detected in the data. In fact, the total means for the cued-recall test improved slightly over time (whereas those of the multiple-choice test deteriorated slightly). It seems counter-intuitive, however, that scores could improve in either of the test formats between sittings, as one would expect a certain amount of forgetting over time. Moreover, the improvement in performance was observed in the cued-recall test, a format where this apparent anomaly in performance can not be accounted for by fluctuations in the participants' guesswork.

A possible reason for this may lie in the fact that the participants took exactly the same two tests in both sittings, and from taking the first multiple-choice test (which provided clues to the meanings of the items) were given a degree of assistance for the second sitting of the cued-recall test. This would explain why there should be improvement in only one of the testing formats between sittings. Such problems, although impractical in this study, may have been avoided if different participants had been used for different test dates; i.e. adopting a between-participants rather than

within-participants factor.

A further possibility that may account for the improvement in scores is that, as noted above, after the first sittings of the tests some of the students were observed discussing their answers with their peers. There is evidence to suggest that when learners are eager to know the meanings of words (such as after a test), the words have a greater perceived saliency and are consequently more likely to be remembered (Brown, 1993: 265). Although steps were taken to counter this (and eliminate the papers of students who were known to have done this), some students may have continued discussing the test items outside the classroom and not declared this on their questionnaire feedback. So there remains the possibility that experimental conditions were broken for some of the participants, which could weaken the validity of the second set of test results.

High Variability

A striking aspect of the data was the high variability in the test scores. One possibility that may account for this lies in the measure that was adopted to control for the participants' L2 ability. Students' STEP test results were used, but this is a measure of overall proficiency and may mask significant individual differences in the learners' capacity to memorise words.

Also, it was discovered at the marking stage that the participants adopted rather different approaches to answering test questions. Whereas some participants attempted to answer every question, others only attempted a very limited number of responses. It is likely, for the multiple-choice paper, that this led to a wider range of scores than might have been expected.

Low Scores

A further striking aspect of the data was the low scores in all conditions. An explanation for this may lie in the difficulty of expecting second language learners to learn a list of (minimally contextualised) words in a relatively short period of time. Lexical items are learned as a result of repeated exposure to the items in meaningful contexts over an extended period of time (Nagy, 1997: 73). Thus, with only a single period of exposure afforded for the participants to learn the words and only one example sentence provided showing the words in context, it is perhaps unsurprising that the scores were rather low.

In addition, there are many factors that contribute towards complete knowledge

of a lexical item, such as its coreness, collocations, syntactic behaviour, register characteristics and so on. This study was restricted to examining a very shallow (principally semantic) level of item knowledge, and thus can only partially inform about what actually goes on in 'real' language learning where repeated exposure and diverse contextualisation leads to a richer level of item knowledge.

Implications and Conclusions

This study investigated the effectiveness of two variations of the keyword method compared to an 'own' strategy control condition for learning foreign language vocabulary items with a group of experienced Japanese learners of English. The participants comprised second year undergraduate university students and the experiment employed a design mirroring self-study conditions.

The principal finding of this study was that, both one week and three weeks after treatment, the results on subsequent tests of vocabulary retention and recall showed a tendency favouring the participants in the control condition, whom it was subsequently discovered had all used rote learning techniques to memorise the treatment items. It was also discovered that, of the two keyword groups, results were slightly superior in the group that had been supplied with a keyword by the researcher, compared to the group that had been asked to generate its own keywords. None of these tendencies, however, was found to be statistically significant.

The final part of the research question examined the rate of forgetting of the treatment items between the two test dates. No clear pattern could be found in the data.

Salient methodological differences prevent accurate comparison with previous research. However, an implication of these results is that they further support the findings of McDaniel *et al.* (1987), Brown and Perry (1991) and Rodriguez and Sadoski (2000) which showed that the keyword method (used in isolation) may not be an effective facilitator of long-term vocabulary retention.

