

目 录

第1章 英语词汇学简论	1
1. Why Study Words	1
2. The Lexicon	6
第2章 英语词汇学研究的新进展	16
1. Lexicology and Translation Studies	16
2. Recent Development of Lexicology	24
第3章 英语词汇的构词研究	29
1. Compound Words, Blends and Phrasal Words	29
2. Morphology and Mind	38
第4章 英语词汇的聚合关系与组合关系	43
1. Lexical Relations	43
2. Words in Combination	51
第5章 英语词汇的历史研究	66
1. A History of the English Lexicon	66
2. The Early Modern English Lexicon	77
第6章 英语新词研究	83
1. The FUDGE Factors	83
2. How New Words Are Created	93
第7章 英语词汇的国别研究	102
1. Vocabulary of New Englishes	102
2. Models of English	105

第 8 章 英语词汇的社会研究	114
1. Gender, Language and Attitudes	114
2. Sociolinguistic Stimuli to Development of the English Lexicon ...	118
第 9 章 英语词汇的文化研究	126
1. Key Words and Culture	126
2. Applied Cultural Linguistics	134
第 10 章 英语词汇的语用研究	141
1. Five Theses of Lexical Pragmatics	141
2. Towards the New Linguistic Discipline of Lexical Pragmatics ...	145
第 11 章 语篇分析与英语词汇研究	154
1. Lexical Patterning in Discourse	154
2. Discourse and Semantics	162
第 12 章 语料库语言学与英语词汇研究	181
1. Corpora in Lexical Studies	181
2. Future Prospects in Corpus Linguistics	184
第 13 章 认知语言学与英语词汇研究	189
1. Introducing Cognitive Linguistics	189
2. Cognitive Semantics	201
第 14 章 词典与英语词汇研究	215
1. The Dictionary	215
2. Meaning and Definition	226
第 15 章 英语词汇教学研究	236
1. Teaching Vocabulary	236
2. Planned Vocabulary Teaching	247
第 16 章 英汉词汇对比研究	255
1. Methodology of Contrastive Functional Analysis	255
2. Semantic Primes within and across Languages	264
Bibliography	270

英语词汇学简论

1.

Why Study Words

From Katamba, Francis. 2005. *English Words* (2nd ed.). London: Routledge. pp. 1-4.

1. Introduction

1. 1 Why study words

Imagine a life without words! Trappist monks opt for it. But most of us would not give up words for anything. Every day we utter thousands and thousands of words. Communicating our joys, fears, opinions, fantasies, wishes, requests, demands, feelings — and the occasional threat or insult — is a very important aspect of being human. The air is always thick with our verbal emissions. There are so many things we want to tell the world. Some of them are important, some of them are not. But we talk anyway — even when we know that what we are saying is totally unimportant. We love chitchat and find silent encounters awkward, or even oppressive. A life without words would be a horrendous privation.

It is a cliché to say that words and language are probably humankind's most valuable single possession. It is language that sets us apart from our biologically close relatives, the great primates. (I would imagine that many a chimp or gorilla would give an arm and a leg for a few words — but we will probably never know because they cannot tell us.) Yet, surprisingly, most of us take words (and more generally language) for granted. We cannot discuss words with anything like the competence with which we can discuss fashion, films or football.

We should not take words for granted. They are too important. This book is intended to make explicit some of the things that we know subconsciously about words. It is a linguistic introduction to the nature and structure of English words. It addresses the question "what sorts of things do people need to know about English words in order to use them in speech?" It is intended to increase the degree of sophistication with which you think about words. It is designed to give you a theoretical grasp of English word-formation, the sources of English vocabulary and the way in which we store and retrieve words from the mind.

I hope a desirable side effect of working through *English Words* will be the enrichment of your vocabulary. This book will help to increase, in a very practical way, your awareness of the relationship between words. You will be equipped with the tools you need to work out the meanings of unfamiliar words and to see in a new light the underlying structural patterns in many familiar words which you have not previously stopped to think about analytically.

