7.5

How Large Can a Receptive Vocabulary Be?

ROBIN GOULDEN, PAUL NATION, and JOHN READ

Victoria University of Wellington

Studies of vocabulary size based on dictionary sampling have faced several methodological problems. These problems occur in trying to answer the following three questions: (1) How do we decide what to count as words? (2) How do we choose what words to test? (3) How do we test the chosen words? The present study attempts to overcome these problems and checks in several ways to see if the problems have been overcome. The results indicate that what were previously thought of as conservative estimates of vocabulary size are likely to be the most accurate. These estimates suggest that well-educated adult native speakers of English have a vocabulary of around 17,000 base words. This represents an acquisition rate of around two to three words per day.

1. WHY STUDY VOCABULARY SIZE?

In countries like Great Britain, the United States, Canada, Australia, and New Zealand, learners of English as a second language attend the same schools and universities as native speakers of English and follow the same classes. An obvious and striking difference between these two groups is in the size of their vocabulary. A young immigrant might arrive at school knowing almost no English words at all. A foreign student might enter an English medium university knowing only a few thousand words of English. What is the size of the vocabulary learning task facing such learners? How can they best be helped to manage it? How long will it take them to perform at a level comparable to that of native speakers? Diller (1978) put this last question another way: 'How long does it take to produce a bilingual?'

In one of the few studies in this area, Jamieson (1976) found that the rate of vocabulary increase was similar for five- and seven-year-old native speakers of English and learners of English as a second language who were in the same school system. This meant that the second language learners' rate kept pace with that of their native-speaking peers. However, the second language learners did not bridge the gap between their vocabulary size and that of the native speakers that already existed when they entered the school system.

In a more recent study, Cummins (1981) showed that the gap can be narrowed, but only after a number of years of English-medium education. He analysed the vocabulary test results of 1,210 foreign-born ESL students in the Toronto school system in relation to their age on arrival and length of residence in Canada. On the average, he found that children who had immigrated at the age of six or later took five to seven years to achieve scores that approximated those of native-born students at their grade level. As Cummins points out, this can have serious consequences for immigrant students because of the extent to

Applied Linguistics, Vol. 11, No. 4 © Oxford University Press 1990

which vocabulary knowledge influences performance on verbal IQ tests. These students may be wrongly diagnosed as having low academic potential on the basis of IQ tests administered to them during their first five years of residence in the host country.

Cummins relates his findings to a more general distinction between basic interpersonal communicative skills (BICS) and cognitive-academic language proficiency (CALP). According to this analysis, a major problem in assessing immigrant ESL learners is that they may rapidly acquire fluency in BICS within one or two years of arrival and thus give the impression that they have achieved a level of competence in the language equivalent to that of their native-speaking peers, whereas tests of vocabulary and reading comprehension reveal that they lack the underlying proficiency (or CALP) that is required to cope with academic study through the medium of their second language. Therefore, in this context, measures of vocabulary size—particularly the size of academic vocabulary—are important indicators of the ability of second language learners to achieve academic success.

Another perspective on the effect of inadequate vocabulary knowledge on academic performance is provided by Corson (1983), who suggests that the significance of the gap may be one not so much of size as of the type of vocabulary involved. His studies with native-speaking students of different social class backgrounds show 'a lexical bar' existing between the very frequent, largely monosyllabic, mainly Anglo-Saxon vocabulary of conversational English and the Graeco-Latin vocabulary of the English of academic study. The size and nature of the gap between learners on one side of the lexical bar and those on the other would have a direct influence on the type of help teachers would provide. If the gap was small for a particular group of learners, it might be more effective to encourage the learning of particular sets of words, like the university word list (Xue and Nation 1984). If the gap was large, concentration on vocabulary learning strategies and procedures would be more effective.

Investigations of vocabulary size then are of interest to teachers of English as a second language in that they can provide information about the size of the task facing second language learners and the most suitable ways of dealing with the task.

2. PREVIOUS STUDIES

Investigations of vocabulary size, particularly of native speakers, have been published since early this century. The most notable feature of these investigations is the enormous divergence among the results. The various estimates of the number of words known by college students and adults, for example, range from 3,000 words to 216,000 words (Fries and Traver 1960; Lorge and Chall 1963; Diller 1978).

These results are more striking if they are turned into daily estimates. Diller (1978), for example, estimates that secondary school children learn 20,000 words a year. This works out to be a rate of around 60 words a day! The lower estimates of vocabulary size, for example Diack (1975), convert to around three

to five words a day. Clearly, direct teaching of vocabulary will be more significant if the lower estimates are nearer to the truth. If the higher estimates are more accurate, then other approaches are needed.

These divergences between the estimates are the result of methodological problems involved in measuring vocabulary size (Thorndike 1924; Lorge and Chall 1963). By finding satisfactory solutions it is possible to reinterpret previous studies and develop tests that will give reliable and valid estimates. These problems occur in trying to answer the following three questions:

1 How do we decide what to count as words?
2 How do we choose what words to test?
3 How do we test the chosen words?

3. HOW DO WE DECIDE WHAT TO COUNT AS WORDS?

Studies of the size of receptive vocabulary have been based either on frequency counts or on dictionaries. The pros and cons of these methods have been discussed elsewhere (Lorge and Chall 1963; Nagy and Anderson 1984). In the case of dictionary-based studies, the estimation is done in the following way. First an estimate is made of the number of words in the dictionary. Then a representative sample of these words is drawn and learners are tested on their knowledge of each word. The proportion of words that they know in the test is taken as the proportion of words they know in the whole dictionary. For example, Diller (1978) assumed that the largest Webster's dictionary contained 450,000 entries, which is the figure given in the preface of the dictionary. He selected a sample of 1,000 words to be tested. When he administered the test to high-school seniors, the median score was 480 out of 1,000, or 48 per cent. He thus calculated that their total vocabulary size was 216,000 words: 48 per cent of the 450,000 entries in the dictionary.

For such a calculation to be accurate, it is essential to estimate the number of words in the dictionary in the most reliable and valid way. This involves two types of decision. First, what is a word and what is not a word? For example, will abbreviations, names of persons, geographical place names, other names, prefixes, and suffixes be counted as words or not? Lorge and Chall (1963) showed that over 30 per cent of the basic entries in the Funk and Wagnall (1937) unabridged dictionary were of these types. Second, which words are to be considered as base words and which derived words? Clearly, in an estimate of vocabulary size, regarding *mending*, *mend*, *mends*, *mended* as different words will lead to inflated estimates. Similarly, regarding *govern*, *misgovern*, *government*, *governor*, and *ungovernable* as different words will also inflate the estimates.

