A Review of Research into Vocabulary Acquisition through Dictionary Use

Part 2: Incidental Vocabulary Acquisition through Dictionary Use

James Ronald

(Received on October 14, 2003)

Introduction

In this second part of a critical review of L2 vocabulary acquisition through dictionary use we will look at studies in which the purpose of researchers has been to investigate incidental vocabulary acquisition; vocabulary acquisition that takes place in circumstances in which the dictionary users' primary purpose is something other than vocabulary acquisition. In other words, as Laufer and Hill (2000) put it, it is "a by-product of another activity, such as reading or communication, without the learner's conscious decision, or intention, to learn the words". The issue is more complex than this statement may suggest, not least because the learner's decisions or intentions may not coincide with researchers' assessments, but for now it would be valid to suggest that none of the studies reviewed here investigates the effect of vocabulary learning activities on vocabulary learning.

For most of the studies, as we shall see, dictionary use and vocabulary acquisition took place in the context of reading comprehension texts given to students of English as a foreign language. Two other contexts for dictionary use are also investigated: reading and translation. The three contexts investigated — reading, reading comprehension, and translation — are important areas for many dictionary users in everyday life. Implicitly, this

may suggest that researchers are investigating everyday dictionary use in natural circumstances. In reality, however, there are many obstacles to monitoring natural dictionary use, and even greater difficulties in evaluating the effect of dictionary use on vocabulary acquisition in these circumstances.

There has been an increasing number of studies in this particular field over the past dozen years, perhaps because of the apparently fairly direct relevance that the results of these studies may have on pedagogical lexicography and language learning pedagogy. In this review we will not attempt to include all studies in the field. Rather, we will take a selection of six studies with different contexts or approaches and review them in detail. These will be studies by Krantz (1991), Bogaards (1992), Luppescu and Day (1993), Hulstijn, Hollander and Greidanus (1996), Aizawa (1999), and Laufer and Hill (2000). The studies will be reviewed in groups, according to the context of dictionary use: reading, translation, and reading comprehension. This paper will conclude by summarizing the achievements of the studies reviewed and suggest what may be yet to be achieved in this field.

Reading

We will look at two studies into the effect on reading on vocabulary acquisition: a little known large-scale study by Krantz (1991) and a smaller but often cited study conducted by Luppescu and Day (1996).

Krantz (1991) *Learning vocabulary in a foreign language: a study in reading strategies*

Krantz's impressive study investigates how Swedish learners of English acquire vocabulary through extensive reading with the aid of monolingual

or bilingual dictionaries.

Summary

For this study, English Department university undergraduates read a geography textbook written in English. While doing this, they were allowed to use a computerized dictionary: either a monolingual English dictionary or a bilingual English-Swedish dictionary. Krantz summarizes his objectives in four research questions. The first two are to what extent L2 vocabulary learning results from encounters with unknown words in context and to what extent it results from a combination of encounters in context and dictionary use. The two other questions are which of these approaches is more effective in different conditions, and which type of dictionary, monolingual or bilingual, is more efficient for the group of learners investigated.

Fifty-two Swedish English Department undergraduate students took part in this study. They were divided into two equivalent groups, a Monolingual Dictionary group and a Bilingual Dictionary group. As the students studied Economic Geography in addition to English, an English textbook on this subject was chosen as a suitable reading text. The text contained just over 50,000 tokens. Computerised dictionaries were chosen to keep an accurate record of the subjects' dictionary use.

In addition to the reading and accompanying dictionary use, a vocabulary test was made of 148 words which occur in the text. The main criterion for inclusion of test items was that most of them would probably be unknown to most of the subjects. 32% of items occur only once in the text, almost 26% occur twice, and 10% three times. The remaining 32% occur four or more times. In the tests, conducted before and after the reading, subjects were asked to supply the meaning of the test items, in Swedish, English or by

any other means. For eighty of the test items, only the English word was presented, and for the remaining 68 items, the words were presented in a sentence written to provide as little contextual support as possible.

In reporting the various results of the study, we will begin with the differences between the vocabulary pretest and posttest for the subjects as a whole. We will then look at the two groups' scores. After this we will consider the effect of dictionary lookups on the acquisition of test items.

Overall, the subjects each learned an average of 23 test items, with numbers of targeted words learned ranging between 9 and 40. Comparing the lowest scorers for the pretest with the highest scorers, the top 10 subjects learned an average of 24 test words while the bottom 10 learned an average of 17 words. For the top 10 subjects, the 24 test items represent 29% of previously unknown items while the bottom 10's 17 words only represents 11% of items unknown in the pretest. It is also worth noting that about 14% of words correctly identified in the pretest were not correctly answered in the posttest. We will consider below why this may be.

Raw gains for the two groups are almost identical: 23 items for the Monolingual Dictionary group and 22.7 for the Bilingual Dictionary group. However, in the pretest the monolingual dictionary group had correctly identified more of the test items: 39, as opposed to 30.4 for the bilingual dictionary group. This means both that the Monolingual Dictionary group started with a higher vocabulary than the Bilingual Dictionary group and that they learned a greater proportion of previously unknown words. The author suggests that as the Monolingual Dictionary group is stronger, the small differences in test scores can be attributed to this cause, and that in terms of vocabulary gains, the two groups are equivalent. As for the direct effect of dictionary lookup, as we shall see, there are differences between the two groups.

