Teaching and learning vocabulary

Preface

The idea behind this book is that a systematic, principled approach to vocabulary development results in better learning. The material in the book draws heavily on the large amount of research, experimentation, and classroom experience of teachers and researchers over the last 100 years, which is reflected in the bibliography of over 900 items.

There are two main ideas running through the book. The first is that information about the statistical nature of vocabulary provides valuable guidelines for the teaching and learning of vocabulary. The most important distinction here is between the small number of high-frequency words, which all deserve a lot of attention, and the very large number of low-frequency words, which require the mastery of coping strategies like guessing from context. The second idea is that an analysis of the tasks facing the learner in acquiring vocabulary provides information about how learning can be made more efficient and effective. In particular, this involves looking at the learning burden of words and what it means to know a word.

These two ideas are related to each other. Differences in the distribution of certain words are reflected in the learning task for those words. Thus, although most chapters can be read separately, it is worth following the information on language and learning in the earlier chapters through into the application of this information for teaching and learning in the chapters on the skills of listening, speaking, reading, and writing.

This book is intended for teachers of English as a second or foreign language, and it answers questions about vocabulary teaching and learning that teachers typically ask. Because the majority of such teachers are themselves not native speakers of English I have tried to avoid technical vocabulary and have tried to express myself as directly and clearly as possible. I find vocabulary teaching and learning a fascinating area. I hope you do too.

Acknowledgments

Chapter **1**

Introduction

This chapter looks at the following ideas: Should vocabulary be taught? What are the possible approaches to vocabulary learning? Then the following points are considered: What vocabulary do my learners need to know? How will they learn their vocabulary? How can I test to see what my learners need to know and what they now know? The answers to these questions will help teachers plan the vocabulary component of their courses.

SHOULD VOCABULARY BE TAUGHT?

It is not difficult to find language teachers who think that vocabulary can be left to take care of itself, and there is some experimental evidence to support that position (Elley & Mangubhai, 1981). What this book tries to show, however, is that there are very strong reasons for a systematic and principled approach to vocabulary by both the teacher and the learners. Let us look at these reasons. First, because of the considerable research on vocabulary we have good information about what to do about vocabulary and about what vocabulary to focus on. This means that our vocabulary work can be directed toward useful words and can give learners practice in useful skills. We can thus feel confident that learners will get a good return for the effort that they put in. Second, one of the aims of this book is to show that there is a wide variety of ways for dealing with vocabulary in foreign or second language learning. Some teachers do not feel happy using some of these ways—for example, getting learners to study words out of context, or interrupting learners' reading to get them to guess at an unknown word in context. Dissatisfaction with one approach to vocabulary should not result in ignoring all the other ways of helping learners enrich their vocabulary. It is important that when a teacher chooses or rejects a way to deal with vocabulary this choice or rejection should be based on a good understanding of the way of dealing with vocabulary, the principles behind it, and its theoretical and experimental justification. For example, many teachers too quickly dismiss the

approach of getting learners to study lists of words out of context. For a teacher faced with learners with a small vocabulary who wish to go on to academic study in a few months' time, this approach is very effective. Moreover, there is a very large amount of experimental research showing the effectiveness of such an approach and providing useful guidelines on how to go about it (Nation, 1982). Later in this chapter we will look at four broad approaches to vocabulary and at the reasons which would make a teacher choose one over another.

The third reason for having a systematic and principled approach to vocabulary is that both learners and researchers see vocabulary as being a very important, if not the most important, element in language learning. Learners feel that many of their difficulties in both receptive and productive language use result from an inadequate vocabulary.

Simply increasing learners' vocabulary without giving attention to putting this knowledge to use may not be effective, but getting learners to do language tasks when their vocabulary is inadequate for the task is a frustrating experience. Research on readability (Chall, 1958; Klare, 1974–1975) stresses the importance of vocabulary knowledge in reading, as does research on academic achievement (Saville-Troike, 1984).

Finally, giving attention to vocabulary is unavoidable. Even the most formal or communication-directed approaches to language teaching must deal with needed vocabulary in one way or other. Let us now look at the possible ways.

APPROACHES TO VOCABULARY LEARNING

It is useful to make a distinction between direct and indirect vocabulary learning. In direct vocabulary learning the learners do exercises and activities that focus their attention on vocabulary. Such exercises include word-building exercises, guessing words from context when this is done as a class exercise, learning words in lists, and vocabulary games. In indirect vocabulary learning the learners' attention is focused on some other feature, usually the message that is conveyed by a speaker or writer. If the amount of unknown vocabulary is low in such messages, considerable vocabulary learning can occur even though the learners' attention is not directed toward vocabulary learning. Krashen (1981a) calls this the input theory of language learning. He believes that certain conditions must apply for such learning to occur. First, the learners must be interested in understanding the message. From the point of view of vocabulary learning, this interest creates a need to understand the unknown words in the message. Second, the message should contain some items that are just outside the learners' present level of achievement. These items, however, should be understandable from the context in

which they occur. This includes both language and nonlanguage contexts. The chapter on simplification in this book looks at ways of making unknown items understandable. Third, the learners should not feel worried or threatened by their contact with the foreign language.

One of the basic ideas in this book is that there is a place for both direct and indirect vocabulary learning. Opportunities for indirect vocabulary learning should occupy much more time in a language learning course than direct vocabulary learning activities. This is in fact just another way of saying that contact with language in use should be given more time than decontextualized activities. The range of contextualized activities of course covers the range of the uses of language. As long as suitable conditions for language learning apply, such as those described by Krashen (1981a), then indirect vocabulary learning can take place.

