Frequency and the lexis of low level EFL texts

JAMES MILTON & PANOS VASSILIU

University of Wales Swansea, UK

ABSTRACT
Computer-based Yes/No tests that estimate vocabulary size work well at intermediate and higher levels but appear to underestimate the true vocabulary of low level learners. It is thought this may be due to the thematic content of elementary teaching books which require the use of infrequent vocabulary not included in frequency-based tests. An examination of 3 beginner EFL texts confirms that they do include large quantities of highly infrequent vocabulary and the selection of this vocabulary is idiosyncratic. As a result, a test based on the most frequent lexis according to general frequency lists may test less than half of the vocabulary to which learners have been exposed. The delivery of two tests, one based on frequency and the other on course book vocabulary, to low level learners confirmed that frequency-based tests can underestimate learners’ vocabulary knowledge by more than 50%.

1. INTRODUCTION
The work described in this paper derives from two sources. One is the work carried out at the university of Swansea by the Vocabulary Research Group. This has a number of areas of interest under the general umbrella of vocabulary research, and one of these is the development of tests of vocabulary knowledge and the estimation of vocabulary size. Such tests have been based, amongst other things, on word frequency counts and while they appear to work successfully at intermediate level and above, they are not so successful when used with low level learners. The second lies in the European funded project VIIRLAN which involves the development of internet-based communicative tools for low level language learners. In order that such tools could be accessible by the widest number of users, the language and specifically the lexis of low-level books was examined and the internet materials were designed to match the language
which these texts yielded. The unusual characteristics of the low level texts may suggest a reason why tests based on frequency counts do not work well amongst young beginners. This possibility is tested in this paper.

2. ESTIMATING VOCABULARY SIZE

In 1988 Meara and Jones published a vocabulary-based placement test which was different from the usual tests of this type. Rather than using the traditional Multiple Choice testing method, they adopted a Yes/No method based on signal detection theory. It is a word recognition task in which learners are presented with a test word and asked whether or not they recognise the word. Learners answer Yes if it is recognised and No if it is not recognised. A task of this kind has a number of benefits over Multiple Choice testing, i.e. it tests only the word that is wanted. On the other hand, in Multiple Choice tests, the learners are also tested on all the words and structures that appear in the context of the text word as well as on the distractors. This may challenge the validity of the test. Yes/No tests are also quicker, they can test a large number of words in a short time span, and this should mean that any conclusions drawn from the results of such a test are more reliable. A fair question at this point would be: “What if the testee pretends to know all of the words, or guesses at unknown words?” This is where signal detection theory is applied. One third of the words tested are not real English words but are pseudo-words, designed to resemble real words. The purpose of these false words is “to estimate how far the testee is just guessing when he claims to know one of the real words” (Meara 1990:83). A preliminary estimate of words known within a field can be gained from the score on the real words; this can then be adjusted according to the amount of guessing. Research in the area of Yes/No tests (Buxton 1986, Anderson and Freebody 1983) has found that there is a satisfactory correlation between results of Yes/No tests and Multiple Choice tests.

By testing a learner’s knowledge of vocabulary across a sample of the first ten thousand word frequency bands, this test was able to derive a profile of the learner’s knowledge and so make an estimation of what appears to be the vocabulary size. This is possible because vocabulary learning in a foreign language appears to follow a broadly predictable course where the more frequently a word occurs in a language, the earlier it is likely to be learned. Estimates based on the most frequently occurring vocabulary in a language should, in theory, enable a good estimate of overall vocabulary size to be made.

Such a theory ought to imply that highly frequent vocabulary tends to be learned first although this hypothesis was not testable using Meara and Jones’s Eurocentre’s Vocabulary Size Test (1988). Meara and Jones (1990) acknowl-
edge that the test will not work well at low level but presume the problem is one of sampling. Although Yes/No methodology can test a comparatively large sample, if the learner’s knowledge is limited, then even that sample may not be adequate to accurately estimate the learner’s knowledge.

For this reason, some attempts were made to overcome this problem and test more frequent vocabulary more intensively. One is by the design of the LLEX test (Meara 1995), which is based on the Threshold level materials (Van Ek and Trim 1990). This seems to have certain advantages but does not permit an estimate of overall vocabulary size. Another attempt has been made by Al-Hazemi (1993) who explicitly attempts to assess the vocabulary level of his low level students using test vocabulary from the first 1000 most frequent vocabulary items in Nation’s (1984) Word Lists. The results he got, however, were not those that he expected. The scores were very low indeed and thus he believed they underestimated the vocabulary knowledge of his students.