Word forms and learning

Language teachers should be interested in learning, and decisions about what is included in a word family and what is not should be based on learning

requirements. If a learner knows *govern* and is familiar with the prefix *mis-*, then *misgovern* requires little if any additional learning. For this reason *govern* and *misgovern* can be regarded as one word, or as one base word and a derived word. *Row* (in *a row of chairs*) and *row* (in *row a boat*) are homographs. They have the same form but quite different meanings. Knowing one meaning of *row* does not mean that minimal effort is needed to learn the other, so they would be regarded as two base words. This criterion of additional learning has implications for vocabulary teaching as well as for counting words. It is in the teacher's and learners' interests to reduce the number of words there are in English. This is done in two ways: by directing attention to the underlying concepts of words, and by making learners familiar with word parts. For example, the *head* forms in *the head of a match*, *the head of the school*, and *the head on your body* are homographs which can be considered as three different words or one word. In order to consider them as one word we need to show that there is one concept of *head* that is included in those three uses. If we can do this, then we have made learning *head* more interesting and economical and have reduced the number of different words in English. Similarly, explaining word meanings by using the meanings of the parts of the word reduces the number of different words by making new items appear as recombinations of already known parts.

Dictionaries tend to increase the number of words in a language rather than decrease them. For example, in some dictionaries words of a similar meaning but a different part of speech are listed separately. Thorndike (1924) and Williams (1932) found that the failure of vocabulary researchers to take account of homographs was a major sampling error in vocabulary size estimates. Derived forms of words with meanings closely related to a base word are often listed separately. Clearly it is not sufficient to rely solely on the dictionary to distinguish different words. Different entries and different words are not the same. Some information provided by the dictionary is of value, particularly for classifying abbreviations, proper nouns, compound words, archaic or obsolete words, dialect words, or slang. Decisions about base and derived words, and homographs with unrelated meanings, however, need to be made using outside criteria. In this way more realistic estimates of dictionary size and thus vocabulary size can be made.

Word types in Webster's Third New International Dictionary

In an attempt to evaluate previous estimates of vocabulary size and to provide a basis for the development of a test, a study was made of *Webster's Third New International Dictionary* (1961). This dictionary was chosen because it is the largest non-historical dictionary of English.

The preface to the dictionary says that it has a vocabulary of over 450,000 words. As noted above, Diller (1978) used this figure in his test of vocabulary size. It is only possible to get near a figure of 450,000 by counting every bold-faced entry in the dictionary. Many of these occur in the same line, for example, *batten*, *battened*, *battening*, *battens*, and represent inflected forms of

alternative spellings. If same line entries are excluded, the dictionary contains around 267,000 entries (Dupuy 1974).

The words were classified into the following categories: base, derived, proper words (including compound proper words), compound words, and others. 'Others' include symbols, prefixes and suffixes, letters, abbreviations, alternative spellings, archaic words, and dialect words. Compound words consist of two or more words separated by a space or a hyphen (*blue earth*, *blue-eyed*). Proper words are those in which the dictionary indicates as being usually, often, or sometimes capitalized (resedaceae *n.pl, cap*; puebla *adj. usu. cap*). Derived words are defined generally as words which require minimal extra learning. A corresponding base word must occur as a main entry in the dictionary. The meaning of the derived word must be clear from the meaning of the parts that make up the word or involve the minimum of extra learning. In some cases it was necessary to define as 'derived', words which had one derived meaning out of many, since in a test such words would have to be marked correct if the derived meaning was given as an answer. Irregular inflectional forms are included as derived words. Words consisting of common prefixes, for example, *re-*, attached to base words are marked 'derived'.

Nagy and Anderson (1984) set up a scale to distinguish base and derived words. The distinction between Nagy and Anderson's levels two and three was applied when distinguishing 'base' from 'derived'. All entries which were not included in the previous categories were, by elimination, 'base words'. Some entries with prefixes were classified 'base' although a specialized knowledge of the prefix, for example, *pterygo-*, may allow the meaning to be derived from the meaning of the parts. A word was classified as a base word if it was the least inflected form of a group of related words. When there was a choice between items of roughly similar length or inflection then nouns or verbs were counted as the base rather than adjectives or adverbs, for example, *congruity* would be considered as the base over *congruent*, *congruous*, *congruism*.

Considerable care was taken over the distinction between base and derived. The meaning relationship and the amount of extra learning required were always the main considerations. To clarify the classification of items further, here is a running set of items from one of the dictionary samples.

[1]cow	This is the first in a set of homographs so it is acceptable. It was classified as a base word.
crampfish	This is classified as a base word because there is no hyphen or space between *cramp* and *fish* and the meaning is not easily discovered from its parts.
[2]credit	This is ignored because it is a subsequent homograph and the next item *creditability* is chosen. It is classified as derived because its meaning is clearly related to a base entry, *credit*.
crile	The dictionary indicates that this is a dialect word *chiefly Scot* so it is classified under 'others'.

crop-bound	This is classified as a compound word.
crouchant	This means 'crouching' and is classified as derived.
cry-	This is a prefix and is classified under 'others'.

Note that wherever possible the criteria for deciding between the various types of words were based on forms or information provided in the dictionary. As we shall see in the next section of this article, there is sufficient evidence to show that the criteria were consistently applied.

4. HOW DO WE CHOOSE WHAT WORDS TO TEST?

In studies of vocabulary size, words have most commonly been selected by means of a spaced sampling procedure. This involves working through the dictionary from a randomly determined starting point taking words at a specified interval, for example, the first word on every fifth page, or every twelfth column. However, as Lorge and Chall (1963) pointed out in their analysis of the Seashore and Eckerson (1940) test, sampling from a dictionary needs to take account of the large amount of space given to high-frequency words. Furthermore, many entries in the dictionary are actually subsequent homographs of a previous entry. These represent high-frequency, often polysemous words which also have a large number of derived forms associated with them. For example, there are six entries in *Webster's Third* for *fly*, and its derivatives, including *fly-away*, *fly-blown*, *flier*, and *flying* constitute another large set of entries which occupy more than a page. Because such words occupy so much space and account for so many entries, the chance that they will be chosen in any spaced sample is high, and as a result a test based on this kind of sample will contain too many high-frequency items which will inflate the estimate of the learners' vocabulary size.

