1 *Introduction*

- What is a word?
- What does it mean to know a word?
- How many words are there in English?
- How many of these words do I know?

> The White Rabbit put on his spectacles. "Where shall I begin, please your Majesty?" he asked. "Begin at the beginning," the King said, very gravely, "and go on till you come to the end: then stop."
>
> – Lewis Carroll, *Alice's Adventures in Wonderland,* p. 106

The advice given in this quote from *Alice in Wonderland* seems to be appropriate for an introductory text, so to start at the beginning we must consider what we mean by vocabulary. The first idea that probably springs to mind is *words,* a formulation that is admirably adequate for the layperson. But for anyone interested in exploring the subtlety and magic of lexis, the term *word* is too general to encapsulate the various forms vocabulary takes. Consider the following items:

die
expire
pass away
bite the dust
kick the bucket
give up the ghost

The six examples are synonymous, with the meaning "to die." (*Synonyms* are words that have approximately the same meaning.) However, they are made up of from one to four words. *Die* and *expire* are single words, *pass away* could probably best be described as a phrasal verb, and the last three are idioms. (An *idiom* is a string of words which taken together has a different meaning than the individual component words. Similarly, a *phrasal verb* is made up of a verb plus one or more other words, which also has an idiosyncratic meaning compared to the component words.) Thus there is not necessarily a one-to-one correspondence between a meaning and a single word. Very often, in English at least, meanings are represented by multiple

words. To handle these multiword units, the term *lexeme* (also *lexical unit* or *lexical item*) was coined. These three interchangeable terms are all defined as "an item that functions as a single meaning unit, regardless of the number of words it contains." Thus, all of the six examples above are lexemes with the same meaning.

In addition to the possible lack of correspondence between individual words and individual meanings, the term *word* also has difficulties with the various grammatical and morphological permutations of vocabulary. It is not all that clear whether *walk, walked, walking,* and *walks* should be counted as a single word or four. Likewise, are *stimulate, stimulative,* and *stimulation* the same word? In these examples, there is a *base, root,* or *stem* word that is the simplest form of that word. To this stem, affixes are added. If the purpose of the affixes is grammatical, then the resulting word is called an *inflection. Walked, walking,* and *walks* are inflections of the root word *walk.* However, if the affixes change the word class of a stem, the result is a *derivative.* Thus *stimulative* (adjective) and *stimulation* (noun) are derivatives of *stimulate* (verb). It is clear that although these words have different *orthographic* (written) shapes, they are closely related in meaning. Sets of words like these are referred to as *word families.* A word family is usually held to include the base word, all of its inflections, and its common derivatives. The term *lemma* is more restricted and includes only the base word and its inflections (Nation & Waring, 1997). This terminology allows us to get around the potential ambiguity of *word,* and to speak of vocabulary in more precise terms when necessary. Not only is this expedient, but there is evidence that the mind groups the members of a word family together, giving a psychological justification for using word families as a unit for counting and teaching (Nagy et al., 1989). (To enhance the accessibility of this book, I will use the term *word* unless more precise terminology is required to make a point.)

These distinctions may seem a bit trivial, but they are essential if we are to answer interesting questions such as "How many words are there in English?" and "How many words do native speakers know?" Scholars have produced widely varying answers to these questions, mainly because they used different definitions of what counted as a word. Let us look at these questions in a bit more depth, because the answers will determine to a large extent how we conceptualize and teach vocabulary.