Vocabulary teaching can fit into a language learning course in any of four ways. Most courses make use of all four, but the amount of time spent on each of these ways depends on the teacher's judgment in relation to a large number of factors, such as the time available, the age of the learners, the amount of contact with English outside school hours, and the teacher's theory of how language is best learned. The four ways described below are listed from the most indirect to the most direct.

1. Material is prepared with vocabulary learning as a consideration. The most common examples of this are the preparation of simplified material and the careful vocabulary grading of the first lessons of learning English. To an observer of such an English course it might appear as if no attention is being given to vocabulary, but in fact the selection and grading of vocabulary has been given a lot of attention before the course begins.

2. Words are dealt with as they happen to occur. This means that if an unknown word appears in a reading passage, the teacher gives some attention to it at the moment it causes a problem. A lot of vocabulary teaching is done in this way. Although the selection of vocabulary seems unplanned, the way it is treated need not be. Teachers may follow principles when dealing with such words. For example, they draw attention to the underlying concept of the word rather than just giving a contextual definition. They point out regular features of the spelling and grammar so that the learning of this word will help the learning of other words. They focus attention on the learning burden of the word, and they carefully avoid "unteaching." They consider the frequency and usefulness of a word when deciding how much time to spend on it.

3. Vocabulary is taught in connection with other language activities. For example, the vocabulary of a reading passage is dealt with before the learners read the passage. Sarawit (1980) describes the teaching of vocabulary a few days before a type of debate. Through direct teaching and reading the learners become familiar with the topic vocabulary before

they need to use it in the formal speaking activity. Johnson (1980) suggests similar preteaching as a preparation for listening exercises which give further practice in the vocabulary. Another possibility is to have vocabulary exercises following reading or listening texts. "Find the words in the passage which mean. . . " is the most common example of this. In all the activities described in this part, the teaching of vocabulary is directly related to some other language activity.

4. Time is spent either in class or out of school on the study of vocabulary without an immediate connection with some other language activity. For example, time is spent on learning spelling rules or on activities like dictionary use, guessing words, the use of word parts, or list learning. This time can be spent on activities involving the whole class as in learning mnemonic techniques, using pair or group work as in paraphrase activities or combining arrangement exercises, or individually as in the use of vocabulary puzzles or code exercises. Such vocabulary work can have the aim of establishing previous learning or increasing vocabulary so that future language use can go more smoothly.

POINTS TO CONSIDER IN VOCABULARY LEARNING

To put any of these four ways into practice a teacher needs to answer three questions.

1. What vocabulary do my learners need to know?
2. How will they learn this vocabulary?
3. How can I test to see what they need to know and what they now know?

Let us preview some of the answers to these questions.

What Vocabulary Do My Learners Need to Know?

Later in this book we divide vocabulary into three groups—high-frequency words, low-frequency words, and specialized vocabulary. Teachers need to decide which of these three groups contains the words that their learners need. This is an important decision because it will affect the way the words are dealt with for learning, it will affect the amount of learning expected, and it will affect the type of learning, receptive or productive, needed. The goals that teachers or learners set for learning English will affect the way vocabulary is selected. If, for example, the goal is to read, then following a word list like that of the Longman Structural Readers, the Newbury House Readers, or Collins

English Library will give access to a large amount of ready-made graded reading material, thus increasing vocabulary knowledge and reading skill at the same time. If the learners aim to cover listening, speaking, reading, and writing, then a multipurpose list like the General Service List, or the Cambridge English Lexicon will be more suitable. If the learners want to read university texts, then the university word list is a suitable goal. Estimates of vocabulary size, particularly of native speakers of English, can also help in setting goals, particularly in schools where there are both native speakers and second language learners.

As we look at each of the skill areas of listening, speaking, reading, and writing we will give information about vocabulary size.

How Will They Learn This Vocabulary?

First, we need to consider what kind of learning is required. In the chapter on what is involved in learning a word a distinction is made between receptive and productive learning. Receptive learning involves being able to recognize a word and recall its meaning when it is met. Productive learning involves what is needed for receptive learning plus the ability to speak or write needed vocabulary at the appropriate time. If learners study English in order to be able to read and understand lectures, a receptive knowledge of vocabulary is sufficient. If learners need to cover the whole range of language skills, then a productive vocabulary of around 3000 base words and a larger receptive vocabulary is needed. It is also important to consider previous learning of the mother tongue, other languages, or earlier lessons in English when selecting and teaching vocabulary, and to avoid "unteaching," particularly as a result of organizing and presenting vocabulary. If productive learning is important, then the development of the quality of learning a small vocabulary is important. Intensive practice in using vocabulary in speech and/or writing is therefore a useful activity. If receptive learning is important, then quantity of vocabulary is the main goal. Techniques which give familiarity with a large number of words are needed. As learners read and listen, the quality of knowledge of these words will develop without further attention from the teacher.

Often a teacher may wish to give particular attention to a word. Most of the procedures and techniques for this take words out of context and focus attention on them. These techniques are useful when for a variety of reasons words deserve particular attention. Here are some of these reasons.