3. THE QUESTIONS ADDRESSED IN THIS PAPER

Were the tests simply inaccurate, or do the results relate to the model of learning the tests are based on? Do students depart from the model of learning as presented by Meara (1992), that is, acquiring more frequent vocabulary prior to infrequent? It must be remembered that learners acquire their knowledge from course books and not from frequency lists. Beginners’ course books ought to contain infrequent vocabulary to provide content, so the theory that learners will learn exclusively the most frequent vocabulary first is at least questionable. This would call into question the construct validity of Al-Hazemi’s tests which may explain his poor results. This is an area which deserves investigation. This research aims to:

- investigate the lexis of low level course books and test the hypothesis that such books contain large amounts of infrequent vocabulary.
- investigate the vocabulary knowledge of low level learners and test the hypothesis that this knowledge is not fully tested by frequency-based vocabulary tests.

4. THE LEXICAL CONTENT OF LOW LEVEL TEXTS

There are good reasons to assume that the lexical content of beginner course books is very different from the material frequency lists derive from. Frequency lists are typically taken from adult materials such as newspapers, novels and spoken corpora. This happens because such lists do not want to focus on a narrow group of language users, and, therefore, include a variety of text types. Typically, however, the text types they favour are drawn from areas of adult interest such as politics, finance, literature and work.
Young learners, on the other hand, have entirely different interests. They are not so concerned with politics or finance, or with stories from novels, and beginner texts attempt to reflect this difference. A review of 16 beginner texts undertaken for VIRLAN revealed that the most frequently occurring themes included animals, the home, celebrations, body parts, museums and dinosaurs. These texts included vocabulary items which one might not otherwise expect to encounter. Vocabulary such as lion, kangaroo, platypus and acrobat are examples. The language of instruction itself is also a feature of these texts. This includes vocabulary items such as read, write, listen and speak which would be quite frequent, but also tick, draw and mark which are not.

It must be remembered too that frequency lists are not conveniently organised according to themes or ideas, and if a teaching text is to have any coherent thematic and linguistic content then the vocabulary will have to be selected from a range of frequency bands. A beginner unit which focuses on animals, for example, could be expected to include highly frequent lexis such as articles and the verb to be, but also infrequent lexis in the names of the animals themselves, such as lion or even duck billed platypus. It is plausible, therefore, that for these reasons early learners may not follow the frequency model but may be exposed to and learn highly infrequent vocabulary.

In order to examine and quantify the effect which this range of frequency has on vocabulary learning among beginners, three course books, currently used for instruction in Greece to first year EFL students, were examined in detail. The books were Sam on Channel 9 (Webster and Cobb 1983), referred to as textbook A, Splash! 2 (Abbs, Warrall and Ward 1993), referred to as textbook B, and Round up (Pagoulatou-Vlachou 1994), referred to as textbook C. The contents of these books were scanned and examined with a concordancer in order to see the vocabulary burden of each book separately, and of all the three books together. The vocabulary content and word frequencies could then be compared between the three books and with results drawn from general corpora.

This exercise produced a corpus of 27,517 words and 2,040 unlemmatised types. Textbook A contained 909 types, textbook B contained 964 types, and textbook C contained 1175 types. This material included, as might be expected, personal names and other material not relevant to this study and these were extracted from the corpus reducing it to 1710 types. This is a very large figure indeed bearing in mind Milton and Meara’s (1998) finding which suggests that about 4 words per contact hour are learned during language tuition. An average learner in this study, therefore, might be expected to learn approximately 400 words.
A comparison of the lexical content of the three books reveals the amount of common vocabulary. Since they were all designed for the same level, and writers at this level acknowledge the need to include the most frequent words in a language (for example, Rixon 1990:5), it might be expected that a large proportion of the content would be the same. Comparison shows that there was relatively a small amount of common vocabulary. This can be seen in Figure 1 where only 328 types appear to be common in all the three course books. Over half of all types in the corpus - 456 in text A, 239 in text B, and 302 in text C - appear in one text only.