Positive results have been obtained for the keyword method in long-term retention, however, notably in the Merry (1980) and Avila and Sadoski (1996) studies. Aside from the fact that neither of those studies involved Japanese learners, an important distinction between the present study and those studies lies in the nature of the control condition. In the Merry (1980) experiment, inexperienced 11 year-old learners were used, and in the Avila and Sadoski experiment (1996) the control group

participants were assigned to a learning condition rather than allowed to use their own preferred strategies. In the present study, the participants were experienced learners of English who were allowed to use their own strategies. The results of this study suggest that the comparative success of the keyword method in long-term retention may be linked to the proficiency of the learners, in that experienced learners may already have in place effective memorisation strategies suited to their own cognitive styles (which may match or out-perform the keyword method).

Criticisms have been made regarding the applicability of the keyword method. Notably that it is of little help in productive acquisition (Hulstijn, 1997: 207), that it is too time-consuming and complex to be useful for large numbers of words (McCarthy, 1990: 118), and that it can only be used effectively for certain types of words (Cohen, 1987: 52). For Japanese EFL learners, given the acoustic requirements of the technique, and the limited phonetic inventory of the Japanese language by comparison to English, it is possible that phonological considerations further limit its applicability.

It may also be that the technique is only effective as a resource for beginner level students. This position can be reconciled with the positive results obtained for the Linkword beginners' course in both short-term (Gruneberg and Jacobs, 1991; Gruneberg and Sykes, 1991) and long-term retention (Beaton *et al.*, 1995: 118).

Unlike many previous keyword studies (including the Linkword course materials), in this study, a deliberate decision was taken not to incorporate explicit imagery instructions. It was felt that this would more closely mirror a genuine self-study situation where, upon encountering a new lexical item, learners draw upon their own cognitive resources to memorise it. The conclusions of this study would be strengthened if it could be said with confidence that the participants in the keyword groups had been able to apply the technique effectively. The participants' reactions, however, indicate that this was not always the case and that there were problems in the students' ability to find (amongst other things) imagery links. These problems may have been reduced or avoided if a more substantial period of training had been afforded the participants.

The cultural aspect of this study cannot be ignored. It was discussed in the above that an underlying reason affecting the disappointing results of the keyword participants may not have been that there were problems with the keyword method itself, but instead, with the learners' culturally determined expectations regarding its implementation. It would appear that even though the learners in the present

study may have skills suited to the application of the keyword method, they may not have been suited to the creativity requirements of generating their own visual links to connect the keyword to its referent. The present results have implications for classroom practitioners in that they demonstrate the need to work with learners' culturally bound expectations regarding the learning experience.

One of Atkinson's seminal questions was concerned with whether or not it is better for the researcher to supply imagery instructions (1975: 825). For these learners, it would appear that their culturally bound expectations demand it. It would also appear, in answer to another of the issues raised by Atkinson (1975: 824), that for these learners, it is better for the researcher to supply the keyword. A future research possibility, in the context of Japanese learners of English, may be to compare an imagery and keyword supplied condition with an 'own' strategy condition for their effects in facilitating long-term retention.

Appendix A

(i) Sample of the instruction materials that were presented to group B. For Groups A and C, keywords were not supplied

1. Booze (n.) *Keyword:* buu buu

English: (inf.) alcoholic drink

Japanese: sake

Example: Have we got enough booze for the party tonight?

2. shattered (adj.) *Keyword:* shatto auto

English: (inf.) very tired

Japanese: tsukareta

Example: By the time I got home I was shattered.

(ii) Sample Questions taken from the cued-recall test

Vocabulary Test

Please translate these words into Japanese. Write in romaji.

(n.) = noun (meishi), (vb.) = verb (doushi), (adj.) = adjective (keiyoushi)

Example: English: book

Japanese: hon

1. English: fib (n.)

Japanese:

2. English: stagger (vb.)

Japanese:

(iii) Sample Questions taken from the multiple-choice test

PART A Which of the following words best fit the sentences 1) to 8). Write the letter in the space.
You can use the same word more than once.

(a)robber	(b) delay	(c) coat	(d)witness	(e)boom	(f)grub	(g)debate
(h)barney	(i)junk	(j)racket	(k)stink	(l)fib	(m)booze	(n)packet

Example: It was really cold today so I wore a(e).....

