Editors' introduction

Conventions, terminology and an overview of the book

Over the last 20 years vocabulary research has grown from a 'Cinderella subject' in foreign language teaching and research, to achieve a position of some salience. Vocabulary is now considered integral to just about every aspect of language knowledge. With this development have come standard and widely used tests, such as vocabulary size and lexical richness measures, and very commonly accepted metaphors, such as 'a web of words' to describe the mental lexicon. Less widely known outside academic circles, however, is the extensive work on learners' lexis and the utility, reliability and validity of the tests we use to measure and investigate vocabulary knowledge and growth. Vocabulary is a lively and vital area of innovation in academic approach and research. The penalty we pay for working in so vital a subject area is that even recent, and excellent, surveys of the field are rapidly overtaken by new ideas, fresh insights in modelling and testing, a healthy re-evaluation of the principles we work under, and an ever-growing body of empirical research. The intention of this volume, therefore, is to place in the hands of the reader some of these new ideas and insights. It brings together contributions from internationally renowned researchers in this field to explain much of the background to study in this area, and reconsider some of the ideas which underpin the tests we use. It introduces to a wider audience the concerns, new approaches and developments in the field of vocabulary research and testing.

To place these ideas in context, and to provide a point of entry for non-specialists in this field, this introduction will survey the conventions and terminology of vocabulary study which, if you are not familiar with them, can make even simple ideas impenetrably difficult. The background this introduction provides should allow the chapters which follow to be placed in context and help to explain why the concerns they address are of importance to researchers. The second half of this introduction provides summaries of the chapters.

Conventions and terminology

What is a word?

One of our colleagues used to begin lectures on vocabulary learning by asking his audience how many words they thought they knew in English. Most people had no idea of course, and had to guess, and the answers they suggested varied enormously – from 200 words to many millions. These extremes are unusual but in truth it was a question without a clear answer, because the answer depends on what you mean by a word and therefore what your unit of counting is. According to context and need, researchers can consider *types*, *tokens*, *running words*, *lemmas*, and *word families* as words.

In one sense it is obvious what a word is. Words are the black marks you are reading on this page and you know when one word ends and another one begins because there are spaces between words. There are occasions when it is appropriate to use a definition of this kind in making word counts, for example, in counting the number of words in a student's essay or the number of words in the huge corpus that a researcher will collect so that they can use real examples of word use. When counting words in this way we often refer to them as *tokens* so it is clear what we are talking about. Sometimes we also refer to *running words* with much the same meaning, for example, if you consult a dictionary corpus you may be presented with the information that the word *maunder* occurs on average only once every several million running words.

In addition to knowing the number of words in a text or a corpus, researchers sometimes want to know the number of *different* words that occur in a given text. The terms *tokens* and *types* are used to distinguish between these two ways of counting. *Tokens* refers to the total number of words in a text or corpus while *types* refers to the number of different words. In the sentence:

The cat sat on the mat

there are six *tokens* (a total of six words), but the word *the* occurs twice so there are only five *types*.

But there are problems even with a catch-all definition of this kind. How do you count contractions such as *don't*, *it's* or *won't*? Should they be counted as single words or two? Is the number at the top of this page a word or not? Are the names we have put on the title page of this book words? And if you are counting words in speech rather than writing, how do you count the *ums* and *ers* which always occur? Practice can vary according to the needs of the researcher but often,

numbers, proper nouns and names, and false starts and mistakes are excluded from word counts.

Once you start counting the number of words a person knows more difficulties raise their heads. If a student learns the verb *to work*, for example, this will involve learning the form *works* for use with the third person singular in the present simple tense, the form *worked* for use in the simple past, and *working* for use with continuous tenses. The question arises whether the learner has learned one word or four here. These inflections or changes to the root form of the verb are highly regular and can be applied to most verbs in English. Provided a few simple rules of grammar are known, learners only need to learn a new root form to have these other forms at their disposal and available for use. It is often convenient, therefore, to think of all these word forms as a single unit since they do not have to be learned separately by the learner; learning the root form means all the others can be deduced from it and will therefore also be known. This has the profound advantage of reducing the numbers of words we have to work with in describing vocabulary knowledge to manageable levels: to a few thousand or tens of thousand instead of hundreds of thousands. A collection of words such as *to work*, *works*, *working*, *worked*, comprising a root form and the most frequent regular inflections, is known as a *lemma*. Where a noun has a regular plural formed by adding *-s*, as in *orange* and *oranges*, for example, these two words would also form a single lemma. In most word-frequency counts and estimates of learners' vocabulary sizes, the lemma is used as the basis of counting, and *work, works, working* and *worked* would be counted as just one lemma. Rather confusingly, lemmas are often called words, and researchers are not always consistent in their use of terminology. In both Nation's vocabulary level's test (1983) and Meara and Milton's *X-Lex* (2003a) word knowledge is tested in what are called 1,000-word frequency bands. In fact, the researchers used lemmatised word lists and these should have been referred to as 1,000-lemma frequency bands.

