Beyond A Clockwork Orange: Acquiring Second Language Vocabulary through Reading

Marlise Horst *Concordia University, Montreal*, **Tom Cobb** *Université du Québec à Montréal*, **Paul Meara** *University of Wales, Swansea*.

This replication study demonstrates that second language learners recognised the meanings of new words and built associations between them as a result of comprehension-focused extensive reading. A carefully controlled book-length reading treatment resulted in more incidental word learning and a higher pick-up rate than previous studies with shorter tasks. The longer text also made it possible to explain incidental learning growth in terms of frequency of occurrence of words in the text. But the general frequency of a word was not found to make the word more learnable. Findings also suggested that subjects with larger L2 vocabulary sizes had greater incidental word learning gains. Implications for incidental acquisition as a strategy for vocabulary growth are discussed.

INTRODUCTION

In first language acquisition research, it is well established that reading is one of the main ways of learning new words, and that people who do more reading know more words (Sternberg 1987a, West & Stanovich 1991). Reading is important for first language development and it is assumed to be important for second language development as well. Language teachers believe that extensive reading helps their students acquire new vocabulary, and second language acquisition researchers have determined that learning new words from reading should be possible (Krashen 1989, Wodinsky & Nation 1988). But as learners read, does word learning occur to any practical extent? And, given a choice of methods, is reading extensively more effective than direct vocabulary instruction, as Krashen (1989) has argued? It is important to establish what extensive reading can actually accomplish in the way of imparting new vocabulary knowledge.

Unfortunately, the experimental support for incidental vocabulary acquisition through reading in a second language is weak and plagued by methodological flaws. Furthermore, the research has done little to explain how acquisition occurs. Teachers need substantive answers to questions like: What kinds and amounts of reading facilitate incidental vocabulary acquisition? What makes a learner a good incidental

MARLISE HORST is a teacher trainer at the TESL Centre at Concordia University in Montreal. She is also a doctoral student in a distance education program at the Centre for Applied Language Studies at the University of Wales in Swansea. Her research interest is incidental vocabulary acquisition.

TOM COBB recently completed his Ph.D. in educational technology at Concordia University. His area of interest is vocabulary learning through computer concordancing. Presently he is involved in developing a system of computerized language tests at Université du Québec à Montréal. He enjoys running with the Montreal Hash Harriers.

PAUL MEARA directs research in second language vocabulary acquisition at the Centre for Applied Language Studies at the University of Wales. Current projects include investigating models of the lexicon and the acquisition process and developing vocabulary-size tests. He also plays third bassoon in the City of Swansea Concert Band.

acquirer? Is there a particular stage at which learners are most likely to benefit? This study addresses both methodological and explanatory issues. It attempts to make a clearer, more convincing case for vocabulary learning through reading, and to go beyond this to consider factors that may affect the learning process.

The first study claiming to show that second language vocabulary learning occurs incidentally through reading is a well known experiment by Saragi, Nation and Meister (1978). They tested native speakers of English who had read Anthony Burgess's *A Clockwork Orange* on their understanding of many of the Russian-based slang words that occur in the novel. They found that the subjects were able to correctly identify the meanings of most these *nadsat* words, especially the frequently occurring ones, in a surprise multiple-choice test. But it seems strange to equate the circumstances of this study with second language learning. Here, native speakers of English used contexts which they must have fully understood to infer, for example, that *droog* meant *friend;* but making such connections is probably much harder for readers in a foreign language for whom many words in the context may be unknown or only partially known.

The mean number of words subjects acquired in the experiment was 68.4, amounting to about three quarters of the 90 words tested. But replications of this study with second language learners have not managed to reproduce these impressive results (see Table 1 below). For instance, Pitts, White and Krashen (1989) report a mean score of just two *nadsat* words correctly identified after subjects read *A Clockwork Orange* for an hour and took a test on 30 items. Other studies using a Clockwork methodology (Day, Omura & Hiramatsu 1991, Hulstijn 1992) report similar gains of just one, two or three words. Dupuy and Krashen (1993) report a larger gain of almost seven words, but this higher than usual result may have little to do with reading since their experiment also involved viewing a video.

Taken as a whole, these L2 reading studies indicate a rate of roughly one word correctly identified in every twelve words tested. Such small gains are not surprising because opportunities to read and encounter new words were limited in the experiments; none of the reading treatments lasted more than an hour, and most were much shorter. And, in contrast to the 60,000-word novel that the subjects of the Saragi et al study read over a period of days, the longest L2 reading task amounted to 6,700 words (Pitts et al 1989), and others such as Hulstijn's (1992) 907-word task were far shorter. The amount of reading that actually transpired is probably much less than text lengths indicate since there was no strict control on whether subjects completed reading tasks. Pitts et al report that over 50 per cent of their subjects failed to finish; the other studies may have suffered in the same way to an unknown degree in an unknown number of cases.

