

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)

Introduction

Oh, no. Not another book on vocabulary. What makes this one any different?

As Meara initially noted in 1987 (and which has since become something of a cliché), the field of vocabulary studies is now anything but a neglected area, and the mushrooming amount of experimental studies and pedagogical and reference material being published is enough to swamp even lexical specialists trying to keep abreast of current trends. Included in this recent flood have been a number of books focusing on various aspects of lexis, for example: *The Nature of Vocabulary Acquisition* (McKeown and Curtis, 1987; L1 acquisition), *Words in the Mind* (Aitchison, 1987; the mental lexicon), *Teaching and Learning Vocabulary* (Nation, 1990; pedagogy), *The Bilingual Lexicon* (Schreuder and Weltens, 1993; psycholinguistic aspects), *Second Language Reading and Vocabulary Learning* (Huckin *et al.*, 1993; guessing from context), and *Second Language Vocabulary Acquisition* (Coady and Huckin, 1997; research studies). Given this amount of rather disparate material, we felt there was a need for a single volume that presented a broad view of the ‘state of the art’ in vocabulary studies as it currently exists.

Preview of the chapters

The book is in three parts. This division is not arbitrary, but corresponds to what we feel are the three main strands that contribute to an applied linguistic theory of vocabulary. Firstly, vocabulary must be defined and described: what is the nature of the linguistic data we are dealing with? Is it sufficient to equate ‘vocabulary’ with single words? What about idioms and other multi-word phenomena? From where does the evidence come? From textual corpora? If so, then from written, spoken or both? No book on vocabulary could ignore the basic

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)*Introduction*

descriptive issues. But vocabulary is also acquired, and is stored in and retrieved from the human mind. Thus no theoretical framework is complete without consideration of questions of acquisition and mental processing. Finally, our applied linguistic purpose is a pedagogical one, and pedagogical questions relating to materials, methods and syllabuses complete the framework for an applied linguistic perspective of lexis. Description, acquisition and pedagogy thus form the natural divisions for the present book. Their order should not be taken as suggesting that we consider any one more important than the other two: they complement one another, give rationale to one another and mutually explicate areas of difficulty for the applied linguistic researcher whose ultimate purpose is the integration of a better description, a deeper understanding of the mental lexicon and a more informed pedagogy.

Recurring themes in the book

This book includes a broad selection of vocabulary topics. But this diversity does not mean that each topic can be isolated and dealt with on its own. On the contrary, true understanding of any individual topic can only be gained by understanding others which are related. For example, one could not have a good understanding of vocabulary teaching without taking on board the points raised in Nick Ellis' chapter. In fact, nearly all of the topics are related to each other in some way. This being the case, there are several themes that run through the various chapters. These underlying currents are important because they add coherence to the discussions and represent some of the most significant ideas included in this book.

L1 words and acquisition – L2 words and acquisition

There is seldom a one-to-one relationship between L1 and L2 words, and the processes of learning an L1 and an L2 are potentially different because of age, cognitive maturity, the way a society categorizes the real world, etc. Nevertheless, a learner's L1 is one of the most important factors in learning L2 vocabulary. The L1 will determine whether a majority of L2 words are easy or difficult, and whether whole new knowledge systems (new alphabets, new sounds and sound combinations, new syntactic notions like articles, phrasal verbs, or case endings) have to be mastered. If the L1 and L2 are similar, there is a much higher likelihood that the initial mapping of the new L2 word will simply be the relabelling of an L1 word, rather than the addition of a totally new conceptual unit. Of course, this relabelling will eventually have to be

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)*Recurring themes in the book*

adjusted towards the exact properties of the L2 word, but it does have the advantage of being initially easy. Various chapters will touch upon the relationship between the L1 and the L2, but we should remember that it is an influence that is almost impossible to escape when dealing with almost *any* aspect of L2 vocabulary.

Deeper processing of words enhances their learning

The more cognitive energy a person expends when manipulating and thinking about a word, the more likely it is that they will be able to recall and use it later. This idea was first formalized as the Depth (or Levels) of Processing Hypothesis (Craik and Lockhart, 1972; Craik and Tulving, 1975). The implications extend to pedagogy, suggesting that exercises and learning strategies which involve a deeper engagement with words should lead to higher retention than 'shallower' activities.