For the student of language, words are a very rewarding object of study. An understanding of the nature of words provides us with a key that opens the door to an understanding of important aspects of the nature of language in general. Words give us a panoramic view of the entire field of linguistics because they impinge on every aspect of language structure. This book stresses the ramifications of the fact that words are complex and multi-faceted entities whose structure and use interacts with the other modules of the grammar such as PHONOLOGY, the study of how sounds are used to represent words in speech, SYNTAX, the study of sentence structure, and SEMANTICS, the study of meaning in language.

In order to use even a very simple word, such as *frog*, we need to access various types of information from the word-store which we all carry around with us in the MENTAL LEXICON or DICTIONARY that is tucked away in the mind. We need to know:

[1.1]

- (i) its shape, i.e. its PHONOLOGICAL REPRESENTATION /frɒg/ which enables us to pronounce it, and its ORTHOGRAPHIC REPRESENTATION *frog*, if we are literate and know how to spell it (see the Key to symbols used on page xix);
- (ii) its grammatical properties, e.g. it is a noun and it is countable — so you can have one *frog* and two *frogs*;
- (iii) its meaning.

But words tend not to wear their meaning on their sleeve. Normally, there is nothing about the form of words that would enable anyone to work out their meaning. For, typically, the relationship between a LINGUISTIC SIGN like the word *frog* and its meaning is ARBITRARY. Other languages

use different words to refer to this small tailless amphibian. In French it is called (*la*) *grenouille*. In Malay they call it *katak* and in Swahili *chura*. None of these words is more suited than the others to the job of referring to this small reptile.

And of course, within a particular language, any particular pronunciation can be associated with any meaning. So long as speakers accept that sound-meaning association, they have a kosher word. For instance, *convenience* originally meant “suitability” or “commodiousness”, but in the middle of the nineteenth century a new meaning of “toilet” was assigned to it and people began to talk of “a public convenience”. In the early 1960s the word acquired the additional new meaning of “easy to use, designed for hassle-free use” as in *convenience food*. ...

Let us now consider one further example. All competent speakers of English know that you can add *-s* to a noun to indicate that it refers to more than one entity. So, you say *cat* when referring to one and *cats* if there is more than one. If you encountered in the blank in [1.2a] an unfamiliar word like *splet* (which I have just made up), you would automatically know from the context that it must have the plural form *splets* in this position since it is specified as plural by *all*. Further, you would know that the plural of *splet* must be *splets* (rather than *spletren* by analogy to *children* or *spleti* by analogy to *stimuli*). You know that the majority of nouns form their plural by adding the regular plural suffix or ending *-s*. You always add *-s* unless express instructions are given to do otherwise. There is no need to memorise separately the plural form of most nouns. All we need is to know the rule that says “add *-s* for plural”. So, without any hesitation, you suffix *-s* to obtain the plural form *splets* in [1.2b]:

[1.2]

- a. We put all the big _____ on the table.
- b. We put all the big *splets* on the table.

The study of word-formation and word-structure is called MORPHOLOGY. Morphological theory provides a general theory of word-structure in all the languages of the world. Its task is to characterise the kinds of things that speakers need to know about the structure of the words of their language in order to be able to use them to produce and to understand speech.

We will see that in order to use language, speakers need to have two types of morphological knowledge. First, they need to be able to analyse existing words (e.g. they must be able to tell that *frogs* contains *frog* plus *-s* for plural). Usually, if we know the meanings of the elements that a word contains, it is possible to determine the meaning of the entire word once we have worked out how the various elements relate to each other. For instance, if we examine a word like *nutcracker* we find that it is made up of two words, namely the noun *nut* and the noun *cracker*. Furthermore, we see

that the latter word *cracker* is divisible into the verb *crack* and another meaningful element *-er* (roughly meaning “an instrument used to do X”), which, however, is not a word in its own right. Numerous other words are formed using this pattern of combining words (and smaller meaningful elements) as seen in [1.3]:

[1.3]

[[tea]_{Noun} - [strain-er]_{Noun}]_{Noun}
 [[lawn]_{Noun} - [mow-er]_{Noun}]_{Noun}
 [[can]_{Noun} - [open-er]_{Noun}]_{Noun}

Given the frame [[_____]_{Noun} - [_____ er]_{Noun}]_{Noun}, we can fill in different words with the appropriate properties and get another compound word (i.e. a word containing at least two words). Try this frame out yourself. Find two more similar examples of compound words formed using this pattern.