	Saragi et al (1978)	Pitts et al (1989) exp 1	Pitts et al (1989) exp 2	Day et al (1991) exp 1	Day et al (1991) exp 2	Hulstijn (1992) exp 1	Dupuy & Krashen (1993)
Subjects	20 NS	35 NNS	16 NNS	89 NNS	200 NNS	65 NNS	42 NNS
Reading treatment	60,000 words	6700 words	6700 words	1032 words	1032 words	907 words	video + 15 pages ? words
Time for reading	? days	60 mins	40 mins	30 mins	30 mins	? mins	40 mins
No. and type of items	90 nadsat	30 nadsat	28 nadsat	17 English	17 English	12 Dutch	30 French
Test Type	MC	MC	MC	MC	MC	state meaning	MC
Words learned, mean no.	68.4	1.8	2.4	1.1*	3.0*	0.9	6.6*
Words learned, mean %	75	6	9	6	18	8	22
Approx. pick up rate	3 of 4	1. of 17	1 of 12	1 of 15	1 of 6	1 of 13	1 of 5

NS = native speaker; NNS = non-native speaker; MC = Multiple Choice; * = gain established by comparison to a control group

Table 1: Overview of Clockwork Orange experiment and replications

In addition to the limited opportunities to pick up new words in reading treatments, there is also the problem that short tests presented limited opportunities for subjects to demonstrate what they might have learned. None of the experiments test more than 30 items. On the other hand, in experiments where real words were used instead of *nadsat* items (e.g. Day et al 1991), tests may have overestimated learning, since results may include as 'growth' unspecifiable numbers of previously known words. Still another reason to question these already questionable findings is the fact that gain scores in several of the studies are based on comparing experimental groups to controls which may or may not have been comparable.

Small incidental learning gains are to be expected since studies of first language learners have found pick-up rates to be low. Nagy, Herman and Anderson (1985) determined that for school-age children learning English as their first language, the chance that a reading encounter with a new word will result in the learner being able to answer a multiple-choice question about it correctly is less than one in ten.

So the one in a dozen rate established by the *nadsat* series could be a reasonable estimate of the word learning that second language learners manage to achieve. The problem is not with the size of the finding (though it does raise questions about the efficiency of extensive reading as a vocabulary teaching technique) but with weaknesses in the studies that establish it. Gain scores of just two or three words in methodologically flawed experiments hardly amount to convincing evidence of learning second language vocabulary incidentally through reading.

The case for incidental vocabulary acquisition clearly needs more substantive support, and the experiment reported here attempts to provide it by expanding the reading treatment, testing more words, and exercising tighter experimental control. But even if this results, as expected, in a greater amount of vocabulary growth, will this advance our understanding of the incidental effects of reading? Meara (1997: 113) has suggested that research in the *Clockwork Orange* mode is like "planting seeds in a plot in order to confirm that they will grow into flowers." When the daisies appear, the growth hypothesis is experimentally confirmed, but very little can be said about how or why it happened.

The original *Clockwork Orange* study did attempt an explanation, however. It related incidental word learning growth to numbers of occurrences in the text, but none of the replication studies with second language learners have pursued the frequency (or any other) explanation. The experiment reported here considers incidental vocabulary growth in relation to text frequency and two other possible explanations, general frequency in the language as a whole and subjects' prior vocabulary size.

RESEARCH QUESTIONS

In this study, second language learners read all of a 109-page book, a simplified version of Thomas Hardy's *The Mayor of Casterbridge* (Jones 1979) over a ten-day period and had many opportunities to acquire new words in the process. Whether or not this occurred was explored in two ways. It was expected that subjects would recognize more definitions of words after reading the 21,232-word text, and would also be able to make more meaning associations between words.

Words occurring more often in the text were expected to be learned more than less frequent ones; that is, it was expected that text frequency would play a facilitating role, as the Saragi et al (1978) study found. The frequency of words in the language as a whole was also investigated; Brown (1993) found overall frequency to be a better predictor of incidental vocabulary growth than frequency in the specific texts her subjects read. The third explanatory variable was learner vocabulary size. It was assumed that knowing more words would assure better global comprehension of the text and, as a result, more incidental word acquisition. Laufer (1989, 1992) found evidence of a strong relationship between measures of learner vocabulary size and text comprehension.

