Open Linguistics Series

Series Editor: Robin Fawcett, Cardiff University

This series is 'open' in two senses. First, it provides a forum for works associated with any school of linguistics or with none. Most practising linguists have long since outgrown the unhealthy assumption that theorizing about language should be left to those working in the generativist-formalist paradigm. Today large and increasing numbers of scholars are seeking an understanding of the nature of language by exploring one or other of the various cognitive models of language, or in terms of the communicative use of language – or both. This series is playing a valuable part in re-establishing the traditional 'openness' of the study of language. The series includes many studies that are in – or on the borders of – various functional theories of language – and especially because it has been the most widely used of these, Systemic Functional Linguistics. The general trend of the series has been towards a functional view of language, but this simply reflects the works that have been offered to date. The series continues to be open to all approaches – including works in the generativist-formalist tradition.

The second way in which the series is 'open' is that it encourages works that open out 'core' linguistics in various ways: to encompass discourse and the description of natural texts; to explore the relationships between linguistics and its neighbouring disciplines such as psychology, sociology, philosophy, and cultural and literary studies; and to apply it in fields such as education, language pathology and law.

Relations between the fields of linguistics and artificial intelligence are covered in a sister series, *Communication in Artificial Intelligence*. Studies that are primarily descriptive are published in a new series, *Functional Descriptions of Language*.

Words, Meaning and Vocabulary

An Introduction to Modern English Lexicology

Howard Jackson
Etienne Zé Amvela

Words, Meaning and Vocabulary

Open Linguistics Series

This series is 'open' in two senses. First, it provides a forum for works associated with any school of linguistics or with none. Most practising linguists have long since outgrown the unhealthy assumption that theorizing about language should be left to those working in the generativist-formalist paradigm. Today large and increasing numbers of scholars are seeking an understanding of the nature of language by exploring one or other of the various cognitive models of language, or in terms of the communicative use of language – or both. This series is playing a valuable part in re-establishing the traditional 'openness' of the study of language. The series includes many studies that are in – or on the borders of – various functional theories of language – and especially because it has been the most widely used of these, Systemic Functional Linguistics. The general trend of the series has been towards a functional view of language, but this simply reflects the works that have been offered to date. The series continues to be open to all approaches – including works in the generativist-formalist tradition.

The second way in which the series is 'open' is that it encourages works that open out 'core' linguistics in various ways: to encompass discourse and the description of natural texts; to explore the relationships between linguistics and its neighbouring disciplines such as psychology, sociology, philosophy, and cultural and literary studies; and to apply it in fields such as education, language pathology and law.

Relations between the fields of linguistics and artificial intelligence are covered in a sister series, *Communication in Artificial Intelligence*. Studies that are primarily descriptive are published in a new series, *Functional Descriptions of Language*.

Words, Meaning and Vocabulary

An Introduction to Modern English Lexicology

Howard Jackson
Etienne Zé Amvela

Contents

Preface

This introduction to the words and vocabulary (lexicology) of English arises from courses that Etienne Zé Amvela and Howard Jackson have taught over many years at their respective universities, the former in Cameroon (Yaoundé and Buea) and the latter in Birmingham, UK (Birmingham Polytechnic/University of Central England). The study of vocabulary, by contrast with syntax, has been a rather neglected pursuit in English over recent years, and there are few textbooks in the area. It is our hope that this work will go some way towards redressing the balance.

Etienne Zé Amvela has been responsible for the first four chapters and Howard Jackson for the last four, but we have commented extensively on each other's work, and we hope that the book reads as a seamless whole.

The book is aimed at students of English language/linguistics, taking courses in the analysis and description of the English language, possibly with little prior knowledge of linguistics. However, we do advise students to have a good up-to-date dictionary to hand, preferably of the desk-size or collegiate type. In order to encourage interaction with the material discussed in the book, each chapter is interspersed with exercises, some of which require dictionary consultation. A key to the exercises is provided at the end.

Howard Jackson
Etienne Zé Amvela

To Hilary and Mama Régine

1   What is Lexicology?

This book is about English lexicology. But before we begin to discuss the various facets of the subject, we need to suggest a definition of the term lexicology (1.1). The rest of the chapter will discuss lexicology as a level of language analysis (1.2), and explore the structure of the English vocabulary (1.3).

1.1 Lexicology defined

At this early stage, a definition of lexicology is best considered as a working tool for a better understanding of subsequent chapters. In fact, we believe that this whole book is an answer, or at least a partial answer, to the fundamental question, 'What exactly is lexicology?' We shall not have completed our definition until we reach the end. Even then, we cannot claim to have said everything about lexicology.

For the purpose of an introductory textbook of this nature, lexicology may be defined as the study of lexis, understood as the stock of words in a given language, i.e. its vocabulary or lexicon (from Greek *lexis*, 'word', *lexikos*, 'of/for words'). This working definition shows that the notion of 'word' is central in the study of lexicology. However, 'word' itself needs to be defined and discussed as a technical term. This is done in Chapter 3. Since our main focus in this chapter is on the definition of lexicology, and in order to avoid a lengthy digression, we use 'word' somewhat loosely, in the usual traditional sense of a sequence of letters bounded by spaces. A comparison of the words 'vocabulary', 'lexis', and 'lexicon' would show that the three items may be considered more or less synonymous. However, it must be added that the first is more colloquial, the third more learned and technical, and the second may be situated half-way between the other two. A distinction must, nevertheless, be drawn between the terms 'vocabulary', 'lexis' and 'lexicon' on the one hand, and 'dictionary' on the other. While each of the first three may refer to the total word stock of the language, a dictionary is only a selective recording of that word stock at a given point in time.

Lexicology deals not only with simple words in all their aspects, but also

with complex and compound words, the meaningful units of language. Since these units must be analysed in respect of both their form and their meaning, lexicology relies on information derived from morphology, the study of the forms of words and their components, and semantics, the study of their meanings. A third field of particular interest in lexicological studies is etymology, the study of the origins of words. However, lexicology must not be confused with lexicography, the writing or compilation of dictionaries, which is a special technique rather than a level of language study.

To avoid possible confusion and in order to introduce some of the technical terms we need in our discussion of lexicology, we shall examine the four related fields mentioned above, viz. morphology, semantics, etymology and lexicography. Finally, we shall discuss lexicology as a level of language analysis.

EXERCISE 1/1

Examine the following definitions of 'lexicology'. What do they agree on as the scope of lexicology? And where do they disagree?

1. An area of language study concerned with the nature, meaning, history and use of words and word elements and often also with the critical description of lexicography. (McArthur, ed. 1992)

2. The study of the overall structure and history of the vocabulary of a language. (*Collins English Dictionary* 1998)

3. A branch of linguistics concerned with the meaning and use of words. (*Longman Dictionary of the English Language* 1991)

4. The study of the form, meaning, and behaviour of words. (*New Oxford Dictionary of English* 1998)

1.1.1 Morphology

Morphology is the study of morphemes and their arrangements in forming words. Morphemes are the smallest meaningful units which may constitute words or parts of words. They are 'smallest' or 'minimal' in the sense that they cannot be broken down further on the basis of meaning, as Katamba (1994: 32) puts it: 'morphemes are the atoms with which words are built'. They are 'meaningful' because we can specify the kind of relationship they have with the non-linguistic world.

Consider the following items: cat, child, with, sleeping, armchairs, farmer. A close examination shows that *cat, child* and *with* cannot be analysed further into meaningful units. However, *sleeping, armchairs* and *farmer* can be analysed as 'sleep + ing', 'arm + chair + s', and 'farm + er'.

