Testing and Teaching Vocabulary

Paul Nation
Victoria University
Wellington, New Zealand

How many words do our learners know? How can we help them increase their vocabulary? This article presents one way of giving a partial answer to these questions. Firstly, it looks at ways of counting words and the results of some counts. It is necessary to do this for two reasons, to make it clear what is meant by vocabulary size, and to use statistical information about vocabulary to guide teaching and learning strategies. Secondly, the article describes the making of a vocabulary test and how the results of the test can be used. The appendix contains the test.

Statistical information about vocabulary

We can count the number of words in a book in three ways. (1) If we want to know how long the book is, we can count how many words there are on a typical page and multiply that by the number of pages. So if each page contains around three hundred words and the book consists of 200 pages, we can estimate that the book is sixty thousand words long. To put it another way, we can say the book contains 60,000 *running words*. (2) If we want to know how many words you need to understand in order to be able to read the book, we must count the words in a different way. First, we have to decide what we mean by the term *word*. Are *mend, mends, mended, mending* one word or four words? Are *branch* (of a tree) and *branch* (of a bank) one word or two? Second, we have to go through the book and make a list of all the different words that occur in the book. If the book contains 60,000 running words, we will probably find that it contains approximately nine thousand different words. (3) If we want to do a frequency count of the words in the book, we make a list of all the different words in the book and we count how often each one occurs. The most frequent word will probably be *the*. In 60,000 running words it will occur about 4,200 times. Many words will occur only once. The word *know* will probably occur about 42 times. If it occurs 4,200 times in 60,000 running words, it accounts for 7% ($\frac{4,200}{60,000} \times \frac{100}{1}$) of the running words in the book. If we add the frequency of the ten most frequent words together we will find that these ten words account for almost 24% of the running words in the text. Table 1 gives the typical figures for a collection of texts consisting of five million running words. We can draw some very useful conclusions from such information.

Table 1

Figures based on a count of 5,000,000 running words (Carroll et al 1971)

Different words	% of running words
86,741	100
43,831	99
5,000	89.4
3,000	85.2
2,000	81.3
100	49
10	23.7

(1) A small number of different words accounts for a very large proportion of the running words in a book. For example, if you know 81% of the 2,000 most frequent words of English, you can expect to meet these high frequency words in almost any book we read. The most useful of these words can be found in *A General Service List of English Words* by Michael West (Longman 1953). They form the basis of most elementary and intermediate English courses and many series of simplified reading books. Here are the first 10 words from the *General Service List*: *a, able, about, above, abroad, absence, absolutely, accept, accident, according (to)*.

(2) A very large number of different words occurs only once or twice in a particular text. Most of these low frequency words are very important for the message carried by the text. However, any time spent teaching or learning them is not always a good use of time. This is because we are not likely to meet them again very soon. By the time we meet them again we may have already forgotten them. There are thousands of low frequency words in English. One estimate is that there are approximately 600,000 English words. If three to five thousand of these are high frequency words, then the other 595,000 are low frequency words. We can see in Table 1 that to know the last 1% of the running words it would be necessary to learn almost 43,000 different words.

(3) Table 1 does not show us one further important group of words. This could be called specialized vocabulary. If we choose a specialized area of the language, for example the language of university textbooks, we find that certain words occur quite frequently in that specialized area but

not very frequently in the language as a whole. For example, the words *complement*, *emerge* and *generate* occur frequently in university textbooks but do not occur very frequently in other texts. Similarly certain words occur frequently when we talk or write about sport, but do not occur very frequently in other fields. There have been several studies of the specialized vocabulary of university textbooks but there has not been much study of the vocabulary of other specialized areas (Campion and Elley 1971, Lynn 1973, Praninskas 1972).

We have looked at three groups of vocabulary, high frequency vocabulary, specialized vocabulary, and low frequency vocabulary. Each of these groups consists of particular words. Each group also requires particular teaching and learning strategies.

The vocabulary test contained in the appendix to this article can be used to decide where learners should be given help with vocabulary learning. The test is divided into five levels. The 2,000 and 3,000 word levels contain high frequency words. Because each of the words at this level occurs frequently it is worth spending class time on them. The university word level represents one type of specialized vocabulary. The 5,000 word level is on the boundary of high and low frequency words. The 10,000 word level contains low frequency words.

Let us look first at how the test was made and how to use it. Then we will look at how to use the results to guide vocabulary teaching and learning.