Size of the English vocabulary

Reports of the size of the English language in the popular press have a very wide range: from 400,000 to 600,000 words (Claiborne, 1983, p. 5), from

a half million to over 2 million (Crystal, 1988, p. 32), about 1 million (Nurnberg & Rosenblum, 1977, p. 11), and 200,000 words in common use, although adding technical and scientific terms would stretch the total into the millions (Bryson, 1990). This discrepancy is due largely to differing definitions of a word, and so a study attempted to produce a more reliable estimate by using word families instead of words as the unit of counting. Goulden, Nation, and Read (1990) counted the number of word families in *Webster's Third New International Dictionary* (1963), which is one of the largest nonhistorical dictionaries of English. Dictionaries such as this obviously cannot contain every current word family, but they are still the best resource available, and therefore estimates of the number of words in a language have usually been based on them. After excluding entries such as proper names and alternative spellings, Goulden et al. found that the dictionary contained about 54,000 word families. This is a huge number of items (remember that each word family contains several words), and so we as teachers must give up on the idea of ever teaching all of them to our students in a classroom situation. Only a fraction are likely to be acquired through formal study, leaving the pedagogical implication that any others will have to be acquired through simple exposure to the language or not acquired at all. This puts a premium on nonteaching activities that can bolster exposure to a language, with reading being an especially important source.

How many words do native speakers know?

Mastery of the complete lexicon of English (and probably any other language) is beyond not only second language learners but also native speakers. Still, the amount of vocabulary the average native speaker acquires is prodigious. This is shown by studies that have estimated that English native-speaking university graduates will have a vocabulary size of about 20,000 word families (Goulden et al., 1990; D'Anna, Zechmeister, & Hall, 1991). Nation and Waring (1997, p. 7) review vocabulary size studies and conclude that

> the best conservative rule of thumb that we have is that up to a vocabulary size of around 20,000 word families, we should expect that [English] native speakers will add roughly 1,000 word families a year to their vocabulary size. This means that a [L1] five year old beginning school will have a vocabulary of around 4,000 to 5,000 word families.

This would be consistent with a 20-year-old university student having 20,000 word families. In contrast to the impossibility of learning every word

in English, these figures indicate that building a native-sized vocabulary might be a feasible, although ambitious, undertaking for a second language learner.

Let us put the scope of this task into perspective. Imagine learning 15,000 to 20,000 telephone numbers. For each of these numbers you must remember the person and address connected with those numbers. This might be somewhat analogous to learning all of the various kinds of lexical knowledge attached to each word. Then, because these are word families and not single words, you would have to learn not only the single number, but also the home, work, and facsimile variants. Of course, vocabulary and phone numbers are not directly comparable, but the example does indicate the magnitude of achievement in learning a such a vocabulary.

Indeed, learning language is probably the most cognitively (mentally) challenging task a person goes through. But whereas the grammar of a language is largely in place by the time a child is 10 years old (Crystal, 1987, p. 243), vocabulary continues to be learned throughout one's lifetime. This is because the grammar of a language is made up of a limited set of rules, but a person is unlikely to ever run out of words to learn.

The complex nature of vocabulary

The mechanics of vocabulary learning are still something of a mystery, but one thing we can be sure of is that words are not instantaneously acquired, at least not for adult second language learners. Rather, they are gradually learned over a period of time from numerous exposures. This incremental nature of vocabulary acquisition manifests itself in a number of ways. We have all had the experience of being able to recognize and understand a word when we see it in a text or hear it in a conversation, but not being able to use it ourselves. This common situation shows that there are different degrees of knowing a word. Being able to understand a word is known as *receptive knowledge* and is normally connected with listening and reading. If we are able to produce a word of our own accord when speaking or writing, then that is considered *productive knowledge* (*passive*/*active* are alternative terms).

The assumption is that people learn words receptively first and later achieve productive knowledge. This generally seems to be the case, but in language learning there are usually exceptions. An example of knowing a word productively (at least in speaking mode) but not receptively in the written mode happened to me with a word connected with law. I had often

heard and verbally used a word describing the formal charging of a criminal with a crime or offense. I never had the occasion to write this word, although I assumed from its pronunciation (ɪn ˈdaɪt) that the spelling was "indite." At the same time I had occasionally seen the word *indict.* I did not know what it meant, but assumed that it rhymed with *predict.* It was only later that I figured out that *indict* was the spelling for the word I had used for years to talk about law.