Second, speakers need to be able to work out the meanings of novel words constructed using the word-building elements and standard word-construction rules of the language. Probably we all know and use more words than are listed in dictionaries. We can construct and analyse the structure and meaning of old words as well as new ones. So, although many words must be listed in the dictionary and memorised, listing every word in the dictionary is not necessary. If a word is formed following general principles, it may be more efficient to reconstitute it from its constituent elements as the need arises rather than permanently commit it to memory. When people make up new words using existing words and word-forming elements, we understand them with ease — providing we know what the elements they use to form those words mean and providing the word-forming rules that they employ are familiar. This ability is one of the things explored in morphological investigations.

In an average week, we are likely to encounter a couple of unfamiliar words. We might reach for a dictionary and look them up. Some of them may be listed but others might be too new or too ephemeral to have found their way into any dictionary. In such an event, we rely on our morphological knowledge to tease out their meanings. If you heard someone describe their partner as “a great list maker and a ticker-off”, you would instantly know what sort of person the partner was — although you almost certainly have never encountered the word *ticker-off* before. And it is certainly not listed in any dictionary. The *-er* ending here has the meaning of “someone who does whatever the verb means”. Given the verb *tickoff*, a *ticker-off* must be a person who *ticks off*. Similarly, if you know what established words like *handful*, *cupful* and *spoonful* mean, you are also able to figure out the meanings of novel words like *fountain-penful* (as in *a fountain-penful of ink*) or *hovercraftful* (as in *hovercraftful after hovercraftful of English*).

shoppers returned from Calais loaded down with cigarettes, cheese and plonk). Virtually any noun denoting a container can have *-ful* added to it in order to indicate that it is “full of something”.

To take another example, a number of words ending in *-ist*, many of which have come into use in recent years, refer to people who discriminate against, or hold negative views about, certain less powerful subgroups in society, e.g. *racist*, *sexist*. Anyone who knows what *racist* and *sexist* mean, given the right context should have no difficulty in understanding the nature of discrimination perpetrated by people who are described using the novel words *ageist*, *sizist* and *speechist*. *Ageism* is discrimination on grounds of (old) age — for instance, denying employment to people over the age of 60; *sizism* is discrimination (usually against fat people) on grounds of size and *speechism* is discrimination against people with speech impediments like stuttering.

Did you notice how I exploited your tacit knowledge of the fact that words ending in *-ist* and *-ism* complement each other? You were glad to accept *ageism*, *sizism* and *speechism* because you know that corresponding to an adjective ending in *-ist* there will normally be a noun ending in *-ism*. This is important. It shows that you know that certain word-forming bits go together — and others do not. I suspect that you would reject putative words like **agement*, **sizement* and **speechment*. (An asterisk is used conventionally to indicate that a form is disallowed.) In word-formation it is not a case of anything goes.

A challenging question which morphology addresses is, “how do speakers know which non-occurring or non-established words are permissible and which ones are not?” Why are the words *fountainpenful*, *hovercraftful* and *speechist* allowed while **agement*, **speechment* and **sizement* are not?

Morphological theory provides a general theory of word-formation applicable to any language but, as mentioned earlier, this book focuses on word-formation in English. Its objective is to provide a description of English words designed to make explicit the various things speakers know, albeit in an unconscious manner, about English words. The emphasis will be on the description of English words rather than the elaboration of morphological theory. So, data and facts about English words are brought to the fore and the theoretical and methodological issues are kept in the background for the most part. The use of formal notation has also been kept to a minimum in order to keep the account simple.

2.

The Lexicon

From Singleton, D. M. 2000. *Language and the Lexicon: An Introduction*. London: Arnold. pp. 1–10.