The items *cat, child, with, sleep, -ing, arm, chair, -s, farm,* and *-er* are all morphemes. Some are simple words such as *cat, child, with, sleep, arm, chair,* and *farm,* while others are only parts of words such as *-ing, -s,* and *-er.* But both types meet our definition of morpheme. On the one hand, they are minimal, since they cannot be broken down into further meaningful units; on the other hand, they are meaningful, because we can establish a stable relationship between each item and the non-linguistic world of experience. For example, the references of *cat, farm* and *chair* can be explained by pointing or acting out the meaning as in 'This is a chair', or 'That is a farm', 'It is a domestic animal that goes "miaow", "miaow"'. The meaning of *with* may be given as 'in company of', 'in antagonism to'; that of *-s* as 'plural'; while that of *-er* may be expressed as follows: '*-er* combines with the preceding lexical item to designate things or persons with a function describable in terms of the meaning of the preceding morpheme'. For example, the meaning of *-er* in *farmer* and *dreamer* is describable in terms of those of *farm* and *dream* with which the morpheme *-er* is combined.

Morphemes that can occur alone as individual words are 'free' morphemes. Those that can occur only with another morpheme are 'bound' morphemes. Thus, the morphemes 'cat', 'chair', 'farm' are free, while '-ing', '-s', and '-er' are bound, indicated by the hyphen (-). Any concrete realization of a morpheme in a given utterance is called a 'morph'. Hence, the forms *cat, chair, farm, -ing, -s,* and *-er* are all morphs. Morphs should not be confused with syllables. The basic difference between the two is that while morphs are manifestations of morphemes and represent a specific meaning, syllables are parts of words which are isolated only on the basis of pronunciation.

An examination of a number of morphs may show that two or more morphs may vary slightly and still have the same meaning. For example, the indefinite article may be realized either as *a* or as *an*, depending on the sound (not the letter) at the beginning of the following word. Morphs which are different representations of the same morpheme are referred to as 'allomorphs' of that morpheme (from Greek *allo* 'other' and *morph* 'form'). For example:

> *a* context vs. *an* index
> *a* battle vs. *an* apple
> *a* union vs. *an* onion.

The last pair of words deserves some comment. Its members begin with *u* and *o*, which are classified as vowel letters. However, while *union* begins with the same sound as *yes* which is treated as a consonant, *onion* begins with the same sound as *onwards*, which is a vowel; hence '*a* union' vs. '*an* onion'.

The use of 'vs.' (versus) highlights the point that where the allomorph *an* occurs, its counterpart *a* cannot occur and vice versa. They are therefore mutually exclusive and are said to be in complementary distribution. It should be pointed out that as a descriptive term, 'distribution' refers to the total set of distinct linguistic contexts in which a given form occurs,

sometimes under different morphological shapes. For example, the distribution of the indefinite article described above may be defined as: *a* before consonant sounds (e.g. *a* battle) and *an* before vowel sounds (e.g. *an* apple).

We now turn our attention to the relation between morphology on the one hand, and simple, complex and compound words on the other. Simple words such as *door, knob, wild, animal* are all free morphemes. They are therefore morphologically unanalysable. Complex (or derived) words such as *spoonful, wildish, reanimate, mentally, farmer* are formed from simpler words by the addition of affixes or some other kind of morphological modification. The limiting case for complex words is that of zero modification or conversion as in *answer, call* and *question*, which may be either nouns or verbs, or *clean, dirty,* and *dry,* which may be either adjectives or verbs, without the addition of further sounds/letters. Compound words, or simply compounds, are formed by combining two or more words (free morphemes) with or without morphological modification, e.g. *door-knob, cheeseburger, pound saver, wild-animal-tamer.* It should be pointed out that the distinction between word compound (solid and hyphenated) and phrasal compound (open) is not very clear in English. This fact is reflected by the inconsistency with which spaces and hyphens are used with compounds in written English.

This brief discussion shows the importance of morphology in lexicology. In fact, the construction of words and parts of words, and the distinction between the different types of words are all based on morphological analysis. As will be seen later in Chapter 4, morphology is particularly relevant in the discussion of word formation.

1.1.2 Semantics

Semantics is generally defined as the study of meaning. Its aim is therefore to explain and describe meaning in natural languages. The term 'meaning' is used here in the ordinary, non-technical sense, without reference to any particular theoretical framework. Most linguists agree that meaning pervades the whole of language. However, they are not always unanimous on the terms to be used in the discussion of semantics. For our purpose in this book, we adopt the terminology presented and the theoretical distinctions made by Jackson (1988: 244–7) in his brief treatment of semantics.

To highlight the pervasive nature of meaning, Jackson states that if we are to talk about semantics at all, then we should identify several kinds of semantics: pragmatic semantics, which studies the meaning of utterances in context; sentence semantics, which handles the meaning of sentences as well as meaning relations between sentences; lexical semantics, which deals with the meaning of words and the meaning relations that are internal to the vocabulary of a language. Semantics is usually approached from one of two perspectives: philosophical or linguistic. Philosophical semantics is concerned with the logical properties of language, the nature of formal

theories, and the language of logic. Linguistic semantics involves all aspects of meaning in natural languages, from the meaning of complex utterances in specific contexts to that of individual sounds in syllables.

Consequently, since semantics covers all aspects of human language, it must be considered not only as a division of lexicology, but also as part of phonology, syntax, discourse analysis, textlinguistics, and pragmatics. But for our purpose in this book, it is enough to assume that lexical semantics is relevant to lexicology.

It will also be useful to introduce two terms which belong more to the area of sentence semantics, but which are equally relevant to our discussion of lexicology, viz. 'acceptability' and 'meaningfulness'.

'Acceptability' and 'meaningfulness' are distinct but related concepts. They are important in our discussion of lexicology because we may have utterances that are meaningless but acceptable, while others may be meaningful but unacceptable. Consider the following:

That woman is a man.
That doll is a bomb.
That walking-stick is a gun.

They may be considered meaningless in the sense that a human being cannot be both 'a woman' and 'a man' at the same time. Similarly, it may be argued that an object cannot be 'a doll' and 'a bomb', just as the same object cannot be simultaneously 'a walking-stick' and 'a gun'. But with a bit of imagination, one can think of contexts where such utterances, and others like them, can be considered acceptable. For example, in a play, a character may be a man biologically and play the role of a woman; in a film, an actor could be carrying a doll or a walking-stick which in fact could be deadly weapons such as a bomb or a gun. To paraphrase Leech (1969: 13), the 'effective message' in all such utterances is: 'What appears as an "x" is in fact a "y".'

There are other types of meaningless utterance that may be acceptable for various reasons. Some may involve 'slips of the tongue', 'typographical errors', 'sarcasms', 'different figures of speech', etc. Others may be considered deviations from the norm of the language under study. Still others may have different origins or justifications. For example, if a person who has a bad cold and a completely blocked nasal cavity says 'It's *dice beeting* you', after he/she has just been introduced to someone, this utterance may be considered meaningless, strictly speaking. However, the 'effective message' it conveys in this context would be something like 'It's nice meeting you but I have a bad cold.'

The important point here is that there are several factors that contribute to the meaningfulness and the acceptability of utterances. As opposed to utterances that are meaningless but acceptable, others are meaningful but unacceptable. The latter category includes assertions that are false because of our knowledge of the real world, rather than for purely semantic reasons. Consider the following:

Crocodiles can fly.
The basket ate the vegetables.
John's behaviour pleased the *bananas*.