Guessing from context vs. explicit teaching

There has been a long-running debate about which of these two methods of learning vocabulary is most important, although we feel that it has not necessarily been a useful one. We believe we should not be thinking in terms of better/worse or either/or, but rather we should see the two methods as complementary. As Ellis (2.2) illustrates, some aspects of vocabulary learning are more amenable to conscious learning than others. Similarly, Nation and Waring (1.1) show that some words make more sensible candidates for explicit teaching than others. What we should move towards is a realization of the benefits of both methods and an attempt to combine them for maximum results. Explicit teaching can be a very good first introduction to a word; after this, the context encountered when reading can lead to new knowledge of its collocations, additional meanings, and other higher level knowledge. In addition, repeated exposure from reading will help to consolidate the meaning(s) first learned.

Additionally, explicit teaching is probably essential for the most frequent words of any L2, since they are prerequisites for language use. The learning of these basic words cannot be left to chance, but should be taught as quickly as possible, because they open the door to further learning. McCarthy and Carter's chapter (1.2), for example, highlights a number of spoken language words which emerge as being indispensable for basic communication. Less frequent words, on the other hand, may be best learned by reading extensively, since there is just not enough time to learn them all through conscious study. So a well-considered

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)

Introduction

vocabulary learning programme will eventually include both methods, with each lending its own strengths.

Word knowledge

Many of the chapters assert that knowing a word requires more than just familiarity with its meaning and form. Nation (1990: 30–33) and Richards (1976) describe the various kinds of word knowledge necessary to master a word completely, including knowledge of its orthographical and phonological form, meanings, grammatical behaviour, associations, collocations, frequency and register. This listing can be very useful for the discussion of the types of word knowledge beyond meaning; however, it should be remembered that it is descriptive, not explanatory. The list gives us an inventory of ideal native-like knowledge, but it does not tell us *how* this knowledge is acquired. We have not yet reached a state where we can describe how each kind of word knowledge is acquired, or how the acquisition of one type affects the learning of another (although some tentative steps have been taken in this direction; see Schmitt and Meara, in press).

When considering this list, it is also important to remember that the categories of word knowledge are separated for convenience's sake; in reality, the different categories are interrelated, and it is probably best to think of vocabulary knowledge as an integrated whole from which we can artificially separate various kinds of word knowledge for research or discussion purposes.

Importance of word form

Research has shown that the eye samples almost every word when reading, refuting top-down models which suggested that 'higher level' knowledge allowed the guessing and skipping of many words. Reading appears to be much more of a bottom-up process than previously thought, with top-down knowledge (in an L2 context) largely filling in for deficiencies in word recognition automaticity. Not only do we have to recognize the orthographical forms of words, but we have to do it in an automatic and relatively error-free way in order to have anything like fluent reading (see Carrell *et al.*, 1988, for more on reading research). If mastery of a word's form is crucial for reading, then the implication is that it is also crucial to incorporate an emphasis on word form into vocabulary teaching. However, it is not clear how this can best be done, for Ellis (2.2) suggests that word form is largely implicitly acquired. In any case, word form is important for all L2 learners, and as

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)

Recurring themes in the book

Ryan (2.5) shows, for certain learners it can be one of the greatest stumbling blocks to successful language acquisition.

Lexical chunks

Much of the discussion in this book highlights the fact that vocabulary items are often not single orthographic units, but rather multi-word units. Moon's chapter (1.3) shows the importance and prevalence of these multi-word items, and McCarthy and Carter's chapter (1.2) notes how the most frequent spoken words are prone to forming everyday chunks (e.g. *I see, you know*). Where their data come from computer corpora, there is also basic agreement that multi-word units have psycholinguistic reality as well. Lewis (1993) suggests that the mind uses 'chunks' of prefabricated language because these are easier to process and use than an equivalent number of individual words that have to be strung together via syntactical rules. Likewise, Peters (1983) believes that much of vocabulary learning occurs as these unanalysed chunks are segmented into their component words. Lexical chunking is likely to become increasingly discussed as we become more aware of its importance in the psycholinguistic functioning of the mental lexicon.

Inevitably in a collection such as this, a balance must be struck between uniformity and consistency of approach and the fact that all the authors are well-known as individual scholars working in separate areas that have developed their own ways of discussing issues. We therefore decided to allow each author to retain his/her own 'voice' and individual style in order to produce a more illuminating and engaging text. As you progress through this book, we hope that you learn as much from it and enjoy it as much as we did in putting it together. Any remaining shortcomings you may perceive as you go along are our responsibility.

Norbert Schmitt
Michael McCarthy
University of Nottingham
November 1996

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)

Part I Vocabulary and description

1.1 Vocabulary size, text coverage and word lists

Paul Nation

Victoria University of Wellington

Robert Waring

Notre Dame Seishin University

How much vocabulary does a second language learner need?