- 1). "That's such a! That toilet needs to be cleaned."
- 2). There will be a 20 minute before the plane takes off
- 3). "I really would like something to drink. Have you got any in the house?"
- 4). Johnny didn't do his homework. But he said he did do his homework. He told a

Appendix B

Analysis of Variance (ANOVA) Tables

Table a,
One-way analysis of variance for the effect of Group on test performance

	Source	SS	DF	MS	F	p
Cued-recall (after 1 week)	Between Groups	14.741	2	7.370	1.937	0.166
	Within Groups	91.333	24	3.806		
	Total	106.074	26			
Multiple-choice (after 1 week)	Between Groups	4.963	2	2.481	0.608	0.553
	Within Groups	98.00	24	4.083		
	Total	102.963	26			
Cued-recall (after 3 weeks)	Between Groups	7.185	2	3.593	1.046	0.367
	Within Groups	82.444	24	3.435		
	Total	89.630	26			
Multiple-choice (after 3 weeks)	Between Groups	17.556	2	8.778	1.862	0.177
	Within Groups	113.111	24	4.713		
	Total	130.667	26			

(SS = sum of squares, DF = degrees of freedom, MS = mean sum of squares, p = significance/ F ratio)

Table b,
Two-way Analysis of Variance for the effect of Group and Test on memory for Vocabulary

Source of Variation	SS	DF	MS	F	p
Within + Residual	294.94	24	12.29		
Group	41.72	2	20.86	1.70	0.204
Within + Residual	89.94	72	1.25		
Test	72.33	3	24.11	19.30	0.000**
Group by test	2.72	6	0.45	0.36	0.900

(SS = sum of squares, DF = degrees of freedom, MS = mean sum of squares, p = significance/ F ratio)

p<0.001**

Bibliography

- Atkinson, Richard C. (1975) 'Mnemonotechnics in second-language learning' *American Psychologist* 30, 821-828.
- Atkinson, Richard C., and Michael R. Raugh (1975) 'An Application of the Mnemonic Keyword Method to the Acquisition of a Russian Vocabulary' *Journal of Experimental Psychology: Human Learning and Memory* 104:2, 126-133.
- Avila, Enrique, and Mark Sadoski (1996) 'Exploring New Applications of the Keyword Method to Acquire English Vocabulary' *Language Learning* 46:3, 379-395.