![Venn Diagram](image)

Figure 1. Vocabulary common to texts A, B and C

It would appear that the choice of vocabulary content in these books is highly idiosyncratic. This may be characteristic of texts at this level since it has been noticed in other studies of beginner foreign language teaching texts. In Milton and Vincent Benn’s (1933) study of texts for the teaching of French at elementary level, a total of 6,135 different words were used in the 29 course books examined, but only 19 words were common to all the textbooks. Meara (1996) also notices a similar trend when examining the word load encountered in the Spanish elementary courses.

5. THE FREQUENCY OF LEXIS IN LOW LEVEL TEXTS

It has been pointed out that the lexis of beginner texts may include large quantities of words such as those associated with animal and other themes which, according to general frequency lists, would be highly infrequent. It is suggested too that highly frequent words in normal frequency lists, such as those associated with government and politics, are likely to be absent from beginner texts. In
order to test the degree to which the language of text books A, B and C contained unusual vocabulary, the corpus was lemmatised producing a new corpus of 1396 types. This was then compared with frequency information taken from Nation (1984). In this analysis level 0 is structural or functional vocabulary such as articles, pronouns and prepositions. Level 1 is the 1000 most frequently occurring lemmas, excluding structural vocabulary, and thus is the most frequently occurring lexical or content vocabulary. Level 2 is the next 1000 most frequently occurring lemmas, and level 3+ is any lemma not included in the other three levels. The results of this analysis are shown in Table 1.

Table 1. Low level text vocabulary content divided by frequency range.

<table>
<thead>
<tr>
<th>Nation's levels</th>
<th>level 0</th>
<th>level 1</th>
<th>level 2</th>
<th>level 3+</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of lemmas</td>
<td>8</td>
<td>40</td>
<td>6</td>
<td>46</td>
</tr>
<tr>
<td>no. of lemmas</td>
<td>105</td>
<td>558</td>
<td>84</td>
<td>649</td>
</tr>
</tbody>
</table>

It will be seen from this data that the beginner EFL materials examined contain very large percentages of infrequent vocabulary. Some 649, or 46%, of the lemmas in this corpus fall outside level 0, 1 and 2 and this appears a very high proportion. It is the largest of the four groups presented here. However, the percentage in Nation's level 1 is also large, but it is marginally smaller than level 3+. Despite the high levels of infrequent vocabulary, the texts examined also display, as might be expected, a tendency to include a large proportion of the most frequent lexical vocabulary.

The percentage in Nation's level 0 appears very small given that this group contains by far the most frequently occurring lemmas in English. However, this might be explained by a beginner text's predisposition to a limited range of short and simple structures. Once again, this finding is not unique. Milton and Vincent Benn (1933) noted a similar trend in the French elementary texts they examined and conclude that a relatively small amount of words taught belong to the first 1000 most common vocabulary which in their analysis would include structural and functional lexis. This finding supports the idea that young EFL learners are exposed to significant proportions of infrequent vocabulary at the very outset of learning.

It is worth noting that a very high proportion of infrequent vocabulary in a text is often considered a characteristic of specialised and technical text (Milton and Hales 1996:72). Milton and Hales demonstrate that specialised texts, such as car manuals, have proportionately high levels of infrequent vocabulary (Nation level 3+) and low levels of frequent vocabulary (Nation level 0 and 1).
when compared with unspecialised text types as illustrated by corpora taken from Marie Claire and The Guardian. This profile technique draws results from 1000 word samples of running text and calculates the proportion of each text which falls into nations levels 0, 1, 2 and 3+. Could it be that the language of low level foreign language course books is a type of specialised or technical text like the language of car mechanics? In order to test this a vocabulary profile for the low level texts, drawn from a standard length 1000 word sample of running text, was examined and compared with samples of other types of text drawn from Milton and Hales (1996:79). The results are shown in Table 2.

This table is reassuring in certain ways in that it appears that the course books are not so very odd after all. The proportion of structural vocabulary, Nation’s level 0, in the course books which is 22% is clearly more like the unspecialised text types such as Marie Claire magazine and The Guardian, rather than the highly technical car manual. While the range of structures may be limited in elementary texts, they must be frequently repeated to give this sort of figure. More than half the vocabulary in the course books’ running text is from the most frequent band of lexical vocabulary, Nation level 1. This is slightly higher than Marie Claire, which was thought to be at the easy end of public reading and much higher than either The Guardian or the car manuals. This finding suggests that while the number of lemmas found in beginners’ texts is unexpectedly low, each lemma is more often repeated than in texts such as Marie Claire and The Guardian. This might be expected in a teaching text.

