

Introduction

This book is about the teaching and learning of vocabulary, but the teaching and learning of vocabulary is only a part of a language development programme. It is thus important that vocabulary teaching and learning is placed in its proper perspective.

Learning goals

Vocabulary learning is only one sub-goal of a range of goals that are important in the language classroom. The mnemonic LIST is a useful way of remembering these goals that are outlined in Table 0.1. L = Language, which includes vocabulary; I = Ideas, which cover content and subject matter knowledge as well as cultural knowledge; S = Skills; and T = Text or discourse, which covers the way sentences fit together to form larger units of language.

Although this book focuses on the vocabulary sub-goal of language, the other goals are not ignored. However, they are approached from the

Table 0.1 *Goals for language learning*

General goals	Specific goals
Language items	pronunciation vocabulary grammatical constructions
Ideas (content)	subject matter knowledge cultural knowledge
Skills	accuracy fluency strategies process skills or subskills
Text (discourse)	conversational discourse rules text schemata or topic type scales

2 *Learning Vocabulary in Another Language*

viewpoint of vocabulary. There are chapters on vocabulary and the skills of listening, speaking, reading and writing. Discourse is looked at in Chapter 6 on specialised uses, and pronunciation, spelling and grammar are looked at in relation to vocabulary knowledge in Chapter 3.

The four strands

The approach taken in this book rests on the idea that a well-balanced language course should consist of four major strands (Nation, 2007; Nation and Yamamoto, 2011). These strands can appear in many different forms, but they should all be there in a well-designed course.

Firstly, there is the strand of learning from comprehensible meaning-focused input. This means that learners should have the opportunity to learn new language items through listening and reading activities where the main focus of attention is on the information in what they are listening to or reading. As we shall see in the following chapter, learning from meaning-focused input can best occur if learners are familiar with at least 98 per cent of the running words in the input they are focusing on. Put negatively, learning from meaning-focused input cannot occur if there are lots of unknown words.

The second strand of a course is the strand of meaning-focused output. Learners should have the chance to develop their knowledge of the language through speaking and writing activities where their main attention is focused on the information they are trying to convey. Speaking and writing are useful means of vocabulary development because they make the learners focus on words in ways they did not have to while listening and reading. Having to speak and write encourages learners to listen like a speaker and read like a writer. This different kind of attention is not the only contribution that speaking and writing activities can make to language development. From a vocabulary perspective, these productive activities can strengthen knowledge of previously met vocabulary.

The third strand of a course is one that has been subject to a lot of debate. This is the strand of language-focused learning, sometimes called form-focused instruction. There is growing evidence (Ellis, 2005; Williams, 2005) that language learning benefits if there is an appropriate amount of usefully focused deliberate teaching and learning of language items. From a vocabulary perspective, this means that a course should involve the direct teaching of vocabulary and the direct learning and study of vocabulary. As we shall see, there is a very large amount of research stretching back to the late 19th century which shows that the gradual cumulative process of learning a word can be given a strong boost by the direct study of certain features of the word.

The fourth strand of a course is the fluency development strand. In the activities which put this strand into action learners do not work with new language items. Instead, they become more and more fluent in using items they already know. A striking example of this can be found in the use of numbers. Learners can usually quickly learn numbers in a foreign language. But if they go into a post office and the clerk tells them how much the stamps they need are going to cost, they might not understand because the numbers were said too quickly for them. By doing a small amount of regular fluency practice with numbers (the teacher says the numbers, the learners write the figures), the learners will find that they can understand one-digit numbers said quickly (1, 7, 6, 9) although they have trouble with two-digit numbers said quickly (26, 89, 63, 42) or three-digit numbers (126, 749, 537, 628). A little further practice will make these longer numbers fluently available for comprehension. If a course does not have a strong fluency strand, then the learning done in the other three strands will not be readily available for normal use.

In a language course, these four strands should get roughly the same amount of time. That means that no more than 25 per cent of the learning time in and out of class should be given to the direct study of language items. No less than 25 per cent of the class time should be given to fluency development. If the four strands of a course are not equally represented in a particular course, then the design of the course needs to be looked at again.

These four strands need to be kept in mind while reading this book. Where recommendations are made for direct vocabulary learning, these should be seen as fitting into that 25 per cent of the course which is devoted to language-focused learning. Seventy-five per cent of the vocabulary development programme should involve the three meaning-focused strands of learning from input, learning from output and fluency development.