1. The word is very frequent and very important for the learners.
2. The word causes particular difficulty.

3. The word is needed for another activity, such as a game, a reading or listening exercise, a talk.
4. The word contains features of regular patterns. Knowledge of these patterns will help learners master other words more easily.

Most vocabulary learning, however, will happen when the learners use the language for other purposes. The aim of techniques which take vocabulary out of context is to speed up the vocabulary learning process. This means that such techniques should occupy only a small part of any language learning program. Later in the book we will look at ways in which simplification can help the learning of words in context by placing them in a meaningful context, a context which can be understood. Activities using simplified material are very important in most language learning programs and deserve a lot of time. The use of simplified reading materials provides learners with an excellent opportunity to continue learning vocabulary through reading outside class time. If learners have access to tape recorders or language laboratories, then there is the opportunity for a lot of vocabulary learning through listening to simplified material. Real vocabulary learning comes through use, both receptive use and productive use. Teachers can help the process along by drawing attention to particular words, by teaching strategies for learning vocabulary, and by providing simplified material, but meeting the words in a variety of contexts and having to use some of them to express new ideas provide the most important opportunities for vocabulary learning.

There are strategies that learners can use to cope with new vocabulary. By far the most important strategy is guessing from context. It is worth spending a lot of class time on this strategy until learners have mastered it. It is the most powerful way of dealing with unknown words. Other strategies like paraphrasing, using word parts, dictionary use, and mnemonic techniques make learners independent of the teacher. They are tools that can be used as a part of usual language activities or can be used outside class time. Guided practice with these strategies encourages learners to use them and gives learners the skill to use the strategies effectively.

It is useful to make a distinction between increasing vocabulary and establishing vocabulary. Increasing vocabulary means introducing learners to new words and thus starting their learning. Establishing vocabulary means building on and strengthening this initial knowledge—that is, encouraging the knowledge of particular words to develop and expand. For example, teachers might increase learners' vocabulary by getting them to learn lists of words. They might establish this knowledge by getting learners to do lots of reading of material that contains those words first learned in lists. Or they might establish this knowledge through puzzles and games involving the vocabulary.

An important teaching principle is that the old material in any lesson is the most important. There are several reasons for this. First, the old material has already had some teaching and learning time invested in it. If the material is not given any further attention, then this material will be forgotten and all the previous investment is wasted. Further attention to the material will bring it closer to the threshold of learning, after which no further attention is needed. Second, the old material is likely to benefit most from further attention because the learners have already had some experience of it and it is at this stage that generalizations about it by the teacher will be meaningful to the learners. Such generalizations at the time the material is first introduced have little significance for the learners because they have no past experience to relate it to. Third, because the learners have already mastered some features of the material, they are able to give their attention to contextual features of the learning rather than the initial decontextualized features.

These reasons apply particularly to vocabulary learning. The effort given to the learning of new words will be wasted if this is not followed up by later meetings with the words. Understanding the concepts behind words and the types of collocations that they have is best done when it builds upon previous experience of the words. Once learners have a basic meaning for a word, they can give attention to what words it collocates with and the patterns it occurs in. Thus, there are good reasons for letting learners meet new vocabulary incidentally and directing attention to it after several meetings.

As you work your way through this book, you should try to answer the following questions.

Increasing Learners' Vocabulary

1. Do you expect most learning to occur during or outside class?
2. Will you be concentrating on receptive or productive vocabulary learning?
3. Which skills—listening, speaking, reading or writing—will involve most of your vocabulary learning program?
4. Will you be aiming at quality of vocabulary learning or quantity?
5. Will your focus be on particular words or on strategies?
6. How much emphasis will you give to decontextualized vocabulary learning?
7. Will most of your class work involve teacher-directed class activities, group or pair work, or individual work?

When you have answered these questions, go back over them again, listing the techniques, strategies, and procedures you will use. Say where you will get the necessary material.

Establishing Vocabulary   What techniques, strategies, and activities will you use only to establish vocabulary that your learners have already met? When you answer the question, consider the questions in the section on increasing learners' vocabulary.

Individual Needs   What techniques, strategies, and activities will you use to help learners who know less vocabulary than others in your class?

How Can I Test to See What My Learners Need to Know and What They Now Know?

Before testing it is important to be clear about why the learners are being tested and what the information will be used for. Here are some of the reasons for testing.

1. To find learners' total vocabulary size.
2. To compare vocabulary knowledge before and after the course.
3. To keep a continuing check on progress.
4. To encourage learning by setting short-term goals.
5. To see the effectiveness of your teaching.
6. To investigate learning.

Testing usually has two effects: (1) It provides information for the teacher and learners, and (2) it influences the teacher's and learners' attitudes. That is, testing can encourage learning and arouse interest in it. It can also be discouraging and give rise to types of learning that do not have useful long-term results. Usually an open discussion between the teacher and learners of goals and the reasons for testing will avoid these problems. Similarly, attention to individual goals and the use of self-motivating records like graphs which are under the learners' control will have good effects.

Before a test is made, it is good to consider how the information gained from the test will be used. Will it be used to make decisions about goals, to guide teaching, to get the learners to keep a record of progress, to form the basis of a discussion with the learners to help them guide their learning, or to help other teachers?

One of the aims of a testing program is to see how the initial learning of particular words is made richer and more secure. Much of this information will come not from vocabulary tests but from thoughtful observation of the learners' speaking and writing and from talking to and studying individual learners. The introspective approach of Cohen and Hosenfield (1981) is especially useful here.