<table>
<thead>
<tr>
<th></th>
<th>level 0</th>
<th>level 1</th>
<th>level 2</th>
<th>level 3+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course books</td>
<td>22</td>
<td>54</td>
<td>3</td>
<td>20</td>
</tr>
<tr>
<td>Marie Claire</td>
<td>25</td>
<td>50</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>The Guardian</td>
<td>21</td>
<td>38</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td>Car Manual</td>
<td>10</td>
<td>35</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

Nation’s level 2, lexical vocabulary in the frequency range 1001 to 2000, is almost non-existent in course books and much lower than any of the other three text types. While this might be expected in a beginner text, it is interesting to note that many learners display a vocabulary profile with a marked dip in knowledge at this level. It may be that this level of vocabulary is under-represented in text books. The figures suggest that the words from this group are repeated less often within a text than the words from Nation’s levels 0 and 1.

The least frequent vocabulary in the course books, level 3+, is on a par with The Guardian which is thought to be at the difficult end of general public read-
ing. However, it is not nearly so high as the proportion of this vocabulary in the car manuals. This must mean that in beginners’ texts the infrequent lemmas are repeated less often within the text than are lemmas of the most frequent lexical levels. It is worth considering the possibility that the infrequent vocabulary may be there to provide thematic content and is not considered by the writers a major object for learning. This seems doubtful, however, especially in the light of Milton and Benn’s findings that in course books all the vocabulary provided is to be learned. Nonetheless, with vocabulary from Nation’s level 1 repeated by the text so much more often than infrequent vocabulary, better learning in this area would be expected. Perhaps a frequency based test of vocabulary knowledge is not so bad an idea after all.

6. THE ACCURACY OF LOW LEVEL VOCABULARY TESTS

The previous section has shown clearly that elementary course books are highly idiosyncratic in their choice of vocabulary. Amid such diversity, how is the vocabulary knowledge of low level learners to be assessed? It would seem that no single test could hope to provide an accurate assessment. Any single test is likely to include words not covered by the learners and omit words they have been taught. It is likely to be inaccurate in measuring vocabulary size as a result. This may well be what is happening with Al-Hazemi’s low level test.

The previous section has also shown that low level texts display high proportions of infrequent vocabulary. This vocabulary will not be present in tests selected only from the more frequent vocabulary bands such as Al-Hazemi’s (1993) low level test. He may well have been correct, therefore, in thinking that his test considerably underestimated his learners’ knowledge since this conclusion suggests his students had knowledge which was not tested and so was not included in his estimates. A comparison of Al-Hazemi’s test corpus with the corpus described in this study reveals the potential scale of the problem. The two corpora are compared in Figure 2.

It would seem from this data that if Al-Hazemi’s low level tests were administered to a group taught by the books examined in this study, then they would test less than half the vocabulary to which the learners were exposed. Only 558 of the 1396 lemmas taught would be tested, almost 40%. This finding does not, of course, challenge the reliability or concurrent validity of the test he devised. Both of these things he examined quite rigorously. However, as Al-Hazemi himself concluded, the test may fail in its intention to measure vocabulary size. Much will depend on whether the greater repetition amongst the items in Nation’s level 1 vocabulary will result in significantly better learning in this area. The vocabulary not included in Al-Hazemi’s test, infrequent vocabulary, was not so often repeated and may not have been learned anyway.
Figure 2. Comparison of textbook corpus with Nation's 1000 word list.

In order to examine the effect this difference will have on the assessment of vocabulary uptake by elementary learners, the learners who used the books examined in this study were tested. Two sets of tests were administered. The first was composed of tests based on a random selection of words from the corpus that the students had been taught. The second were those devised by Al-Hazemi and based on words taken from the first 1000 most frequent vocabulary of Nation's (1984) Word Lists. Both were Yes/No tests of the type described at the beginning of this paper. Both tests were administered at the end of the teaching year when the learners had completed the work in their course books.