The four strands apply generally to a language course. In this book we will look at how vocabulary fits into each of these strands. It is worth stressing that the strands of meaning-focused input and output are only effective if the learners have sufficient vocabulary to make these strands truly meaning focused. If activities which are supposed to be meaning focused involve large amounts of unknown vocabulary, then they become language focused because much of the learners' attention is taken from the message to the unknown vocabulary. Similarly, fluency development activities need to involve little or no unknown vocabulary or other language items, otherwise they become part of the meaning input and output strands, or language-focused learning.

4 *Learning Vocabulary in Another Language*

Main themes

A small number of major themes run through this book, and these are first dealt with in Chapters 2, 3 and 4. Firstly, there is the cost/benefit idea based on the results of word frequency studies. Its most important application is in the distinction between high-frequency and mid- and low-frequency vocabulary and the different ways in which teachers should deal with these types of vocabulary. The cost/benefit idea also applies to individual words in that the amount of attention given to an item should be roughly proportional to the chances of it being met or used again, that is, its frequency.

Secondly, there is the idea that learning a word is a cumulative process involving a range of aspects of knowledge. Learners thus need many different kinds of meetings with words in order to learn them fully. There is to date still little research on how vocabulary knowledge grows and how different kinds of encounters with words contribute to vocabulary knowledge. In this book, knowing a word is taken to include not only knowing the formal aspects of the word and knowing its meaning, but also being able to use the word.

Thirdly, there is the idea that teachers and learners should give careful consideration to how vocabulary is learned, in particular, the psychological conditions that are most likely to lead to effective learning. Because these conditions are influenced by the design of learning tasks, quite a lot of attention is given to the analysis and design of vocabulary-learning activities.

The audience for this book

This book is intended to be used by second and foreign language teachers. Although it is largely written from the viewpoint of a teacher of English, it could also be used by teachers of other languages.

This book is called *Learning Vocabulary in Another Language* partly in order to indicate that most of the suggestions apply to both second and foreign language learning. Generally the term **second language** will be used to apply to both second and foreign language learning. In the few places where a contrast is intended, this will be clear from the context.

The first and the second editions

“I’ve got the first edition. Is it worth buying the second edition?” – this is a question I expect to be asked, so here is my answer.

Yes. Most of the changes in the second edition are the result of a large amount of research which has appeared since the first edition was published in 2001. By my rough calculation, over 30 per cent of the research on vocabulary that has appeared in the last 110 years was published in the last eleven years. Teaching and learning vocabulary, particularly for foreign and second language learners, is no longer a neglected aspect of language learning. So, if you don't buy the second edition you will be out of date by eleven years and at least 30 per cent of the field. On a rough estimate, at least one-fifth of the book is new material.

There were also errors in the first edition, largely because of a lack of research on the relevant areas. Some of that research has now been done, much of it by my students, colleagues and friends, and a few people who fit two or all of those categories.

I am also pleased to note that my thinking has changed on some issues in the teaching and learning of vocabulary, largely as a result of research findings and my own experience and thinking. These include the idea of mid-frequency vocabulary, largely as a result of research on word lists and testing native speaker vocabulary size. I also now feel that I am beginning to understand what collocations are. I am also becoming more sceptical of the value of vocabulary teaching, largely because of its necessarily limited scope and limited effectiveness.

When working on this second edition, I often wondered if the field of teaching and learning vocabulary is now so vigorous and large that it is beyond the scope of any one book and certainly one person. If you have already bought this book, then I hope I am wrong and you have got your money's worth.

Changes in the second edition

One of the changes in Chapter 1 is because of Chung and Nation's (2003, 2004) research on technical vocabulary. In the first edition I got this completely wrong, saying that about five per cent of the running words in a technical text would be technical vocabulary. In fact research showed that it was closer to 20 to 30 per cent of the running words. The second major change in Chapter 1 is as a result of the development of the lower-frequency word family lists based on the British National Corpus. At the time of writing, these lists now go up to the 24th one-thousand word lists, and the development of these lists has meant that we can do much more detailed analysis of texts and their vocabulary demands, as well as develop more soundly based vocabulary size tests. This research has highlighted the idea of mid-frequency words (Schmitt and Schmitt, 2012). At the time of writing

Cambridge University Press

978-1-107-62302-6 - Learning Vocabulary in Another Language: Second Edition

I. S. P. Nation

Excerpt

[More information](#)6 *Learning Vocabulary in Another Language*

the first edition, Coxhead's (2000) work on the Academic Word List was just being completed. The research just made it into the first edition, but in this second edition it is given the additional attention it deserves.