Within the test, for each student, Krantz identifies that targeted words were subject to one of two different learning conditions: words that were encountered only in the text and words that were encountered in the text and looked up. For previously unknown words, 9% were learned through reading only and 10% where dictionaries were also used.

Comments

Krantz' research offers many valuable insights into the role of dictionaries in L2 vocabulary acquisition. As we review the methods employed and results obtained through this study, we will consider the natural circumstances in which reading and dictionary use take place, the method employed to obtain valuable data about learner dictionary use while reading, and the large number of items from the text that are targeted for investigation in the study. We will also reflect on how data from this study may help shed light on the nature of the mental lexicon.

As the researcher points out, we should not be surprised if a set reading text in a foreign language course is read with two purposes and that these will affect reading behaviour: to understand the content of the text and for L2 language improvement. It is only with a recognition of these combined purposes that the reading condition can be judged in terms of natural behaviour in the specific context. In Krantz' study, there was no announcement of a comprehension or vocabulary test that would follow the reading so in this respect there was no imposed purpose for reading the text. On the other hand, the reading of the text in a monitored location with the use of computerised dictionaries does create an environment in which the importance of careful reading is evident. The dictionary use under investigation in this study can be understood, then, as reflecting typical dictionary use of a particular type: for careful reading in an academic context by highly

motivated learners of English using electronic dictionaries.

Natural dictionary use behaviour for reading long L2 texts is usually to use dictionaries very little. In this study, there was a considerable amount of dictionary use. This averaged 263 lookups per student during the reading, ranging between 58 and 641 lookups, with reading times between seven and eighteen hours. Readers using the bilingual dictionary consulted their dictionary almost twice as often as those using the monolingual dictionary: an average of 353 lookups for the Bilingual Dictionary group as compared to an average of 180 for the Monolingual Dictionary group.

Three or possibly four factors may help account for the generally high levels of dictionary use by subjects in both reading conditions. One, as noted by Knight (1994), is the relative speed and ease of consulting electronic dictionaries. Another consideration is the appeal of the new technology that the electronic dictionaries represented for the subjects. A third factor may be the monitored reading environment, which promotes careful reading, a part of which includes increased attention to unknown words. Finally, in this environment, advanced learners might make more use of dictionaries than less able learners would. Although in terms of unknown words they would have less need of dictionaries than lower level proficiency learners, their dictionary use should be more efficient; each lookup would be faster and interrupt the flow of the reading less than would be the case for less able learners. Although this may not seem to be borne out by the lookup data for the subjects of this study — the bottom ten subjects used the dictionary more than twice as often as the top ten subjects — we need to remember that all the subjects in Krantz's study may be described as advanced level, motivated learners: Swedish learners who chose to major in English at university and who volunteered to assist with this study.

A further impressive aspect of this research is the exceptional record of dictionary use it provides. Although, as we have discussed above, the use of electronic dictionaries will affect dictionary use in some respects, it does provide an ideal means of keeping an accurate yet non-intrusive record of dictionary use. This makes possible a much more accurate indication of how actual dictionary use affects vocabulary acquisition for the words that were known to have been looked up, and even allows for calculations of the relationship between time spent on lookups and learning. The advantages of this electronic recording of dictionary use are clear, especially with studies of larger numbers of readers, when we note that in many studies not using this technology, the only comparison available is of overall test results for readers with different reading conditions who may or may not have used the dictionaries at their disposal for some of the test items.

Also of particular note is the large number of test items included in this study. This is especially valuable with investigations into vocabulary acquisition resulting from extensive reading through which tens of thousands of words are encountered. Even the 148 test-words drawn from the text, a very large number when compared with the 20 or 30 items in other studies, by no means represent a comprehensive survey of unknown words in the text. With an average of 35 of the test-words correctly identified in the pre-test, this leaves 113 test items unknown to the average subject. When we compare this to the estimated average of 507 unknown word types in the text, the previously unknown test-words represent under a quarter of the estimated average unknown words in the text. Overall, though, using this number of items does provide a much more accurate and reliable picture of vocabulary acquisition through dictionary use than do studies with smaller numbers of test items.

While there are a number of valuable aspects to this study, one piece of

the data reveals problems with the assumptions upon which this, and much other research, is based. Of the average 35 words items correctly identified in the pretest, 14% (5 items) could not be correctly identified in the posttest. Compared to the average gains of around 23 items between the pretest and the posttest, this figure is not large but it does reveal a challenge to the view that vocabulary development proceeds uniquely forward. There are various possible explanations for these items correctly identified in the pretest and not recognized in the posttest. It could be argued that in the period of a few days between the two tests, the subjects forgot some of the words they had known. This theory relies on the subjects' failing to encounter, or notice, these words in the text. Another possibility is that the contexts or definitions of these items were misleading and challenged the subjects' prior understanding of the meanings of the words. While these factors may play a part in the apparent loss of vocabulary knowledge, two further causes appear much more likely: the instability of much of a language learner's word knowledge and the instability of language learners' confidence about word knowledge. That 14% of "known" items should became "unknown" a few days later could be seen as reflecting the unstable nature of our mental lexicon, perhaps amplified by the use of a test in which there is no way of indicating partial knowledge.