We may use different criteria to account for such utterances. For example, they may be explained by logical argument to highlight the contradictions, inconsistencies or incompatibilities in the message. From a syntactic point of view, such utterances are treated as errors in predication, meaning that the subject or object noun phrases are syntactically unsuitable to the corresponding verb phrases. Hence, the subjects *crocodiles* and *the basket* are syntactically unsuitable to the verb phrases *can fly* and *ate* respectively. Such examples point to the fact that all of syntax, semantics and lexicology contribute to a comprehensive study of language.

1.1.3 *Etymology*

A third field which should be of particular interest in lexicological studies is etymology, which may be defined as the study of the whole history of words, not just of their origin. The term 'etymology' was coined by the Stoics, a group of Greek philosophers and logicians who flourished from about the beginning of the fourth century BC. They noticed a lack of regularity in the correspondences between the forms of the language and their respective contents. In other words, they found no necessary connection between the sounds of the language on the one hand and the things for which the sounds stood on the other. Since they were convinced that language should be regularly related to its content, they undertook to discover the original forms called the 'etyma' (roots) in order to establish the regular correspondence between language and reality. This was the beginning of the study known today as etymology.

One of the difficulties faced by etymological studies is that some words are not etymologically related to ancient forms. It is therefore difficult to establish and indicate their origins. Consequently, the forms from which such words are said to derive can only be produced by analogy. Another difficulty is that while it is possible to specify the exact time when some terms entered the language, for example through borrowing, it is clearly impossible to say exactly when a form was dropped, especially since words can disappear from use for various reasons.

The most crucial difficulty faced with etymological studies is that there can be no 'true' or 'original' meaning, since human language stretches too far back in history. To paraphrase an example given by Palmer (1981: 11), one may be tempted to say that from the etymological point of view the adjective *nice* really means 'precise' as in 'a *nice* distinction'. But a study of its history shows that the word once meant 'silly' (Latin *nescius*, 'ignorant'), and earlier, it must have been related to *ne*, 'not' and *se*, probably meaning 'cut'. The form *se* is also used in the Modern English words *scissors* and *shears*. But at this level of analysis, one would still be left with the Latin items *ne*,

'not' and *se*, 'cut', the origins of which are still unknown. In other words, no matter how far back one goes in history, one cannot expect to reach the beginning of time. So, the Stoics' quest has proved fruitless.

As suggested in our definition, etymological information goes beyond the origin of the word. It also makes reference to cognates (i.e. words related in form) in other languages. Furthermore, in the case of borrowed words, it gives the source language, together with the date when the borrowing took place. Finally, it supplies any other information on the previous history of the word. In dictionary entries, such information is contained traditionally in square brackets. The amount of detail provided in etymologies varies from one dictionary to another. But in spite of its potentially wide range of coverage, etymological information is generally scanty in most monolingual dictionaries. In fact, it is totally absent from both bilingual and learners' dictionaries, presumably on the grounds that it is not helpful to language learners. However, as pointed out by Jackson (1988: 175), it could be argued that 'knowledge of etymology may help some learners to understand and retain new vocabulary items'.

Before we close this brief discussion of etymology, we should mention the expression 'folk etymology'. It is a historical process whereby speakers who cannot analyse an obscure form replace it with a different form which is morphologically transparent. Gramley and Pätzold (1992: 31) give good examples of this process with the noun 'bridegroom' and the verb 'depart'. In Middle English, the original spelling of the first word was 'bridegome' (*bride*, 'bride'; *gome*, 'man'). But the second element ceased to be understood and was altered to *groom* to make *bridegroom*. The etymology of *depart* is more complex. Initially, its use was restricted to wedding ceremonies to mean 'separate' in the expression 'till death us *depart*'. Later, the verb became obsolete and was analysed as *do* and *part*, hence the corresponding Modern English expression 'till death *do* us *part*'. Although a few other examples could be given, it must be acknowledged that folk etymology is not a very productive process in Modern English. See Chapter 2 for a discussion of the origins of English words; and Chapter 4 for the various word-formation processes.

EXERCISE 1/2

Etymology can be a fascinating area of study. Look up the etymology of *bridegroom* in your dictionary. Does it mention the folk etymology, i.e. the change from 'gome' to 'groom'?

Now look up the etymology of *adder*. Is there any folk etymology at work here?

Finally, look up the etymology of *snake*. What does your dictionary say is the origin? And does it give cognates?

1.1.4 Lexicography

The fourth and last field which is of special interest in lexicological studies is lexicography, which has already been defined as a special technique, the writing and compilation of dictionaries. This definition may be considered rather restrictive. In its widest sense, lexicography may also refer to the principles that underlie the process of compiling and editing dictionaries. Some of those principles are clearly lexical or lexicological in nature, while others stem from the specific domain of book production and marketing. But lexicographical compilation may be considered as derived from lexicological theory (Jackson 1988: 248). It is in this sense that lexicography can be regarded as 'applied lexicology'.

However, it should be acknowledged that it is only in recent years that the link between lexicography and linguistics has been clearly established. For example, the accuracy and consistency in the transcription of words and the adoption of a 'descriptive' as opposed to a 'prescriptive' approach to lexicography are direct applications of linguistic principles. It may be argued that initially, lexicography developed its own principles and tradition independently of linguistics in general; but this is no longer the case. In fact, since current dictionaries are compiled mainly by lexicographers who have been trained in linguistics, one should expect a more direct and more substantial input from lexicology. However, it should be pointed out that lexicology is not the only branch of linguistics which provides an input to lexicography. Clearly, morphology, syntax and phonology do. And sociolinguistics, too, contributes, not only in the study and selection of the language variety to be used in the dictionary, but also in the inclusion of information on style and registers. For a detailed discussion of lexicography as applied lexicology, see Chapter 8.

1.2 Lexicology as a level of language analysis

Lexicology is only one possible level of language analysis, others being phonology, morphology, syntax and semantics. Although an attempt may be made at treating any of these levels in isolation, it must be said that none of them can be studied successfully without reference to the others. All these different levels of analysis interact with one another in various ways, and when we use language, we call on all simultaneously and unconsciously. We briefly discussed morphology and semantics in 1.1; here we consider the relation of lexicology to phonology and syntax.

1.2.1 Lexicology and phonology

It may be thought at first sight that phonology does not interact with lexicology in any significant manner. But a close analysis will reveal that in many cases, the difference between two otherwise identical lexical items can be reduced to a difference at the level of phonology. Compare the pairs of words p*ill* and b*ill*, shee*p* and shi*p*, mea*t* and mea*l*. They differ only in one

sound unit (the position of which has been shown in each word) and yet the difference has a serious effect at the level of lexicology. As suggested by these examples, the sounds responsible for the difference may occur anywhere in the structure of the word; i.e. at the initial, medial, or final position. In some cases, the phonological difference does not involve discrete sound units but 'suprasegmental' or 'prosodic' features such as stress; and yet, even such differences are enough to differentiate otherwise identical items; e.g. *ex'port* (verb), vs. *'export* (noun). Note that the symbol (') is placed immediately before the syllable which receives primary stress.

Compounds provide another good example to show the relevance of phonology in lexicology. At first sight, the process of compounding may be viewed as a simple juxtaposition of two words. Thus, *green* and *house* may be put together to form *greenhouse*, 'a glass house for growing plants'. But such an analysis would be superficial, since the same items can be put together in the same order to produce *green house*, 'a house that is green'. The major difference between the two utterances is a matter of stress, which is a phonological feature. But this feature is enough to distinguish compounds from noun phrases containing the same words. Compare the stress pattern of the compound nouns in (a) and the corresponding adjective plus noun constructions in (b).