In a spaced sampling procedure, the amount of space taken up by a high-frequency word is less of a problem than the number of entries devoted to it. Dictionaries vary in the extent to which they create separate entries for homographs rather than incorporating them in a single entry. A comparison of the number of homographic entries for ten words (*collective*, *crab*, *feather*, *fish*, *grow*, *hum*, *lop*, *man*, *presage*, *slough*) in six dictionaries showed that for eight of the ten words *Webster's Third* had more or the same number of homographs as other dictionaries, and in the remaining two cases the *Shorter Oxford English Dictionary* (1973) had more. So homographic entries are as much, or more, of a problem for spaced sampling from Webster's compared with other dictionaries.

In the present study, three spaced samples were taken from *Webster's Third*. For Sample 1, the first new entry in every eleventh column was selected, unless it was a homograph of a previous entry, in which case it was ignored and the next item in the column was selected. For example, ^{2}*appassionato* on page 103 would not have been selected because it is a subsequent homograph of ^{1}*appassionato*. The next item *appaumé* would have been selected instead. The purpose of not choosing the subsequent homographs was to reduce, if not eliminate, the over-representation of high-frequency words in the sample. The

interval of every eleventh column was used because it had been calculated that this would yield a sample containing about 700 words altogether and 200 base words. When the words in the sample were classified into the various categories identified above, it was found that the sample contained only 164 base words. Therefore, a second sample was drawn in the same way, except that the sampling interval was every tenth column. (As it turned out, this still produced only 181 base words.)

For Sample 3, it was decided to include subsequent homographs if they were the first new entry in the column but to keep a tally of them so that they could be treated as a separate category of items. In this case the sampling interval was again every eleventh column.

Table 1 gives the composition of the three samples. The subsequent homographs in Sample 3 have been excluded here to allow a more direct comparison with the results of the other two samples. The percentages in the table show a high degree of consistency in the results obtained from the three samples.

Table 1: Results of three spaced samples of Webster's Third

Classification	*Sample 1*	*Sample 2*	*Sample 3*	*Total*
Base	164 (22.7%)	181 (22.8%)	147 (26.1%)	492 (23.9%)
Derived	201 (27.9%)	227 (28.6%)	156 (27.8%)	584 (28.1%)
Proper	61 (8.5%)	50 (6.3%)	60 (10.7%)	171 (8.5%)
Compound	223 (30.9%)	252 (31.8%)	147 (26.1%)	622 (29.6%)
Others	72 (10.0%)	83 (10.5%)	52 (9.3%)	207 (9.9%)
Total	721	793	562	2,076

We now have to deal with the question of what proportion of the total number of words in the dictionary are subsequent homographs. In Sample 3 there were 156 of them, which represents nearly 22 per cent of the sample. However, in another study which involved two spaced samples and three samples of all the words on a set of consecutive pages, Goulden (1984) arrived at the more conservative estimate that 15 per cent of the entries in *Webster's Third* were subsequent homographs. At least two thirds of them did not differ substantially in meaning from the first entry. For example, *green* has four entries. The first three have the same meaning but are different parts of speech, while the fourth entry is a dialect use meaning 'to yearn, to long for'. Therefore, in the present study the figure of 15 per cent has been used.

The next step is to translate the percentages from the samples in Table 1 into estimates of the number of entries in each category in the dictionary, based on the confirmed estimate that the total number of entries is 267,000. First, the figures in Table 1 need to be adjusted to incorporate subsequent homographs as a separate category. If we take it that 15 per cent of the total entries belong in this new category, that means that there are 40,050 of them, i.e. 15 per cent of

267,000. When the 40,050 subsequent homographs have been subtracted from the total, the remaining 226,950 words can be assigned to the other categories according to the percentages in Table 1. For example, base words comprised 23.9 per cent of the three samples combined, which gives us an estimate of 54,241 base words in the whole dictionary (i.e. 23.9 per cent of 226,950). This represents 20.3 per cent of the total of 267,000 entries. The complete set of estimates is given in Table 2.

Table 2: Proportions and totals of types of entries in Webster's Third

Word type	*Estimated number of entries*	*Percentage of entries*
Base	54,241	20.3
Derived	63,773	23.9
Proper	19,291	7.2
Compound	67,177	25.2
Others	22,468	8.4
Subsequent homographs	40,050	15.0
Total	267,000	100.0

Sampling: derived words

Because the distinction between base words and derived words is so important for this study, it is worth checking that the ratio of base to derived is valid. Table 3 gives the relative base and derived figures for three different studies of vocabulary in addition to the present study of *Webster's Third*. In the case of the Nagy and Anderson (1984) study, which was based on the American Heritage word frequency count (Carroll, Davies, and Richman 1971), the percentages in the table were calculated by re-arranging their categories to match those used in the present study. The Seashore and Eckerson (1940) study was based on Funk and Wagnall's (1937) unabridged *New Standard Dictionary of the English Language*. Seashore and Eckerson followed the dictionary maker's classification of base and derived words. The figures for the *Shorter Oxford Dictionary* were obtained by the present authors in a sample of words from the *English Word Speculum* (Dolby and Resnikoff 1967).

Table 3 shows that the classification into base and derived in the *Webster's Third* study gets considerable support from other studies. The exception is the *Shorter Oxford* study. A comparison of entries in *Webster's Third* and the *Shorter Oxford* shows that the *Shorter Oxford* includes some derivatives as part of the main entry while *Webster's Third* is more likely to list them as separate entries.

A study of *Webster's Third* by Dupuy (1974) provides some confirming evidence for the proportion of words classified as compound words and proper words. After we adjusted Dupuy's percentage to account for the different criteria he used to determine his total figure (he deliberately excluded

Table 3: Relative proportions of base and derived words in several studies

Word type	*Nagy and Anderson (1984)*	*Seashore and Eckerson (1940)*	Shorter Oxford English Dictionary	Webster's Third
Base	48%	45%	54%	46%
Derived	52%	55%	46%	54%

subsequent homographs and prefixes from his total), it was found that compound words made up 29 per cent of the total and proper words made up 9 per cent of the total. These figures are very close to our figures of 25.2 per cent and 7.2 per cent.