Once the purposes and uses of testing have been decided, the next points to consider are what kind of knowledge is to be tested, and what type of test is most suitable. The chapter on what is involved in learning a word can provide very useful guidelines on what to test, and looking at the learning burden table in that chapter can clarify ideas about what to test.

Chapter 2

The Goals of Vocabulary Learning and Vocabulary Size

This chapter answers several questions. How many words does a second language learner need? This is answered by looking at the vocabulary of native speakers and frequency counts of vocabulary in texts. The next question is: How do we choose what vocabulary to teach? This is answered by looking at frequency counts and word lists. The final section looks at setting goals for your learners.

HOW MANY WORDS DOES A SECOND LANGUAGE LEARNER NEED?

There are two ways of answering this question. One way is to look at the vocabulary of native speakers of English and consider that as a goal for second language learners. The other way is to look at the results of frequency counts and the practical experience of second language teachers and researchers and decide how much vocabulary is needed for particular activities.

The Vocabulary of Native Speakers

The most notable feature of estimates of the vocabulary size of native speakers of English is that there is enormous variation in the estimates. Fries and Traver (1960) and Lorge and Chall (1963) discuss some of these widely varying estimates. Recent research (Nagy & Anderson, 1984; Goulden et al., in press) suggests that estimates of around 20,000 words for undergraduates are most likely to be correct. This suggests that first language learners add between 1,000 and 2,000 words per year to their vocabulary, or 3 to 7 words per day (Table 2.1).

A study of a young second language learner by Yoshida (1978) found that the learner had about 260 to 300 words in his productive vocabulary after seven months of exposure to English. His main contact with English

Table 2.1   VOCABULARY SIZE OF NATIVE SPEAKERS

Age in years	Vocabulary size	Investigator
1.3	235	Kirkpatrick
2.8	405	Kirkpatrick
3.8	700	Kirkpatrick
5.5	1,528	Salisbury
6.5	2,500	Termon and Childs
8.5	4,480	Kirkpatrick
9.6	6,620	Kirkpatrick
10.7	7,020	Kirkpatrick
11.7	7,860	Kirkpatrick
12.8	8,700	Kirkpatrick
13.9	10,660	Kirkpatrick
15.0	12,000	Kirkpatrick
18.0	17,600	Kirkpatrick

Source: Fries and Traver (1960, p. 49); also in Seashore and Eckerson (1940).

was two to three hours a day of nursery school, and his parents did not speak English at home. Tests showed that his receptive vocabulary was about 2.2 times his productive vocabulary, which would give him a receptive vocabulary of about 1000 words in a year.

Tests on learners of English as a foreign language in India and Indonesia (Barnard, 1961; Quinn, 1968) have shown that learners have a 1000- to 2000-word vocabulary after a five-year period of four or five English classes a week.

If second language learners are in the same school system as native speakers of English, they need to match the native speakers' rate of vocabulary learning and make up for the difference in English vocabulary that existed when the second language learners entered the system. So, for example, if a second language learner enters the school system knowing almost no English at the age of 5, he needs to learn, say 2500 words, plus another 1000 words a year.

Jamieson (1976) found that the English vocabulary levels of 5- and 7-year-old Tokelauan children in the New Zealand school system were two years behind their native-speaking classmates. The rate of increase of vocabulary was similar for Tokelauan children and native speakers of English. This suggests that Tokelauan children may continue to be about two years behind native speakers in vocabulary as they progress through the school. So, the difference in size between a native speaker's vocabulary and a second language learner's vocabulary can be quite large, usually several thousand words. Although second language learners may be able to match native speakers' rate of vocabulary learning, they may need special help to overcome the difference in vocabulary size.

Frequency Counts of Vocabulary in Texts

Let us look at a short text written for young native speakers and see what information it can give us about the vocabulary that a learner would need to read it. The text has been marked to show two kinds of words. The unmarked words are all in the most frequent 2000 headwords of English. These are the high-frequency words of English. Each of these words occurs often in the material we read or listen to. They also have a wide range—that is, they occur in many different kinds of material on many different topics. The underlined words are the low-frequency words of English. Although they may occur several times in a particular text, in our daily use of language most of them are not likely to occur often.

Finding an Ammonite   by Graeme Stevens

Two hundred and fifty million years ago, during the age of the dinosaurs, much of what is land today was covered by warm, shallow seas. Most of New Zealand was then under water.

Among the creatures which lived in the seas were ammonites. They had heads and tentacles, like their modern squid and octopus relatives, but were attached to coiled shells. Some were small, some were two metres or more wide. Their food was small fish, shrimps and crayfish, but they too were preyed upon by the giant sharks and sea-lizards which lived then. One ammonite shell has been found which has puncture marks from sixteen bites. A mosasaur—a kind of sea-lizard—had tried to get at the animal inside.

Where the sea has gone back from the land, people find fossils of the shells and bones of the prehistoric sea creatures. There have been fossils or ammonites found in New Zealand, although they have been less common here than in North America and Europe. This is probably because the seas that lapped on to New Zealand in those times were cooler than those of North America and Europe. Until fairly recently, the total number of ammonites found in New Zealand would hardly be enough to fill a large cabin trunk—even though geologists and rock hounds have looked quite thoroughly for them. Most ammonites found here have been quite small. A size close to that of a softball (seven to ten centimetres) is most common.