In summary, the main questions under investigation were as follows:

1.   Does reading a simplified novel lead to increased word knowledge?
2.   Are words that occur more frequently in the text more likely to be learned?
3.   Are words that occur more frequently in the language at large more likely to be learned?
4.   Do learners with larger vocabulary sizes learn more words?

METHOD

The 34 subjects in the quasi-experimental study were students in an intensive English program at Sultan Qaboos University in Oman. These low-intermediate learners were members of two intact classes in a 14-week reading course designed to prepare them for the Cambridge Preliminary English Test (1990), henceforth referred to as PET. To ensure that the subjects read all 21,232 words of the simplified *Mayor of Casterbridge* text, a rather unorthodox strategy was used: subjects followed along in their books while the entire text was read aloud in class by the teacher. This proved to be a valuable way of controlling important aspects of the experiment. Careful attendance records were kept over the six classroom sessions of about an hour each that were needed to complete the book. This means that it is possible to say with confidence that subjects were exposed to the entire text. One student who missed three of the sessions and another who missed two were withdrawn from the study. The remaining 34 appeared to be absorbed by the story of secret love, dissolution and remorse, and tears were shed for the mayor when he met his lonely death at the end.

Reading aloud created the circumstances for incidental acquisition by precluding opportunities for intentional word learning. The reading focused subjects' attention on the events of the story and allowed the text itself (and a few pictures) to function as support for learning new words, but the pace did not allow for looking words up in dictionaries. The texts of *The Mayor of Casterbridge* were distributed to the students at the beginning of each session and collected afterwards, so that few words could be looked up or studied at home. To deal with the considerable pressure to allow dictionary use or explain words during the classroom sessions, the students were told that they could have the books once the story was finished. It was suggested that they could circle any problem words as they occurred and look them up later, and the students appeared to be satisfied with this compromise.

The entire text of *The Mayor of Casterbridge* was typed into a computer in order to identify the words to be tested and their frequency in the text. The novel is one of a series of simplified classics published by Nelson for learners of English who know approximately 2000 basewords. It was assumed that while most of the text would be made up of these high-frequency items, there would also be a substantial number of low-frequency words that occurred often, were well supported by the

text, and were unlikely to be known to the subjects. To locate such words a computer program called *EspritDeCorpus* (Cobb 1994) was used to identify all items not on the Cambridge PET list of 2387 high-frequency words of English (based on Hindmarsh 1980). Since the PET word list was systematically studied in another part of the subjects' course, it was important to exclude these words.

After proper nouns were removed from the list of words the computer analysis had identified as non-PET items, 222 basewords remained, ranging in frequency from 1 to 17 occurrences. Two thirds of these words occurred only once and were rejected as being too infrequent to be good candidates for incidental learning over a ten-day period. A few of the remaining 75 words, e.g. *furmity* and *skimmity*, met the criteria of being unlikely to be already known and unlikely to be encountered anywhere else. However, most of the words turned out to be far more common items, e.g. *dusk* and *harvest*. It became clear that any sizable list of test items would have to include words that some subjects would probably know already. Eight of the non-PET words occurred seven times or more in the text and all of these were included on the test. A further 37 items were chosen at random from the other frequency levels so that middle and low frequency levels were also represented and there was a range of opportunities for the hypothesized frequency factor to act.

Two tests of knowledge of the items were prepared, a 45-item multiple-choice instrument which required subjects to recognize a correct definition for each word, and a 13-item word-association test that required making a meaning link between two words by rejecting a third odd one out. The association test was based on a model developed by Read (1993) and modified by Vives (1995); three native speakers were found to concur completely on the words that did not belong in the sets. Sample questions are shown in Figure 1 below.

Circle the letter of the correct meaning:	Circle the one that does not belong:
1. carriage	1. sorrow
A. you ride in it	suffering
B. confident feeling	stare
C. fight, argument	
D. diary, notebook	2. carriage
	flame
2. companion	procession
A. business	
B. music program	3. affair
C. you put clothes in it	folk
D. friend	relative

Figure 1: Sample test questions, two types

The two measures were administered as a pre-test about a week before the reading-aloud sessions commenced. It was assumed that this time lapse would allow the items to be forgotten to the extent that they would not be immediately recognized as testing points when they were encountered in the story. This seems to have been effective; in a discussion held after the post-test, students were surprised to learn that the tested words had occurred repeatedly in the *Mayor of Casterbridge*. Their response also suggests that any word learning that occurred was implicit and incidental.