Making the test

Each section of the test consists of six words and three definitions. This type of item was chosen because it was easy to make and easy to mark, provided very low chances of guessing correctly, tested a large number of words in a short time and allowed learners to make use of whatever knowledge they had of the meaning of a word. The words in each section of the test were chosen so that they would be representative of all the words at that level. No capital letter words, for example the names of people, countries or cities, were chosen. Words like *homesick* or *complexity* which were closely related to words of a higher frequency level were not chosen. Although only 18 words are matched at each level, in fact 36 words at that level are tested. This is because the distractors in the test are not meanings but words.

People who know the words at one level well can do that level very quickly because there are only a few things to read. A native speaker did the whole test in five minutes and got full marks. Usually

a maximum of 50 minutes should be allowed for sitting the test. Most people will need less than this.

The items put together in each section were not related in their meanings. For example, in an early draft of the test *static* (which was to be matched with the definition *not moving or changing*) and *reliable* were in the same section. Because *reliable* has a slight connection with *not moving or changing*, it was replaced by another word. Thus if people sitting the test had a rough idea of the meaning of a tested word they should have been able to match it correctly. This came through very strikingly in the results of one person during the trials of the test. This person's scores for the five sections were 12, 14, 5, 15, 6. The surprising feature of these scores is the relatively high score, 15, on the university word list section. Most other people sitting the test gained lower scores as they moved from one level to another. This person's high score on the university level occurred because his mother tongue was Spanish and he was able to use this knowledge to guess the large number of words derived from Latin at that level. This partial knowledge of the words was sufficient to allow correct guessing. Thus someone's score on the test should not be considered as a conservative estimate. This also indicates that the test is not suitable for learners whose mother tongue is a language which has been strongly influenced by Latin. The definitions in the test use words from a higher frequency level than the tested words. The words from the 2,000 word level use words in the first 1,000 words of English. The words at the 3,000 word level are defined by words in the *General Service List*.

Using the test

The instructions should not require any explanation but people sitting the test should be helped with the instructions if this is necessary.

When marking the test give one mark for each correct matching of a word and its definition. The test can be objectively marked by matching a slip of paper containing the correct answers with the spaces provided for the learners to write their answers. On an average it takes two minutes to mark and add the score of one test. Record the scores for each of the five sections of the test. These are more useful than the total score for the test. If someone scores 12 or less out of 18 in a section of the test then it is worth helping them study the vocabulary at that level. The chances of guessing are low and someone's score on the test can be taken as a close approximation to the proportion of words in the test that they know. A score of 12 out of 18 indicates that approximately one-third of the words at that level are not known. Thus there will be at least 200-300 words worth studying at that level. Table 2 gives guidelines about how to help learners study the vocabulary at the various levels.

Table 2

Vocabulary level	Type of vocabulary	Learning required to increase vocabulary knowledge at each level
2,000 word level	*The General Service List.* The vocabulary of simplified reading books.	1. Learning lists of words based on the Longman Structural Readers Lists or *The General Service List* 2. Intensive and extensive reading of simplified reading books 3. *Advanced English Vocabulary*, Workbook 1 (Barnard, 1972)
3,000 word level	A basis for beginning to read unsimplified texts.	1. Intensive reading of a variety of texts 2. Extensive reading of the Bridge Series
5,000 word level	A wide vocabulary.	1. Training in guessing words in context 2. Wide general reading — novels, newspapers, university texts etc. 3. Intensive reading of a variety of texts 4. *Advanced English Vocabulary*, Workbooks 1 and 2
the university word level	The specialized vocabulary of university texts.	1. Learning words in the University Word List 2. Intensive reading of university texts 3. *Advanced English Vocabulary*, Workbooks 2 and 3 4. Learning prefixes and roots
10,000 word level	A large wide vocabulary.	Activities similar to the 5,000 word level, combined with learning prefixes and roots.

The direct teaching of vocabulary through class teaching and in- dividualized exercises is appropriate for most high frequency words.

The suggestions in Table 2 for increasing knowledge of high frequency words are based on the idea that high frequency words are worth individual attention and thus activities such as learning lists of words and vocabulary study using books like Barnard (1972) are appropriate. Teachers who are doubtful about getting learners to study vocabulary lists should read Nation (1982) which reviews experimental research on list learning. Direct teaching of vocabulary is also appropriate for high frequency words. Besides learning words by direct study of vocabulary, large amounts of words can be learned by meeting them incidentally in context. Thus extensive reading of simplified texts and extensive listening activities are an essential part of a vocabulary learning programme.