(a) *Compound*	(b) *Noun Phrase*
'blackboard	,black 'board
'blackbird	,black 'bird
'greyhound	,grey 'hound
'White House	,white 'house

Note: (') = main or primary stress, and (,) = secondary stress.

As a general rule, the primary stress falls on the first word of the compound as in '*blackboard* (a dark smooth surface in schools for writing on with chalk). The same rule applies to the rest of the words in (a). However, in the noun phrases in (b) like ,*black 'board* (as opposed to any board that is painted green, red or yellow) both words can potentially receive stress. All the examples in (b) are phrases, not words. Consequently, they function as units at the syntactic level. They are therefore an indication that stress, which is a phonological feature, has a direct influence on syntax. The relation between lexicology and syntax is further explored in 1.2.2. This is a clear illustration of the interdependence of phonology, lexicology and syntax. However, since language is so complex that it cannot be studied all together, we must consider each level as if it were autonomous. But it must always be remembered that such an approach is only methodological and does not always reflect the way language operates when it is used by its speakers.

1.2.2 Lexicology and syntax

We use the term 'syntax' to refer to the particular knowledge which enables us to assemble words when we construct sentences. Syntax is also responsible, at least in part, for an appropriate understanding of the sentences we hear and those we read. That is, syntax is concerned with the relationships between words in constructions and the way these words are put together to form sentences.

As a basic assumption, we believe that we might know the meanings of all the words in a large English dictionary and still be quite unable to speak or understand the language. Consequently, to say that someone speaks English, or that they 'know' English, amounts to saying that they have somehow acquired a set of rules, among which are the rules of syntax, that enables them to produce English sentences as needed. The rules also enable them to understand the sentences of another person speaking the language. However, unless they have some special training in linguistics, the speaker and hearer cannot talk confidently about the nature of such rules.

In 1.2.1 above we saw how semantics and phonology are both relevant in any serious study of lexicology. We shall now investigate the relationship between lexicology and syntax. We assume that, although these two levels of language analysis are comparable, they may also be kept distinct. One argument in favour of the distinction between syntax and lexicology is the observation that a given sentence may be syntactic but unacceptable from the lexical point of view. One such example is the famous sentence 'Colorless green ideas sleep furiously', proposed by Chomsky (1957). This sentence is built according to the rules of English syntax but it is unacceptable on lexical grounds. If a sentence can satisfy the rules of syntax but be unacceptable lexically, this is perhaps an indication that the rules of syntax are different from those of lexicology; consequently, the two levels are distinct. In fact, the problem of the distinction between lexicology and syntax may be reduced to the distinction between sentences that are unacceptable on syntactic grounds and those that are deviant from the lexical point of view.

The essential difference between syntax and lexicology is that the former deals with the general facts of language and the latter with special aspects. It is in the main a question of general versus particular. Syntax is general because it deals with rules and regularities that apply to classes of words as a whole, whereas lexicology is particular because it is concerned with the way individual words operate and affect other words in the same context. Although borderline cases do exist in both lexicology and syntax, e.g. in the case of 'grammatical' or 'function' words (1.3.4), the distinction between the two levels is fairly clear.

At first sight, it may be thought that when judged in terms of how deviant they are, lexical restrictions are generally not a matter of well-established rules but of tendencies. In other words, it may be assumed that if asked whether or not a given lexical association is acceptable, one cannot answer

by a categorical 'yes/no'; one is more likely to give an answer of the nature 'more/less', or 'it depends on the context'. Such an assumption is, however, an oversimplification. In syntax as well as in lexicology, there are cases of deviation which may be answered by yes/no, and others that can be answered only by more/less, though a 'yes/no' answer is more likely in syntax than in lexicology. For example, a sentence such as 'Sophisticated mice prefer to eat red elephants', though undoubtedly syntactic is lexically doubtful because it does not correspond to our experience of the world. Judging from our present knowledge of the natural world, the acceptability of this sentence is not a matter of 'more/less' but of a categorical 'no'. However, a sentence such as 'The flower gracefully walked away' may seem odd in the sense that 'flowers' are not normally associated with 'walking' but with some imagination, we can picture a context in which this sentence, which is already acceptable syntactically, is also acceptable lexically.

Similarly, some sentences are clearly ungrammatical, while others are clearly well-formed syntactically. For example, a sentence such as 'Did it he and I' is clearly deviant only on syntactic grounds and it could be corrected simply by changing the word order into 'He and I did it'. But there are also marginal cases such as 'Give it to whomever wants it' versus 'Give it to whoever wants it', where English speakers are not unanimous as to which alternative is grammatical. As a final observation, it should be pointed out that some sentences, such as 'Did it John and the table', are deviant on both syntactic and lexical grounds.

EXERCISE 1/3

Consider the following 'deviant' sentences. In which of them would an alternative selection of words (lexis) make an improvement, and in which does the arrangement of the words (syntax) need to be adjusted?

1. Visitors are aggressively requested to remove their shoes before leaving the temple.

2. You put can table the the on bread you bought have.

3. All mimsy were the borogroves and the mome raths outgrabe.

4. Anyone lived in a pretty how town, with up so many bells floating down.

5. Off you go, up the apples and pears and into uncle ned.

1.3 The structure of English vocabulary

As used in this book, the terms vocabulary, lexis and lexicon are synonymous. They refer to the total stock of words in a language (see 1.1).

Assuming the existence of some internal organization of vocabulary, the basic question we want to address here may be formulated as follows: 'How is the total stock of English words structured and organized?'

There have been a number of attempts to discover some of the general principles on which vocabulary is organized. These attempts have focused on three main areas: that of individual words and their associations, that of semantic or lexical fields, and lastly that of word families. We shall discuss these three approaches in turn. Finally, we shall also mention the notion of 'word class' as a way of accounting for the structure of the English vocabulary.

1.3.1 The word and its associative field

According to this approach, every word is involved in a network of associations which connect it with other terms in the language. Some of these associations are based on similarity of meaning, others are purely formal (i.e. based on forms), while others involve both form and meaning. In de Saussure's graphic formula, a given term is like the centre of a constellation, the point where an infinite number of co-ordinated terms converge. De Saussure tried to represent these associations in the form of a diagram using the French word *enseignement* ('teaching') (de Saussure 1959: 126). In Figure 1.1, we have not only modified the form of the diagram (by giving one word a more central position) but have also used English words to suit our discussion here. In this diagram, four lines of association radiate from the noun *lecturer*: (1) connects it with the verb forms *lectured* and *lecturing* by formal and semantic similarity based on the common stem *lecture*; (2) connects it with *teacher* and *tutor* by semantic similarity; (3) associates it with *gardener* and *labourer* because they all have the suffix *-er* forming agent nouns from verbs; (4) associates it with the adjective *clever* and the inflected adverb *quicker* by accidental similarity in their endings. The use of an arrow and that of *etc.* at the end of each line of associations suggests that the line has no limit and that an infinite number of words can be added to those suggested in the diagram.

Figure 1.1 also shows, among other things, that any word chosen from a given context will suggest other words to us, because they either resemble or differ from each other in form, meaning or both. Such relations are referred to as 'paradigmatic'. They are called relations 'in absentia', because the terms involved consist of a word present in the utterance and others that are not actually in the same utterance but that are substitutable for it in that context. For example, 'difficult' is paradigmatically related with 'easy', 'funny', 'silly', etc. in expressions such as 'an easy question', 'a funny question', 'a silly question', etc. Similarly, 'question' is paradigmatically related with 'problem', 'word', etc. in expressions like 'a difficult problem', 'a difficult word', etc.