As we shall see, few other studies succeed in producing comparable data as regards the relationship between dictionary use and vocabulary acquisition, and that this is especially admirable in the context of extensive reading.

Luppescu and Day (1993) *Reading, dictionaries, and vocabulary learning*

This study asks what effect dictionary use has on L2 vocabulary acquisi-

tion, as opposed to text comprehension. Often cited in subsequent studies, it has, as we shall see, many problems of its own.

Summary

In this experiment, 303 Japanese university students were asked to read a short story of 1853 tokens written in English. 148 of the students were not allowed access to a dictionary while reading and 145 were allowed to use to use their English-Japanese dictionaries while reading. There was no time limit for the reading. Immediately following the reading, all the subjects were given a vocabulary test in which knowledge of 17 words from the short story was tested. The test was a multiple choice type, with the choices for each question being the test answer, three distractors, and an *I don't know* option. Two points were awarded for a correct answer, one point for *I don't know* and no points for a wrong answer.

Scores for the group permitted to use dictionaries were significantly higher than for the group that was not. The chances of subjects allowed to use dictionaries getting a right answer were 1.86 times greater than for the no-dictionary group, although there was a wide variation among test items. Two further statistics presented were that the no-dictionary group chose the *I don't know* response almost half the number of times, and that the group permitted the use of dictionaries took almost twice as long to complete the reading.

Comment

This experiment raises various questions relating to the investigation of L2 vocabulary acquisition through reading and dictionary use. Three aspects of this paper are of particular interest: the challenge of investigating the effects of dictionary use in a relatively natural environment, the compa-

rability of the two groups of subjects, and the quantity and quality of the test items used to measure vocabulary development through reading and dictionary use.

One point the authors make is that their aim is to investigate natural dictionary use. As such, although they divided the subjects into two groups, a group with dictionaries and a group without dictionaries, they could not dictate that subjects in the with-dictionary group had to use their dictionaries, let alone tell the subjects which words they should look up. Neither did they feel that any observation or monitoring of dictionary use would be sufficiently non-invasive as to leave the subjects' dictionary use behaviour unaffected. This means that all they are able to state with authority is that one group had possession of dictionaries and one group did not, and that the group with dictionaries spent almost twice as long reading the short story as the group without. It is, reasonably, inferred that the extra time was spent on dictionary use. Other than this, the main, again indirect, evidence of dictionary use is that the group with dictionaries had better test scores than the group without; since the two groups were believed to be the same except for the possession or not of dictionaries, it is assumed that this must account for the difference in test scores. As for whether the test items had been looked up in the subjects' dictionaries, the only indication of this, again circumstantial, is of a negative kind. The group without dictionaries gave more correct answers for words with the most confusing dictionary entries: those with a large number of different senses listed in the dictionary and for which it would be harder to locate the sense used in the story.

The inferences made above about dictionary use accounting for the differences between the groups' results can only be justified if we are sure that the two groups are the same in all respects other than whether or not they

had access to a dictionary during the reading. For this experiment there was no pre-test or other data to show that the two groups were equally proficient in English either generally or in terms of vocabulary. We are told that the groups are made up of whole classes of students and that students had been assigned to these classes solely on the basis of their surnames. The authors claim that the assigning of whole classes of students to the two groups can be seen as equivalent to randomly assigning individual students to the two groups. For a test of grammatical knowledge, this reasoning may be valid, but in terms of vocabulary it is quite conceivable that different classes of students may have studied different material with different vocabulary, some of which may be included in the 17 test items. Further, attitudes towards study, such as how much time and effort students are willing to devote to a task, may often be largely shared among students in a class, especially for tasks conducted in the class such as the reading task in this experiment.

Related to the above questions about the comparability of the two groups is the small number of test items. Initially, there were 17 words in the short story that were chosen as test items, although data from two of the items were later excluded from analysis. This left 15 items, some of which are relatively high frequency items (*happen, appear, worse*) and some which are known in Japanese as loanwords from English (*slide projector, clear*). The problem with these two types of words is that there is a reasonable likelihood of their having been encountered in some classes but not in others. This alone may cause a significant difference between the two groups' results to be recorded.

Finally, there are also some problems with the answers and distractors for the test items. For the target word *chant*, two of the choices offered might be acceptable: *speak* and *sing*. With *clear*, for which the word oppo-

site in meaning was requested, three of the four choices may be acceptable: *dirty*, *vague*, and *dull*.   As for *happen*, the correct answer is *occur*, a markedly less frequent word than the target word.   In any test there may be one or two items with which there are problems.   Here, as many as seven items are unsatisfactory in some way, accounting for almost 50% of those used for analysis.   With such a small total   number of test items, there is a greater likelihood of problems with individual items having a considerable effect on overall results.

The approach employed by Luppescu and Day in this study does appear to offer real insights into vocabulary development through dictionary use in a relatively natural environment.   However, questions about the comparability of the experimental groups, the small number of test items, and the fact that a large proportion of these items are faulty in some way, means that the claims made in this study remain largely unsupported.