In order to investigate the possible role of vocabulary size in incidental learning from reading, the Levels Test (Nation 1990) was administered at the 2000, 3000 and 5000 frequency levels. According to this instrument, the average knowledge of the 5000 most frequent English basewords was estimated at 2071 words (sd = 560), and the average knowledge of the 2000 most frequent words was 1203 (sd = 348). These averages suggest that the choice of a reader at the 2000-baseword level was roughly on target, although some in the group must have found it challenging. The considerable variance in scores means that there was ample opportunity to observe effects of the hypothesized vocabulary-size factor.

How Many Words Were Picked Up?

The pre-test mean of 21.64 on the multiple-choice measure indicated that almost half of the 45 target words were already known in the group. In other words, although individuals differed with respect to which items they already knew and how many, in the group as a whole, an average of about 23 words remained available for possible incidental acquisition. This figure defines the amount of growth that could occur, unlike earlier studies using real word, non-*nadsat* targets (e.g. Day et al 1991) where the extent to which subjects were being tested on items they already knew was not strictly controlled.

Given the history of small findings and low pick-up rates, a learning gain of one or two of the 23 words might be expected. However, as shown in Table 2 below, the post-test average was found to be 26.26, indicating a mean gain of about five words (with considerable variance). A *t*-test for paired samples showed that this pre-post test difference was significantly greater than chance. The knowledge gain of five of the 23 means that about 22 per cent of the words that could have been learned were learned; in other words, there was an average pick-up rate of about one new word in every five — considerably more than the one in a dozen of the *Clockwork Orange* replications.

	Pre-test	Post-test	Mean gain
Mean	21.64	26.26	4.62
Sd	6.45	6.43	4.08
t(33) = 5.81; p<0.05			

Table 2: Word knowledge results: 45-item multiple choice test, n=34.

Performance on the word association test also improved significantly. Before reading the text, the subjects made an average of 5.53 correct associations (of a possible 13). The post-test figure was 6.71 indicating a gain of 1.18 associations or about 16 per cent (see Table 3). In fact, this fairly modest difference is more substantial than it appears since each item reflects knowledge of three of the targeted *Mayor of Casterbridge* words. Also, it represents a different, possibly more complex type of word knowledge than recognition of a correct multiple-choice definition.

	Pre-test	Post-test	Mean gain
Mean	5.53	6.71	1.28
Sd	2.22	2.22	2.33
t(33) = 2.95; p<0.05

Table 3: Word knowledge results: 13-item word association test, n=34.

These findings offer conclusive evidence that small but substantial amounts of incidental vocabulary learning can occur as a result of reading a simplified novel. As expected, a longer reading treatment produced more evidence of word learning – more seeds were planted and more daisies blossomed. The main difference between this study and its predecessors is the higher pick-up rate; more seeds than usual sprouted in this particular plot. The study now goes on to consider possible explanations for this growth, that is, to investigate whether frequent encounters and vocabulary size can help account for why it occurred.

THE TEXT FREQUENCY FACTOR

To examine the relationship between the number of times a word appeared in *The Mayor of Casterbridge* and the extent to which that word was learned through reading, each of the 45 words in the experiment was assigned a frequency rating and a learning gain score. Frequency ratings, which were determined by the computer analysis discussed above, ranged from 2 to 17 occurrences.

A particular word was considered to have become better known if more subjects could identify its meaning on the multiple-choice post-test than had been able to on the pre-test. Pre- and post-test differences are shown in the absolute gain column in Table 4, where *magistrate* would appear to be the most learned word with a gain of 12 correct identifications, and *furmity* a close second with 11. But absolute gains do not take into account the fact that words varied in the extent to which they were already known to the subjects. In the case of the unusual word *furmity*, for instance, the low pre-test score indicates that there was a large pool of subjects who could possibly learn it, but *magistrate* turns out to have been known to more of the group, leaving a more limited space for new growth.

To take varying opportunities for growth into account, a relative gain percentage was calculated according to a method devised by Shefelbine (1990). For each word,

relative gain was determined by expressing absolute gain in terms of the word's availability for learning in the group of 34 subjects. The following formula was used:

$$\text{Gain} = [(\text{post} - \text{pre})/(34 - \text{pre})] \times 100$$

The formula's ability to capture growth in a way that absolute gains cannot is illustrated by the results for the words *lean* and *trade*. In the case of *lean*, there were nine correct responses on the pre-test and 15 on the post-test, so the word registered an absolute gain of six additional correct identifications. The word *trade* also registered a gain of six with a pre-test score of 23 and a post-test score of 29; so in absolute terms, growth on these two items is identical.