Specialized vocabulary can be treated in much the same way as high frequency vocabulary because it is frequent within a special- ized area. Its frequency justifies attention to individual words as a part of a vocabulary learning programme. Because many words in the university word list are of Latin derivation, learning prefixes and roots is a useful aid to learning.

Individual low frequency words do not deserve teaching time unless they contain useful prefixes or roots or are an example of some other regular feature that will help vocabulary learning in general. The strategy of guessing words using context clues is particularly useful and is worth spending time on in class (Clarke and Nation 1980, Honeyfield 1977, Nation 1980).

The basic idea behind the vocabulary test is that the statistical distribution of vocabulary should guide the teaching and learning strategies. The test samples various frequency levels and provides information about where learners need to increase their vocabulary. Teachers can then direct the learners to the appropriate vocabulary and use suitable teaching and learning strategies.

References

Barnard, Helen. 1972. Advanced English vocabulary. Mass.: Newbury House.

Campion, Mary E. and Elley, Warwick B. 1971. An academic vocabulary list. Wellington: NZCER.

Carroll, J.B., Davies, P. and Richman, B. 1971. The American Heritage word frequency book. New York: American Heritage.

Clarke, D.F. and Nation, I.S.P. 1980. Guessing the meanings of words from context: strategy and techniques. System 8:3. 211-220.

Honeyfield, J. 1977. Word frequency and the importance of context in vocabulary learning. RELC Journal 8:2. 35-42.

Kucera, H. and Francis, W.N. 1967. A computational analysis of present day American English. Providence: Brown University Press.

Lynn, Robert W. 1973. Preparing word lists: a suggested method. RELC Journal 4:1. 25-32.

Nation, I.S.P. 1980. Strategies for receptive vocabulary learning. Guidelines 3. 18-23.

———. 1982. Beginning to learn foreign vocabulary: a review of the research. RELC Journal 13:1. 14-36.

———. 1983. Teaching and learning vocabulary. Wellington: English Language Institute.

Praninskas, Jean. 1972. American university word list. London: Longman.

Thorndike, E.L. and Lorge, I. 1944. The teacher's word book of 30,000 words. New York: Columbia University Teachers' College.

West, Michael. 1953. A general service list of English words. London: Longman, Green.

Appendix 1

Name:

A VOCABULARY LEVEL TEST

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

1 business
2 clock
3 horse
4 pencil
5 shoe
6 wall

—— part of a house
—— animal with four legs
—— something used for writing

You answer it in the following way.

1 business
2 clock
3 horse
4 pencil
5 shoe
6 wall

__6__ part of a house
__3__ animal with four legs
__4__ something used for writing

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are *business*, *clock*, *shoe*.

Try to do every part of the test.