Dictionary use while translating

Despite the widespread use of dictionaries by people doing translating, the only paper relating to vocabulary acquisition in this context is that by Bogaards reviewed below.   It is a very simple paper in some respects, but it also reveals two very simple but often overlooked facts about dictionary use: that the usefulness of dictionary types depends on the task for which they are being used, and that comprehension and correct use of previously unknown words are no guarantee of their retention.

Bogaards (1992) *Dictionnaires pédagogiques et apprentissage du vocabulaire*

This is a relatively small study which addresses the usefulness of different types of dictionaries in translation, and the effect on L2 vocabulary

acquisition of using these dictionaries. This focus on translation as the context of dictionary use is especially valid when we reflect on how much dictionary use, both in and outside of formal learning contexts, takes place as part of translation work..

Summary

In this study, Dutch university students of French were asked to translate a Dutch text into French with the aid of various types of dictionary. Altogether, 42 students completed the required tasks of translating the text followed, two weeks later, by a test of 17 less common words which appear in the text. The students were put into four groups according to the kind of dictionary provided during the translation task: a bilingual (Dutch-French) dictionary group, a (French) monolingual learner dictionary group, a standard French dictionary group, and a no dictionary group. Students using dictionaries were asked to keep a record of which words they had looked up. A further 14 students served as a kind of Control group, taking the vocabulary test but not having translated the text.

It is not indicated how the experimental groups were formed. Summaries of student grades for previous work indicate that the groups are similar but that they are not the same; this suggests that these were already pre-existing groups of students. As for the 17 test items, they are French translation equivalents of Dutch words in the text, described as not normally being in the productive vocabulary of the type of learner taking part in the study. The text to be translated was a 150-word Dutch passage, judged to be relatively easy with the exception of the 17 targeted words. The vocabulary test, conducted two weeks after the text translation task, consisted of the list of the 17 isolated Dutch words, which the subjects were asked to translate into French.

There are two sets of results: for the translation of the targeted words in the text and for the translations of the same words as isolated test items. In addition, within these results, scores are provided for words which had been looked up and words which had not. For the bilingual dictionary group, 12.0 of the 17 targeted words were looked up, with 10.3 of these correctly translated and a further 3.2 of the 4.6 words not looked up being correctly translated. This compares with an average of 7.6 targeted words looked up by the learner dictionary group, only 3.6 of which were translated correctly, with 4.0 of the 9.3 words not looked up being correctly translated. For the standard French dictionary group, too, success rates were low: an average of under 5.9 words were looked up, with 2.4 of these correct. Of the 11.1 words not looked up, 5.6 were correctly translated. As for the no dictionary group, 5.6 of the 17 words were correctly translated. This compares with totals of correctly translated words of 13.5 for the bilingual dictionary group, 7.2 for the learner dictionary group, and 7.5 for the standard French dictionary group.

In terms of targeted words correctly translated two weeks later, the figures for the three dictionary groups are much closer. They range from 7.6 correct words for the standard French dictionary group to 8.2 for the bilingual dictionary group, and 8.8 for the monolingual learner dictionary group, with an average of 5.0 words for the Control group. We will now go on to consider what these figures may tell us about the effect of the three types of dictionary on vocabulary acquisition.

Comment

This study, at first glance, appears very straightforward both in terms of the procedure employed and with regard to the results gained through the study. There are problems with the small numbers of targeted words, with

the small numbers of subjects in each group and, quite possibly, with the composition of the groups, but at least the purposes and the procedures employed seem to investigate dictionary use and vocabulary acquisition in a careful and coherent way. Bogaards sums up the findings by pointing out that the bilingual dictionary was most useful as an aid to translation and that the usefulness of monolingual dictionaries depends on the type of word being investigated, but that there was no clear advantage for either dictionary type as regards vocabulary retention.

The answer to the question about which type of dictionary is most useful for translation does seem self-evident. After all, a bilingual dictionary provides translation equivalents for looked up words while a monolingual dictionary does not. The use of a monolingual target language dictionary in translating from the native language will be indirect, involving looking up related words and finding the unknown target language equivalent, pointless if such a search is in vain, or for confirmation of lexical knowledge rather than providing information about L2 words which are new to the user.

As a simple comparison of results from the translation and from the test indicate, benefits for vocabulary acquisition of different types of dictionaries are less clear-cut, and harder to interpret. For the bilingual dictionary group, compared with the words correctly translated in the text only a little over 60% were correctly translated in the test. For the learner dictionary group, this figure is over 110%, and for the standard French dictionary, just over 100%.

The use of translation as a means of investigating vocabulary acquisition brings various problems with it. If we reflect that words other than the specified targeted French words are acceptable answers in the translated text and in the test itself, we may ask what is being tested and what the

results may tell us. While a bilingual dictionary will provide the translation equivalent for unknown L2 words, the absence of one may force learners to use known L2 vocabulary which approximates in meaning to the L1 word. A monolingual dictionary will serve the purpose of confirming whether these known words may be acceptable substitutes for the unknown translation equivalent. As far as the vocabulary test is concerned, for test items with previously unknown translation equivalents, bilingual dictionary group subjects will either be recalling the looked up translation equivalent or, for the first time, trying to think of a satisfactory equivalent. For the two monolingual dictionary groups, the task is much closer to that undertaken during the translation, this time without the aid of context but with the aid of the memory of the experience two weeks earlier. We should also not underestimate the motivation of language learners to consult their bilingual dictionaries after the translation task to check whether their guesses or impressions were correct.