1. About 87 percent of the words in the text are high-frequency words. So with a vocabulary of just 2000 words a learner could read 87 percent of the words in the text. These 2000 high-frequency words of English are clearly very useful and important for a learner of English. Any time spent learning them will be well repaid because they cover a lot of text and will be met often.

2. The remaining words in the text can be divided into two groups. One group is made up of words like *ammonite, fossil, squid, dinosaur.* These words relate closely to the topic of the text and are important for any learner continuing to study in this area. We can call this group "technical words." Within a certain topic or subject area they may occur several times, but they are unlikely to occur in texts outside that subject area.

3. The other group is made up of words like *lapped, cabin, hounds.* They are clearly not technical words in the Ammonite text. It would be easier for second language learners if they were replaced by high-frequency words. It is very unlikely that they will occur again in the book containing this text. These are the low-frequency words of English.

Let us now look at another text. This time it is one from a secondary school textbook for native speakers. This time three groups of words have been marked. The underlined words are in a list of words which commonly occur in all kinds of university texts and which are not in the first 2000 words of English.

Machines

Since earliest times Man has tried to conserve his energy whenever he has to do some work. By a variety of methods (many still in use today) men cut down on the size of the force, or forces, they had to apply to lift objects. Today Man has become more efficient in designing devices which enable him to transfer energy from one place to another extremely quickly, or to do useful jobs more easily. These devices are called machines. A machine is an energy transformer; it can change energy from one form to another (as an electric motor does) or it can hand energy on from one place to another (as a hydraulic press does).

You will have seen many machines around your home—a crowbar, hammers, a bicycle, wheelbarrow; even a sloping piece of board is a machine. Many other and more complicated machines can be seen in factories — pulley systems, presses, fork-lifts and the bewildering complex, automatic machines, sorting and assembling articles at high speed.

However, no matter how complicated the machine is, its parts, or components, are basically one or other of two simple machines—the lever and the inclined plane. The application of these simple devices has been used by Man for many thousands of years—you may have read how giant inclined planes were used during the construction of the Great Pyramid in Egypt.

1. A machine is an _____(two words).
2. A complicated machine is just various combinations of two basic machines—the _____ and the _____(two words).
3. Into what category of basic machine do each of the following fit: a spade; a car screw-jack; a hydraulic lift; a spanner; a winding road up a mountain; a human arm?

The lever

The lever (or, as you may have called it when you were younger, the see-saw can be used to multiply forces or to produce turning effects (torques).

A screwdriver opening a paint-tin is a good example of a lever in action. The screwdriver is pivoted on a point (called the fulcrum), and you apply a force called the effort at the handle end, while the blade is pressing against the lid which is to be removed—the force caused by the lid being called the load.

The lever's input energy (E.S1) and its output energy (L.S2) are equal; as a machine it merely hands energy on; it neither creates nor destroys any. This does not prevent a lever from being very useful: it changes the force to a more convenient size (E is smaller than L, if S1 is bigger than S2); it alters the direction of the force; and it can transmit a force to any selected point.

If there is friction at the pivot some of the input energy can be wasted as heat.

1. Once again we can see how useful the high-frequency words are. Those words which are not marked in the text make up 87 percent of the words in the text. These frequent words deserve considerable time and attention from both teachers and learners.

2. The underlined words, which are from the university word list (see Appendix 2), make up 8 percent of the words in the text. These words include *conserve, energy, efficient, design, device.* These words are very useful for learners who study in upper secondary school or at a

university or a technical institute. They differ from technical words in that they are common in most kinds of technical writing. Technical words occur only in a very limited area. For learners in secondary school or at a university, the university word list is a very important goal after the first 2000 high-frequency words are mastered. Adding the first 2000 words to the 800 headwords in the university list gives learners coverage of 95 percent of the text, for a total vocabulary of 2800 words.

3. The words in boxes are the technical and low-frequency words in the text. They account for 5 percent of the words in the text. The technical words include *lever, pivot, fulcrum,* and *torques,* among others. Li (1983) examined the low-frequency words in a first-year university science text. It was found that approximately 16 different words per 400-word page were low-frequency words. Eleven of these 16 words could be classed as technical words—well worth learning by any person specializing in that subject. About 2 of the remaining 5 words per page were not technical words but had useful prefixes. So they would not be difficult to learn. The remaining 3 words per page were nontechnical words that were not repeated in the text and that did not contain useful parts. These included words like *arch, arid, gulp, ramify,* and *turbulent.* If we add the 11 technical words to the 2 with useful parts, then 13 of the 16 low-frequency words per page were worth learning by anyone intending to specialize or do further reading in the subject of a particular text. A study by Becka (1972) supports Li's findings.

Most of the words in boxes in the Machines text are technical words.

4. The low-frequency words in the Machines text include *crowbar, bewildering,* and *see-saw.* These words occur only once in the text and they are unlikely to be met again for a very long time. Most words in English are like this. The large Webster's dictionary contains about 128,000 headwords. The high-frequency words and the words in the university list add up to 2,800 words. Most lists of technical vocabulary contain 1,000 to 2,000 entries. This means that the remaining 123,200 words of English will be low-frequency words for a learner of English (Table 2.2).

So, we can describe low-frequency words as being a very large group of words, covering a very small proportion of any text, with each word occurring very infrequently. It is not worth spending time on such

Table 2.2 WORD TYPES AND TEXT COVERAGE

	No. of words	Proportion of text
High-frequency words	2,000	87%
University word list	800	8%
Technical words	2,000	3%
Low-frequency words	123,200	2%
Totals	128,000	100%

words. It is more important to teach learners strategies like guessing from context, or using word parts to deal with these words as they occur.