1. Introduction: The lexicon — words and more

1. 1 Some preliminary definitions

This book is about the lexicon. *Lexicon* is the Anglicized version of a Greek word (*Λεξικόν*), which basically means “dictionary”, and it is the term used by linguists to refer to those aspects of a language which relate to words, otherwise known as its *lexical* aspects. *Lexicon* is based on the term *lexis*, whose Greek meaning is “word”, but which is used as a collective expression in linguistic terminology in the sense of “vocabulary”. The study of *lexis* and the lexicon is called *lexicology*.

In fact, as we shall see in the course of the next 200 pages or so, almost everything in language is related in some way or other to words. We shall also see that, conversely, the lexical dimension of language needs to be conceived of as rather more than just a list of lexical items.

1. 2 Words and language

“In the beginning was the Word ...”

This opening pronouncement of the Gospel of John in the New Testament may or may not be a true claim about the origins of the cosmos. However, if taken as a statement about where our thinking about language started from (and continues to start from) it is hard to fault.

The original version of John’s Gospel was written in Greek, and in this version the term used for “word” is *logos* (λόγος), which, significantly enough, meant (and in Modern Greek still means) “speech” as well as “word”. This kind of association between the concept of word and a more general concept of speech or language is by no means confined to Greek culture. For example, to stay with the Gospel of John for just a little longer, the Latin translation of the above quotation is: “*In principio erat Verbum ...*”, where λόγος is replaced by *verbum*, an expression which, like the Greek term, was applicable to speech as well as to individual words. Thus, for example, one way of saying “to speak in public” in Latin was *verbum in publico facere* (literally, “word in public to make”).

A similar association between “word” and “speech” is to be found in many other languages. For example, this dual meaning attaches to French

parole, Italian *parola* and Spanish *palabra*. Similarly, in Japanese the term *kotoba* (“word”, “phrase”, “expression”) is often abbreviated to *koto* or *goto* and is used as a suffix in expressions referring to speech such as *hitorigoto o iu* (literally, “by-oneself-word say” = “talk to oneself”) and *negoto o iu* (literally, “sleep-word say” = “talk in one’s sleep”); in Swedish the expression *en ordets man* (literally, “an of-the-word man”) is used to refer to a skilled speaker; and in German one way of saying “to refuse someone permission to speak” translates literally as “from someone the word to remove” — *einem das Wort entziehen*. In English, too, the association between the word and language in use is very much a feature of the way in which linguistic events are talked about in ordinary parlance, as the following examples illustrate:

That traffic warden wants a word with you.

A word in the right ear works wonders.

When you are free for lunch just say the word.

The Prime Minister's words have been misinterpreted by the media.

The wording needs to be revised.

Nor is it particularly surprising that words should loom so large in people’s understanding of what language is. After all, words are vital to linguistic communication, and without them not much can be conveyed. For instance, a visitor to a Spanish-speaking country anxious to discover where the toilets are in some location or other may have a perfect command of Spanish pronunciation and sentence-structure, but will make little progress without the word *servicios* (in Spain) or *sanitarios* (in Latin America).

It needs perhaps to be added that awareness of words is not limited to literate societies. The American linguist Edward Sapir, for example, conducted a great deal of fieldwork among Native Americans in the early part of the twentieth century. His goal was the transcription and analysis of Native American languages which had not previously been described. He found that although the Native Americans he was working with were illiterate and thus unaccustomed to the concept of the written word, they nevertheless had no serious difficulty in dictating a text to him word by word, and that they were also quite capable of isolating individual words and repeating them as units.

Interestingly, a child acquiring language appears to develop an awareness of words earlier than an awareness of how sentences are formed. For example, research has shown that children in the age group 2–3½ correct themselves when they make errors with words before they start self-correcting in the area of sentence construction. Thus, examples like the first one given below will begin to appear earlier than examples like the second one cited.

you pick up ... you take her (substitution of *take* for initial word-choice *pick up*)

The kitty cat is ... the ... the spider is kissing the kitty cat's back (reordering of elements in order to avoid the passive construction *The kitty-cat's back is being*

kissed by the spider)

