Teaching and learning vocabulary

Paul Nation
(LALS, Victoria University of Wellington, New Zealand)

This talk describes the different levels of vocabulary and what teachers should do about them. It distinguishes high frequency, academic, technical and low frequency vocabulary. It is suggested that high frequency and low frequency words need to be dealt with in different ways with attention being given to high frequency vocabulary in both meaning focused and language focused ways. From the teacher's perspective, low frequency vocabulary should be focused on using strategies rather than direct teaching. A range of strategies and activities are examined for the teaching and learning of vocabulary.

I. What vocabulary?

This talk looks at the following questions. What vocabulary do my learners need? How should it be dealt with in class? How should it be learned?

Vocabulary can be divided into four levels largely on the basis of how often it occurs in the language (its frequency) and how widely it occurs (its range).

II. High frequency words

The most important group of words is the high frequency words of the language. These words occur very frequently in all kinds of uses of the language. They are needed in formal and informal uses of the language, in speech and in writing, and in novels, conversation, newspapers and academic texts. Most lists of high frequency words consist of around 2,000 word families. If you look at such lists you may notice several things. Firstly, most of the words are quite short. Secondly, 169 of the 2000 word families are function words such as a, at, because, four, I. All the rest are content words (nouns, verbs, adjectives, adverbs). Thirdly, these words are ones that even very young native speakers of English are likely to know. They are very common words that we need every day we use English. A list of high frequency words can be found in the Vocabulary Resource Booklet available free from http://www.victoria.ac.nz/lals/staff/paul nation/nation.aspx.

Figure 1 is a part of a conversation. All of the words which are not marked by bold (AWL) or italics (low frequency) are high frequency words. Notice that a

very large number of the words in the text are high frequency words, that is, words from the most frequent 2,000 words of English. In most texts around 80% or more of the running words are from the most frequent 2,000 words of English. In friendly conversation, over 90% of the running words tend to be from the high frequency words of English.

---

yeah I just had a look on the board over there and it came out

that I was accepted into *chem* one three oh

good

and have you looked at the other what was the other one

oh *info* it's not there

yeah that number's *info* mm

right you'll need to go to

who's who's the main **contact** person on the board

you need to go and see that person whoever it is

oh is there a name on the board   oh yeah

yes and the room number and get that changed to a c

and they will also tell you what *lab* stream you're in too

right yep

and *lab* stream

so get that signed then you'll need a signature from

who's your **major** person

it's gonna be was gonna be **economics** so I'll

have to go up there

oh actually you're a *BA*?

Yeah   (from the Wellington Corpus of Spoken English)

Figure 1: Types of vocabulary in conversation

III. Academic words

When we look at the special purposes for which we use language, it is usually possible to find a vocabulary that consists of words that are not from the most frequent 2,000 words but which are frequent and widely used within that specialized area. It is thus possible to make lists of newspaper vocabulary, the vocabulary of children's books, or the vocabulary of very informal conversation. The most well researched special purposes area for vocabulary is academic

writing. Academic writing includes academic text books like economics or geography texts, academic articles such as articles from journals, and laboratory manuals.

Coxhead (2000) looked at academic texts from four faculty areas Arts, Science, Commerce and Law, each of which included texts from seven different subject areas making a total of 28 different subject areas. Using a computer program (available free at http://www.vuw.ac.nz/lals/staff/paul nation/nation.aspx) she found the words that (1) were not in the mot frequent 2,000 words of English, (2) occurred at least 100 times in her 3,600,000 running word academic corpus, (3) occurred at least ten times in each of the four faculty areas, and (4) occurred in at least 15 of the 28 different subject areas. There were 570 such word families. She divided them into 10 groups (9 with 60 words, and 1 with 30 words). These 570 word families are called the Academic Word List (see Coxhead's home page http://language.massey.ac.nz/staff/awl/index.shtml).

Figure 2 is a part of an academic text. The unmarked words are from the most frequent 2,000 words of English. The words in bold are from the Academic Word List. The technical words are underlined. Typically the academic words make up around 8.5%   10% of the running words in academic texts, that is, ap   ximately one word in every ten comes from the Academic Word List.