- Beaton, Alan et al. (1995) 'Retention of foreign vocabulary learned using the keyword method: a ten-year follow-up' Second Language Research 11:2, 112-120.
- Brown, Cheryl (1993) 'Factors Affecting the Acquisition of Vocabulary: Frequency and Saliency of Words' in Huckin, Haynes and Coady, 263-287.
- Brown, Thomas S., and Fred L. Perry (1991) 'A Comparison of Three Learning Strategies for ESL Vocabulary Acquisition' TESOL Quarterly 25:4, 655-670.
- Burden, Tyler (2002) 'An Investigation into Japanese University Students' Use of Vocabulary Learning Strategies with Reference to the Keyword Method' 明海大学大学院応用語学研究科紀要、応用言語学m. No.4, pp.45-53.
- Coady, James, and Thomas Huckin, eds. (1997) Second Language Vocabulary Acquisition, a rationale for Pedagogy. Cambridge: Cambridge University Press.
- Cohen, Andrew D. (1987) 'The Use of Verbal and Imagery Mnemonics in Second-Language Vocabulary Learning' Studies in Second Language Acquisition 9, 43-62.
- Cohen, Andrew D., and Edna Apeh (1980) 'Retention of Second-Language Vocabulary over time: Investigating the Role of Mnemonic Associations' System 8, 221-235.
- Coleman, Hywel, ed. (1996) Society and the Language Classroom. Cambridge: Cambridge University Press.
- Craik, Fergus I. M., and Robert S. Lockhart (1972) 'Levels of Processing: A Framework for Memory Research' Journal of Verbal Learning and Verbal Behaviour 11, 671-684.
- Craik, Fergus I. M., and Endel Tulving (1975) 'Depth of Processing and the Retention of Words in Episodic Memory' Journal Of Experimental Psychology: General 104:3, 268-294.
- Elhelou, Mohamed-Wafaie A. (1994) 'Arab children's use of the keyword method to learn English vocabulary words' Educational Research 36:3, 295-302.
- Ellis, Nick C., and Alan Beaton (1993) 'Psycholinguistic Determinants of Foreign Language Vocabulary Learning' Language Learning 43:4, 559-617.
- Gruneberg, M. M., and G. C. Jacobs (1991) 'In defence of Linkword' Language Learning Journal 3, 25-29.
- Gruneberg, Michael, and Robert Sykes (1991) 'Individual differences and attitudes to the keyword method of foreign language learning' Language Learning Journal 4, 60-62.
- Huckin, Thomas, Margot Haynes, and James Coady, eds. (1993) Second language reading and vocabulary learning. New Jersey: Ablex Publishing Corporation.
- Hulstijn, Jan H. (1997) 'Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications' in Coady and Huckin, 203-224.
- LoCastro, Virginia (1996) 'English language education in Japan' in Coleman, 40-58.
- McCarthy, Michael (1990) Vocabulary. Oxford: Oxford University Press.
- McDaniel, Mark A., Michael Pressley, and Paul K. Dunay (1987) 'Long-Term Retention of Vocabulary After Keyword and Context Learning' Journal of Educational Psychology 79:1, 87-89.
- McKeown, M. G., and M. E. Curtis, eds. (1987) The Nature of Vocabulary Acquisition. New Jersey: Lawrence Erlbaum Associates.
- Merry, R. (1980) 'The Keyword Method and Children's Vocabulary Learning in the Classroom' British Journal of Educational Psychology 50, 123-136.
- Miura, Osamu, and Akira Kaneko (1997) Goro awase ei-tango. Tokyo: Lion-sha.
- Nagy, William (1997) 'On the role of context in first- and second-language vocabulary learning' in Schmitt and McCarthy, 64-83.

- O'Malley, J. Michael (1987) 'The Effects of Training in the Use of Learning Strategies on Learning English as a Second Language' in Wenden and Rubin, 133-144.
- Oxford, Rebecca L. and Robin C. Scarcella (1994) 'Second Language Vocabulary Learning among Adults: State of the Art in Vocabulary Instruction' System 22:2, 231-243.
- Paivio, Allan (1991) 'Dual Coding Theory: Retrospect and Current Status' Canadian Journal of Psychology 45:3, 255-287.
- Pressley, Michael, Joel R. Levin, and Harold D. Delaney (1982) 'The Mnemonic Keyword Method' Review of Educational Research 52:1, 61-91.
- Pressley, Michael, Joel R. Levin, and Mark A. McDaniel (1987) 'Remembering Versus Inferring What a Word Means: Mnemonic and Contextual Approaches' in McKeown and Curtis, 107-127.
- Raugh, Michael R. and Richard C. Atkinson (1975) 'A Mnemonic Method for Learning a Second-Language Vocabulary' Journal of Educational Psychology 67:1, 1-16.
- Reischauer, Edwin O. (1990) The Japanese Today. 4th ed., Tokyo: Tuttle.
- Rodriguez, Maximo and Mark Sadoski (2000) 'Effects of Rote, Context, Keyword, and Context/Keyword Methods on Retention of Vocabulary in EFL Classrooms' Language Learning 50:2, 385-412.
- Schmitt, Norbert (1997) 'Vocabulary learning strategies' in Schmitt and McCarthy, 199-227.
- Schmitt, Norbert and Michael McCarthy, eds. (1997) Vocabulary: Description, Acquisition and Pedagogy. Cambridge: Cambridge University Press.
- Shiraishi, Eiki (1999) Gogen no image de super Anki-hou. Tokyo: Juken Kenkyu-sha.
- Tinkham, Thomas (1989) 'Rote learning, Attitudes, and Abilities: A Comparison of Japanese and American Students' TESOL Quarterly 23:4, 695-698.
- Wenden, A. and J. Rubin, eds. (1987) Learner Strategies in Language Learning. Cambridge: Prentice Hall.