With regard to the specialist study of language, this too has been highly word-centred. For instance, in *phonology*, under which heading fall both the sound-structure of languages and the study of such sound-structure, a major focus of attention is the identification of sound distinctions which are significant in a particular language. Anyone with any knowledge of English, for example, is aware that in that language the broad distinction between the “t-sound” and the “p-sound” is important, whereas no such importance attaches to the distinction between an *aspirated t* (i.e. a t-sound pronounced with a fair amount of air being expelled) and an *unaspirated t* sound (i.e. a t-sound pronounced without such a voluminous expulsion of air). This last distinction is, in English, determined simply by the particular environment in which the t-sound occurs; thus, *aspirated t* occurs at the beginnings of words like *ten*, *tight* and *toe*, whereas *unaspirated t* occurs after the s-sound in words like *steer*, *sting* and *stool*. Phonologists talk about environmentally conditioned varieties of the t-sound in a given language as belonging to or being realizations of the /t/ *phoneme*, and label them as *allophones* of the phoneme in question. (Notice that the convention in linguistics is for phonemes to be placed between slashes /t/ -, whereas allophones are placed between square brackets — the transcription of the aspirated allophone of /t/, for example, being [t^h]).

To return to the role of words in all this, one of the crucial tests for phonemic distinctions is that of lexical differentiation — that is, the test of whether a particular sound distinction differentiates between words. This can be tested by use of minimal pairs — pairs of words which differ in respect of just one sound (*pin/tin*, *top/tot*, *gape/gate* etc.). Distinctions between sound segments which serve to differentiate between words in this way — such as the difference between the English p-sound and the English f-sound — are called *phonemic* distinctions, whereas distinctions between sound segments which do not differentiate between words — such as degrees of aspiration of English consonants are described as *non-phonemic*. It should be noted, incidentally, that in other languages (such as Sanskrit and its modern descendants) the distinction between aspirated and unaspirated consonants, which in English is merely allophonic, is as important in differentiating between words as the distinction between /p/ and /t/ in English.

There are other ways of studying the sounds used in human languages — ways which do not need to refer to phonemes and hence have no particular connection with lexical issues. For example, it is perfectly possible, without getting involved in questions of word differentiation and without any regard to the semantic implications of using one sound rather than another in a particular language, to study the acoustic properties of human speech (in terms of the physics of sound) or the physiological aspects of speech production (the interplay of the lips, the tongue, the vocal cords etc.).

These kinds of phenomena and their investigation go under the heading of *phonetics*. The Greek root *phoné* (“sound”, “voice”) is shared by both *phonetics* and *phonology*, but whereas phonology deals with the sound systems of individual languages (and any universal organizational principles which may emerge from such investigations) — and in doing so uses lexical differentiation as an important reference point — phonetics is concerned with speech sounds without reference to linguistic system and meaning. Thus it can be said that what differentiates phonology from phonetics is an interest in lexical differentiation in the above sense.

At the grammatical level too, the distinction between two major areas of interest essentially revolves around words — although in a somewhat different manner. Grammar has traditionally been seen as having two branches — *syntax* and *morphology* — and in both cases the very definition of terms is lexically based. Thus, the term *syntax*, a derivative of the Greek word *syntaxis* (“putting together in order”), denotes the whole range of regularities which can be observed in the combination of sentence components (and the study of such regularities), and it turns out that these components are largely identifiable as words and groups of words. For example, the distinction between the syntax of statements in English (e.g. *John can swim.*) and the syntax of questions (e.g. *Can John swim?*) is, at least from one perspective, a distinction between different ways of ordering words. The term *morphology*, for its part, owes its origins to the Greek root *morphé* (“form”, “shape”) and denotes the internal structure of words (and the study thereof) — that is, how words are built up out of basic units (known as *morphemes*) which may or may not be capable of standing alone as words in their own right (e.g. *un-just-ly*, *de-nation-al-ize*, *re-en-act-ment* etc.).