Word frequency and sampling

The three *Webster's Third* samples contain a total of 492 base words. In order to make a one-percent sample of the base words in *Webster's Third*, 50 more words were selected by means of spaced sampling. These were added to the list, making a total of 542 base words (one percent of the 54,241 in Table 2).

It was necessary to check that the procedure of ignoring subsequent homographs during sampling achieved the intended result of producing a sample from *Webster's Third* which contained a suitable number of high frequency words. This was done by seeing how many items in the sample occurred in Thorndike and Lorge's (1944) word frequency list. Before the check could be made, the Thorndike and Lorge list itself was analysed to find the number of base words it contains. Table 4 gives the results. The base words in Thorndike and Lorge are actually under-represented in the one-percent sample of Webster's. A total of 139 Thorndike and Lorge words was expected but only 123 words from the list occurred in the Webster's sample. This under-representation was evenly spread between the first 10,000 and the second 10,000 levels of Thorndike and Lorge.

Table 4: Types of words in Thorndike and Lorge's (1944) list of 30,000 words

Type	*Items in sample*	*Percentage of total*	*Estimated totals in Thorndike and Lorge*
Base	139	46.3	13,900
Derived	118	39.3	11,800
Proper words	39	13.0	3,900
Abbreviations, etc.	4	1.3	400
Total	300	100.0	30,000

It is important that the high-frequency words are not over-represented or under-represented because these are the words most likely to be known. To take account of this, 16 words at suitable levels were randomly selected from Thorndike and Lorge and added to the Webster's list. Seventeen other items were dropped from among the lowest frequency words to keep it as a one per cent sample of *Webster's Third*.

The 542 words were now a representative sample of the base words in *Webster's Third*.

Bringing Webster's Third *up to date*

Webster's Third was published in 1961. The most recent addition to it, entitled *9,000 Words* (1983), includes most of a previous addendum, *6,000 Words*, new material for the 1981 addendum, and items added since then. An analysis was made of the first word on every second page of *9,000 Words*, starting with page two and omitting pages where a new letter of the alphabet began (a one-percent sample of the 9,000 words). Twelve of the 90 words in the sample had been marked by the compilers as new senses of words already in *Webster's Third* (for example, *bent* = dishonest, *hawk* = one who wants a war, *pill* = contraceptive pill). One of the 90 was a homograph with a meaning unrelated to the form already in the dictionary (*camp* as an adjective). Fifteen of the 90 words were derivatives of words already in *Webster's Third* (for example, *antiepileptic*, *brassware*, *buttondown*, *corotate*). Three were proper words (for example, *Chinese fire drill*), and 30 were compound words (for example, *black nationalist*, *command module*, *credibility gap*). The remaining 29 words were classified as base words. They included *aeroplankton*, *aliesterase*, *anamestic*, *astrochemistry*, and *barf*. Nineteen of them were clearly technical terms. All of the 29 words were added to the Webster's list, making a total of 571 words.

Table 5: Word types in 9,000 Words

New base words	29	32%
Homographs of previous entries in *Webster's Third*	13	15%
Derived words	15	17%
Proper words	3	3%
Compound words	30	33%
Total	90	100%

Words excluded from testing

Proper words, compound words, derived words, and a variety of items classified under 'others' have been excluded from the test list created from this study. There are four reasons for this exclusion.

(1) Dictionaries differ considerably in the way they deal with compound words and proper words, both in their policies of inclusion and place of inclusion. For

example, if we look at page 197 of the *Shorter Oxford*, which includes the items from *bittacle* to *black*, and compare the same range of items in *Webster's Third*, we find that *Webster's Third* includes 36 compound words which do not occur at all in the *Shorter Oxford*. The *Shorter Oxford* includes six compound words which are not in *Webster's Third*. All of the *Shorter Oxford* compound words are listed under the headword.

Page 365 of the *Shorter Oxford*, which runs from *coindicate* to *cole-tit*, contains no capitalized proper words, while the equivalent range in *Webster's Third* contains 16 proper words. The *Shorter Oxford*'s attitude to proper words is clear from the following quotation from page x of the Preface:

> In addition to and interpenetrating the common vocabulary with all its ramifications and outliers, there is a vast number of proper names, which either themselves acquire connotative value or give rise to derivatives which take their place among the ordinary words of the language.

Clearly, proper words are not seen as being ordinary words. Lorge and Chall (1963) have a similar opinion.

(2) Derivatives, abbreviations, alternative spellings, inflected words, some proper words, and a large proportion of compound words do not represent a significant extra learning burden. The distinction between derivatives and base words was made with this as a guide. About one-fifth of the proper words and three-quarters of the compound words can be classified as derivatives of these two classes. This means that in *Webster's Third* base proper words account for about 16,000 of the 267,000 entries, and base compound words (compound words where the first member of the compound does not have a separate entry in the dictionary) account for about 17,000 entries. Table 6 lists the non-derived entries in *Webster's Third*. Of these 113,161 base items, which all

Table 6: Base words and other types of base items in Webster's Third

Word types	*Estimated number of entries*	*Percentage of entries*
Base words	54,241	20.3
Base proper words	16,000	6.0
Base compound words	17,000	6.4
Other		
abbreviations	9,900	3.7
word parts	2,670	1.0
Subsequent homographs with unrelated meanings	13,350	5.0
Total	113,161	42.4

represent new learning for a native speaker, only a one per cent sample of the 54,241 base words was chosen for testing.

(3) It is important that the sample of words to be tested represents the words in the dictionary as closely as possible. Therefore the larger the sample, the better it is. With a one per cent sample of *Webster's Third* and *9,000 Words*, the test list contains 571 items. The addition of other word types would increase the number of items in the test list to an unmanageable level.

(4) The main aim of this study is to determine the potential size of a native speaker's vocabulary. Because the frequencies of the various word types in *Webster's Third* are provided here, it is possible to make rough estimates of other groupings of word types by adding appropriate proportions of the desired word types to the base scores.

5. HOW DO WE TEST THE CHOSEN WORDS?

Now that we have a representative sample of words in the language, we need a method of finding out how many of the words people know. Before discussing how to use our *Webster's Third* sample for this purpose, let us look at the approach to vocabulary testing adopted by Diack (1975) in his book *Word-power: Your Vocabulary and its Measurement*. Diack's work is interesting not only because of the testing procedure he used, but also because his way of sampling the words seems to produce relatively conservative estimates of vocabulary size that are much closer to our estimates than those of other investigators.