Interesting as the notion of paradigmatic relations may be, it has one major difficulty: every other word in the language either resembles or fails
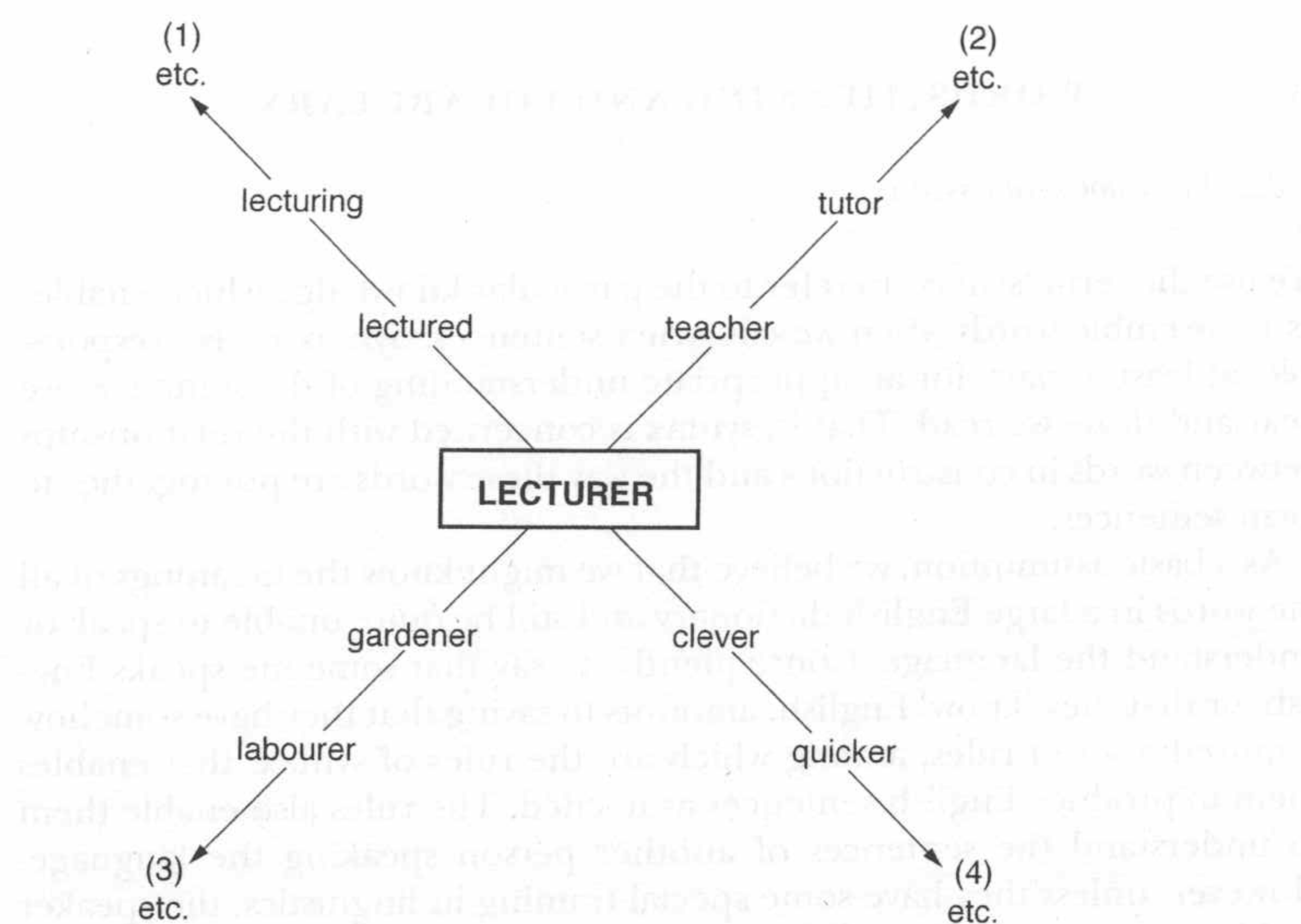
Figure 1.1 The word and its associative field

to resemble in form or meaning any given word. This difficulty notwithstanding, the notion has proved useful in language teaching and in the arrangement of words in a thesaurus. It has also been used to establish cross-references in the definitions of words in dictionaries.

This important notion was first introduced by de Saussure under the label 'associative relations'. But the term 'paradigmatic' has been substituted at the suggestion of the Danish linguist Hjelmslev (1963). Another notion which was also introduced by de Saussure was that of 'syntagmatic' relations. They are called relations 'in praesentia' since the words involved are actually co-occurent items. For example, the word *difficult* is syntagmatically related with the article *a* and the noun *question* in the expression 'a difficult question'.

Paradigmatic and syntagmatic relations may be represented in a diagram as in Figure 1.2. This shows that every word may be considered in terms of two dimensions or axes of structure. The 'horizontal' or syntagmatic and the 'vertical' or paradigmatic. It is precisely in terms of syntagmatic and paradigmatic relations that the meaning of English words can be determined. However, our interest in constructions larger than the word is secondary, though see 5.6 on collocation.

1.3.2 Lexical fields

Some isolated attempts have been made to study the structure of some semantic or lexical fields, such as the hierarchy of military ranks, numerals,

Syntagmatic (sequence)

The	first	question	was	difficult

Paradigmatic (substitution)

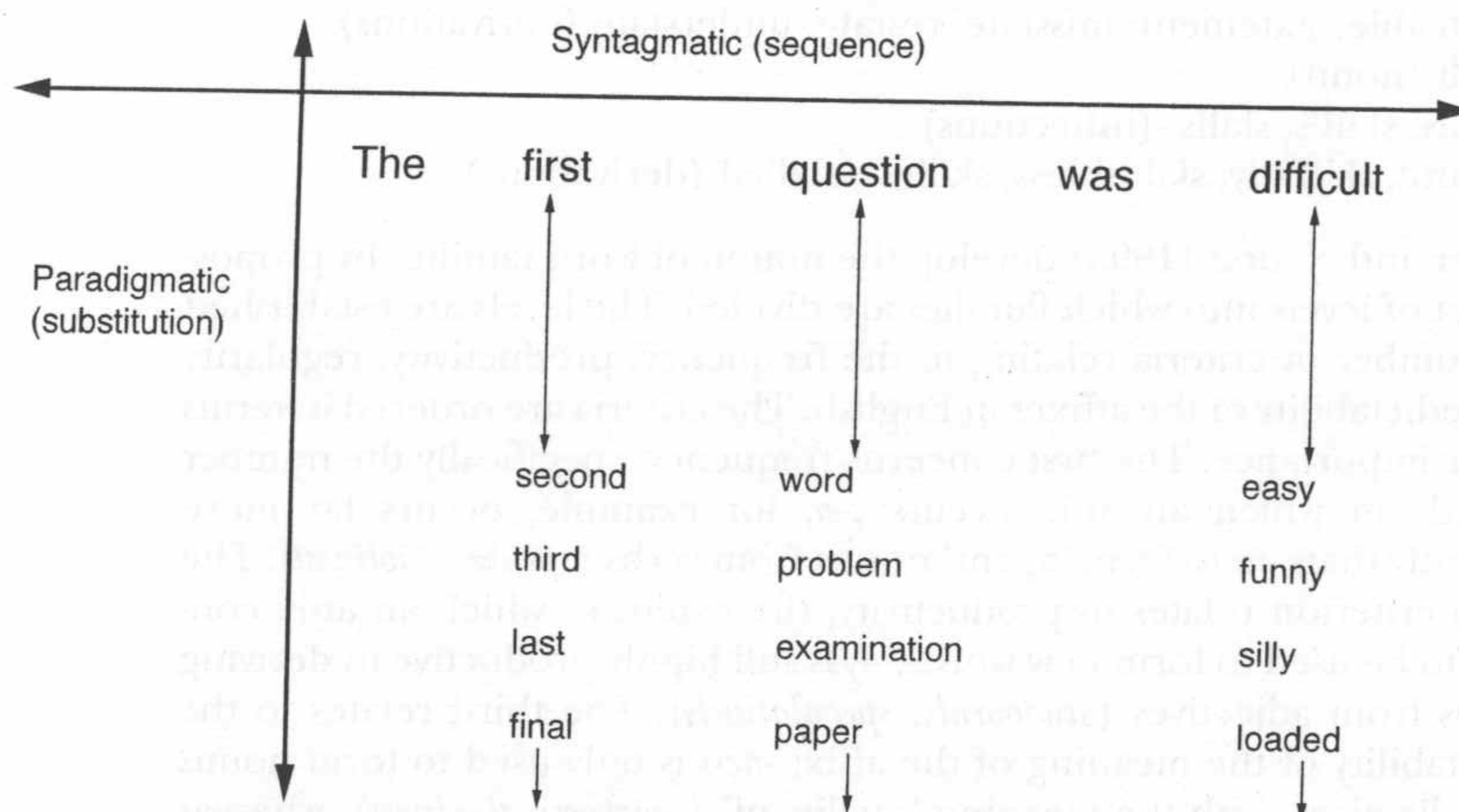| second | word | easy |
third	problem	funny
last	examination	silly
final	paper	loaded