A third area right at the heart of linguists’ interests, namely *semantics* — that is, the domain of meaning (and its investigation) — is also very much bound up with words. Although the coverage of the term *semantics* (from the Greek *sēma* — “sign”, “signal”) extends well beyond the limits of the lexicon, and semanticists certainly do not confine their attention to the meanings of individual words, the lexical level of meaning has always been the starting point for semantic study and theorizing, and remains a focus for debate. Thus, for instance, there is continuing discussion over whether the meaning of a word like *man* should be seen as an aggregate of the relations between *man* and words such as *animal*, *woman*, *child* etc., whether it should be treated as decomposable into smaller atoms of meaning (human, male, adult etc.), whether it should be envisaged as some kind of idealized or stereotypical mental image against which actual instances of men are compared, or whether all three approaches should be integrated in some way.

1.3 What's in a word?

Although, as is clear from the above, the word is central to the way in which non-specialists and specialists alike think about language, defining what a word is poses a problem or two. To begin with, what we mean by *word* will depend very much on whether we are talking about actual occurrences of any items that might qualify or whether we are intent on grouping or classifying items in some way or other.

To illustrate this, let us begin by looking at the chorus from the Beatles' song *She Loves You*:

She loves you, yeah, yeah, yeah;
She loves you, yeah, yeah, yeah;
She loves you, yeah, yeah, yeah, yeah.

How many words are there in these three lines? If we take actual occurrences of any items — word *tokens* — as the basis of our count, we shall come up with 6 words in the first line, 6 in the second, and 7 in the third. That is 19 overall. On the other hand, if we base our count on word *types* — items with different identities — the overall figure for the entire extract will be just 4 (*she, loves, you* and *yeah*). Similarly, the phrase *going, going, gone* will be considered a three-word expression on a count of tokens but will be considered to contain only two words (*going* and *gone*) on a count of types.

In another sense of *word*, the sequence *going, going, gone* may be thought of as containing just one word — the verb *go*, represented by two of its forms (*going* and *gone*). This approach to the notion of the word — seeing it as a “family” of related forms or as an abstract unit which is realized by one or other of these forms as the linguistic environment demands — calls to mind the concept of the phoneme and its allophones (see above). This linkage with the phoneme idea is expressed terminologically: the notion of the word as a family of forms or as an abstract unit is captured in the term *lexeme*, while a lexeme's concrete representatives or realizations are referred to as *word-forms*. When we want to refer to a given lexeme in, for instance, a dictionary-entry, we typically do so using just one of its various forms, and the choice of this form, known as the *citation form*, is determined by convention, which varies from culture to culture and language to language. ...

We can also see words in different perspectives according to the particular level of linguistic classification we are applying. For example, if we look at the English word *thinks* from the point of view of the English orthographic (spelling) system we shall see it as a series of letters — *t + h + i + n + k + s*; if we consider it as a phonological entity we shall perceive it as a sequence of phonemes — /θ/ + /i/ + /ŋ/ + /k/ + /s/ — one of which, /θ/, corresponds to the letters *th* in the English writing system; if we view *thinks* in grammatical terms, we shall focus on the fact that we have before us

the third-person singular present form of a verb; and if we approach it as a carrier of meaning, we shall be led to relate it to (among other things) the synonyms which can replace it in different contexts, for example:

I think/believe I can do it.

The philosopher's task is to think/cogitate.

I'll think about/consider your suggestion.

Mention of meaning brings us to the distinction which has been drawn between what are termed *content words* (also called *full words* or *lexical words*) and *form words* (otherwise known as *grammatical words*, *empty words* or *function words*). Words described as *content words* are those which are considered to have substantial meaning even out of context, whereas words described as *form words* are those considered to have little or no independent meaning and to have a largely grammatical role. Some examples of content words are: *bucket, cheese, president*; some examples of form words are: *a, it, of*. This distinction is not unproblematic, since many so-called form words — such as prepositions like *around* and *towards* and conjunctions like *although* and *whereas* — are clearly far from empty of semantic content. In any case, we need to be careful with the idea of “semantic content”. We have to keep in mind that it is a metaphor, and that people not words are the sources of meanings, even if words are used as instruments to signal such meanings. Actually, a more satisfactory way of distinguishing between content words and form words is in terms of set membership: grammatical words belong to classes with more or less fixed membership (at least during any individual speaker's lifetime), while content words belong to open classes whose membership is subject to quite rapid change, as new terms come into being and others fall into disuse.