Diack prepared a set of 50 tests of 60 words each, based on an analysis of the *Concise Oxford*, *Everyman's English Dictionary*, and *Chambers' Twentieth Century Dictionary*. He also reported making use of the Thorndike and Lorge (1944) word list. The tests are divided into six frequency levels; each level represents 6,000 words and so each test item represents 600 words. Here is an example.

Test Number 24

Level 1	Level 2	Level 3
1 *asphalt*	11 *asterisk*	21 *absorbent*
2 *carol*	12 *centigrade*	22 *carnage*
3 *desert*	13 *density*	23 *deluge*
4 *encyclopaedia*	14 *estuary*	24 *eliminate*
5 *oblong*	15 *negative*	25 *negotiate*
6 *paragraph*	16 *perforated*	26 *parole*
7 *rafter*	17 *radius*	27 *recalcitrant*
8 *scale*	18 *section*	28 *rudimentary*
9 *scarcity*	19 *solitary*	29 *stringent*
10 *trapeze*	20 *superfluous*	30 *translucent*

Level 4	Level 5	Level 6
31 *acrimony*	41 *anabaptist*	51 *alburnum*
32 *bauxite*	42 *chiaroscuro*	52 *cacique*
33 *cachet*	43 *dragoman*	53 *dunnage*
34 *denouement*	44 *eidolon*	54 *enclitic*
35 *egregious*	45 *nenuphar*	55 *niello*
36 *obeisance*	46 *parallax*	56 *paraheliotropism*
37 *paradox*	47 *parang*	57 *radula*
38 *rationale*	48 *recalescence*	58 *rocambole*
39 *sacrosanct*	49 *rococo*	59 *surrebutter*
40 *zany*	50 *subpoena*	60 *talus*

Based on his extensive use of the tests with various groups in Britain, Diack reports that people's vocabulary level depends primarily on their level of education, but it is also influenced by 'wide reading and wide contacts with life' (1975: 12). He says that adults with a secondary education are generally at Level 3 (12,000–18,000 words). To reach Level 4 (18,000–24,000 words) normally requires either a university education or a lively mind and a wide range of reading. Those at Level 5 (24,000–30,000 words) are 'among the most widely read in the country. They are to be found in the top echelons of their professions or heading in that direction' (op. cit.: 11). Diack doubts the reliability of the tests beyond Level 5 but he suggests that there are only a few people at Level 6.

A comparison of five of Diack's tests with the Thorndike and Lorge list reveals that 7 or 8 of the 10 words at Level 4 in Diack's tests are in the Thorndike and Lorge list. Level 5 in Diack's test includes 4 words out of 10 which are in Thorndike and Lorge. According to Diack, Level 4 represents a vocabulary of 24,000 words. Apparently when dividing words into levels, Diack relied on the frequency levels in Thorndike and Lorge, and he did not take account of the large number of derived words in Thorndike and Lorge which are already represented by base words in the list. Our analysis of Thorndike and Lorge showed that, of the 30,000 items in the list, only 13,900 are base words. To adjust scores on Diack's test on the basis of our Thorndike and Lorge analysis, it is therefore necessary to multiply a learner's score on Levels 1 to 4 by 0.46 or 9/20, or roughly, to reduce it by half.

So, the fault of Diack's tests is that each of his 6,000 word levels does not represent 6,000 words because: (1) he did not accurately estimate the number of base words he was working with and did not make a random sample with appropriate precautions, and (2) he was misguided by the frequency levels in Thorndike and Lorge because he did not allow for derivatives and proper words in that list. However, because his estimate of the number of base words was near a sensible figure, the results of his tests are not wildly misleading, as other studies have been.

Let us now focus on the testing procedure that Diack sets out in his book. It is designed for self-assessment, along the lines of a 'Test Your Own IQ' book. Readers work through one of the tests, beginning at Level 1. When they come to

words that they do not know or are not sure of, they write down the numbers of those words and continue until they have written ten numbers. Then they stop and go back over the last five words that were thought to be known. For each of these five words, some kind of definition has to be written:

> You can show your knowledge of the word by giving a synonymous word or phrase, by using it in a sentence that demonstrates your knowledge, or you can do it by diagram or sketch. (Diack 1975: 6)

The answers are then checked with a dictionary. If all or most of the five words are correct, it is assumed that all of the previous words that were thought to be known were in fact known and so the score is simply the total number of words known (after subtracting any of the five words that were wrong). In order to improve the reliability of the procedure, Diack recommends that people should take at least three of the tests and use the average score as the basis for estimating the size of their vocabulary, which is calculated by multiplying the average score by 600.

Diack's procedure is an example of the most straightforward method of measuring knowledge of words: a 'yes/no' or checklist approach, in which the respondents are simply asked to indicate whether they know each word or not. This method has been used in vocabulary research for a long time, especially in studies of the vocabulary size of school children (for example, Sims 1923; Tilley 1936), although there have been continuing doubts about its validity. There are at least two questions that need to be addressed in using this method:

(1) *What do we mean by 'knowing a word'?* The investigator has to indicate simply and clearly to the respondents how they should decide whether they know a word or not. Does it mean that one has seen the word before, without perhaps knowing what it means? Or is it necessary to be able to give a dictionary-type definition of the word? What if one thinks that one knows the meaning but is not sure? Diack required his respondents to be fairly sure that they knew at least one meaning of the word, though he allowed them to express their knowledge in various ways; they did not have to compose a dictionary definition of the word.

(2) *How can we detect whether the respondents have a tendency to overrate their knowledge of the words?* Even with a clear explanation of the criterion for knowing a word, it is likely that some respondents will tend to say 'yes' to more words than they really know. In recent years several researchers (Anderson and Freebody 1983; Nagy, Herman, and Anderson 1985; Meara and Buxton 1987; Meara and Jones 1987) have found a way of detecting this by including a good proportion of plausible non-words in their vocabulary lists. If respondents say that they know a number of the non-words, this indicates that they tend to overrate their vocabulary knowledge, and their scores can be reduced accordingly. This procedure has been used in studies with schoolchildren and second language learners, but such a check may not be so necessary in the case of adult native speakers. In Diack's tests, the last five 'known' words are

checked. If all or most of them are wrong, he suggests that further checking of earlier words should be done in order to get a more realistic score. In addition, such a person should take some more of the tests and be more conservative in judging which words she or he knows.