As with many other studies, the small number of test items leaves us with as many questions as answers. The author suggests that success in finding or retaining unknown words depends on the type of word, but with so few targeted words of any type, this study offers little support for this theory. The focus on translations is attractive and useful, and does highlight the insufficiency of monolingual dictionaries for the purpose, but it also brings problems of its own. On the other hand, the surprisingly low retention rate for bilingual dictionary users does serve as a reminder that comprehension of L2 words, as demonstrated through translation, cannot automatically be equated with their retention.

Reading comprehension

We now come to three studies drawn from an increasingly large number

of such studies over the past few years that use reading comprehension as the task for which dictionaries are used and through which vocabulary is acquired. As we shall see, in this specifically language learning context, the distinction between intentional and incidental learning becomes fuzzy, at best. Reading and translation may be ends in themselves but what is the purpose of reading comprehension in a language class other than language learning? We will consider this and other questions as we review the studies by Hulstijn, Hollander and Greidanus (1996), Aizawa (1999), and Laufer and Hill (2000).

Hulstijn, Hollander and Greidanus (1996) *Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses, dictionary use, and reoccurrence of unknown words*

This carefully constructed study investigates the influence of different text conditions and differing occurrences of targeted words on L2 vocabulary acquisition.

Summary

A total of 78 advanced learners of French at three Dutch universities took part in this research. They were asked to read a short story in French, 1306 words in length, following which they were told they would be given a comprehension test. Sixteen words were selected as targeted items from this text, which was edited so that eight of the words occur once in the text and eight occur three times. The subjects were divided into three comparable groups and for each the text reading condition was different: reading the text with L1 marginal glosses, reading the text while being able to consult a bilingual French-Dutch dictionary, or reading the text without access

to glosses or a dictionary.

Twenty-five minutes was allowed for reading the text, following which three tests were administered. The first asked the subjects to indicate which of 32 words in the test they recalled having seen in the text, and to write the meaning of these words. The group with dictionaries were also asked to indicate which words they had looked up. In the second test the subjects were asked which of the 16 targeted words they had been familiar with prior to reading the text. For the third test, the targeted words were presented in the context, of a few lines, in which they had appeared in the text and asked to provide the meanings of the words.

There are four main types of data presented: the subjects' self-reported preknowledge of the test items, the Dictionary group subjects' reporting of their dictionary use, the test results of the subjects in the three reading conditions, and the test results for the words occurring once or three times in the text. For the latter two sets of results, there are results for the tests for the words in isolation and presented in context.

Of the items reported as previously known, only those for which a correct answer was given were counted: out of 16 targeted words, for the Marginal Glosses group an average of 0.7 words were already known, for the Dictionary group 0.1 words and for the Control group 0.4 words. Reported dictionary use by the Dictionary group subjects was very low: 4 subjects did not use their dictionary at all, while for the remaining 20 subjects targeted words were only looked up a total of 38 times, an average of 1.9 targeted words looked up by each of these subjects.

In the retention tests, the Marginal Glosses group's scores were significantly higher than the other two groups' scores, both for test words in isolation and in context, for words occurring once or three times in the text, and for whole points only or whole and partial points. There was no significant

difference in the results of the Dictionary group and the Control group but, as we will discuss below, analysis of scores for individual items do reveal interesting data about the effect of dictionary use.

Comment

This study illustrates how even a carefully planned and executed study can be marred by the unpredictability of L2 readers' dictionary use. The research does provide interesting data about the effect of frequency of occurrence of words on their retention and of the effect of providing L1 marginal glosses, but little relating directly to the effect of dictionary use. This is because very little dictionary use took place: overall, less than one in ten targeted words were looked up. Although knowledge of just one or two key words in a text may affect comprehension of the whole text, in this case, at least, the very selective dictionary use does not appear to have aided vocabulary learning in any global way. We will review what effect of dictionary use there was before going on to consider why dictionary use was so limited.

Only when there is analysis of scores for items that were looked up do we see what effect there may be for dictionary use. For the 38 lookups of targeted words, there was a retention score of 59% for words presented in isolation, as compared with an average of 8.75% for the Dictionary group responses overall and 26.25% for the Marginal Glosses group. For individual items the scores are still higher; the most looked up word, *pépinière*, was looked up by 17 subjects, 15 of whom gave a correct or partially correct response. This represents an 88% rate of retention for these subjects, as opposed to 70% for the Marginal Glosses group. These figures should, however, be treated with caution. While we may accept that the three original groups of subjects are equivalent, a subgroup distinguished by its

dictionary use will not be. The high levels of standard deviation for the means of test scores alert us to the wide variation among subjects within each group.