Figure 1.2 Syntagmatic and paradigmatic relations

colour and kinship terms. Most of these attempts are directly or indirectly connected with what has now come to be known as the theory of 'semantic fields' or 'lexical field theory'. 'Semantic field' or 'semantic domain' are used alternatively for the terms 'lexical field' or 'lexical set'. Crystal (1995: 157) defines a semantic or lexical field as a 'named area of meaning in which lexemes interrelate and define each other in specific ways'. For example, the lexical field of 'kinship terms' comprises the lexemes: father, mother, son, daughter, cousin, nephew, uncle, aunt, grand-father, grand-mother, etc.

Field theory was first put forward by a number of German and Swiss scholars in the 1920s and 1930s. However, according to Lyons (1977: 250) its origin can be traced back at least to the middle of the nineteenth century and more generally to the ideas of Humboldt and Herder. We shall not attempt a comprehensive treatment of field theory, nor shall we try to review the considerable body of descriptive work based on it. Instead, we shall discuss briefly how the theory may be used to account for the general organisation of the English vocabulary (see also 5.5.3).

According to lexical field theory, the vocabulary of a language is essentially a dynamic and well-integrated system of lexemes structured by relationships of meaning. The system is changing continuously by the interaction of various forces such as the disappearance of previously existing lexemes, or the broadening or narrowing of the meaning of some lexemes. The system is mainly characterized by the general–particular and part–whole relationships, which hold not only between individual lexemes and the lexical field within which they are best interpreted, but also between specific lexical fields and the vocabulary as a whole. One of the early theorists, Jost Trier, puts it like this: 'Fields are living realities intermediate between individual words

and the totality of the vocabulary; as parts of a whole, they share with words the property of being integrated in a large structure and with the vocabulary the property of being structured in terms of smaller units' (quoted in Lyons 1977: 253). For example, the lexical field of 'colour terms' includes the lexemes: black, white, red, green, yellow, blue, orange, etc. Similarly, the lexical field of colour terms, together with those of kinship terms, military ranks, vehicles, among others are only parts of the whole English vocabulary. Furthermore, the general lexeme *red* for instance may in turn be considered a lexical field (or sub-field) within which the particular lexemes *scarlet*, *crimson*, *vermillion*, etc. may best be interpreted. These characteristics of field theory may be illustrated by Figures 1.3a and 1.3b.

As we would expect, the lexicologist who attempts to assign all the words in English to lexical fields is bound to face a number of difficulties. According to Crystal (1995: 157) these difficulties are of three kinds. First, some lexemes tend to belong to fields that are vague or difficult to define. For instance, it is not obvious to which field the lexemes *noise* or *difficult* should be assigned. Secondly, some may validly be assigned to more than one field, e.g. *orange*, either to the field of 'fruit' or to that of 'colour'; *tomato* as 'fruit' or 'vegetable'. The last difficulty concerns the best way to define a lexical field in relation to the other fields on the one hand, and its constituent lexemes on the other. To use Crystal's examples (1995: 157), one may ask such questions as the following: is it more illuminating to say that *tractor* belongs to the field of 'agricultural vehicles', 'land vehicles', or just 'vehicles'? Should the lexeme *flavour* be a member of the lexical field of 'taste' or is *taste* a lexeme in the lexical field of 'flavour' or should both be treated as lexemes in a broader semantic field such as 'sensation'?

The existence of these difficulties is a pointer to the fact that the English vocabulary is not made up of a number of discrete lexical fields in which each lexeme finds its appropriate place. Furthermore, it provides an additional proof that language usually cannot be analysed into well-defined and watertight categories. The analyst will always have to take a number of difficult decisions in order to accommodate fuzzy cases. However, it must be acknowledged that large numbers of lexemes can in fact be grouped together into fields and sub-fields in a fairly clear-cut manner. This is an indication of the usefulness of field theory in lexicology (see further 5.5.3).

1.3.3 Word families

The treatment of words in terms of 'word families' is a common approach in the lexicology of French (e.g. Lehmann and Martin-Berthet 1997: 109). This section, however, follows a recent treatment of English in such terms (Bauer and Nation 1993). Words are grouped into 'families' on the basis of their morphology, both their inflections and their derivations. A family consists of a base form, its possible inflectional forms, and the words derived from it by prefixation and suffixation (see Chapter 4), e.g.
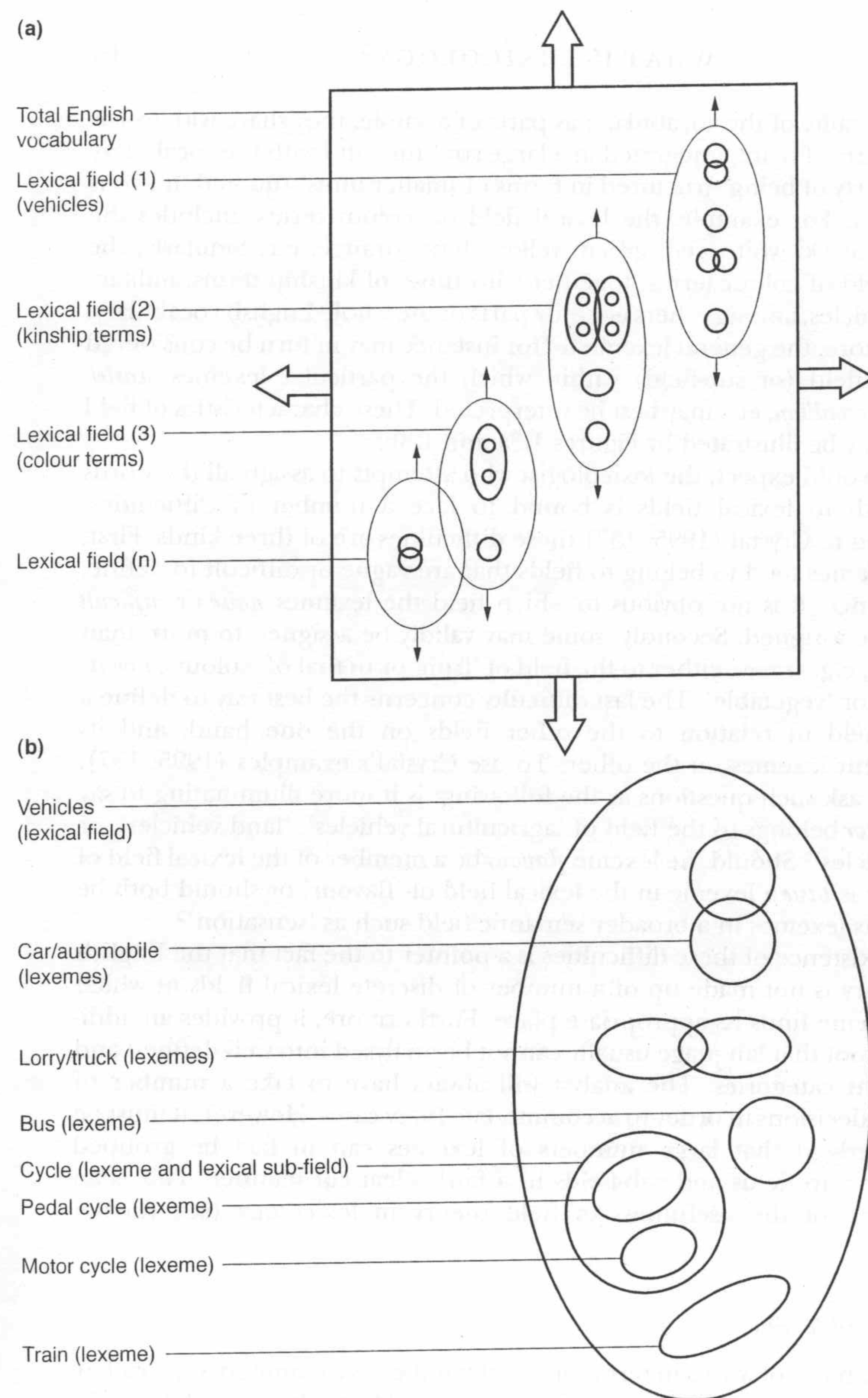
(a)