So far we have looked at short texts. Let us now look at a word-frequency count of much larger pieces of English. We can count the number of words in a book in three ways. (1) If we want to know how long the book is, we can count how many words there are on a typical page and multiply that by the number of pages. So if each page contains around 300 words and the book consists of 200 pages, we can estimate that the book is 60,000 words long. To put it another way, we can say the book contains 60,000 *running words.* (2) If we want to know how many words you need to understand in order to be able to read the book, we must count the words in a different way. First, we have to decide what we mean by the term *word.* Are *mend, mends, mended, mending* one word or four words? Are *branch* (of a tree) and *branch* (of a bank) one word or two? Grouping vocabulary under headwords is an attempt to increase the coverage of high-frequency vocabulary. The implication is that learning a word involves learning its derived and inflected forms as well. Second, we have to go through the book and make a list of all the different words that occur in the book. If the book contains 60,000 running words, we will probably find that it contains approximately 9,000 different words. (3) If we want to do a frequency count of the words in the book, we make a list of all the different words in the book and we count how often each one occurs. The most frequent word will probably be *the.* In 60,000 running words it will occur about 4,200 times. The word *know* will probably occur about 42 times. Many words will occur only once. *The* is a very useful word. If it occurs 4,200 times in 60,000 running words, it accounts for 7 percent of the running words in the book. If we add the frequency of the 10 most frequent words together, we will find that these 10 words account for almost 24 percent of the running words in the text. Table 2.3 gives the typical figures for a collection of texts consisting of 5 million running words.

Table 2.3 FIGURES BASED ON A COUNT OF 5 MILLION RUNNING WORDS

Different words	% of running words
86,741	100
43,831	99
5,000	89.4
3,000	85.2
2,000	81.3
100	49
10	23.7

Source: Carroll, Davies, and Richman (1971).

We can see that the figures from the Carroll et al. count, and indeed from all other frequency counts of English, support our findings from short texts. What the table does not show is that 40.4 percent of the different words in the Carroll et al. count occurred only once. That means that if learners read the 5 million running words (the equivalent of about 80 books each 200 pages long), they would meet about 35,000 different words that occurred only once. In the Carroll et al. count, these "one-timers" included words like *malformation, malignancy, maligned, malingering, Malinovski, Popeye, popguns, popinjays, popish,* and *poplin.* Clearly, low-frequency words present a major difficulty for a learner of English.

Before we summarize the features of the various types of words, let us look at their origins. Most low-frequency vocabulary comes to English from Latin and Greek, often through French. Roberts (1965) gives the following proportions: About 44% of the first 1,000 words of English come from French, Latin, or Greek. This rises to a little over 60% in the second 1,000 words, and for the rest of the first 10,000 words of English it stays around 66 percent. Thus, about two-thirds of the low-frequency words of English come from French, Latin, or Greek.

High-frequency vocabulary consists mainly of short words which cannot be broken into meaningful parts. Low-frequency vocabulary, on the other hand, while it consists of many thousands of words, is made from a much smaller number of word parts. The word *impose*, for example, is made of two parts, *im-* and *-pose*, which occur in hundreds of other words—*imply, infer, compose, expose, position.* This has clear implications for teaching and learning vocabulary and we will look at this in more detail in a later chapter. Table 2.4 summarizes the features and their implications for teaching of the different types of vocabulary described in this chapter. These implications will be dealt with in more detail in later chapters of this book.

HOW DO WE CHOOSE WHAT VOCABULARY TO TEACH?

Frequency Counts

We can get information about which words will be most useful for learners of English by looking at frequency counts of vocabulary. Usually a vocabulary count is done by making a list of the words in a particular text or group of texts and counting how often and where they occur. Some of the more recent counts have used computers to list the words and count their frequency.

Word-frequency counts can help teachers and course designers in several ways. They can help a teacher develop a feeling about which

Table 2.4 TYPES OF VOCABULARY, THEIR FEATURES, AND THE IMPLICATIONS FOR TEACHING AND LEARNING

Type of vocabulary	Number of words	Frequency	Coverage of text	Origins	Implications for teaching and learning
High-frequency words	2,000	Occur frequently in all kinds of texts	About 87% of the running words in a text	About half are from Latin, French, or Greek	Spend a lot of time on these words. Make sure they are learned.
Academic vocabulary	800	Occur frequently in most kinds of academic texts	About 8% of the running words in academic texts	About two-thirds are from Latin, French, or Greek	If learners are in upper secondary school or in tertiary education, spend a lot of time on these words. Make sure they are learned. Learning the subject involves learning the vocabulary. Subject teachers can deal with the vocabulary, but the English teacher can help with learning strategies.
Technical vocabulary	About 1,000 to 2,000 for each subject	Occur, sometimes frequently, in specialized texts	About 3% of the running words in a specialized text		
Low-frequency words	About 123,000	Do not occur very frequently	About 2% or more of the words in any text		Teach strategies for dealing with these words. The words themselves do not deserve teaching time.

words are useful and should be given attention and which are infrequent. They can provide a principled basis for developing word lists for teaching, for designing graded courses and reading texts, and for preparing vocabulary tests. Where frequency counts give information on range, they are also useful for developing specialized word lists. Several early frequency counts are described in Fries and Traver (1960). *The Teacher's Word Book of 30,000 Words* by Thorndike and Lorge (1944) is the most widely known. It has been used as the basis for vocabulary selection for many English courses and many series of simplified reading books.