But when these absolute gains are considered in terms of the growth that was possible in the group, a very different picture emerges. In the instance of *lean*, only nine of the 34 subjects had identified it correctly on the pretest, so there remained a rather large group of 25 who had not, and the gain of six amounts to improvement in only a quarter of the cases where change could have occurred (6/25 x 100 = 24%). In the case of *trade*, however, there were 11 subjects who could not identify it on the pre-test, so the gain of six indicates a greater change; i.e. there was improvement in over half of the cases where change was possible (6/11 x 100 = 55%). Relative gain percentages for each of the 45 items are listed in the fifth column of Table 4 below. Figures must be seen as approximate as there was a role for guesswork in both pre- and post-test scores for an item.

The correlation between the number of times each word occurred in the book and relative learning gains was found to be 0.49 (cf. 0.34 in Saragi et al, 1978). This confirms a role for frequency of occurrence in the text in incidental learning of second language vocabulary but it also shows that other factors are involved.

Generally, the text frequency data suggest that sizable learning gains can be expected to occur consistently for items that are repeated eight times or more. With fewer than eight repetitions, growth is much less predictable and the role of other factors becomes more apparent. There was a large amount of variation in gains on words that were repeated five times or less, including some instances of negative gains. The negative gain figures are based on small pre-post differences, and may not be very meaningful given the role for guesswork in the data (although forgetting or unlearning of words is possible, of course).

The word that turns out to have been most learned was *ma'am* (relative gain = 100%), a word that occurred in the text 17 times. High text frequency also coincided with high learning gains in the cases of *wheat* (12 occurrences) and *trade* (8 occurrences), which were both learned in over half of the instances where learning was possible. High scorers *grave* and *magistrate* do not stand out for their frequent

	M of C freq	pre-test	post-test	absolute gain	relative gain (%)	gen freq (H'marsh)
ma'am	17	25	34	9	100	5
hay	17	16	22	6	33	4
furmity	12	5	16	11	38	8
wheat	12	25	30	5	56	3
whisper	11	19	24	5	33	3
trade(n.)	8	23	29	6	55	4
grain	7	13	17	4	19	3
witness	7	13	16	3	14	4
skimmity	6	5	10	5	17	8
stare	6	14	18	4	20	4
maid	6	13	17	4	19	4
burst	6	11	15	4	17	4
entirely	6	10	13	3	13	3
dusk	6	6	9	3	15	5
treat(v.)	6	20	21	1	7	5
relative(n.)	6	32	33	1	50	4
magistrate	5	11	23	12	52	8
awkward	5	13	21	8	38	5
sorrow	5	17	25	8	47	5
suffer(ing)	5	19	26	7	47	3
attempt	5	9	16	7	28	8
lean (v.)	5	9	15	6	24	4
affair	5	10	14	4	17	3
grave	5	29	33	4	80	4
folk	5	12	15	3	14	8
inquire	5	16	17	1	6	4
willing	5	10	9	-1	-4	3
confuse	5	30	29	-1	-25	4
procession	5	6	4	-2	-7	3
harvest	5	12	10	-2	-9	3
widow	4	22	29	7	58	8
carriage	4	22	26	4	33	4
dull	4	14	17	3	15	5
cheek	4	25	27	2	22	4
ancient	4	26	28	2	25	4
wealth	4	31	32	1	33	3
weary	4	8	9	1	4	8
fellow	4	18	18	0	0	8
image	4	10	9	-1	-4	4
effect	3	11	17	6	26	4
companion	3	15	19	4	21	4
swear	3	17	20	3	18	6
flame	3	14	16	2	10	4
expression	2	18	19	1	6	4
root	2	26	23	-3	-38	3

Table 4: Text frequencies, learning scores, general frequencies

occurrence in the text (5 times), but both were pictured in the book, which may have made them salient for learning. Brown (1993) found such "conceptual gap filling" — the matching of a written form to a previously encountered visual concept — to have a powerful influence on incidental word learning in her study of video and text input. Many of the words with high gain scores can be categorized as concrete nouns, and this may have contributed to their learnability as Sternberg (1987b) found in his work with intentional learning from context.

The Overall Frequency Factor

To see if overall frequency in the language was a predictor of incidental word learning, each word was also assigned a general frequency rating (see Table 4 above). This was a level number in a scheme by Hindmarsh (1980) that identifies seven frequency levels: words in the 2200-most-frequent category are rated Level 1, less frequent words Level 2, and so on. *Mayor of Casterbridge* words that could not be found on the Hindmarsh list were assumed to be less frequent than the top Level 7 and were assigned a frequency rating of 8.

