

# LEARNING VOCABULARY

I.S.P. Nation,  
English Language Institute, Victoria University.

Statistical information about vocabulary provides a useful basis for determining the strategies teachers and learners should follow when approaching vocabulary learning. This article looks at this statistical information as it applies to English and then describes the practical implications of this information.

## Vocabulary distribution

Estimates of vocabulary distribution are contained in frequency counts which are based on samples of the language in use, for example textbooks, novels, newspapers, and records of spoken English. Recent frequency counts have made use of computers (Carroll 1971, Kucera and Francis 1967), and they suffer from limitations because of this. The older frequency counts (Thorndike and Lorge 1944) were done manually and in many ways provide more useful information for teaching purposes. All of these frequency counts give similar information about vocabulary distribution.

When researchers do a frequency count they must make decisions about how they will define what is a word and about what material they will choose to count. When they have made the necessary decisions then they begin counting. While counting they must distinguish between **types** and **tokens**. If we were counting tokens, each time the word **a** for example occurred we would count it as a new word. So on a page we might find ten or fifteen tokens of **a**. If we were counting types the word **a** would only be counted once. Each additional occurrence of **a** would not be considered as a new word but simply as another occurrence of the same type. In our count we might also include **a, A, an, An** as all being occurrences of the same type. Three pages of an ordinary book usually contain around a thousand tokens which consist of around two hundred and forty types. Some of these types such as **a, the, of, it, is (or be)** are represented by many tokens. That is, they have a very high frequency. Other types such as **additional, moreover, frequency** may be represented by only one token. That is, they have a low frequency.

Most frequency counts are based on a sample of at least one million words. The Thorndike and Lorge list is based on eighteen million words. These counts reveal similar findings.

1. A small number of types are very frequent.
2. Some types are not very frequent in the language as a whole but in specialized areas they are quite frequent.
3. A very large number of types are very infrequent.

Let us now look at each of these three classes of words in turn and see what strategies should be used to deal with them.

## High frequency words

The ten most frequent word types of English make up 25% of the tokens of any text longer than 500 words. The following table is based on the Carroll et al (1971) count of over a million tokens.

Tokens	Types
99.0%	43,831
95.0%	9,786
89.4%	5,800
87.6%	4,000
85.2%	3,000
81.3%	2,000
74.1%	1,000
66.6%	500
49.0%	100
23.7%	10

Notice that learning the first 1,000 words of English gives a good return by including 74% of the tokens in a text. Learning the second 1,000 words includes an additional 7%. Learning the third 1,000 includes only an additional 4%. It is clear from these figures that there is a small amount of vocabulary which occurs very frequently and is worth spending time on. For English, **The General Service List** (West, 1953) is a useful source of information about these words.

Research by Eaton (1940) reveals that similar lists for some other European languages contain words with meanings similar to those in **The General Service List**.

While low frequency words are often met only in reading, high frequency words occur in the receptive activities of reading and listening, and also in the productive activities of writing and speaking. For this reason when learning high frequency words it is worth giving attention to the spoken and written forms of the words, the grammar and collocations of the words, and their meaning including any limitations on their appropriateness. Here are some techniques to do these things.

1. **Spoken form:** The teacher writes a list of words on the blackboard and then the learners take turns in saying the words that they feel might cause them difficulty. The teacher reads a simplified story to the class and writes on the blackboard any words which the learners might have difficulty in recognizing as they listen. This is done with as little interruption to the story as possible.
2. **Written form:** The teacher says, "You know how to spell **produce, out** and **dance**, so you should be able to spell **pronounce**." The teacher repeats the word while the learners try to write it. This use of analogy is a very important way of making learners aware of spelling rules. The teacher writes the letters of an irregularly spelled word in the wrong order on the blackboard, for example "g..s..n..i". The teacher says the word several times, "Sign, sign . . ." and the learners rearrange the letter to spell the word correctly.

3. **Grammar and collocation:** Using analogy and model sentences the teacher helps learners make sentences using the new word.

"The rain prevented him from coming."  
"Lack of money . . ."

The teacher gets the learners to work in groups to think of as many collocations as they can for a given word. Finally all the collocations are put on the blackboard for the whole class to study.

4. **Meaning:** When dealing with high frequency words which are worth spending time on, teachers should use techniques which keep the meaning of the words away from the learners for a while so that they have the opportunity to give motivated attention to the words. Here are some such techniques. The teacher puts up 12 pictures illustrating the meaning of 12 new words. The teacher says one of the words several times as one learner tries to guess, by pointing, which picture the word refers to. This continues with plenty of repetition of the earlier words until all the words are matched. The teacher explains the meaning of a new word using several sentences. The first sentences do not give much indication of the meaning of the word but they provide opportunity for repeated attention to its form.

There is a lot of blood in the world.

Every person has blood.

We need blood to stay alive.

Blood is a part of our bodies.

Blood is red or blue.

Blood is like water.

If we cut ourselves, blood runs out from the cut.

What is blood?

The learners answer by translating the new word into their mother tongue, by pointing to a picture, by drawing a picture or by giving an explanatory sentence.