It is useful to distinguish two different but related meanings of '**interaction** '. First, **interaction** can be viewed as the social behaviour that **occurs** when one person **communicates** with another. **Interaction** in this sense is *interpersonal.* It can **occur** face-to-face, in which case it usually takes place through the *oral* **medium,** or it can **occur** as **displaced** activity, in which case it generally **involves** the written **medium.** In some sense, *oral interpersonal* **interaction** is basic to human **communication,** as all **communities,** whether *literate* or not, *engage* in it. Also, it **constitutes** the **primary** purpose for our species - **specific** *language* **capacity** and the means by which it has developed both *photoelectrically* and *ontogenetically.* This book will concern itself mainly with *oral* **interaction** of the *interpersonal* kind. Second, **interaction** can **occur** inside our minds, both when we *engage* in the kind of '*private speech*' discussed by *Vygotsky* (1978), and, more *covertly,* when different *modules* of the mind **interact** to **construct** an understanding of or a **response** to some **phenomenon.** In reading, for example, we draw **interactively** on our ability to *decode* print, our stored knowledge of the language we are reading and the content *schemata* through which our knowledge of the world is organized.

> **Interaction** of this kind, then, is _intrapersonal._ This book will also consider _intrapersonal_ **interaction** and its relation to second language (L2)learning.

Figure 2: Types of vocabulary in academic texts

Academic words do not occur so often in other kinds of language use. Less than 2% of the running words in conversation are from the Academic Word List. The words in the Academic Word List are very important for learners who will use English for academic study either in upper secondary schools (senior high school) or in universities or technical institutes.

IV. Technical words

Academic words can occur in all kinds of academic subject areas like Botany, Politics, Accounting, or Family Law. There are even more special purposes words and these are the words that are very common in one particular area, such as the vocabulary of Physics or the vocabulary of Computing. When we see these words we can usually guess what subject area they come from if we know a little bit about that subject area. What subject areas do you think these words come from?

cardiovascular, ligament, costal, cartilage, jugular

ROM, mouse, pixel, cursor

The first group are from Anatomy (Medicine would have been a correct answer), and the second group are from Computing. Most technical words occur only in a specialized area, but some technical words can occur in other areas, some with the same meaning and some with different meanings. _Bypass_ and _neck_ occur very frequently in medicine and are technical words in that area. They also occur outside that area.

In a specialized text, such as an Economics textbook, technical vocabulary occurs much more frequently than in texts in other areas and than in a collection of specialized texts. In Anatomy around 30% of the running words are technical words (Chung and Nation, 2003, 2004). That is, almost one word in every three. Some of these technical words are also in the 2,000 high frequency words but have a specialized use in a particular area. Some may be from the Academic Word List, but many are likely to be peculiar to that subject area. In Figure 2, the technical words are underlined.

Technical words are clearly very important for anyone who specializes in a particular area. There have not been many statistical studies of technical vocabulary, but at a best guess, it would seem that at least 20% of the running

words in most technical texts are likely to be technical words. We do not know a lot about technical vocabularies but they probably range in size from around 1,000 words to 5,000 words.

V. Low frequency words

So far we have looked at three groups of words   high frequency words, academic vocabulary, and technical vocabulary. The last group we will look at is the biggest group by far, the low frequency words. Low frequency words are a diverse group. They include words that were not quite frequent or wide range enough to get into the high frequency words (*abort, absorb, accelerate, accent, accusation, acid, acre*), technical words from other areas (one person's technical vocabulary is another person's low frequency vocabulary), and words that just occur rarely. If only single word items are counted, English probably has a low frequency vocabulary of over 100,000 word families. Most native speakers probably have a vocabulary of around 20,000 word families. Have a look at a very large dictionary like *Webster¡s 3ʳᵈ International Dictionary* or *The Shorter Oxford English Dictionary* to see the low frequency words.

Figure 3 is a part of a newspaper text. The high frequency words are unmarked, the academic words are in bold, the low frequency words are in italics.

Typhoon Dianmu, which influenced weather conditions on the *peninsula* over the weekend, brought heavy rain that left some 400 people homeless and caused the blocking of roads and railways.

Though the rain stopped Monday afternoon, the rainy season is *forecast* to start Friday.

Dianmu struck western *Japan*, but produced over 300 *millimeters* of *rainfall* in some **regions** of Korea, particularly in the southern parts of the *peninsula*.