Chapter 2, 'Knowing a word', includes recent research on the relationship between first language (L1) and second language (L2) vocabulary storage. Chapter 3 includes a description of Technique Feature Analysis first introduced in Nation and Webb (2011a). Chapter 4 on listening and speaking includes recent work on vocabulary learning through lectures and learning in interactive activities. Chapter 5 on reading and writing is largely reorganised, and there is much more on glossing because of the growth in research on electronic glossing. It also includes recent corpus and experimental work on text coverage as well as recent studies of learning from graded readers and reading fluency. Chapter 6, 'Specialised uses', now has critiques of the academic word list, recent work on technical vocabulary and a section on content-based vocabulary teaching. Chapter 7, 'Vocabulary-learning strategies', includes recent research on strategy training and strategy use. The research on strategy use is now becoming more rigorous with less dependence on questionnaires. Chapter 8 contains recent research on guessing. Chapter 9, 'Word parts', has only very few changes. The changes in Chapter 10 are largely due to the growth in electronic dictionaries. Chapter 11, 'Deliberate learning from word cards', includes recent research on whether expanded spacing is better than even spacing within a learning session. It also includes criteria for evaluating flashcard programmes (Nakata, 2011). It also includes what I consider to be the most significant recent research finding in the field of vocabulary learning, namely that rote learning results in both implicit and explicit knowledge (Elgort, 2011) and thus the learning/acquisition distinction is not relevant for vocabulary. Chapter 12 on finding and learning collocations is almost completely rewritten. For me this was the most unsatisfactory chapter in the first edition. I now feel I am beginning to see how the work on collocations fits together, largely by separating the types of criteria used to classify collocations into criteria of form, meaning, function and storage. I have kept a few small sections but have taken a new approach to the chapter. There has been a large amount of research on collocations and some of it is very innovative. However there is still a need for clear definitions of what kind of units are being investigated and following these definitions closely when doing the research. Chapter 13 on testing changes the table of test sensitivity to agree with Laufer and Goldstein's (2004) findings. There is now more on the Word Associates Test, and there is also recent research on vocabulary size, including (Biemiller, 2005) findings with

Cambridge University Press

978-1-107-62302-6 - Learning Vocabulary in Another Language: Second Edition

I. S. P. Nation

Excerpt

[More information](#)

Introduction 7

L1 learners and research on the *Vocabulary Size Test*. Chapter 14 on planning has very few changes.

There is now an international community of vocabulary researchers and I am grateful to them for the knowledge, support and encouragement they have given me in the preparation of this book and in my research and writing.

Since I wrote my first book, *Teaching and Learning Vocabulary* (Nation, 1990) and the first edition of this book, another generation of vocabulary researchers has appeared. Although this is still a relatively small group, it is made up of very productive researchers who have identified a range of useful research focuses and who persist in exploring and refining research in those chosen areas. It is also notable that recently two books focusing on the research methodology of vocabulary studies have appeared (Nation and Webb, 2011; Schmitt, 2010). Research on vocabulary is clearly alive and well.

I am very grateful to Norbert Schmitt, Pavel Szudarski, Suhad Sonbul and Laura Vilkaite for comments on a draft of this book. Their insightful comments led to significant improvements in the book.