In fact, experience with both the Diack tests and our sample (see Appendix 1) shows that the problem words form only a small proportion of the words being tested. If the words in the sample are arranged in order from most frequent to least frequent, there will be a certain number of words at the beginning that are all known by a native speaker, while many of the low-frequency words will be totally unfamiliar. For example, Diack (1975: 20) reports that, when he had to give his tests to adults within a limited time, he would ask them to begin at Level 3, because he knew that he could assume knowledge of the first 20 words. On the other hand, he found that very few people knew more than the odd word at Level 6. Similarly, everyone knows the very common words at the beginning of our list, but we expect that well over half of the words in the sample will be completely unknown to native speakers. Thus, the doubtful words are those in the transitional section of the list between the two extremes. Any check of understanding is most profitably done on words in this intermediate section of the list.

To prepare the *Webster's Third* sample for testing, the words were put in order of frequency. For the more common words, this was based on Thorndike and Lorge (1944). Next came words that were not listed in Thorndike and Lorge, but which could be found in *Webster's Collegiate Dictionary* (1979). The remaining words (not in either Thorndike and Lorge or *Webster's Collegiate*) were placed at the end of the list. As a result, some words which are relatively familiar to native-speaking adults in New Zealand came well down the list among much less common words. Therefore words such as *footage*, *glaucoma*, *cockup*, *geisha*, *weta*, *golliwog*, *ravioli* and *detente* were moved to higher positions on the list in recognition of their frequency in contemporary usage.

The complete list was then divided into two parallel halves and tested with a small group of native speakers. They were asked to look at the words in the lists and indicate with a tick those words that they felt confident they knew and to put a question mark next to those they were not sure about. Their responses showed that a large number of the words were not known. On the basis of their responses the complete list was divided into two sections, one section of 250 words containing the most frequent words and the known low-frequency words. This was later subdivided into five equivalent tests each containing 50 words (see the Appendix). The remaining 221 words in the *Webster's Third* sample are included in the Appendix in a list titled 'Words in the *Webster's Third* sample that are not likely to be known'.

If each of the fifty-item tests are used separately each item represents 500 words, so the number of items known in a test should be multiplied by 500 to get a total base vocabulary size score. If all five tests are sat, then the scores for all the tests should be added together and multiplied by 100. People gaining a score

above 15,000 words on these tests should also check the list of words not likely to be known. Each item represents 100 words.

The procedure for taking the five tests is fairly closely modelled on that of Diack. One difference in our test is the use of the question mark by respondents to mark words that they are not sure about. This was adopted partly for psychological reasons (many respondents find it more satisfying to have that response available) and also to guide the respondents to doubtful words when they check the list a second time.

The five fifty-item tests were given to a group of 20 native speakers, all of whom were university graduates over the age of 22. Their average scores on the five tests ranged from 13,200 to 20,700 with a group average of 17,200.

6. CONCLUSION

The present study leads us to the following conclusions:

(1) Using the largest non-historical dictionary available and allowing for some recent additions to the language, there are less than 58,000 base words to draw on for a vocabulary test. The majority of these base words are unknown to most native speakers of English.

(2) If proper words, compound words, affixes, and homographs with unrelated meanings are also included as base items, then there are just over 110,000 base items in *Webster's Third*. Clearly, estimates of the vocabulary size of adult native speakers which credit them with vocabularies of 216,000 words (Diller 1978) or 80,000 words (Miller and Gildea 1987) are greatly inflated. It is more likely that the average educated native speaker has a vocabulary of around 17,000 base words and has acquired them at the average rate of about two to three words per day.

(3) If native speakers do in fact acquire vocabulary at this relatively slow rate, it would seem that for second language learners direct teaching and learning of vocabulary is a feasible proposition. This could be done in a variety of ways (Nation 1990). Because around 66 per cent of the low frequency words of English come from French, Latin, or Greek, the use of word parts as mnemonic devices is a useful procedure. Similarly research on the keyword mnemonic techniques (Pressley *et al.*, 1982) has demonstrated its value as a strategy for improving direct vocabulary learning. There is a wide range of vocabulary learning activities using words in sentence contexts. These include exercises using code, selective cloze exercises, and the use of texts with glossaries or direct teacher explanation. These activities, combined with large amounts of extensive reading accompanied by practice in guessing words from context, would allow second language learners to develop their vocabulary at a rate well above that of most native speakers.

(Revised version received February 1989)

REFERENCES

Anderson, R. C. and **P. Freebody.** 1981. 'Vocabulary knowledge' in J. T. Guthrie (ed.): *Comprehension and Teaching: Research Reviews*. Newark: International Reading Association.

Anderson, R. C. and **P. Freebody.** 1983. 'Reading comprehension and the assessment and acquisition of word knowledge.' *Advances in Reading/Language Research* 2: 231–56.

Campion, M. E. and **W. B. Elley.** 1971. *An Academic Vocabulary List*. Wellington: NZCER.

Carroll, J. B., P. Davies, and **B. Richman.** 1971. *The American Heritage Word Frequency Book*. New York: Houghton Mifflin.

Clarke, D. F. and **I. S. P. Nation.** 1980. 'Guessing the meanings of words from context: strategy and techniques.' *System* 8: 211–20.

Corson, D. J. 1983. 'The Corson measure of passive vocabulary.' *Language and Speech* 26: 3–20.

Cummins, J. 1981. 'Age on arrival and immigrant second language learning in Canada: a reassessment.' *Applied Linguistics* 2: 132–49.

Diack, H. 1975. *Test Your Own Wordpower*. St. Albans: Paladin.

Diller, K. C. 1978. *The Language Teaching Controversy*. Rowley, Mass.: Newbury House.

Dolby, J. L. and **H. L. Resnikoff.** 1967. *The English Word Speculum.* The Hague: Mouton.

Dupuy, H. J. 1974. *The Rationale, Development and Standardization of a Basic Word Vocabulary Test*. Washington, DC: US Government Printing Office.

Fries, C. C. and **A. A. Traver.** 1960. *English Word Lists*. Ann Arbor: George Wahr.