Munkyong, in North Kyonsang Province, reported a total of 366mm of rain from Friday through Monday. *Chongju* received 332 *millimeters* of *rainfall*, *Chechon* 321, *Taejon* 296, *Tonghae* 300, *Hampyong* 248, and *Chonju* 243, according to the *Korea Meteorological* **Administration**.

The rain *inundated* 222 homes, leaving 522 people homeless nationwide. It also swept over a total of 29,794 *hectares* of *farmland* and caused power failures to 26,800 houses across the nation, according to the Central *Disaster* and Safety *Countermeasures Headquarters.*

> Two people lost their lives while four others went missing.
>
> Some parts of the railroads between *Chongnyangni* in *Seoul* c *Kangnung* were *inundated* around 7 p.m. Sunday delaying trains for one to eight hours. Rail services recovered at 8 a.m. the following day.
>
> The rain stopped Monday, and clear skies are *forecast* for today. But it will start to rain again in *Cheju* Island from Thursday and nationwide from Friday, with the beginning of the *seasonal* rainy *spell*.
>
> (From The Korea Times on the web)

Figure 3: Different types of vocabulary in a newspaper text

In friendly conversation about 5% of the running words are low frequency words, in newspapers about 10%, and in academic texts about 10%.

We have not dealt with one important group of words in this description of vocabulary levels   proper nouns. Proper nouns are the names of people,   :es and products. They are partly signalled by being written with an initial capital letter. In a mixed corpus, they make up about 4% 5% of the running words. It might be argued that proper nouns are the technical words of novels and newspaper reports. In the coverage figures reported here we have assumed that proper nouns do not require previous learning in the second language but this may be a dangerous assumption.

VI. Deciding which words are high frequency, academic, technical or low frequency

Researchers studying the occurrence of vocabulary have used frequency, range and dispersion data to help decide how vocabulary should be classified. Here is data on two words from the British National Corpus.

```
evidence        NoC    %       215     100     0.93
Unix    NoP    %       42      34      0.28
```

The British National Corpus (BNC) consists of 100,000,000 running words of English of which 10% are spoken English and 90% written English. The figures are   from   Leech,   Rayson   and   Wilson   (2001; http://www.comp.lancs.ac.uk/ucrel/bncfreq/flists.html).

We can see that the word *evidence* had a frequency of 215 per 1,000,000 running words. It also has a wide range because the range figure of 100 indicates that it occurred in all of the 100 subdivisions of the corpus (each one containing 1,000,000 running words). The last figure is the dispersion figure. This indicates how similar

the frequency figures were in each of the 100 subdivisions. If the word occurred only once or twice in some divisions and over 100 times in others, this dispersion figure would be low (see *colitis* below with 0.35). If it occurred with roughly the same frequency in each of the subdivisions, the dispersion figure would be high (0.93 for *evidence*). We can now use the ideas of frequency, range and dispersion to describe the nature of the different kinds of vocabulary. High frequency words have very high frequencies of occurrence, they are wide range, and they have high dispersion.

two   Num   %   1563   100   0.97

Academic words have high frequencies of occurrence and reasonable dispersion in a wide range of academic subject areas.

Technical words have much higher frequencies of occurrence, limited range, and low dispersion.

colitis   NoC   %   10   14   0.35

If you want to check texts you are interested in to see which words are high frequency, academic or low frequency, then you can go to these web sites

Frequency Level Checkers
                    http://language.tiu.ac.jp/flc/
                    http://www.er.uqam.ca/nobel/r21270/textools/web_vp.html
                    http://www.edict.com.hk/textanalyser/
                    http://www.missouri.edu/~youmansc/vmp/index.shtml
or use the RANGE program mentioned above.

VII. How should the various types of words be dealt with in class?

The high frequency words of English have the following characteristics.
1   Each high frequency word occurs very often so the effort of learning it will be repaid by plenty of opportunities to meet and use it.
2   The high frequency words are useful no matter what use is made of English. They are important for both receptive and productive use, for both oral and written use, and for both formal and informal use. They are useful when reading newspapers, novels and academic texts. They are useful when writing assignments, writing letters, taking part in conversation, and giving formal talks.
3   Because of their frequency and wide range they make up a very large proportion of the running words in all kinds of texts and language use.