One might expect that common words like *willing* and *harvest*, which are among the 2200 most frequent words of English according to the Hindmarsh scheme, would probably have been encountered often enough in other parts of the subjects' English coursework for the additional exposure in *The Mayor of Casterbridge* (5 times each) to have pushed them over the edge into the 'known' category. But these items, both of which were unknown to most of the group at the beginning of the experiment, failed to register learning gains, suggesting that high frequency words were not necessarily learned more readily. Statistical analysis confirmed this impression; the Pearson product-moment coefficient for overall frequency ratings and word learning gains was found to be 0.14. It seems likely that these learners have not had enough general exposure to English language input for repetition effects to accumulate and bring high-frequency words to the verge of being known.

So far, this study has looked mainly at words and text. Findings suggest that learners are more likely to pick up words that are repeated often in a text but frequency in the language does not appear to be a relevant factor. The investigation now turns to the learners themselves.

The Rich Get Richer?

It was predicted that subjects who knew more words generally would find it easier to understand the text and learn new words from it than subjects with smaller vocabulary sizes. To investigate the relationship between individual vocabulary size and learning gains, a relative gain percentage was calculated for each subject. This was expressed as the difference between pre- and post-test scores on the 45-item multiple-choice instrument, following the method used earlier to determine gains for words. The formula was as follows:

$$\text{Gain} = [(\text{post} - \text{pre})/(45 - \text{pre})] \times 100$$

Again, this captures growth in a way that absolute gain scores cannot, as the case of two subjects who both registered gains of six words illustrates. For one, the pre-test showed that 17 of the 45 words were not known to her, so a gain of six indicates that she learned about one third ($6/17 \times 100 = 35\%$) of the words that were available for her to learn. For the other whose pre-test score indicated that 31 words were unknown, the room for growth was larger so the gain of six is less impressive; there was improvement in only one fifth of the instances where improvement could have occurred ($6/31 \times 100 = 19\%$).

The Pearson product-moment coefficient for the correlation between relative gains and scores on the 2000-level vocabulary size test was 0.31. The correlation to scores on the 5000-level test amounted to 0.36. These figures suggest that prior vocabulary knowledge played a role in facilitating incidental acquisition of new vocabulary, but the relationship was not strong. Nonetheless, there was a tendency in the data towards larger and more consistent incidental learning gains for subjects who scored 1444 or higher on the 2000 level vocabulary-size test.

DISCUSSION

The notion of a connection between the number of words in a text already known to a learner and the number of new words that he or she will be able to pick up remains compelling despite the lack of strong evidence for it in this experiment. One reason why more convincing support failed to emerge may be that measures were not sufficiently sensitive. It seems likely that the Levels Test (Nation 1990) did not provide very precise information about words in the *The Mayor of Casterbridge* that were known to the students, and that the 45-item multiple-choice test did not offer the opportunity to demonstrate all the incidental growth that had actually taken place. If subjects had been tested on all the words in the novel, or at least a much greater sample of them, the relationship between what was previously known and what they were able to learn through reading might have become more clear.

But, as Meara (1997) points out, it is probably wrong to expect a linear relationship between the two variables because of the ways particular readers interact with particular texts. If a learner's vocabulary is small, he or she may simply not know enough of the words in the text to be able to infer the meanings of unfamiliar words. However, if a learner's vocabulary is large, learning gains may also be small because there are few new words available in the text to learn. Incidental uptake may also be low in learners with large vocabularies due to an effect Mondria and Wit-de Boer (1991) have observed: when surrounding contexts are easy to understand, new words are often not noticed (and hence not learned).

In this study of low proficiency learners, growth seems more likely to have been limited by knowing too few words than by knowing too many. Laufer (1982, 1989) claims that readers need a sight recognition of at least 95 percent of the words in a text for it to be comprehensible enough for meanings of unknown words to be inferred. It is difficult to quantify precisely what proportion of words was known to the subjects as they read the simplified text, but their mean score of 1200 on the test of the 2000 most frequent words suggests that it may have been below Laufer's 95 percent criterion. Even though the list of the 2000 most common word families Nation used to devise the Levels Test (Nation 1990) may not entirely coincide with the list of 2000 high frequency basewords used to write the simplified novel, there is probably enough overlap to safely conclude that the match of text and reader was less than perfect in the study, at least for the purposes of incidental vocabulary acquisition.

CONCLUSIONS

One way of further improving the methodology of this type of study would be to test much larger numbers of potentially learnable words in order to ensure that subjects have ample opportunity to demonstrate incidental gains. Studies testing learners on their knowledge of all (or most) of the words that occur in texts before and after they read them may be able to specify an accurate method for predicting incidental growth, so that learners can be matched to texts for maximum word learning effect. Since such research calls for extensive testing and careful control over the reading of lengthy texts, it may be difficult to implement with large groups of subjects. A more feasible alternative might be the case-study approach advocated by Meara (1995).