A further point is that, as the researchers note, for four of the Marginal Glosses group, retention performance for words in context was in the order of 13 or more wholly or partly correct items out of 16. We are, perhaps tellingly, not told how many lookups were made by individuals in the Dictionary group, but it may well be that the most able learners were also the most avid dictionary users. It may be this that is being reflected by the analysis of scores for items that were looked up. Two further factors complicate this issue. If we assume that all the glosses were actually referred to, the incidental vocabulary learning burden for the Marginal Glosses group can be said to be all 16 words, while for the Dictionary group it may be the 1 to 5 or 6 targeted words looked up by the learners. We should not be surprised that the few looked up words are learned better than the larger number of items with glosses — if, that is, the glosses were actually consulted. It is assumed that they were, and the test scores seem to confirm this, but there is no external evidence of the extent to which this did take place.

The second question we need to ask is why dictionary use was so limited. The main reason proposed by the researchers is that the students in the Dictionary group did not feel that looking up words would help them with the comprehension questions they were expecting. Other contributory factors mentioned include the advanced level of the learners, their natural aversion to using dictionaries, the length of the text, and the trouble of using paper dictionaries as compared to electronic dictionaries. Two factors not mentioned but perhaps also relevant are the limited time available for dictionary use and the reading environment. Two-thirds of the stu-

dents in the same room as the Dictionary group did not use dictionaries at all; this may have created an atmosphere in which students with dictionaries felt uncomfortable about using them even if they did have time to use them.

In conclusion, while this study does provide us with various insights into dictionary use, it provides little reliable data about its effect on incidental L2 vocabulary acquisition. On the other hand, the much more impressive results for the Marginal Glosses group may tell us as much about dictionary use as about the use of glosses themselves. The test results for this group show that where meanings of unknown words in an L2 text are provided in an easily accessible form, they will be consulted and substantial levels of vocabulary learning may result. Presumably, how accessible the meaning is — as marginal glosses, in an electronic or paper dictionary, in a bilingual or monolingual dictionary — will affect how much the meanings of unknown words are consulted. So, too, will the purpose of reading a text. If, for example, subjects had been told that no test would follow the reading, or that a vocabulary test would follow, we might expect different dictionary use behaviour to result. Although Hulstijn et al.'s purpose through this study was to investigate incidental vocabulary use in a natural L2 reading environment, we can see that in test conditions there is no such thing as a natural reading environment, and that reading followed by a vocabulary test is arguably as *natural* as reading followed by comprehension questions.

Aizawa (1999) *A study of incidental vocabulary learning through reading by Japanese EFL learners*

Summary

This study focuses on the learning of vocabulary from English L2 texts by

Japanese high school students. Although the stated aim is to investigate incidental vocabulary acquisition, target words are underlined and numbered to ensure that they stand out from the text. The purpose of this is to guide learners to words worth spending time on, in the hope of increasing the low level of successful dictionary use reported in his own and others' previous studies. The questions addressed here are whether bilingual dictionary use helps learners read an L2 text more accurately, whether this dictionary use increases the retention of unknown words, and whether vocabulary acquisition levels are higher for more proficient learners both with and without dictionaries.

A total of 308 high school students took part in this study. Following a vocabulary test, the learners were divided into two nearly equivalent groups, one assigned a reading with dictionary condition and one a reading without dictionary condition.

Two texts were selected as reading passages, and both edited to include 12 target words, the meaning of which should be inferable from context. The texts were each accompanied by eight comprehension questions. Forty minutes was allowed for reading the two texts and answering the 16 questions. Following this, after materials had been collected, a surprise 10-minute vocabulary test was conducted. This was a multiple-choice select definition test of the 24 target words. The same test, with test items reordered, was conducted two weeks later.

The results for the reading comprehension questions showed that the Text Only group scored slightly, but significantly, higher than the Dictionary group. For the immediate vocabulary test, the Dictionary group's scores were almost 50% higher than the Text Only group's scores (a mean score of 15.60 as against 10.88), while in the delayed vocabulary test, the Dictionary group still scored significantly higher than the Text Only group,

but with a much reduced difference between the two groups (13.01 as against 11.42). While the Dictionary group's score fell, the Text Only group's score rose. When the groups were divided into higher proficiency and lower proficiency groups, of most note is that for higher proficiency learners there was almost no difference in the delayed vocabulary test scores for the Dictionary group and the Text Only group. We will consider why this may be below.

Comment

Various aspects of the above test results are surprising in some way: the higher reading comprehension scores for the group without access to dictionaries. Given this, the Dictionary group's markedly higher vocabulary test scores are also unexpected. Finally, the rise in the Text Only group's vocabulary test scores in the delayed test also needs further examination.

Aizawa suggests two possible reasons for the comprehension test results: i) that too much time was spent using the dictionary to allow enough time for the comprehension questions or ii) that the subjects were not good or efficient dictionary users. While the first point does seem a likely explanation of these scores, it would be more credible if scores for each question were given to show that Dictionary group subjects answered later questions less well or failed to answer them. As for the second point, the author expands on this by suggesting that learners will often look up a word without trying first to guess its meaning from the context or trying to relate the meaning in the dictionary with that in the text.

Subjects may have looked up targeted words in the text without prior guessing of their meaning, settling for the first sense for polysemous words, but this does not appear to have adversely affected their scores. This may

be because for over 20 of the 24 targeted words, the sense in the text and tested in the vocabulary tests is the first or only sense generally given in English-Japanese dictionaries. This means that without checking that the first sense they encounter matches that in the text, dictionary users' focusing on the first sense will usually be to their advantage for the vocabulary tests. This alone, then, would not account for the Dictionary group's poor comprehension test scores.