So far we have looked at direct ways of teaching high frequency words. Most vocabulary however is learned indirectly by meeting it in spoken or written use. Reading or telling stories to learners, providing plenty of opportunities to get information from English and providing an abundance of interesting simplified material for reading are very important and effective ways of increasing vocabulary. Because high frequency words are so important it is worth keeping a running record or check of the learning of these words.

## Special purpose vocabulary

There have been several studies of the vocabulary of university textbooks (Praninskas 1972, Campion and Elley 1971). These studies find that certain vocabulary occurs relatively more frequently in university textbooks than it does in the language as a whole. These counts of university textbooks assume a base vocabulary and count words not in the base vocabulary. Praninskas assumed **The General Service List** and Campion and Elley assumed the first 5,000 words of the Thorndike and Lorge count. The Praninskas list contains a little over 500 base words or 1,347 base words and derivatives. The Campion and Elley list contains 500 words. The words in these lists are important for learners of English who intend to do university study. They can be learned in the following ways.

1. The vocabulary can be learned in lists. This is best done on small cards with the English word on one side and the mother tongue translation on the other. Although this way of learning vocabulary is viewed with disfavour by some language teachers, over seventy years of research has shown that large amounts of vocabulary can be learned in a very short time in this way. This type of learning should be seen only as gaining preliminary knowledge which will act as a basis for a broader understanding of the vocabulary which will come from meeting it in a variety of contexts. Mnemonic techniques like the keyword technique can be used to speed up the learning of lists of words.
2. Learners can study prefixes, roots and suffixes which will help them remember the meanings of the words. Prefixes are usually more useful for this than roots or suffixes.
3. Learners should do large amounts of intensive reading of university or other specialized texts.

## Low frequency vocabulary

High frequency vocabulary and special purpose vocabulary make up a small percentage of the word types of a language. Most of the vocabulary of English occurs very infrequently. In any text of a thousand words or more about 40% of the word types occur only once. Clearly the effort involved in learning these words will not have a very valuable return because the words learned might not occur again before they are forgotten. In a written text about one word in every 18 or 20 is a low frequency word. It is not wise to spend too much time on particular low frequency words but it is better for learners to have strategies for dealing with them.

1. The most important way of dealing with low frequency vocabulary is to guess its meaning using context clues. These clues include the part of speech of the word, its immediate grammatical and collocational context, and the relationship of the sentence or clause containing the unknown word with neighbouring clauses or sentences. Learners given training in the strategy can guess about 80% of the unknown low frequency words.
2. If the low frequency words follow regular spelling or word building patterns they are worth spending time on. The effort spent in learning them will also help the learning of the useful patterns which apply to many other words. This can be done by keeping a note of the low frequency words that occur in the lessons and at the end of the week writing them on the blackboard. Then a variety of exercises can be done with the words. Here are some examples.
  - i. Each learner chooses a word to pronounce and then the learners take turns in saying their chosen word.
  - ii. The teacher asks the learners to help her break some of the words into parts, give the meaning of the parts, and relate these meanings to the meaning of the word.
  - iii. The learners work together to recall or guess collocations for each word.
  - iv. The teacher tells the learners to look at the words for one minute and then she rubs out the words and the learners have to use their memory to write out the list again.
 All these exercises are simply ways of focusing the learners' attention on the new vocabulary that occurred during the week so it is more likely to be remembered.

This article has attempted to show that statistical information about vocabulary has implications for teaching and learning vocabulary. One of the principles underlying this is that learning items deserve attention not because of their difficulty but because of their usefulness. Individual high frequency words occur often in many different contexts and thus deserve attention. Low frequency words occur rarely but make up a significant part of a text. They require generalizable strategies which will allow learners to deal with them as they occur.

## Bibliography

- Campion, M.E. and Elley, W.B. **An Academic Vocabulary List** NZCER, Wellington (1971)
- Carroll, J.B., Davies, P. and Richman, B. **The American Heritage Word Frequency Book** American Heritage Pub. Co., New York (1971)
- Eaton, Helen S. **An English-French-German-Spanish Word Frequency Dictionary** Dover Publications, New York (1940)
- Kucera, H. and Francis, W.N. **A Computational Analysis of Present-day American English** Brown Univ. Press, Providence (1967)
- Praninskas, Jean **American University Word List** Longman, London (1972)
- Thorndike, E.L. and Lorge, I. **The Teacher's Word Book of 30,000 Words** Teachers' College, Columbia Univ. (1944)
- West, M. **A General Service List of English Words** Longman, Green & Co., London (1953)

I.S.P. Nation  
English Language Institute  
Victoria University

## Further Reading

**Learning Vocabulary 1982**, a 61 page publication, is available from the Secretary, English Language Institute, Victoria University, Wellington. In it, Paul Nation writes about recent research into foreign language vocabulary learning. He deals with results of international research into such imponderables as the place of the first language, learning through context, lists, mnemonics etc. There are also more details of vocabulary learning strategies including a check-list for guessing meaning through context. Up to this point (page 22), the publication will be of interest to teachers of all languages.

The second part deals with English and included are comprehensive tables of prefixes, suffixes and roots. These are aimed at more advanced levels but teachers of E.S.L. at all levels will find the content useful for reference.

Roger Stokell