Some estimates of a speaker's vocabulary size, however (for example, Goulden, Nation and Read's (1990) estimate of 17,000 words for educated native speakers of English) use a larger unit still and are actually estimates of the number of *word families* a person knows. The forms of a word which can be included in a lemma are fairly limited. But words often have lots of other forms which are clearly related to the root form. The lemma *work*, for example, includes *working*, *works* and *worked* but does not include *worker* although this is obviously a derived form which is very closely

related. The lemma *govern* would include *governs*, *governing* and *governed* but not *governor* or *government*. Closely related words like this would be called a *word family*. Clearly, estimates of size based on the *lemma* and on the *word family* will be quite different.

At first sight this may appear confusing and quite unnecessarily complex. Certainly, researchers often contribute to the confusion both by being unclear as to the units they use, and by adopting idiosyncratic definitions. The divisions between a word, a lemma and a word family are not entirely arbitrary, however, and are based on Bauer and Nation's (1993) frequency-based groupings of affixes in English. *Lemmas* will generally be words made by using affixes from the top three groups, and *word families* from the top six. Thus, *lemmas* would include only the most common affixes and would not generally involve changing the part of speech from that of the head word, while a *word family* would be much more inclusive. The *lemma* of a word such as *establish*, for example, would include *establishes*, *establishing*, and *established* but not *establishment* which would change the part of speech and includes a suffix at Level 4 in Bauer and Nation's hierarchy, while the *word family* would include *establishment* and many other words using less frequent affixes such as *interestablishment* or *antiestablishment*. Further, this hierarchy of word units is not the product of whim on the part of researchers but rather a result of the need to reduce the figures we work with to manageable proportions. In measuring distance we use millimetres, centimetres, metres and kilometres, to name just a few, according to the size of what is being measured, and in measuring vocabulary we are behaving no differently.

What is 'knowing a word'?

If defining a word has presented problems, then deciding when a word is actually known is no easier. There are a number of qualities which might be included in the definition of *knowing* and this has been added to over the years. Nation's list, in Table 1, is the latest and most comprehensive incarnation.

Depending on how you define *knowing*, you will have very different ideas about what constitutes a learner's knowledge of words, and statistical counts of a learner's vocabulary size will then also vary according to the definition of *knowing* used. Perhaps the most basic, catch-all definition would be simple, passive, word recognition; the learner recognises the form of a word and that it is a word rather than a meaningless jumble of symbols. This aspect of knowing is clearly identified in Nation's table. There are several tests (e.g. Meara

Table 1 What is involved in knowing a word? (from Nation, 2001: 27)

Form	spoken	R	What does the word sound like?
		P	How is the word pronounced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concepts and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this word make us think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of word occur with this one?
		P	What words or types of words must we use with this one?
	constraints on use	R	Where, when and how often would we meet this word?
		P	Where, when and how often can we use this word?

R = receptive, P = productive.

and Jones's EVST, 1990; Meara and Milton's *X-Lex*, 2003a) which use this definition of *knowing*. In principle, a calculation made using this definition will surely include every other kind of knowledge since, presumably, a learner could not reasonably use, attach a meaning to or find a correct collocation for something they do not even recognise as a word. Most of the tests we use to calculate vocabulary size are based on written forms of knowledge and these predict a range of reading- and writing-based language abilities as well, but the ability to recognise or use the spoken form of a word is much less well investigated. Interestingly, initial results from studies using phonologically based vocabulary size tests (Milton, 2005)

suggest that aural word recognition predicts oral proficiency particularly well. This ties in with Daller and Huijuan Xue's chapter in this volume (Chapter 8) which addresses the problems of finding a good measure of lexical knowledge to tie in with oral proficiency.

A second very common definition of knowing a word can be found within the 'Meaning' section of Nation's table. This rests on the idea that a word is known if the learner can attach a meaning, such as an explanation or a translation, to a foreign language word. Calculations of vocabulary knowledge and size made on this basis ought to be smaller than those made on the basis of passive word recognition. Every learner must be familiar with the sensation of encountering a word they know they have seen before but cannot, for the moment, attach to a meaning. It seems this aspect of knowledge can be surprisingly fragile in the foreign language learner's vocabulary. The link between form and meaning can disappear quite suddenly and without explanation and, just as suddenly, reappear. The chapters by Meara and Wilks (Chapter 9) and by Schur (Chapter 10) investigate the applicability of various kinds of network theory to vocabulary, and begin to make this kind of phenomenon explicable but, as their chapters show, this work is still in its infancy. It is a phenomenon which also underlies the questions encountered in Chapter 3 by Eyckmans, Van de Velde, van Hout and Boers and by Fitzpatrick in Chapter 6 where differences in translation and receptive test scores challenge easy interpretation.