Goulden, R. 1984. 'Measuring Vocabulary Size.' Unpublished BA Hons. research paper.

Jamieson, P. 1976. *The Acquisition of English as a Second Language by Young Tokelau Children Living in New Zealand*. Unpublished Ph.D. thesis, Victoria University of Wellington.

Lorge, I. and **J. Chall.** 1963. 'Estimating the size of vocabularies of children and adults: an analysis of methodological issues.' *Journal of Experimental Education* 32: 147–57.

Meara, P. and **B. Buxton.** 1987. 'An alternative to multiple choice vocabulary tests.' *Language Testing* 4: 142–51.

Meara, P. and **G. Jones.** 1987. 'Tests of vocabulary size in English as a foreign language.' *Polyglot* 8: 1–40.

Merriam-Webster. 1983. *9,000 words*. Springfield, Massachusetts: Merriam-Webster Inc.

Miller, G. A. and **P. Gildea.** 1987. 'How children learn words.' *Scientific American* 257 3: 86–91.

Nagy, W. E. and **R. C. Anderson.** 1984. 'How many words are there in printed school English?' *Reading Research Quarterly* 19: 304–30.

Nagy, W. E. and **P. A. Herman.** 1984. 'Limitations of vocabulary instruction.' Technical Report, Center for the Study of Reading, University of Illinois 326.

Nagy, W. E., P. Herman, and **R. C. Anderson.** 1985. 'Learning words from context.' *Reading Research Quarterly* 20: 233–53.

Nation, I. S. P. 1990. *Teaching and Learning Vocabulary*. New York: Newbury House.

New Standard Dictionary of the English Language. 1937. New York: Funk and Wagnall.

Pressley, M., J. R. Levin and **H. D. Delaney.** 1982. 'The mnemonic keyword method.' *Review of Educational Research* 52: 61–91.

Seashore, R. H. and **L. D. Eckerson.** 1940. 'The measurement of individual differences in general English vocabularies.' *Journal of Educational Psychology* 31: 14–38.

Shorter Oxford English Dictionary. 1973. Oxford: Oxford University Press.

Sims, V. M. 1923. 'The reliability and validity of four types of vocabulary tests.' *Journal of Educational Research* 20: 91–6.

Thorndike, E. L. 1924. 'The vocabularies of school pupils' in J. Carelton Bell (ed.): *Contributions to Education. Vol. 1*. New York: World Book Co.

Thorndike, E. L. and **I. Lorge.** 1944: *The Teacher's Word Book of 30,000 Words*. New York: Teacher's College, Columbia University.

Tilley, H. C. 1936. 'A technique for determining the relative difficulty of word meanings among elementary school children.' *Journal of Experimental Education* 5: 61–4.

Webster's Collegiate Dictionary. 1979. Springfield, Mass.: Merriam-Webster Inc.

Webster's Third New International Dictionary. 1961. Springfield, Mass.: Merriam-Webster Inc.

Williams, H. M. 1932. 'Some problems of sampling in vocabulary tests.' *Journal of Experimental Education* 1: 131–3.

Xue Guoyi and **I. S. P. Nation.** 1984. 'A university word list.' *Language Learning and Communication* 3: 215–29.

APPENDIX 1

Vocabulary size tests

These are tests to estimate how many words you know. You will find below a list of 50 words which is part of a sample of all the words in the language. The words are arranged more or less in order of frequency, starting with common words and going down to some very unusual ones.

Procedure

1. Read through the whole list.
   Put a tick next to each word you know, i.e. you have seen the word before and can express at least one meaning of it.
   Put a question mark next to each word that you think you know but are not sure about.
   (Do not mark the words you do not know.)

2. When you have been through the whole list of 50 words, go back and check the words with question marks to see whether you can change the question mark to a tick.

3. Then find the last five words you ticked (i.e. the ones that are furthest down the list). Show you know the meaning of each one by giving a synonym or definition or by using it in a sentence or drawing a diagram, if appropriate.

4. Check your explanations of the five words in a dictionary.
   If more than one of the explanations is not correct, you need to work back through the list, beginning with the sixth to last word you ticked. Write the meaning of this word and check it in the dictionary. Continue this process until you have a sequence of four ticked words (which may include some of the original five you checked) that you have explained correctly.

5. Calculate your score for that 50-item test by multiplying the total number of known words by 500. Do not include the words with a question mark in your scoring.

APPENDIX 2

The 577 word *Webster's Third* sample divided into five tests and a list of uncommonly known words

Test 1

1 as	11 abstract	21 aviary
2 dog	12 eccentric	22 chasuble
3 editor	13 receptacle	23 ferrule
4 shake	14 armadillo	24 liven
5 pony	15 boost	25 parallelogram
6 immense	16 commissary	26 punkah
7 butler	17 gentian	27 amice
8 mare	18 lotus	28 chiton
9 denounce	19 squeamish	29 roughy
10 borough	20 waffle	30 barf
31 comeuppance	41 cupreous	
32 downer	42 cutability	
33 geisha	43 regurge	
34 logistics	44 lifemanship	
35 panache	45 atropia	
36 setout	46 sporophore	
37 cervicovaginal	47 hypomagnesia	
38 abruption	48 cowsucker	
39 kohl	49 oleaginous	
40 acephalia	50 migrationist	

Test 2

1 bag	11 avalanche	21 bastinado
2 face	12 firmament	22 countermarch
3 entire	13 shrew	23 furbish
4 approve	14 atrophy	24 meerschaum
5 tap	15 broach	25 patroon
6 jersey	16 con	26 regatta
7 cavalry	17 halloo	27 asphyxiate
8 mortgage	18 marquise	28 curricle
9 homage	19 stationery	29 weta
10 colleague	20 woodsman	30 bioenvironmental
31 detente	41 gamp	
32 draconic	42 paraprotein	
33 glaucoma	43 heterophyllous	
34 morph	44 squirearch	
35 permutate	45 resorb	
36 thingamabob	46 goldenhair	
37 piss	47 axbreaker	
38 brazenfaced	48 masonite	
39 loquat	49 hematoid	
40 anthelmintic	50 polybrid	