Total English
vocabulary

Lexical field (1)
(vehicles)

Lexical field (2)
(kinship terms)

Lexical field (3)
(colour terms)

Lexical field (n)

(b)

Vehicles
(lexical field)

Car/automobile
(lexemes)

Lorry/truck (lexemes)

Bus (lexeme)

Cycle (lexeme and lexical sub-field)

Pedal cycle (lexeme)

Motor cycle (lexeme)

Train (lexeme)

Figure 1.3a Lexical fields in the total vocabulary
Figure 1.3b Example of a lexical field

(a) state (verb)
    states, stated, stating (inflections)
    stateable, statement; misstate, restate, understate (derivations)
(b) skill (noun)
    skills, skill's, skills' (inflections)
    skilful, skilfully, skilfulness, skilless, skilled (derivations)

Bauer and Nation (1993) develop the notion of word families by proposing a set of levels into which families are divided. The levels are established on a number of criteria relating to the frequency, productivity, regularity and predictability of the affixes in English. The criteria are ordered in terms of their importance. The first concerns frequency, specifically the number of words in which an affix occurs; -er, for example, occurs far more frequently than -ist to form 'agent' nouns from verbs (speaker, violinist). The second criterion relates to productivity, the extent to which an affix continues to be used to form new words; -ly is still highly productive in deriving adverbs from adjectives (stubbornly, speculatively). The third relates to the predictability of the meaning of the affix; -ness is only used to form nouns from adjectives, with the meaning 'quality of' (craziness, tiredness), whereas -ist has a number of possible meanings. The remaining criteria concern regularity of spelling and pronunciation (of the base and affix) and regularity of the function of an affix in terms of the word class of the base to which it attaches.

Using these criteria, Bauer and Nation (1993) establish seven levels of family relationship. At the first level, each word form is regarded as a different word; so, there is no family. The second level groups words that have a common base but variant inflectional suffixes (plural and possessive for nouns; present and past tense, and present and past participle for verbs; comparative and superlative for adjectives), e.g.

Noun: road, roads, road's, roads'
Verb: fly, flies, flew, flying, flown
Adjective: great, greater, greatest.

At the third level are added words formed by the addition of 'the most frequent and regular derivational affixes', which are established on the basis of an analysis of a computer corpus (see Chapter 7), viz. -able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-. At level four are added forms with 'frequent, orthographically regular affixes': -al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, -ous, in-. At level five come forms derived with some fifty 'regular but infrequent affixes', e.g. -ary, -let, anti-, sub-. The sixth level has forms derived by 'frequent but irregular affixes': -able, -ee, -ic, -ify, -ion, -ist, -ition, -ive, -th, -y, pre-, re-. Lastly, at level seven are included words formed using classical (Latin and Greek) roots and affixes, e.g. bibliography, astronaut and the common prefixes ab-, ad-, com-, de-, dis-, ex-, sub-. As an illustration, we borrow Bauer and Nation's example of the verb develop, which conveniently has examples at all of the levels two to six:

Level 2: develop, develops, developed, developing
Level 3: developable, undevelopable, developer(s), undeveloped
Level 4: development(s), developmental, developmentally
Level 5: developmentwise, semideveloped, antidevelopment
Level 6: redevelop, predevelopment.

The practical benefits of an analysis of vocabulary in terms of word families, especially one such as Bauer and Nation's (1993) with its carefully differentiated levels, can be found in language teaching and in lexicography. For language teachers there is a system of word building available both to pass on to their learners and to guide them in the selection of reading materials. For lexicographers, such an analysis provides a more secure basis for the treatment of affixes and derived words in dictionaries.

EXERCISE 1/4

For each of the following sets of words, say what the principle is that groups them into a set. Is the principle one of a common meaning, related forms of a lexeme, or something else (please specify)?

1. dawdle, saunter, meander, wander, swan, tootle

2. speak, speaks, spoke, speaking, spoken

3. telephone, dial, number, answer, ring, engaged

4. rich, wealthy, well-off, loaded, affluent, well-heeled, oofy

---

1.3.4 Word classes

The notion of word class may also be used to account for the structure of the vocabulary as a whole. Following an approach that can be traced back to Latin and Greek, traditional grammars of English distinguish eight parts of speech: noun, pronoun, adjective, verb, adverb, preposition, conjunction, and interjection. More modern grammarians have elaborated these parts of speech into further classes. For example, Quirk *et al.* (1985: 67) distinguish the following:

(a) closed classes: preposition, pronoun, determiner, conjunction, auxiliary verb;
(b) open classes: noun, adjective, verb, adverb;
(c) lesser categories: numeral, interjection;
(d) a small number of words of unique function: the particle *not* and the infinitive marker *to*.

The closed classes contain the so-called 'grammatical' or 'function' words,

---

which generally serve the grammatical construction of sentences. They are small classes, with a restricted and largely unchanging membership. The open classes, by contrast, are large, and they are constantly being added to. The members of the open classes are the 'content' words, carrying the main meaning of a sentence; they are the words likely to be retained in a telegram or a headline.

It is assumed that without exception all English words must belong to one or more word classes. However, it is not generally possible to tell which word class a word belongs to, simply by looking at it, though inflections may provide a clue (e.g. a word ending in *-ing* is likely to be a verb, and one ending in *-est* an adjective). But to be sure, we need to study a word's behaviour in sentences. All words that function in the same way are deemed to belong to the same word class. For example, consider the following sentence:

on
The book was *in* the cupboard
under

The words *on*, *in*, and *under* have the same function and express some kind of locational relationship between *the book* and *the cupboard*. Since they behave the same way in the sentence, they belong to the same word class, which we call 'preposition'. The notion of word class is therefore useful because it allows us to make general and economical statements about the way the words of a language behave. However, this notion is more useful in syntax (1.2.2) than in lexicology.