The Thorndike and Lorge count tells us how often each word occurs in 1 million running words of text and gives us some indication of its range. The range of a word is a measure of the number of different types of texts in which a word occurs. Words with a wide range occur in many different kinds of texts and fields of study. The most useful words for our learners are high-frequency words which have a wide range. The Thorndike and Lorge count is still used as a source of information about what words to teach, but it is based on work done over 50 years ago. Roberts (1965, pp. 22–23) describes some of its weaknesses.

More recent counts are those done using computers, namely, Kucera and Francis (1967) and Carroll et al. (1971). These counts give information about the frequency and range of words and, unlike the Thorndike and Lorge count, give a list of the words in order of frequency. The Carroll et al. (1971) count gives detailed information about range and could be very useful in making lists of vocabulary for special subject areas.

Frequency counts provide useful information about the frequency and range of words. But they do not give us enough information and there are several problems associated with them.

1. The most serious problem with word-frequency lists is that certain useful and important words do not occur in the first or second 1000 words. They may occur only at the third, fourth, or fifth thousand word level. Often these useful words are concrete nouns. For example, *soap, a bath, (a) chalk,* and *a stomach* are not in the first 2000 words of Thorndike and Lorge's 1944 list. The words *tidy, stupid,* and *behavior* are not in the first 3000 words of that list. *Damage* is not in the first 2000. If we use frequency as our only principle of selection, we cannot include these words in a beginners' vocabulary of 1000 words.

2. Another problem is that some words that are not suitable for a beginners' vocabulary come in the first 1000 words of most frequency lists. Some examples are *bank* (v.), *bill, company, deal, issue, labor, stock,* and *supply.* (All these words are in the first 1000 words of Thorndike and Lorge's 1944 list. *Chicago, thee,* and *thou* are also in the first 1000.) Some of these words would be frequent in business letters. They would not be very frequent in most other kinds of writing.

3. Word-frequency lists often disagree. Sometimes a word has a high frequency in one list and a fairly low frequency in another list. A person who makes a frequency list has to choose what kinds of writing to count. The choice will affect the results.

4. Usually the order of the words in a frequency list is not the best order in which to teach the words. For example, *his* is the 74th word in one list and *hers* is the 4151st word. If we followed this order, we would not teach the word *hers* until senior high school or university level!

5. Word-frequency lists are not reliable above a certain level. Engels (1968) found in his study of *The General Service List* that words above the 1000-word level did not include many of the words in the texts he examined.

When preparing a word list for learners of English, other criteria in addition to frequency and range need to be used. Here is a list of possible criteria (Richards, 1970).

1. frequency
2. range
3. language needs
4. availability and familiarity
5. coverage
6. regularity
7. ease of learning or learning burden

Word Lists

Some famous word lists have emphasized some of these criteria. Basic English (Richards, 1943) was developed making coverage and ease of learning the main criteria. Coverage is the capacity of a word to take the place of other words (Mackey & Savard, 1967). For example, the word *foot* is a useful word from the point of view of coverage because

1. It can often be used to make a definition of other words, e.g., *Your heel is the back part of your foot.*
2. Its meaning can be extended to replace other words, e.g., *the base of a tree* can be replaced by *the foot of a tree.*
3. It can combine with other words to make new words, e.g., *footpath, football, footstep, foot-and-mouth disease.*

There is one more type of coverage. The word *table* is useful from the point of view of coverage because

4. Its meaning includes the meaning of other words and it can be used instead of them, e.g., *bench, desk, escritoire.*

Mackey and Savard calculated the coverage of French words. They found that words with high coverage were not necessarily the most frequent words. So information about the coverage of words is useful information to add to information about frequency.

West's (1935) definition vocabulary is another good example of a list based on the principle of coverage. The list was made to produce a dictionary where all the definitions were given in the smallest possible vocabulary.

The most famous and most useful list of high-frequency words is West's *General Service List of English Words* (1953). This list is now old (the headwords in the list have not changed since 1936), but it has still not been replaced as a source of useful information about particular words and as a collection of the most important vocabulary for a learner of English. The list contains 2000 headwords. It gives the frequency of the main headword plus the relative frequency of its meanings.

OWN   3244e

own, adj.	*(showing possession)* This is my own house; I don't rent it Use this pen. Thanks, I'd rather use my own	
of your own	Why do you use my pen? Haven't you got a pen (one) of your own (1.3%) [= *for myself*, Cooked my own dinner as the servant was ill, 0.2%]	89%
own, v.	He owns a lot of land round here [own that, own up to = confess, 1.5%]	9%
owner, n., 314e;	ownership, n., 52e	

(GSL, p. 347)

Unlike most other frequency counts, the *GSL* includes forms under the same headword. So the entry for *own* includes *own, owns, owned, owning*, and includes *owner* and *ownership*. Other counts would count and list most of these as different words. In the entry, the parts of speech are distinguished (*own* as an adjective and *own* as a verb) and the various meanings are also distinguished. Each of these distinctions is given a percentage so that the user of the list can easily decide which meaning and use is the most important. It is also possible to decide if *own* as a verb is more frequent than *have* with a similar meaning by calculating 9 percent of 3244 and comparing it with the entry for *have*. Each meaning is also accompanied by examples.