There are three ways of answering this question. One way is to ask ‘How many words are there in the target language?’ Another way is to ask ‘How many words do native speakers know?’ A third way is to ask ‘How many words are needed to do the things that a language user needs to do?’ We will look at answers to each of these questions.

This discussion looks only at vocabulary and it should not be assumed that if a learner has sufficient vocabulary then all else is easy. Vocabulary knowledge is only one component of language skills such as reading and speaking. It should also not be assumed that substantial vocabulary knowledge is always a prerequisite to the performance of language skills. Vocabulary knowledge enables language use, language use enables the increase of vocabulary knowledge, knowledge of the world enables the increase of vocabulary knowledge and language use and so on (Nation, 1993a). With these cautions in mind let us now look at estimates of vocabulary size and their significance for second language learners. Such information will, we believe, help us to outline clear, sensible goals for vocabulary learning.

How many words are there in English?

The most straightforward way to answer this question is to look at the number of words in the largest dictionary. This usually upsets dictionary makers who work with words on a daily basis. They see the vocabulary of the language as a continually changing entity with new words and new uses of old words being added and old words falling into disuse. They also see the problems in deciding if *walk* as a noun is the same

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)

Vocabulary size, text coverage and word lists

word as *walk* as a verb, if compound items like *goose grass* are counted as separate words, and if names like *Vegemite*, *Agnes* and *Nottingham* are to be counted as words. These are all real problems, but they are able to be dealt with systematically in a reliable way.

Two separate studies (Dupuy, 1974; Goulden, Nation and Read, 1990) have looked at the vocabulary of *Webster's Third International Dictionary* (1963), the largest non-historical dictionary of English when it was published. When compound words, archaic words, abbreviations, proper names, alternative spellings and dialect forms are excluded, and when words are classified into word families consisting of a base word, inflected forms, and transparent derivations, *Webster's Third* has a vocabulary of around 54,000 word families. This is a learning goal far beyond the reaches of second language learners and, as we shall see, most native speakers.

How many words do native speakers know?

For over 100 years there have been published reports of systematic attempts to measure the vocabulary size of native speakers of English. There have been various motivations for such studies, but behind most of them lies the idea that vocabulary size is a reflection of how educated, intelligent or well read a person is. A large vocabulary size is seen as being something valuable. Unfortunately the measurement of vocabulary size has been bedevilled by serious methodological problems largely centring around the questions of 'What should be counted as a word?', 'How can we draw a sample of words from a dictionary to make a vocabulary test?', and 'How do we test to see if a word is known or not?'. Failure to deal adequately with these questions has resulted in several studies of vocabulary size which give very diverse and misleading results. For a discussion of these issues see Nation (1993b), Lorge and Chall (1963) and Thorndike (1924).

Teachers of English as a second language may be interested in measures of native speakers' vocabulary size because these can provide some indication of the size of the learning task facing second language learners, particularly those who need to study and work alongside native speakers in English-medium schools and universities or workplaces. At present 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 native speakers will add roughly 1,000 word families a year to their vocabulary size. That means that a five year old beginning school will have a vocabulary of around 4,000 to 5,000 word families. A university graduate will have a vocabulary of around 20,000 word

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)*Vocabulary and description*

families (Goulden, Nation and Read, 1990). These figures are very rough and there is likely to be very large variation between individuals. These figures exclude proper names, compound words, abbreviations and foreign words. A word family is taken to include a base word, its inflected forms and a small number of reasonably regular derived forms (Bauer and Nation, 1993). Some researchers suggest vocabulary sizes larger than these (see Nagy, 1.4), but in the well-conducted studies (for example, D'Anna, Zechmeister and Hall, 1991), the differences are mainly the result of which items are included in the count and how a word family is defined.

A small study of the vocabulary growth of non-native speakers in an English-medium primary school (Jamieson, 1976) suggests that, in such a situation, non-native speakers' vocabulary grows at the same rate as native speakers' but that the initial gap that existed between the two groups is not closed. For adult learners of English as a foreign language, the gap between their vocabulary size and that of native speakers is usually very large, with many adult foreign learners of English having a vocabulary size of much less than 5,000 word families in spite of having studied English for several years. Large numbers of second language learners do achieve vocabulary sizes similar to those of educated native speakers, but they are not the norm.

There is some encouraging news however. A study by Milton and Meara (1995) using the Eurocentres' Vocabulary Size Test (Meara and Jones, 1988, 1990; see also Read, 3.4) shows that significant vocabulary growth can occur if this learning is done in the second language environment. In their study of a study abroad programme of 53 European students of advanced proficiency, the average growth in vocabulary per person approached a rate of 2,500 words per year over the six months of the programme. This rate of growth is similar to the larger estimates of first language growth in adolescence. Although the goal of native speaker vocabulary size is a possible goal, it is a very ambitious one for most learners of English as a foreign language.