It will be seen from Table 3 that the learners scored higher on the test based on the vocabulary of their course books than on Al-Hazemi’s frequency based test. When these figures are extrapolated to arrive at an estimate of vocabulary size, the course book test suggests a figure which is more than double that suggested by Al-Hazemi’s test. The vocabulary size figure of between 700 and 800 words appears high in comparison with other estimates of annual progress (Milton and Meara 1998). The scale of the difference between the two scores also suggests that the increased exposure of the learners to the vocabulary in Al-Hazemi’s test has not significantly benefited learning. The difference between the two test scores is best explained by the conclusion that the learners have gained knowledge of the infrequent vocabulary which was excluded from Al-Hazemi’s test. Interestingly, there is a reasonably good correlation between the scores on the two tests ($r = 0.657$, significant to 0.01) and this suggests that even if a frequency-based test underestimates knowledge, it is still a good indicator of comparative knowledge. It may be underestimating but at least it is underestimating quite consistently.
Table 3. Scores on vocabulary size tests.

<table>
<thead>
<tr>
<th>Test</th>
<th>Target vocab</th>
<th>Mean score</th>
<th>Standard deviation</th>
<th>Vocabulary knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course book test</td>
<td>1396</td>
<td>53.45</td>
<td>16.7</td>
<td>746</td>
</tr>
<tr>
<td>Al-Hazemi's test</td>
<td>1000</td>
<td>34.72</td>
<td>15.35</td>
<td>347</td>
</tr>
</tbody>
</table>

7. CONCLUSION

Two broad hypotheses were outlined for examination in this paper. One is that elementary foreign language text books will, contrary to theory and even the opinion of text book writers, include large quantities of infrequent vocabulary. From the evidence of the three texts examined in this paper this would seem to be the case. In the list of lemmas derived from the course books, nearly half fell in or below the three thousand word frequency band. It appeared too that text book writers are highly idiosyncratic in their choice of infrequent vocabulary and there is no common core of infrequent words for use by beginners at the outset of learning. These conclusions are not new and they mirror almost exactly the findings made by Milton and Vincent Benn (1933) nearly seventy years ago.

It is possible to characterise the lexical structure of elementary texts in the light of the text books studied here and the findings of Milton and Vincent Benn. Elementary texts do include large quantities of infrequent vocabulary but they also include a similar quantity of the most frequent lexical vocabulary, Nation’s level 1. The number of structural words included appears relatively small but this may be explained by the requirement that a beginner text should exemplify only a small range of simple structures. Between the most frequent and least frequently occurring vocabulary groups there appears to be a gap with words from Nation’s level 2 scarcely represented at all. Reassuringly, structural vocabulary from Nation’s level 0 appears to be the most often repeated within a text, and words from Nation’s level 1 are more often repeated within the texts than are words from groups 2 and 3+. In this beginner texts are not totally different, it would seem, from general frequency lists.

As Milton and Vincent Benn point out, there are very practical concerns which arise out of this diversity. If text books are so different, what happens to learners who change school or change class? Previous learning, it would seem, would be almost useless since the students are not exposed gradually to less frequent material after consolidating the frequent vocabulary. Rather, they may learn an idiosyncratic selection of infrequent vocabulary. It must be admitted that to adhere strictly to frequency criteria in the creation of teaching materials is probably impossible. If teaching materials are to have any thematic content at all...
then writers will have to draw on vocabulary from across the frequency ranges. However, the need to reflect the interests of young learners seems to exaggerate this trend. It seems strange, since these problems were appreciated more than 65 years ago, that we seem to have progressed so little in the construction of courses.

From the testing point of view, the question arises as to how it is possible to derive a single test of overall vocabulary ability when the learners have so very little in common in the subject matter they have learned. The attempt at using frequency based materials is one possible solution. This gives rise to the second broad hypothesis examined in this paper, that frequency based tests at low level are likely to underestimate learners’ actual vocabulary knowledge. A comparison of the vocabulary of Al-Hazemi’s low level vocabulary test with the vocabulary to which first year learners in Greece were exposed showed that the test covered less than half of the vocabulary items which learners had acquired. The administration of Al-Hazemi’s test and of a test based on course book vocabulary showed very different scores. It seems that learners have acquired large amounts of the infrequent vocabulary they were exposed to. While the two sets of scores where consistent with each other and correlated reasonably well, it is concluded that Al-Hazemi’s test may have underestimated learners’ vocabulary knowledge by more than 50%. A model of learning that assumes that almost no infrequent vocabulary will be learned at beginner level is almost certainly misleading, therefore. It seems likely too, that any low level test based on frequency data is likely to underestimate vocabulary knowledge in the same way.

REFERENCES