8 *Learning Vocabulary in Another Language***References**

- Biemiller, A. (2005). Size and sequence in vocabulary development, in Hiebert, E. H. and Kamil, M. L. (eds.), *Teaching and Learning Vocabulary: Bringing Research into Practice*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 223–42.
- Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a Foreign Language*, 15, 2, 103–16.
- Chung, T. M., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32, 2, 251–63.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 2, 213–38.
- Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, 61, 2, 367–413.
- Ellis, R. (2005). Principles of instructed language learning. *System*, 33, 209–24.
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54, 3, 399–436.
- Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: A critical investigation of flashcard software. *Computer Assisted Language Learning*, 24, 1, 17–38.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. Boston: Heinle.
- Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1, 1, 1–12.
- Nation, I. S. P. and Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Heinle Cengage Learning.
- Nation, I. S. P. and Yamamoto, A. (2011). Applying the four strands to language learning. *International Journal of Innovation in English Language Teaching and Research*, 1, 2, 1–15.
- Schmitt, N. (2010). *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.
- Schmitt, N. and Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, doi:10.1017/S0261444812000018.
- Williams, J. (2005). Form-focused instruction, in Hinkel, E. (ed.), *Handbook of Research in Second Language Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 671–91.

1 *The goals of vocabulary learning*

The idea behind this chapter is that it is helpful to use frequency and range of occurrence to distinguish several levels of vocabulary. Distinguishing these levels helps ensure that learners learn vocabulary in the most useful sequence and thus gain the most benefit from the vocabulary they learn. Making the high-frequency/mid-frequency/low-frequency distinction ensures that the teacher deals with vocabulary in the most efficient ways.

Counting words

There are several ways of counting words, that is, deciding what will be counted.

Tokens

One way is simply to count every word form in a spoken or written text and if the same word form occurs more than once, then each occurrence is counted. So, the sentence, *It is not easy to say it correctly*, would contain eight words, even though two of them are the same word form, *it*. Words which are counted in this way are called **tokens**, and sometimes **running words**. If we try to answer questions like ‘How many words are there on a page or in a line?’, ‘How long is this book?’, ‘How fast can you read?’ or ‘How many words does the average person speak per minute?’, then our unit of counting will be the token.

Types

We can count the words in the sentence *It is not easy to say it correctly* another way. When we see the same word occur again, we do not count it again. So the sentence of eight tokens consists of seven different words or **types**. We count words in this way if we want to answer questions like ‘How large was Shakespeare’s vocabulary?’, ‘How many

10 *Learning Vocabulary in Another Language*

words do you need to know to read this book?’ or ‘How many words does this dictionary contain?’

Lemmas

Counting *book* and *books* as two different words to be learned seems a bit strange. So, instead of counting different types as different words, closely related words could be counted as members of the same word or **lemma**. A lemma consists of a headword and its inflected forms and reduced forms (*n't*). Usually, all the items included under a lemma are all the same part of speech (Francis and Kučera, 1982). The English inflections consist of plural, third person singular present tense, past tense, past participle, *-ing*, comparative, superlative, possessive (Bauer and Nation, 1993). The Thorndike and Lorge (1944) frequency count used lemmas as the basis for counting, and the computerised count on the Brown corpus produced a lemmatised list (Francis and Kučera, 1982). In the Brown count the comparative and superlative forms were not included in the lemma, and the same form used as a different part of speech (*walk* as a noun, *walk* as a verb) are not in the same lemma. Variant spellings (*favor*, *favour*) are usually included as part of the same lemma when they are the same part of speech. Leech et al. (2001) used similar criteria in their count of the British National Corpus (<http://ucrel.lancs.ac.uk/bncfreq>).

Lying behind the use of lemmas as the unit of counting is the idea of **learning burden** (Swenson and West, 1934). The learning burden of an item is the amount of effort required to learn it. Once learners can use the inflectional system, the learning burden of *mends*, if the learner already knows *mend*, is negligible. One problem to be faced in forming lemmas is to decide what will be done with irregular forms such as *mice*, *is*, *brought*, *beaten* and *best*. The learning burden of these is clearly heavier than the learning burden of regular forms like *books*, *runs*, *talked*, *washed* and *fastest*. Should the irregular forms be counted as a part of the same lemma as their base word or should they be put into separate lemmas? Lemmas also separate closely related items, such as the adjective and noun uses of words like *original*, and the noun and verb uses of words like *display*. An additional problem with lemmas is to decide what is the headword of the lemma – the base form or the most frequent form? (Sinclair, 1991: 41–2).

Using the lemma as the unit of counting greatly reduces the number of units in a corpus. Bauer and Nation (1993) calculated that the 61,805 tagged types (or 45,957 untagged types) in the Brown corpus become 37,617 lemmas, which is a reduction of almost 40% (or 18% for untagged types). Nagy and Anderson (1984) estimated that 19,105