Nation's table of what is involved in knowing a word draws attention to a further distinction, that of *receptive* and *productive* or *passive* and *active* word knowledge: indicated by R and P in column three (see Table 1). The distinction here lies in the difference between the words you can handle in the context of reading or listening to speech, and those you can call readily to mind when you need to speak or write in the foreign language. Usually the additional context information which comes with written or spoken language means that a learner's passive or receptive vocabulary appears to exceed the productive or active vocabulary. The relationship between the two types of knowledge is not clear, and may vary according to a variety of individual learner characteristics or the type of test used. But it is quite extensively researched, going back to Stoddard in 1929. Estimates vary but the range of studies reviewed in Waring (1997) suggest that productive vocabulary size is about 50% of receptive vocabulary size; and presumably one is a subset of the other. There are, of course, methodological problems inherent in measuring these two different kinds of vocabulary in a way which is strictly equivalent and these problems haunt several of the contributors to this

volume such as Richards and Malvern (Chapter 4), and van Hout and Vermeer (Chapter 5). These methods are considered in more detail later on in this introduction.

Other aspects of word knowledge seem much less well researched and standard tests are lacking, in some cases we even lack an agreed approach to testing. For example, in his section on 'Form' (Table 1) Nation suggests that word knowledge can include knowledge at the level of the morpheme. Our concentration on calculating word knowledge using the *lemma* or the *word family* as the basic unit means that our tests cannot tell us about knowledge at this level of detail. But the testing problems experienced by Eyckmans et al. described in Chapter 3, may result to some extent, from learners' abilities to make educated guesses about the meaning of words from their different parts or components. Our concern is that this kind of guesswork may destabilise some tests of vocabulary knowledge and make the scores they produce less useful than we may think they are. Again, knowledge of a word's collocations, connotations and preferred associations is an area where we struggle to find a single, simple way of characterising this knowledge in a way in which it can be usefully quantified and tested. Further, our concentration on tests which use the lemma, and the fact that we often investigate infrequent vocabulary, means that all of the most frequent linking words tend not to be investigated. Such information falls below the radar of the tests we use. Chapters 9 and 10 by Wilks and Meara, and by Schur respectively, are a direct attempt to suggest models of analysis and testing methods which might help fill in these gaps in our knowledge.

What is the lexical space?

It is clear from this discussion that vocabulary knowledge is complex and multi-faceted. The qualities we investigate are not easily described or tested and we tend to resort to analogy and metaphor to try to illuminate the way words are learned and stored. One such idea is that of *lexical space* where a learner's vocabulary knowledge is described as a three-dimensional space, where each dimension represents an aspect of knowing a word (see Figure 1).

In Figure 1 the horizontal axis represents the concept of *lexical breadth* which is intended, in essence, to define the number of words a learner knows regardless of how well he or she knows them. This would include the 'Form' and the *form and meaning* elements of Nation's table. Vocabulary size tests, passive/receptive style tests and translation tests are all tests of lexical breadth, although they may

produce varying estimates of size and knowledge. Chapters 2 and 3 by Milton and Eyckmans et al. respectively, are directly concerned with how to make estimates of vocabulary breadth.

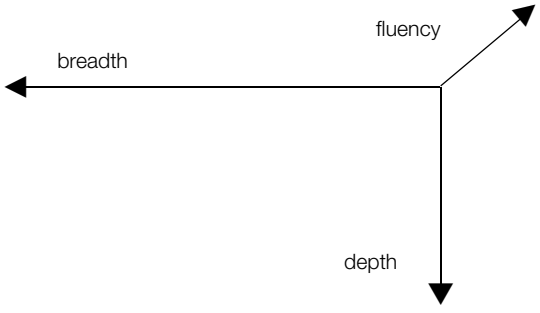

Figure 1 The lexical space: dimensions of word knowledge and ability

The vertical axis in Figure 1 represents the concept of *lexical depth* which is intended to define how much the learner knows about the words he or she knows. This would include the elements of *concepts and referents*, *associations*, *grammatical functions*, *collocations* and *constraints on use* from Nation's table (Table 1). These elements tend to be tested separately, probably because this is a disparate list of word qualities, for which we have not as yet succeeded in pinning down a unifying idea or model which can provide the basis of a comprehensive test of depth. This is not for want of trying, however, and the precise relationship between the lexicon and grammar has been the subject of considerable research (e.g. Hunston and Francis, 2000). This area might properly be the subject of an entire but separate volume. Space in this volume permits only limited reference to this area based on a further metaphor, that of a web of words, which is often used to describe this axis and the way the words interact with each other. Chapters 9 and 10 by Wilks and Meara and by Schur, deal with precisely this issue and investigate the possibility of turning this metaphor into a model of lexical depth which can be empirically tested with real language users. Meara and Wolter (2004) have developed a test which allows learners to activate these webs of grammatical and lexical knowledge so that a score can be assigned to it. At first sight this looks like a promising innovation but it is early days.