How many words are needed to do the things a language user needs to do?

Although a language makes use of a large number of words, not all of these words are equally useful. One measure of usefulness is word frequency, that is, how often the word occurs in normal use of the language. From the point of view of frequency, the word *the* is a very useful word in English. It occurs so frequently that about 7 per cent of the words on a page of written English and the same proportion of the words

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)*Vocabulary size, text coverage and word lists*

in a conversation are repetitions of the word *the*. Look back over this paragraph and you will find an occurrence of *the* in almost every line.

The good news for second language learners and second language teachers is that a small number of the words of English occur very frequently and if a learner knows these words, that learner will know a very large proportion of the running words in a written or spoken text. Most of these words are content words and knowing enough of them allows a good degree of comprehension of a text. Here are some figures showing what proportion of a text is covered by certain numbers of high frequency words.

Table 1 *Vocabulary size and text coverage in the Brown corpus*

Vocabulary size	Text coverage
1,000	72.0%
2,000	79.7%
3,000	84.0%
4,000	86.8%
5,000	88.7%
6,000	89.9%
15,851	97.8%

(taken from Francis and Kucera, 1982)

The figures in Table 1 refer to written texts and are from Francis and Kucera (1982) which is a very diverse corpus of over 1,000,000 running words made up of 500 texts of around 2,000 running words long. As we shall see, the more diverse the texts in a corpus are, the greater the number of different words, and the high frequency words cover slightly less of the text, so these figures are a conservative estimate. The figures in the last line of the table are from Kucera (1982). The *Collins COBUILD English Language Dictionary* (1987) claims that 15,000 words cover 95 per cent of the running words of their corpus. The figures in Table 1 are for lemmas and not word families. (A lemma is a base word and its inflected forms.) Word families would give fractionally higher coverage. Table 1 assumes that high frequency words are known before lower frequency words and shows that knowing about 2,000 word families gives near to 80 per cent coverage of written text. The same number of words gives greater coverage of informal spoken text – around 96 per cent (Schonell, Meddleton and Shaw, 1956). (McCarthy and Carter discuss other differences between spoken and written discourse in the next chapter.)

With a vocabulary size of 2,000 words, a learner knows 80 per cent

Cambridge University Press

978-0-521-58551-4 - Vocabulary: Description, Acquisition and Pedagogy

Edited by Norbert Schmitt and Michael McCarthy

Excerpt

[More information](#)*Vocabulary and description*

of the words in a text which means that one word in every five (approximately two words in every line) are unknown. Research by Liu Na and Nation (1985) has shown that this ratio of unknown to known words is not sufficient to allow reasonably successful guessing of the meaning of the unknown words. At least 95 per cent coverage is needed for that. Research by Laufer (1988a) suggests that 95 per cent coverage is sufficient to allow reasonable comprehension of a text. A larger vocabulary size is clearly better. Table 2 is based on research by Hirsh and Nation (1992) about novels written for teenage or younger readers.

The Hirsh and Nation (1992) study looked at such novels because they might provide the most favourable conditions for second language learners to read unsimplified texts. These conditions could come about because they are aimed at a non-adult audience and thus there may be a tendency for the writer to use simpler vocabulary, and because a continuous novel on one topic by one writer provides opportunity for the repetition of vocabulary. Table 2 shows that under favourable conditions, a vocabulary size of 2,000 to 3,000 words provides a very good basis for language use.

Table 2 *Vocabulary size and coverage in novels for teenagers*

Vocabulary size	% coverage	Density of unknown words
2,000 words	90	1 in every 10
2,000+ proper nouns	93.7	1 in every 16
2,600 words	96	1 in every 25
5,000 words	98.5	1 in every 67

The significance of this information is that although there are well over 54,000 word families in English, and although educated adult native speakers know around 20,000 of these word families, a much smaller number of words, say between 3–5,000 word families is needed to provide a basis for comprehension. It is possible to make use of a smaller number, around 2–3,000 for productive use in speaking and writing. Hazenburg and Hulstijn (1996), however, suggest a figure nearer to 10,000 for Dutch as a second language.

Sutarsyah, Nation and Kennedy (1994) found that a single long economics text was made up of 5,438 word families and a corpus of similar length made up of diverse short academic texts contained 12,744 word families. Within narrowly focused areas of interest, such as in an economics text, a much smaller vocabulary is needed than if the reader wishes to read a wide range of texts on a variety of different topics.