Test 3

1 bird
2 fell
3 improve
4 barn
5 fatigue
6 kettle
7 combat
8 resent
9 redeem
10 hurrah
11 conversion
12 fixture
13 accede
14 avocation
15 calyx
16 conclave
17 hierarchy
18 monologue
19 tamper
20 acanthus
21 blowout
22 crupper
23 gloaming
24 minnesinger
25 perpetuity
26 riffle
27 behindhand
28 embolism
29 angst
30 blowhard
31 devolute
32 envoi
33 golliwog
34 neonate
35 plainchant
36 astrochemistry
37 nondurables
38 carboxyl
39 eyestalk
40 curragh
41 gunlock
42 dipole
43 rigorism
44 localist
45 benchboard
46 stirabout
47 hypothallus
48 doombook
49 paradiplomatic
50 poroplastic

Test 4

1 cool
2 kitchen
3 lead
4 cow
5 frog
6 scent
7 harsh
8 ascertain
9 sprig
10 matron
11 coronet
12 jut
13 amorphous
14 bagpipe
15 choleric
16 crock
17 incumbent
18 offal
19 untoward
20 amphitrite
21 carpel
22 doss
23 havelock
24 nominative
25 pilotage
26 serried
27 blurb
28 scriber
29 appositive
30 capybara
31 directrix
32 footage
33 horseshit
34 nighthawk
35 ravioli
36 aeroplankton
37 tandoor
38 cogito
39 corvette
40 chanterelle
41 hyperthyroid
42 pica
43 immunoassay
44 apertometer
45 scandium
46 gusli
47 chuckie
48 mendeleyevite
49 matelasse
50 slipper

Test 5

1 cotton
2 block
3 precious
4 dig
5 hostile
6 accurate
7 inhabit
8 crook
9 blockade
10 microscope
11 deign
12 marrow
13 armada
14 boomerang
15 chowder
16 earring
17 linguistics
18 radium
19 ventilate
20 asperity
21 centripetal
22 dromedary
23 ideograph
24 nuzzle
25 planking
26 welladay
27 brassie
28 huia
29 baobab
30 chomp
31 doubleheader
32 fusilier
33 interplay
34 nubile
35 repartition
36 cockup
37 saddleback
38 hairspring
39 audivision
40 dactylology
41 isomorphy
42 gaper
43 sextodecimo
44 redact
45 capsulectomy
46 volvulus
47 mancipation
48 exceptionalism
49 parasternum
50 sparrowbill

Words in the *Webster's Third* sample that are not likely to be known

anamnestic
banausic
beechdrops
brachypterous
clinicopathologic
corium
cosmolline
decoupage
didapper
disseminule
enantiomorphism
erythrism
fluerics
greylag
hypermorph
jerkwater
kef
mitogenic
myrmecophagous
octroy
ouabain
pes
pikake
psychrophile
rhus
cerveliere
challoth
chinaball
circumforaneous
cladocera
clowder
coblenzion
compacta
contrist
corynomorpha
cuticula
cytozyme
delomorphous
desmolysis
dissilient
elasmosaur
epornitic
espinillo
eumeces
extramitochondrial
fassaile
fluate
follyer
forepost
frontad
ominate
oxpecker
palar
panela
parageosyncline
patelline
pedalium
perimetrium
phaeomelanin
phenoxybenzamine
phoniatric
pilum
pisote
poachwood
polyacrylate
poppywort
porta
propria
proxenus
punctum
rackabones
rohu
salinella
saprolegniales
scholarch

rubeola
sesamoiditis
smectic
supersedeas
tepa
tragacanthin
tuckahoe
yogh
advertonal
agoura
allopelagic
amphicyrtic
analogon
anthocoridae
arguendo
atacamite
authigenic
baikerinite
barsom
baya
belcher
bigarade
bostryx
boxwork
breastbeam
bungersome
burdenman
cannilan
castorile
catfit
brachium
cascabel
chicalote
crampfish
deepgoing
dioptrics
dispersoid
einkorn
endexine
erythropoiesis
fibroma
glutaraldehyde
hemal
hexamethonium
karyolymph
ketoglutarate
lamin
linoleate
lummox
garefowl
genistin
glia
goldtit
guignolet
halse
hominal
hursinghar
hydrocalumite
hydrorrhea
hypostase
incus
intocostrin
jones
kantiara
kiaat
koombar
leonite
lychnoscope
malma
manroot
melilite
meridienne
metarhodopsin
monochloramine
motacilla
nagaika
neral
nosean
okenite
cheilion
chrysogen
circannual
civilite
cloop
coaxation
cohitre
conducta
copellidine
cotyloid
curuba
cynegetic
deadheart
demivol
diphenoxylate
domnei
dungan
eminento
ervil
sciurus
scudder
sella
shallon
skinball
slipe
snakefish
songman
spikebill
stremmatograph
stylogonidium
symballophone
tanonovicular
theow
tholos
tinsey
tournette
troutbird
twinspur
typicon
vashegyite
vervelle
visceripericardial
whiggamore
wobbulator
xurel
yeara
animalist
baldashin
bargeboard
palilolgy
pectoralis
pepperbush
petrolene
phano
pheoporphyrin
phyllade
pinnaglobin
pleonaste
poikilosmotic
proeutectoid
protenoid
pseudowavellite
quiebracha
roding
rotenoid
sampleite
scallom
sciara

mooneye
notchback
osmol
playa
rachis
rhinencephalon
ruddle
scouse
snowberry
succinate
synaptosome
theca
triazine
uranic
windage
acidophilia
agammaglobulinemea
aliesterase
amgarn
amphiplatyan
anatta
ascidiozoa
aumakua
avahi
axodendrite
banstickle
bauno
beata
biunial
boughpot
breakax
brookweed
buplever
cannabinol
casualism
centumvir
cessionaire
etchant
eusynchite
extrophy
fallowchat
ferritungstite
foedaratus
fonio
frogbit
galenobismutile
gemauve
geta
groundplot
helmetpod
histocyte
howardile
huskanaw
hydrogarnet
insilicate
irreflection
jumpseed
kharmadharaya
koali
kylix
linaloe
lowa
madge
medino
merosome
metasilicate
molave
monorhinic
mouthpipe
naze
norleucine
octanoyl
orseille
ozokerite
scoldenore
sealflower
senam
slipband
slurb
softa
squaloidea
strongback
sucupira
synusia
terna
thermopolymization
tinaja
tiqueur
trochantin
twatchel
tylostyle
vanaprastha
velutino
violescent
weightage
whipster
xanthydrol
xyloketose
yeatmanite