In the light of all that has been said so far in this section, it is hardly surprising that linguists' attempts to provide a general characterization of the word have made reference to quite a wide variety of possible defining properties. The main lines of these different approaches are set out below.

The orthographic approach

In the orthographic approach the word is defined as a sequence of letters bounded on either side by a blank space. This definition works up to a point for languages using writing systems such as the Roman or Cyrillic alphabet, but is not at all useful in relation to languages (like Chinese and Japanese) whose writing-systems do not consistently mark word-boundaries or in relation to language varieties which do not usually appear in written form (e.g. local varieties of Colloquial Arabic) or which have never been written down (e.g. many of the indigenous languages of the Americas). Also, there seems to be something rather odd about defining words in terms of the written medium given that, as we have seen, the word is in no sense a *product* of literacy, and given that, both in the history of human language and

in the development of the individual, written language arrives on the scene well after spoken language. We can note further that defining words in terms of letter-sequences and spaces is very much a form-oriented, token-oriented exercise which takes absolutely no account of more abstract conceptions of the word.

The phonetic approach

Another possible way of trying to define the word is to look for some way in which words might be identifiable in terms of the way they sound — irrespective of the particular sound-systems of specific languages. It might perhaps be imagined, for example, that words are separated from each other in speech by pauses. Alas, life is not that simple! In fact, individual words can rarely be pinpointed in physical terms in the ordinary flow of speech, which is in the main a continuous burst of noise. (Anyone who needs to be convinced of this should tune to a radio station broadcasting in a totally unfamiliar language.) Indeed, the lack of phonetic independence of individual words is precisely what explains linguistic changes such as the loss from some words in English of an initial /n/ (because this was felt to belong to the preceding indefinite article, e.g. *auger* from Old English *nafugar*; *apron* from Old French *naperon*) and the addition of a “stolen” /n/ in some other cases (e.g. *a newt* from *an ewt*; *a nickname* from *an eke-name*). It is, of course, true that pausing is possible between words, and that linguists in the field working on hitherto undescribed languages may sometimes be able to make use of the “potential pause” criterion when gathering data from native speakers — as Sapir did (see above) — but, since speakers do not normally pause between words this criterion has rather limited value.

The phonological approach

At first glance a more promising approach to defining words on the basis of sound is to think in terms of the characteristics of words in particular sound-systems. For example, in some languages — English being a case in point — words tend to have only one stressed syllable, which may occur in various positions (e.g. *renew*, *renewable*, *renewability* etc.). Another instance of a word-related phonological feature is that of vowel-harmony in languages such as Estonian, Finnish, Hungarian and Turkish. In this case the nature of the vowel in the first part of a word determines the choice of vowels in what follows. This is illustrated by the following two Hungarian words: *kegy-etlen-ség-ük-ben* (literally, “pity-less-ness-their-in” = “in their cruelty”) versus *gond-atlan-ság-uk-ban* (literally, “care-less-ness-their-in” = “in their carelessness”). Also to be considered in this context is the fact that in a given language a particular phoneme or combination of phonemes may be found only rarely or not at all in a specific position in the word; for instance, in English /z/ is seldom to be found at the beginning of words and the “ng

sound" (/ŋ/) never occurs at all in this position. One problem with phonological characterizations of words is that, of their very nature, they relate to specific languages or, at best, to specific language-types. Also, such characterizations often have to be seen as descriptions of broad tendencies rather than as absolutely reliable; thus, with regard to stress in English, many units that are recognized as words in that language typically do not actually take stress in ordinary speech, e.g. *and*, *but*, *by*, *if*, *the* — and, on the other hand, in some groups of words which constitute fixed expressions only one main stress is applied in the entire group, e.g. *building worker*, *dancing lesson*, *lifeboat crew*.