This anecdote shows that framing mastery of a word only in terms of receptive versus productive knowledge is far too crude. I had good productive mastery over the spoken form of *indict,* but not over its written form. This suggests that we also need to consider the various facets of knowing a word. Of course, everyone realizes that a word's meaning must be learned before that word can be of any use. In addition, there is the practical matter of mastering either the spoken or the written form of the word before it can be used in communication. A person who has not thought about the matter may believe that vocabulary knowledge consists of just these two facets – meaning and word form. But the potential knowledge that can be known about a word is rich and complex. Nation (1990, p. 31) proposes the following list of the different kinds of knowledge that a person must master in order to know a word.

- the meaning(s) of the word
- the written form of the word
- the spoken form of the word
- the grammatical behavior of the word
- the collocations of the word
- the register of the word
- the associations of the word
- the frequency of the word

These are known as types of *word knowledge,* and most or all of them are necessary to be able to use a word in the wide variety of language situations one comes across. The different types of word knowledge are not necessarily learned at the same time, however. As we have seen, being able to use a word in oral discourse does not necessarily entail being able to spell it. Similarly, a person will probably know at least one meaning for a word before knowing all of its derivative forms. Each of the word-knowledge types is likely to be learned in a gradual manner, but some may develop later than others and at different rates. From this perspective, vocabulary acquisition must be incremental, as it is clearly impossible to gain immediate mastery of all these word knowledges simultaneously. Thus, at any point in time, un-

less the word is completely unknown or fully acquired, the different word knowledges will exist at various degrees of mastery.

Nation's list is convenient in that it separates the components of lexical knowledge for us to consider. But we must remain aware that this is an expedient, and in reality the different kinds of word knowledge are almost certainly interrelated. For example, frequency is related to formality (part of register) in that more frequent words tend to be less formal, and less frequent words tend to be more formal. Thus, greater awareness of formality is likely to be somehow related to awareness of a word's frequency of occurrence, even if this awareness is unconscious. It would therefore be logical to suspect that increasing knowledge of one word-knowledge aspect could help improve knowledge of related aspects. At this point, however, it would still be speculation, as research into these connections is just beginning (e.g., Schmitt & Meara, 1997; Schmitt, 1998b). Therefore, although we can use a word-knowledge perspective to *describe* "what it means to know a word," we will have to wait and see whether it can be used to *explain* lexical acquisition and processing. My own opinion is that word knowledge is a useful framework to discuss vocabulary, and so I have used it as a scaffold in this book to ensure that all of the major vocabulary issues are addressed. Thus, in Chapters 3 to 5, all of the word-knowledge types will be discussed in more detail, hopefully giving you a broad understanding of lexical knowledge.

Summary

In this introduction, I defined several terms that are necessary to discuss vocabulary with precision. I also indicated that languages contain huge numbers of words, something that was probably already obvious from the thickness of your dictionary. Although nobody can learn all of these words, learning the amount of vocabulary a native speaker knows is still an amazing feat. Moreover, the learning process is not an all-or-nothing process in which a word is suddenly and completely available for use. Rather, our knowledge of individual words grows over time, both in our ability to use them receptively and productively and in the different kinds of word knowledge we come to master. With the background knowledge from this chapter in hand, we should be ready to explore the fascinating world of how vocabulary is learned and used. But first we start by considering how people have viewed vocabulary over the ages, and how this has led to our current thinking in the field.

Exercises for expansion

1. Take a text several pages long and choose a few relatively common words. Count how often they occur according to the "word" versus "lexeme" versus "word family" definitions. Is there a great deal of difference in the counts?
2. Make your own estimate of the number of words in a language. Take a dictionary and find the average number of words defined on a page. Then multiply this by the number of pages in the dictionary. From this total, scholars have typically eliminated classes of words such as proper names (Abraham Lincoln) and compound words (dishwasher). Do you agree with this, and should any other classes be disregarded? How does the size of the dictionary affect the total size estimate?
3. To estimate how many word families you know, take this test developed by Goulden et al. (1990).