As far as implications for vocabulary learning are concerned, the experiment makes a stronger case for incidental acquisition than was made in the earlier *Clockwork Orange* replication studies. Subjects who read a full-length book recognized the meanings of new words at a higher rate than in previous studies with shorter texts, and built associations between new words as well. Unlike the same-day findings of earlier experiments, these vocabulary learning results represent knowledge that accumulated and persisted over a period of ten days. It seems likely that other vocabulary learning benefits accrued. For instance, a number of untested words were probably also learned (or partially learned) through exposure to *The Mayor of Casterbridge*, and the quality of vocabulary learning that occurred seems likely to have been high. Cobb (1997) found that encountering new words in multiple contexts resulted in a deeper, more transferrable knowledge of words than the usual strategy of studying short definitions.

The study also points to some ways of enhancing opportunities for vocabulary learning through reading. One area where improvement is possible is text

construction. It was found that frequently repeating items helped ensure that they were picked up, but in *The Mayor of Casterbridge* only six of the words likely to be new to subjects were repeated the eight or more times that turned out to produce sizable and consistent learning gains. Wodinsky and Nation (1988) found a similar lack of repetitions in their analysis of two other simplified novels. To address this, editors or writers of simplified novels can use frequency analysis tools to identify words that already recur in a text. Then additional repetitions can be written in, not so many that the integrity of the text is destroyed, but enough to make more words learnable. This would need to be guided by good information about which words are worth giving this kind of attention, for example the Hindmarsh list (1980) or West's *General Service List* (1953).

Learners can also get the multiple exposures they need through direct vocabulary instruction that complements reading assignments. Texts in computer readable format are increasingly available to teachers, as are frequency analysis tools like *Wordsmith (Scott* 1996). This makes it a simple matter to find out which words will be encountered in a text and how often, and to design vocabulary reinforcement activities accordingly. Or, frequency tools can be placed directly in the hands of learners to useful effect, as Cobb (1997) found in a study of learner concordancing.

But even though it may be possible to develop better resources for incidental learning, the study suggests that extensive reading is not a very effective way for learners who have a mean vocabulary size of around 3000 words to expand their lexicons. After completing the whole 21,000-word book, the subjects in the experiment managed to recognize meanings of an average of only five new words and to make new associations between just three. Also, learning was never fully guaranteed; even with items that occurred eight times or more, gains averaged around 50 percent. In other words, after reading an entire novel and encountering a word many times, only half of the learners who did not already know the word were able to recognize a correct definition in a multiple choice format. In brief, the experiment indicates that teachers of low intermediate learners of English can expect vocabulary growth from reading a simplified novel to be small and far from universal.

In the last two decades, it has often been assumed that incidental acquisition was a sufficient strategy to take care of learner's lexical needs, to the point that explicit vocabulary instruction effectively disappeared from many coursebooks and vocabulary acquisition became "a neglected aspect of language learning" (Meara 1980:221). The present study suggests that the power of incidental acquisition may have been overestimated. The findings support Meara's (1988) argument that since reading in a second language takes a great deal of time, few learners are able to read in sufficient volume to make it the vocabulary enriching experience it has proved to be for first language learners. Nagy, Herman and Anderson (1985) propose that for children learning English as their first language, school reading can account

for the acquisition of thousands of new words each year. Even though the incidental pick-up rate was found to be low, large gains occur, they argue, because children encounter millions of words annually. But this is hardly applicable to beginning second language learners; for the subjects of this study, encountering one million words would entail reading fifty graded readers the size of *The Mayor of Casterbridge* — a worthy but unattainable goal for most learners at this level.

Assuming an optimistic scenario in which reading fifty novels per year was possible, at the rate of five words per novel established in this study, annual gain would amount to only 250 words. At this rate, even if yearly gains increased marginally with increased vocabulary size, it would take many years to acquire incidentally the 5,000 words most frequent word families of English, the figure which has been proposed as the minimum knowledge base needed for learners of English to be able to infer the meanings of new words they encounter in normal, unsimplified texts (Hirsh & Nation 1992, Laufer 1989).