The frequency figure (for *own* it is 3244) gives the number of occurrences in 5 million words. The *GSL* is based on a very large corpus of material.

During the making of the list, factors other than frequency and range were considered. These are described with examples in the introduction to the list (pp. ix–x). The list was not designed as a list of unrelated words but was viewed as a "little language" (Salling, 1959). The makers of the list wanted to be sure that most material could be rewritten within the vocabulary of the list and that the list did not contain unnecessary items. Some low-frequency items were included because there seemed to be no other easy way of expressing a particular idea—for example *whistle, reproduce*. Some frequent items were omitted because other more useful items could easily replace them. West and others produced enormous quantities of simplifed readers and other language learning texts using the *GSL* vocabulary. The Longman Simplifed English Series and the New Method Supplementary Readers, for example, are based on the *GSL*. The *GSL* is certainly tried and tested and it will not be easily replaced.

Most publishers of simplifed reading books have their own graded vocabulary lists. Some of these, like the Longmans, Macmillan Rangers, and Collins lists, are available in a published form. Some other lists are available only to writers of graded readers. All these word lists arrange the vocabulary into steps.

The *Cambridge English Lexicon* (Hindmarsh, 1980) is a graded word list of about 4500 words with over 8000 meanings. The list is a collation of most of the recent word counts, *The General Service List* (West, 1953), the Kucera and Francis count (1967), the Thorndike and Lorge count (1944), and the Carroll et al. count (1971), in addition to many other lists. The words are graded into five levels (Table 2.5).

The meanings of the words are also graded, from 1 to 7. The principles of grading the words and their meanings are not described in the introduction, and it seems that some of the grading was done by intuition rather than using the figures in the *General Service List* (GSL). In general, however, the gradings do correspond to the *GSL*, and where they do not the grading seems reasonable. For some entries the *GSL* gives a more detailed division of meanings; for other entries the *Cambridge English Lexicon* (CEL) gives a more detailed division. A major advantage of the *CEL* over the *GSL* is that it covers a larger vocabulary—

Table 2.5   WORDS IN THE CAMBRIDGE ENGLISH LEXICON

Level	Total	Cumulative total
1	598	598
2	617	1215
3	992	2207
4	1034	3241
5	1229	4470

over a thousand words more. A disadvantage is that it does not list derivatives under the headword as the *GSL* does. So, in the *CEL*, *unpaid* is listed under the letter *U*. In the *GSL*, *pay* includes *pay*, *pay-day*, *unpaid*, and *payment*.

Other word lists, like the *Longman Lexicon of Contemporary English* (McArthur, 1981), provide detailed information about usage with lots of examples. A review of these word lists and others can be found in Fox and Mahood (1982a, 1982b), and Jeffries and Willis (1982).

In this section we have looked briefly at some of the criteria which can be used to make word lists for the preparation of learning material. Typically these word lists are used to make simplified reading material, to design controlled speaking and writing courses, and to produce dictionaries. Although the information from frequency counts provides an important basis for the development of most of these lists, frequency alone is not sufficient. The purpose of the list affects the choice of criteria. Coverage, for example, is an important criterion if a list is going to be used to define the meanings of other words. Regularity is important when learnability is one of the aims of a list.

The lists described in this section are the result of an enormous amount of work. West's (1935) description of his construction of the definition vocabulary is an excellent example of the thought, care, and effort that go into the making of such lists. Their main aim is to make the learning of English more manageable by providing a tried and principled basis for vocabulary selection.

SETTING GOALS FOR YOUR LEARNERS

In this chapter we have looked at the vocabulary of native speakers and information from frequency studies of English. The research on native speakers indicates that second language learners in the same school system as native speakers of English may have to increase their vocabulary by around 1000 words a year, besides making up a 2000- to 3000-word gap, in order to match native speakers' vocabulary growth.

The research on frequency counts gives clear guidelines, particularly for learners of English as a foreign language. They may need a productive vocabulary of around 2000 high-frequency words plus the strategies to deal with the low-frequency words they meet. Learners with special goals, such as study at a university, need to acquire a further 1000 high-frequency words. Several lists of high-frequency words are available for teachers and course designers.

Using this information it is now possible for you to apply it to your learners and your language learning course. Table 2.6 will help you.

Table 2.6   VOCABULARY LEARNING GOALS

	Number of Words	
a. My learners' vocabulary at the beginning of the course		You can find this information by (1) testing your learners, (2) counting the words in their previous textbooks, (3) checking the vocabulary at the beginning of their coursebook with a frequency count such as West (1953), Thorndike and Lorge (1944), Kucera and Francis (1967), Hindmarsh (1980). This will help you decide the word level of the coursebook, (4) getting this information from the introduction to your textbook or from your syllabus notes.
b. Their required vocabulary at the end of the course.		You can decide this by (1) counting the words in the coursebook, (2) looking at the uses they must make of English at the end of the year and deciding how many words they need to know in order to do this, (3) setting a goal that you think they could reach, (4) seeing what your syllabus or coursebook sets as a goal, (5) comparing your learners with native speakers of English of the same age.
c. The number of words to be learned during the course, $c = b - a$.		
d. The number of English lessons, hours, or school weeks in the course.		
e. The number of words to be learned in each lesson, hour, or school week, $e = c \div d$.		