The markedly higher scores for the Dictionary group in terms of vocabulary retention are not in themselves surprising but do contrast with their low comprehension test scores and with those in Aizawa's other study reported above. This might be explained by the underlining and numbering of the targeted items in this study and not in the other. The combination, in this study, of low text comprehension scores and higher vocabulary retention scores seem to confirm that for the Dictionary group the reading task has become a vocabulary learning task. Although the distinction between incidental and intentional vocabulary learning is open to discussion, as we will see in Laufer and Hill's (2000) paper, the behaviour of the subjects in this study suggests that here Aizawa is investigating intentional, not incidental, vocabulary acquisition.

Finally, let us try to account for the rise in the Text Only group's scores for the delayed vocabulary test. The small number of test items, the highly motivated subjects, and the particular rise among more proficient Text Only subjects all suggest that these subjects may well have looked up some targeted words between the two tests, another feature of subjects' behaviour that is often overlooked.

Laufer and Hill (2000) *What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention?*

This study uses a kind of hybrid source of lexical information, half dictionary and half marginal gloss, increasingly common both in computer assisted language learning and in on-screen reading of electronic texts.

Summary

In this study, a short reading text of about 120 words was used, and 12 single-sense words from the text were identified as words that the subjects would be unlikely to know. These were highlighted in the text. Prior to the reading, the subjects were presented, on-screen, with a list of the 12 target words and asked to give the meaning of any words they knew. Following this, the subjects were allowed 10 minutes to read the text on-screen. They were told that a comprehension test would follow the reading, and were encouraged to look up the dictionary entries for the target words. This was done by clicking on the particular highlighted word and choosing whether to have the dictionary entry shown in English or the learner's L1, or whether to listen to the word's pronunciation. There was a test of the 12 target words immediately after the reading; as in the pre-test, the subjects were asked to give the meanings of all words they knew. This, test, however, was written on paper rather than conducted on-screen. Following this test, there was a short comprehension test.

Initially, 97 students of English took part in this study. Of these, data for 25 was rejected as a pre-test indicated that they knew two or more of the target words prior to the study. The focus in this study, then is on the 72 advanced level students of English, 32 Israeli and 40 Hong Kong Chinese,

who knew none or only one of the target words before the study.

The vocabulary post-test results showed that the Israel group gave correct meanings for an average of 4.0 words, while the Hong Kong group gave correct meanings for 7.45 words, representing 33% and 62%, respectively, of the target words. These figures were adjusted to exclude any words correctly identified in the pre-test. The use of dictionary information accessed also varied widely between groups. The Israel group subjects average over 1.5 lookups of some kind per word, while for the Hong Kong group this figure was just over 2.5. For the Israel group, the majority of lookups were for definitions in their L1, while the sources of dictionary information most often accessed by the Hong Kong group were both L1 and L2 entries, with or without other information such as pronunciation of the individual words. The most successful sources of information are the L2 definitions for the Israel group and L1 and L2 entries together for the Hong Kong group. These figures, as we shall see, require more detailed investigation.

Comment

An interesting feature of this study is the similarities it bears with some earlier studies, despite its use of modern technology. These include the small number of target items, the intensive focus on the target words by the subjects, and the high levels of retention of target words. In these respects, it most closely resembles Seibert's 1930 study. We will look here at two main aspects of the paper: the claims regarding incidental vocabulary acquisition and what this study may tell us about the effect of accessing different types of dictionary information.

The authors go to some lengths to claim that their study is investigating incidental vocabulary acquisition, and although the validity of this claim depends to some degree on how incidental is defined, in many respects it

does seem hard to justify. From the design perspective, the pre-test containing only the 12 target items, the highlighting of the items in the text, the use of "dictionaries" containing only these words, the encouragement to look up these words, and the large amount of time made available for reading a very short text combine to give a focus on the target items that may for the subjects have seemed far from incidental. Perhaps more important is the perspective and purposes of the subjects; if for them the main purpose of the task is to increase their vocabulary, it would be hard to argue that vocabulary acquisition is incidental, regardless of the intentions of the researchers. In this study, the behaviour and apparent acquisition levels of the subjects strongly suggest that for them the learning of the target words is a central, not incidental, concern. This is especially true for the Hong Kong group, who averaged over 2.5 lookups per word, often including the pronunciation of the words, and in the post-test were able to give correct meanings of 7.45 of the 12 target words. This dictionary use seems excessive if the goal were only to prepare for a short comprehension test, even given the ease of accessing dictionary information on-screen.

Whether we term the vocabulary acquisition claimed here as incidental or not, we do need to ask why it is apparently so high (without pre-test results we cannot be sure what levels of gains there were in the study) when compared with similar studies. The subjects' intensive focus on the target words is surely an important factor in accounting for the exceptionally high post-test scores for both groups. Clearly, acquisition of the target words was the target aimed for by many of the subjects in this experiment, and they were given ample time to focus on this. The use of a very short text (about 120 words, as compared to over 1,300 words in Hulstijn et al's 1996 study or about 500 words in Knight's 1994 study) will also increase the focus on the target words. A further factor might be the use of an

on-screen pre-test and a paper post-test; especially where different scripts are used, as with Hebrew and Chinese, recording the meaning equivalent using a computer is a more complex and time-consuming task than writing it on paper. This alone may produce higher scores for the post-test than the pre-test.