4   They are a relatively small group of words (2,000) that could be covered in a school teaching programme over three to five years.

For these reasons, high frequency words deserve attention from the teacher. They deserve attention across the four strands of a course including deliberate teaching. The activities in a language course can be classified into the four strands of meaning focused input (learning through message focussed l ning reading), meaning focused output (learning through message focussed sp ing and writing), language focused learning (deliberate teaching and deliberate stud) and fluency development. In a well designed course there should be an even balance of these strands with roughly equal amounts of time given to each strand. The high frequency words are important enough to deserve time in class.

For learners with academic purposes, the academic words are like high frequency words and they deserve similar attention. If learners do not have academic purposes but wish to read newspapers in English, then there is a better return for learning the 570 words in the Academic Word List than learning the third 1,000 words of English.

For learners with academic purposes, technical vocabulary is also important but this is probably best learned while studying the content matter of the particular specialist area. The English teacher can help with this learning by encouraging learners to make connections between technical and non technical uses of the same words, and by using word parts to help remember the meanings of words.

Learners also need to learn low frequency words, after they have a good control of the high frequency, academic and technical words. Teachers however should not spend valuable classroom time teaching low frequency words. Low frequency words have the following characteristics.

1   Each word does not occur very often. For many low frequency words you may need to wait months or years before meeting the same word again.
2   Most low frequency words have a very narrow range. They are not needed in every use of the language.
3   The low frequency words make up a very small proportion of the running words in a text, once proper nouns are excluded usually less than 10% of the running words.
4   They are a very large group of words, over 20,000 for adult native speakers and over 100,000 in the language as a whole (Nation, 2006).

For these reasons, low frequency words do not deserve classroom time. The

teacher should not spend valuable time teaching them. Learners eventually need to learn them, so the teacher should spend time on the strategies needed to deal with these words. The four important strategies are

1   guessing from context,
2   learning from word cards,
3   using word parts,
4   using a dictionary.

These strategies are in order of importance. There are effective and ineffective ways of using these strategies, and there are several skills involved in learning and using them. Learners benefit from getting some training in these strategies. So, when low frequency words come up in class, the teacher should use that as an opportunity to give strategy training. There is also good value in regularly spending time on these strategies. These strategies are very important because they allow learners to deal with the thousands of low frequency words they will meet in their use of English.

From the teacher's viewpoint, the distinction between high frequency and low frequency vocabulary is very important, because the teacher should deal with these two groups of words in different ways. It is thus important for the teacher to know where the learners are at in their vocabulary learning. Do they already know the high frequency words? Do they know the words in the Academic Word List? Should the focus be on high frequency words or low frequency words? There are vocabulary tests to help answer these questions (Schmitt, Schmitt and Clapham, 2001).

We have now looked at the different levels of vocabulary, high frequency, academic, technical and low frequency, and have looked briefly at the kinds of words in each of these levels. We have also looked at the high frequency/low frequency distinction and have suggested that teachers need to deal with these words in very different ways. This knowledge should allow a teacher to plan the broad outlines of a vocabulary development programme that is suited to the needs of a particular group of learners.

① **References**

② Beglar, D. and Hunt, A. (1999) Revising and validating the 2000 word level and the university word level vocabulary test. Language Testing 16, 2: 131-162.

③ Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. Reading in a Foreign Language, 15(2), 103-116.

④ Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. System, 32(2), 251-263.

⑤ Coxhead, A. (2000) A new academic word list. TESOL Quarterly 34, 2: 213-238.

⑥ Leech, G., Rayson, P. and Wilson, A. (2001) Word Frequencies in Written and Spoken English Longman, Harlow.

⑦ Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? Canadian Modern Language Review, 63(1), 59-82.

⑧ Nation, I.S.P. (2001) Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

⑨ Nation, I.S.P.(ed.) (1986) Vocabulary Lists E.L.I. Occasional Publication No.12, Victoria University of Wellington Student Notes Centre.

⑩ Nation, I.S.P. and Beglar, D. (2007) A vocabulary size test. The Language Teacher 31, 7: 9-13.

⑪ Schmitt, N., Schmitt, D. and Clapham, C. (2001) Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. Language Testing 18, 1: 55-88