Traditional lexicology has been concerned mostly with open classes, but more recent descriptions, such as Jackson (1988) and Lipka (1990), attempt to discuss all classes of words. A close examination of the major word classes shows that they all have central and peripheral members and that they overlap. The boundaries between classes are therefore fuzzy. Moreover, a word may belong to more than one word class, e.g. *round* is an adjective in 'a round stone', a preposition in 'round the corner', an adverb in 'they all gathered round', a noun in 'you can buy the next round', and a verb in 'when we round the next bend'.

EXERCISE 1/5

(a) Closed word classes have a restricted membership: list all the personal pronouns in modern English.

(b) Attempt a definition of 'adjective'. Think of the kinds of general meaning adjectives have, their possible forms (inflections), and where they typically occur in sentence structure.

1.4 Summary

This chapter has given us at least a partial answer to the question 'What is lexicology?'. In so doing, it has established the distinction between lexicology as a level of language analysis and related fields such as phonology, morphology, syntax, semantics, etymology, and lexicography. It has also examined the various attempts made to account for the structure of the English vocabulary as a whole. Among other things, the chapter has revealed the central position of the notion of 'word' in lexicology, which we take up in Chapter 3.

2   Where Do English Words Come From?

This question calls for an investigation of the origin of English words. As useful preliminaries, we shall first place English in the context of world languages and discuss the historical development of English vocabulary from the Old English to the Modern English periods. We shall then examine in turn, native English vocabulary, and foreign elements in the English lexis. Finally, we shall discuss the creation of new words, before outlining the most important characteristics of modern English vocabulary.

2.1   The origin of English

The 5000 or so languages of the world can be grouped into about 300 language families, on the basis of similarities in their basic word stock and grammars. One of these families, the Indo-European, includes most of the languages of Europe, the Near East, and North India. One branch of the Indo-European family is called Italic, from which Latin, and later the Romance languages developed. Another is called Germanic, which had three branches: North Germanic, East Germanic, and West Germanic. It is not certain, however, whether North, East and West Germanic corresponded to actual languages.

The North Germanic branch is the linguistic ancestor of modern Scandinavian languages, viz. Danish, Swedish, Norwegian, and Icelandic. The East Germanic branch developed into Gothic, but it eventually died out. The West Germanic branch in turn developed into modern German, Dutch, Frisian, and English. Frisian, spoken in the north-west Netherlands and the island nearby known as Friesland, is considered the closest relative of English. Both languages derive from Anglo-Frisian, just as High and Low German derive from German.

2.1.1 How English came to England

The first people known to have inhabited the land that was later to become England were Celts. They are presumed to have come to the island around

the middle of the fifth millennium BC. Their languages were yet another branch of the Indo-European language family.

The Celtic warriors withstood the initial attempt to add their land to the Roman Empire in 55–54 BC, but a century later they were overwhelmed by a much larger Roman army. Most of the island of Britain was occupied by the Roman legions, government officials, and their households, from about AD 43 until 410. When the Empire began to crumble, the military and government officials withdrew, leaving behind many settlements built around the installations of military government. These bore names such as *Doncaster, Gloucester, Lancaster,* and *Worcester* – all derived in part from the Latin word *castra* 'camp'.

The withdrawal of the Roman troops virtually invited the invasion of the rich lowlands by the Picts and Scots, two tribes in the north of Britain that the Romans had never conquered. The Celts appealed to bands of Germanic warriors from across the North Sea for aid in defending their land. Soon the Germanic tribes – called Angles, Saxons, Frisians, and Jutes – came in ever increasing numbers. The Celts found out too late that their new allies had become their conquerors. Although wars of resistance continued for the next 200 years, some of the Celts were pushed steadily to the fringes of the country, to Wales, Cornwall, Cumbria and the Scottish highlands. Others crossed the English Channel to join their relatives in French Brittany. Those who remained were subjected to the government of the newcomers and became assimilated to them by intermarriage. After a few generations, their identity was lost within Anglo-Saxon society.

2.1.2 *The name of the language*

The Germanic invaders referred to the native Celts as *Wealas* ('foreigners'), from which the name 'Welsh' is derived. The Celts called the invaders *Sassenachs* 'Saxons', regardless of their specific tribes, a practice which was followed by the early Latin writers. By the end of the sixth century, however, the term *Angli* 'Angles' was in use. For example, in 601, a king of Kent, Aethelbert, is called *rex Anglorum* (King of the Angles). During the seventh century, the usual Latin name for the country was *Angli* or *Anglia*. This became *Engle* in Old English, while the name of the language was referred to as *Englisc* (the *sc* spelling representing the sound 'sh'). The word *Englaland* (land of the Angles), which later gave *England*, did not appear until the beginning of the tenth century.

EXERCISE 2/1

Write a definition of the word 'English'. Then compare your version with that given in the Key to Exercises at the end of the book.

2.2 Historical development of English vocabulary

We shall discuss in turn the Old, Middle, Early Modern, and Modern English periods, with special emphasis on their respective characteristics at the lexical level. We shall also highlight the major contributions of each period to the development of English lexis as a whole.

2.2.1 *The Old English period (450–1066)*

The first Old English (OE) manuscripts were simply a few scattered inscriptions written around the fifth and sixth centuries in the runic alphabet brought in by the Anglo-Saxons. These scattered inscriptions give very little information on the language. The literary age began only after the arrival of the Christian missionaries from Rome in AD 597. The first OE manuscripts, dating from around 700, are glossaries of Latin words translated into Old English, and a few early inscriptions and poems. Unfortunately, very little material remains from this period. The most important literary work, which survived in a single copy, was the heroic poem *Beowulf*, written around 1000. But there were also a number of shorter poems concerned with Christian subjects or reflecting Germanic traditions and dealing with topics such as war, patriotism, travelling and celebration.

It is generally acknowledged that most OE texts were written in the period following the reign of King Alfred (849–899), who arranged the many Latin works to be translated, including Bede's *Ecclesiastical History* (Bede 731). But the total corpus is still considered relatively small. Crystal (1995: 10) points out that 'the number of words in the corpus of Old English compiled at the University of Toronto, which contains all the texts, is only 3.5 million – the equivalent of about 30 medium-sized novels'.

Since we are dealing with written records, it is important to say a word on OE letters and spelling. The OE alphabet was very similar to the one still in use today, but the absence of capital letters was a distinctive feature. Furthermore, a few letters had different shapes, while some modern letters were absent, e.g. *j, v, f, q, x* and *z*. Numbers were written only in Roman symbols. Several of the letters were used in combination (digraphs) to represent single sound units, just like modern English *th* and *ea*, as in 'tru*th*' and 'm*ea*t'. There was a great deal of variation in spelling, to such an extent that even with a single scribe in a single place at a single time, there could be variation. The same word could be spelt differently on the same page.

In most general terms, if we allow for the unfamiliar spelling and the unexpected inflections, we still notice a marked difference between the words used in prose and those used in poetic texts. While the majority of words in prose are very close to Modern English, words in poetic texts are different. For example, most of the prepositions and pronouns are identical in form (though not always in meaning): *for, from, in, he, him, his*. Modern English speakers can easily recognise *singan* as 'sing' or *stod* as 'stood'; *onslepte* is quite close to 'asleep', and *geleornode* to 'learned'. Omitting the *ge-*