The semantic approach

If definitions in terms of sound have their limitations, what about definitions in terms of meaning? Might it not be possible, for example, to define words as the basic units of meaning in language? The answer to this question is unfortunately "No". There are admittedly individual units of meaning which are expressed in single, simple words. For example, the English words *ant*, *bottle* and *shoe* are individual and indivisible forms which convey specific individual meanings. However, the relationship between single words and particular meanings is not always quite so straightforward.

Let us consider, for instance the English word *teapot*. This is written as a single item and can be thought of as denoting a single entity, but, on the other hand, it does actually contain two elements which are words in their own right — *tea* and *pot*. Similarly, there are combinations which are not necessarily written as one word, such as: *public house*, *cricket pavilion*, *icecream kiosk* etc. Actually, if we think more carefully about the meanings of such combinations we can recognize the semantic contributions of each individual word, but the image which each combination of words first brings to mind is unquestionably that of a single building or type of building. A further obvious point to be made about the idea of words being minimum units of meaning is that there are actually units *below* the level of the word which function as semantic units. Reference has already been made to the fact that words may contain units that cannot stand alone as words in their own right. For example, the word *un-just-ly* has the word *just* as its core but also contains two elements (*un-* and *-ly*) which are vital to its meaning — *un* meaning roughly "not" and *-ly* meaning something like "in a ... manner".

The grammatical approach

The characterization of the word that seems to be least problematic is that which defines words in grammatical terms. The grammatical approach uses the criteria of "positional mobility" and "internal stability". Words are said to be "positionally mobile" in the sense that they are not fixed to specific places in a sentence. For example, in a sentence like *The cat drowsily stretched*

her elegant forelegs we can re-order the words in various ways without removing or disrupting anything essential.

The cat stretched her elegant forelegs drowsily.

Drowsily the cat stretched her elegant forelegs.

Her elegant forelegs the cat drowsily stretched.

“Internal stability” refers to the fact that within words the order of morphemes remains consistent. Thus the morphemic constituents of, for example, *forelegs* (*fore* + *leg* + *s*) cannot be altered — so that **sforeleg*, **slegfore*, **legfores*, *foresleg* and *legsfore* are not possible versions of the word in question.

Definition of the word as units which are positionally mobile but internally stable works well across languages. However, even this on the whole successful definition needs some qualification. For example, the English definite article *the* would normally be considered a word, but its positional mobility is distinctly limited. That is to say, except when it is being talked about as an object of study (as it is now), it has to be part of a noun phrase, occurring before the noun and any other elements that are included to qualify the noun; thus, *the wolf*, *the large wolf*, *the extremely large wolf*, etc. Interestingly, the words that have such tight restrictions imposed on their possible positions in sentences are typically grammatical words, notably, definite articles (*the*), indefinite articles (*a*, *an*), prepositions (*in*, *on*, *to*, *from*, etc.), which, as we have seen, have traditionally been regarded as lesser species of words, not “full words”.

Defining the word: a summary

Having looked at a number of possibilities for defining the word, then, what can we say about this problem? Well, one thing is clear: there is not just one way of looking at words. We can see them as types or tokens; we can see them as lexemes or word-forms; we can see them as orthographic units, phonological units, grammatical units or semantic units. We can also make a distinction between content words and form words.

Regarding the various approaches to providing a general characterization of the word, it is clear that the grammatical approach in this connection is not only the least problematic but also the one that works best across languages. Phonetic and semantic perspectives offer little in the way of definitional criteria, but they do suggest some procedures which may be of use to the field linguist working with informants. As far as orthographic and phonological approaches are concerned, the criteria which emerge from these approaches apply in different ways and degrees to different languages.

One result of particular sets of criteria operating differently from language to language is that words in one language may have some characteristics which have little or nothing in common with the characteristics of words in another language. For example, a word in Finnish — with

word-stress and vowel-harmony — is rather different from a word in French, a language in which neither word-stress nor vowel-harmony operates. This does not mean, though, that it is inappropriate to use the term *word* in a cross-linguistic context. Finnish words and French words are recognizable on the basis of other criteria — grammatical criteria, the “potential pause” criterion, etc. — which are not tied to any particular language or language-group.