Since most learners have a limited amount of time to devote to second language acquisition, vocabulary growth needs to proceed more rapidly. For learners at the level of the subjects in this experiment, it seems likely that an efficient way to reach the point of lexical independence is through explicit and systematic instruction that focuses on high-frequency vocabulary, a recommendation made repeatedly by Nation (1990). That is not to say that low intermediate learners should never read, but that teaching decisions should be based on an adequate account of what they can gain from their reading. Through reading extensively, they will probably enrich their knowledge of the words they already know, increase lexical access speeds, build network linkages between words, and more, but as this study has shown, only a few new words will be acquired. Therefore, it seems clear that in the early stages of their second language acquisition, learners should direct a considerable portion of their energies to using intentional strategies to learn high frequency vocabulary, in preparation for the day when they will know enough words and can read in enough volume for more substantial incidental benefits to accrue.

In the final analysis, the really interesting question is not whether a small amount of incidental acquisition can be detected, but whether the relative importance of intentional and incidental vocabulary learning strategies can be established for different stages of the language learning process. Determining the point at which the former should give way to the latter remains a challenge for second language reading research.

REFERENCES

Brown, C. (1993). Factors affecting the acquisition of vocabulary: Frequency and saliency of words. In T. Huckin, M. Haynes, & J. Coady (Eds.) *Second language reading and vocabulary learning*. Norwood, NJ: Ablex.

Burgess, A. (1972). *A Clockwork Orange.* Hammondsworth: Penguin.

Cobb, T (1994). *EspritDeCorpus: Computer assisted corpus builder.* Software. Muscat: Sultan Qaboos University.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System,* 25(3), 301-315.

Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language,* 7(2), 541-551.

Dupuy, B., & Krashen, S. (1993). Incidental vocabulary acquisition in French as a foreign language. *Applied Language Learning,* 4(1&2), 55-63.

Hindmarsh, R. (1980). *Cambridge English lexicon.* Cambridge: Cambridge University Press.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 689-696.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud & H. Bejoint (Eds.)*Vocabulary and applied linguistics.* London: MacMillan.

Jones, L. (1979) Simplified version of T. Hardy's *The Mayor of Casterbridge,* 2000 basewords. Walton-on-Thames: Nelson.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *Modern Language Journal,* 73, 440-464.

Laufer, B. (1989). What percentage of lexis is necessary for comprehension? In C. Lauren & M. Nordman (Eds.) *From humans to thinking machines.* Clevedon: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Bejoint (Eds.) *Vocabulary and applied linguistics.* London: MacMillan.

Meara, P. (1980). Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching & Linguistics Abstracts*, 13(1), 221-247.

Meara, P. (1995). Single-study subjects of lexical acquisition. *Second Language Research,* 11(2), i-iii.

Meara, P. (1988). Learning words in an L1 and an L2. *Polyglot,* 9(3), D1-E14.

Meara, P. (1997). Models of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.) *Vocabulary: Description, acquisition, and pedagogy.* Cambridge: Cambridge University Press.

Mondria, J. A., & Wit-de Boer, M. (1991). The effects of contextual richness on the guessability and retention of words in a foreign language. *Applied Linguistics,* 12(3), 249-267.

Nagy, W. E., Herman, P.A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly,* 20(2), 233-253.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Boston: Heinle & Heinle.

Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the Clockwork Orange study using second language acquirers. *Reading in a Foreign Language,* 5(2), 271-275.

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing,* 10(3), 355-371.

Saragi, T., Nation, P., & Meister, G. (1978). Vocabulary learning and reading. *System,* 6, 72-80.

Scott, M. (1996). *Wordsmith.* Software. (Accessible at http:// www.liv.ac.uk / ~ms2928 / worsmit.html/).

Shefelbine, J. L., (1990). Student factors related to variability in learning meanings from context. *Journal of Reading Behavior,* 22(1), 71-97.

Sternberg, R. J. (1987a). Most vocabulary is learned from context. In M. G. McKeown & M. E. Curtis (Eds.) *The nature of vocabulary acquisition* . Hillsdale, NJ.: Erlbaum.

Sternberg, R. J. (1987b). The psychology of verbal comprehension. In R. Glaser (Ed.) *Advances in instructional psychology.* Hillsdale, NJ: Erlbaum.

University of Cambridge (1990). *Preliminary English Test.* Local examinations syndicate: International examinations.

Vives, G. (1995). *The development of a measure of lexical organisation: the association vocabulary test.* Unpublished doctoral thesis, University of Wales, Swansea.

West, M. (1953). *A general service list of English words.* London: Longman, Green and Co.

West, R. F., & Stanovich, K. E. (1991). The incidental acquisition of information from reading. *Psychological Science,* 2(5), 325-329.

Wodinsky, M., & Nation, P. (1988). Learning from graded readers. *Reading in a foreign language,* 5(1), 155-161.