The final axis is that of *fluency* and this is intended to define how readily and automatically a learner is able to use the words they know and the information they have on the use of these words. This might involve the speed and accuracy with which a word can be

recognised or called to mind in speech or writing. It would probably be true to say that we have no widely used or generally accepted test of vocabulary fluency. Some very promising ideas are emerging (for example, Shiotsu, 2001) but it is interesting to note that this field is still somewhat inchoate, so much so that no papers were presented at the Vocabulary Workshop giving rise to this volume.

These three axes define the lexical space and, in principle, it becomes possible to locate a learner's vocabulary knowledge within this space. Some learners may have large vocabularies but are very limited in the speed and ease with which they can recall these words and put them to use communicatively. These learners ought to be placed well along the breadth axis but less far along the fluency or depth axes. Other learners may appear to have different characteristics and possess comparatively few vocabulary resources but considerable fluency in calling these to mind and using them in communication. These learners would occupy a different location in the lexical space, less far along the breadth axis but further along the fluency axis. This way of describing lexical knowledge is both attractive and convenient as it makes it easier to define, briefly, the nature of a test or what defines a learner's knowledge of words. But the notion of lexical space is still fundamentally a metaphor with all the drawbacks that go with that. The nature of the lexicon is not really a three-dimensional space and attempts to turn the metaphor into a detailed model which can be tested empirically run into trouble. The precise nature of the depth axis is a case in point and Read, who uses the term in his (Read, 2000) review of the field, questions the nature of this axis in later work (Read, 2004).

What are the conventional ways of measuring knowledge in this lexical space?

While we lack a comprehensive range of tests across the whole field of vocabulary knowledge, we do have a small number of well-established tests in the area of vocabulary breadth and, more particularly, passive receptive vocabulary knowledge. At first sight, testing how much a person knows from the enormous number of words in the English language (for example) appears a daunting task. There are tens or even hundreds of thousands of words, depending on how you define *word*, potentially available for learners to acquire, and taking a reasonable sample of these words to test a learner's knowledge should be difficult. A learner may only know a few of these words so the task is like searching for a needle in a haystack. Nonetheless, it does appear possible to compile a representative

sample of words and this is because of the way words are used in language. Words do not occur randomly in speech or writing and some occur very much more frequently than others. Thus, verbs such as *make* or *do*, prepositions such as *in* and *on* and pronouns such as *I* or *you* are used a lot by every speaker, while other words such as *anamnestic* and *mitogenic* are very uncommon and might not be used at all even by native speakers except in the most specialised of situations. Where learners are exposed to a new language, therefore, they encounter some words much more often than others, and some words they never encounter at all. Unsurprisingly, learners are more likely to learn the frequent words than the infrequent words, or words so rare they never even see or hear them. Tests such as Nation's Levels Test and Meara and Milton's *X-Lex* take advantage of this reality to produce samples of the most frequent words in a given language to make credible estimates of overall vocabulary knowledge in learners. A good test is possible because it can be focused on areas where learning is likely to occur, rather than on areas where there is no knowledge to be detected. Although these tests work well, the frequency effect is an assumption which does not appear to have been empirically tested and the second chapter in this volume addresses this issue directly, asking not only whether or not the effect really exists, but also how strong it is and whether all learners are affected equally.

The idea of counting the frequency of words in a language tends to be thought of as a recent innovation and something we can now do because we have computers which can process millions of words. But the idea is, in reality, very old and goes back at least to the study of the writings of the Prophet Mohammed in the eighth century. The earliest counts made for pedagogical reasons were made in the 1930s and 1940s and these still provide useful lists, but modern resources such as the Bank of English and the British National Corpus now make very large corpora available to researchers and other organisations and these can be broken down so it is possible to investigate, say the frequencies of only written English or of only spoken English. Modern tests tend to be based on corpora and frequency counts of this kind and, for convenience draw on the most frequent vocabulary only, often in 1,000-word bands.

While the Levels Test and *X-Lex* estimate knowledge within the same area of the lexical space and are based on frequency counts of English, they are nonetheless two very different tests. *X-Lex*, for example, samples the 5,000 most frequent words of English drawing 20 words from each of the five 1,000-word frequency bands within this list and uses this to make an estimate of the number of words