Although the authors suggest that there is little or no correlation between the number of lookups and post-test scores, at least for in-group comparisons, there would clearly be a correlation if lookups and test scores for all 72 subjects were investigated together. Put simply, both the Hong Kong group's lookup rate per word and their average post-test scores are almost twice those of the Israel group. This relationship is confirmed when we consider other research in terms of lookup rates and vocabulary retention. In other studies, with usually no more than one lookup per word, the highest retention rates are between 20% and 25% for looked up words. The only exceptions are studies where there is a deliberate focus on word learning, such as Seibert's (1930) and, arguably, Aizawa's (1999) or Iwai's (2000). In this study, the Israel group's score of 33% for an average of 1.5 lookups per word and the Hong Kong group's score of 62% for an average 2.5 lookups suggest that there is a definite link between lookup rates and acquisition rates, just as there is between occurrence of words in texts and their rate of acquisition.

As for which type of accessed dictionary information is most likely to lead to acquisition, the situation is more complex than that reflected by raw figures showing information types accessed and retention success rate. For example, we are shown that for the Hong Kong group the highest "success rate", of 79%, is for accessing monolingual L2 entries alone. In addition to the information source, we would also have to consider what types of learners might typically access this type of information, and for which types of

word are different types of information typically sought. As far as learner type profiles are concerned, these might include information about L2 language proficiency, skill at using dictionaries, and degree of conscientiousness as a language student. Word types may be categorised according to degree of prior familiar, of perceived relevance to the learner or not, whether they are abstract or concrete, and different parts of speech. No information is available regarding any of these factors which may affect learner dictionary use and acquisition rate.

This study, then, does give an indication of the high levels of acquisition that can be achieved when learners focus intensively on learning a small set of words with the aid of lexical information of the type found in dictionaries. It points towards indicating the effects of different types of information or multiple lookups, but without sufficient data to do more than this. Finally, as far as the effect the technology may have on dictionary use behaviour is concerned, the study does demonstrate how many lookups can be achieved in a short space of time and how students may choose to access and use various combinations of information types.

Conclusion

Four main issues stand out from the studies reviewed above: an insight into the nature of vocabulary, the way in which even the most carefully conducted studies are flawed, questions about how incidental the acquisition investigated here may be, and the use of computers as a means of investigating dictionary use. Very briefly, we will reflect on these four questions.

First, most clearly seen in Krantz' (1991) study, but also undoubtedly a feature of the other studies reviewed, is the phenomenon of apparent vocabulary loss as well vocabulary gain. As suggested above, although other interpretations may be possible, the most likely explanation for this is

that vocabulary items are less like the building blocks we may imagine and more like clouds: less concrete and less permanent than even the words *acquisition* or *attrition* suggest. Secondly, even the most carefully prepared of studies, of those reviewed the studies by Krantz (1991) and Hulstijn et al. (1996), display fairly serious flaws, such as the lack of comparability between experimental groups or the failure to predict very low dictionary use. The third point relates specifically to the claim, explicit or implicit, of investigating natural dictionary use that the studies make. They all take place in an academic context with language learners as subjects, and in at least some of the studies the subjects are encouraged to use dictionaries, target words are highlighted, and excessive time is allowed for the supposed primary task of the experiment. Language learners are undoubtedly an important group of dictionary users but the results of these studies need to be treated with caution by lexicographers.

Finally, the use of electronic dictionaries in various forms in two of the studies reviewed show future possibilities for investigating the effect of dictionary use on vocabulary acquisition. This is especially valuable when we can see that, in certain language learning environments at least, the electronic dictionary is the preferred format for the majority of language learners.

References

Aizawa, K. (1999). *A study of incidental vocabulary learning through reading by Japanese EFL learners.* Unpublished PhD. thesis. Tokyo: Tokyo Gakugei University.

Bogaards, P. (1992). Dictionnaires pédagogiques et apprentissage du vocabulaire. *Cahiers de Lexicologie* 59, 93-107.

Hulstijn, J. H, Hollander, M. and Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: the influence of marginal glosses,

dictionary use, and reoccurrence of unknown words. *Modern Language Journal* 80 (3), 327–339.

Iwai, Y. (2000). *Dictionary use in context and vocabulary retention.* Unpublished M. A. dissertation. Birmingham: University of Birmingham.

Knight, S. (1994). Dictionary use while reading: the effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal* 78 (3), 285–299.

Krantz, G. (1991). *Learning vocabulary in a foreign language: a study of reading strategies.* Gothenburg: University of Gothenburg.

Laufer, B. and Hill, M. (2000). What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? *Language Learning & Technology* 3 (2), 58–76.

Luppescu, S. and Day, R.R. (1993). Reading, dictionaries, and language learning. *Language Learning* 43 (2), 263–287.

Seibert, Louise C. (1930). An experiment on the relative efficiency of studying French vocabulary in associated pairs versus studying French vocabulary in context. *Journal of Educational Psychology* 21 (4), 297–314.