You will find below a list of fifty words that 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 check mark next to each word you know, that is, 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 fifty words, go back and check the words with question marks to see whether you can change the question mark to a check mark.
3. Then find the last five words you checkmarked (i.e., the ones that are farther 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 checkmarked. Write the meaning of this word and check it in the dictionary. Continue this process until you have a sequence of four checkmarked words (which may include some of the original five you checked) that you have explained correctly.
5. Calculate your score for the fifty-item test on the next page by multiplying the total number of known words by 500. Do not include the words with a question mark in your scoring.

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. stationary	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

(Adapted from Goulden et al.)*

Nation and Waring (1997) suggest that an average university-educated English native speaker has a vocabulary of about 20,000 word families. How do you compare? Why do you think you are above or below the figure they mentioned? How accurate do you think this test is? See Chapter 9 for more on this and other vocabulary tests.

4. Consider your own level of knowledge of the words in your lexicon. Listen for words in conversations and watch for words in texts that you understand well but never use yourself productively. Do there seem to be many words like this? Are there any examples of the opposite case where you use them easily when speaking, but have trouble spelling them? Words for which we have these partial states of knowledge are often the rarer ones. Considering that the majority of the words in a language are relatively rare, how would you evaluate the following statement?

   The standard state of vocabulary knowledge, even for native speakers, is partial knowledge.

* From R. Goulden, P. Nation, & J. Read (1990). How large can a receptive vocabulary be? *Applied Linguistics 11,* 358–359. Reproduced by permission of Oxford University Press and the authors.

5. Choose two or three words. List everything you know about these words. Do the same after you have read Chapters 3, 4, and 5. Does the second list indicate a greater awareness of vocabulary knowledge? If so, recommend this book to a friend. If not, try to sell him or her your copy.

Further reading

For receptive versus productive vocabulary: Melka (1997), Meara (1997), Laufer and Paribakht (1998), and Waring (1998).

For the word-knowledge perspective of vocabulary: Richards (1976), Nation (1990), Schmitt (1995a), Schmitt and Meara (1997), Schmitt (1998b), and Nation (1999).

For two good places to begin researching vocabulary on the Internet:
http://www.swan.ac.uk/cals/calsres.html
http://www1.harenet.ne.jp/~waring/vocabindex.html

For bibliographies of vocabulary research: Meara (1983), Meara (1987), Meara (1992).

2 *History of vocabulary in language learning*

- What methodologies have been used to teach second languages through the ages?
- What has been the role of vocabulary in these methodologies?
- What was the "Vocabulary Control Movement"?
- What are some of the notable strands of vocabulary research?

People have attempted to learn second languages from at least the time of the Romans, and perhaps before. In this period of more than two thousand years, there have been numerous different approaches to language learning, each with a different perspective on vocabulary. At times, vocabulary has been given pride of place in teaching methodologies, and at other times neglected. In order to help you better understand the current state of vocabulary studies as discussed in subsequent chapters, this chapter will first briefly review some of the historical influences that have shaped the field as we know it today. (Instead of digressing to explain terminology in this historical overview, key terms are cross-referenced to the page in the book where they are discussed.)

Language teaching methodologies through the ages

Records of second language learning extend back at least to the second century B.C., where Roman children studied Greek. In early schools, students learned to read by first mastering the alphabet, then progressing through syllables, words, and connected discourse. Some of the texts gave students lexical help by providing vocabulary that was either alphabetized or grouped under various topic areas (Bowen, Madsen, & Hilferty, 1985). We can only assume that lexis was considered important at this point in time, as the art of rhetoric was highly prized, and would have been impossible without a highly developed vocabulary.

Later, in the medieval period, the study of grammar became predominant, as students studied Latin. Language instruction during the Renais-