The 2,000 word level

1 original
2 private
3 royal
4 slow
5 sorry
6 total

___ complete
___ first
___ not public

---

1 apply
2 elect
3 jump
4 manufacture
5 melt
6 threaten

___ choose by voting
___ become like water
___ make

---

1 blame
2 hide
3 hit
4 invite
5 pour
6 spoil

___ keep away from sight
___ have a bad effect on
___ ask

---

1 accident
2 choice
3 debt
4 fortune
5 pride
6 roar

___ having a high opinion of yourself
___ something you must pay
___ loud deep sound

---

1 basket
2 crop
3 flesh
4 salary
5 temperature
6 thread

___ money paid regularly for doing a job
___ heat
___ meat

---

1 birth
2 dust
3 operation
4 row
5 sport
6 victory

___ being born
___ game
___ winning

The 3,000 word level

1 administration
2 angel
3 frost
4 herd
5 mate
6 pond

___ managing business and affairs
___ spirit who serves God
___ group of animals

---

1 bench
2 charity
3 fort
4 jar
5 mirror
6 province

___ part of a country
___ help to the poor
___ long seat

---

1 coach
2 darling
3 echo
4 interior
5 opera
6 slice

___ a thin flat piece cut from something
___ person who is loved
___ sound reflected back to you

---

1 marble
2 palm
3 ridge
4 scheme
5 statue
6 thrill

___ inner surface of your hand
___ excited feeling
___ plan

---

1 discharge
2 encounter
3 illustrate
4 knit
5 prevail
6 toss

___ use pictures or examples to show the meaning
___ meet
___ throw up into the air

---

1 annual
2 blank
3 brilliant
4 concealed
5 definite
6 savage

___ happening once a year
___ certain
___ wild

The 5,000 word level

1 alcohol
2 apron
3 lure
4 mess
5 phase
6 plank

___ cloth worn in front to protect your clothes
___ stage of development
___ state of untidiness or dirtiness

1 circus
2 jungle
3 nomination
4 sermon
5 stool
6 trumpet

___ speech given by a priest in a church
___ seat without a back or arms
___ musical instrument

1 apparatus
2 compliment
3 revenue
4 scrap
5 tile
6 ward

___ set of instruments or machinery
___ money received by the Government
___ expression of admiration

1 bruise
2 exile
3 ledge
4 mortgage
5 shovel
6 switch

___ agreement using property as security for a debt
___ narrow shelf
___ dark place on your body caused by hitting

1 blend
2 devise
3 embroider
4 hug
5 imply
6 paste

___ hold tightly in your arms
___ plan or invent
___ mix

1 desolate
2 fragrant
3 gloomy
4 profound
5 radical
6 wholesome

___ good for your health
___ sweet-smelling
___ dark or sad

The university word list level

1 affluence
2 axis
3 episode
4 innovation
5 precision
6 tissue

___ introduction of a new thing
___ one event in a series
___ wealth

1 deficiency
2 magnitude
3 oscillation
4 prestige
5 sanction
6 specification

___ swinging from side to side
___ respect
___ lack

1 configuration
2 discourse
3 hypothesis
4 intersection
5 partisan
6 propensity

___ shape
___ speech
___ theory

1 anonymous
2 indigenous
3 maternal
4 minimum
5 nutrient
6 modification

___ without the writer's name
___ least possible amount
___ native

1 elementary
2 negative
3 static
4 random
5 reluctant
6 ultimate

___ of the beginning stage
___ not moving or changing
___ final, furthest

1 coincide
2 co-ordinate
3 expel
4 frustrate
5 supplement
6 transfer

___ prevent someone from doing something they want to
___ add to
___ send out by force

The 10,000 word level

1 acquiesce
2 contaminate
3 crease
4 dabble
5 rape
6 squint

___ to work at something without serious intentions
___ to accept without protest
___ make a fold on cloth or paper

1 blaspheme
2 endorse
3 nurture
4 overhaul
5 skid
6 straggle

___ give care and food to
___ speak badly about God
___ slip or slide

1 auxiliary
2 candid
3 dubious
4 morose
5 pompous
6 temporal

___ full of self-importance
___ helping, adding support
___ bad-tempered

1 anterior
2 concave
3 interminable
4 puny
5 volatile
6 wicker

___ small and weak
___ easily changing
___ endless

1 dregs
2 flurry
3 hostage
4 jumble
5 saliva
6 truce

___ worst and most useless parts of anything
___ natural liquid present in the mouth
___ confused mixture

1 auspices
2 casualty
3 froth
4 haunch
5 revelry
6 seclusion

___ being away from other people
___ someone killed or injured
___ noisy and happy celebration

Appendix 2

The following table shows the frequency criteria which were used in selecting the words at each frequency level.

Table 3

Word Level	Thorndike & Lorge	General Service List	Kucera & Francis
2,000	A	in *GSL*	—
3,000	30-49	not in *GSL*	within first 6,000 words (15 →)
5,000	14-18	not in *GSL*	within first 10,000 words (8 →)
6,000+	1-13*	not in *GSL*	—
10,000	3	not in *GSL*	within 10,000-16,700 word levels (4-7)

*The words at this level were chosen from Campion and Elley (1971). This list was based on a count of university textbooks and assumed knowledge of the first 5,000 words of the Thorndike and Lorge list. Any words in the university list which had related words within the first 5,000 of Thorndike and Lorge were not chosen for this test.

As Table 3 shows, the words chosen from the Thorndike and Lorge list were checked against the *General Service List* (West 1953), the *Computational Analysis of Present Day American English* (Kucera and Francis 1967) and against related words in the Thorndike and Lorge list. The comparison with the Kucera and Francis list was done to avoid the effect of some of the outdated material used in the Thorndike and Lorge counts. The comparison with the *General Service List* was done because many English courses, sets of reading material and pieces of research are based on this list. It is thus useful to consider the *General Service List* when making recommendations for vocabulary learning.