5 The sounds of language

I take it you already know
Of tough and bough and cough and dough?
Others may stumble but not you
On hiccough, thorough, lough and through.
Well done! And now you wish, perhaps,
To learn of less familiar traps?

Beware of heard, a dreadful word,
That looks like beard and sounds like bird.
And dead: it's said like bed, not bead –
For goodness sake don't call it 'deed'!
Watch out for meat and great and threat
(They rhyme with suite and straight and debt).

T. S. W. (1970)

Imagine that a restaurant manager who has always had trouble with the spelling of English words places an advertisement for a new *SEAGH*. You see the advertisement and your confusion leads you to ask how he came to form this unfamiliar word. It's very simple, he says. Take the first sound of the word *SURE*, the middle sound of the word *DEAD*, and the final sound of the word *LAUGH*. You will, of course, recognize that this form conveys the pronunciation usually associated with the word *chef*.

This tale, however unlikely, may serve as a reminder that the sounds of spoken English do not match up, a lot of the time, with letters of written English. If we cannot use the letters of the alphabet in a consistent way to represent the sounds we make, how do we go about describing the sounds of a language like English? One solution is to produce a separate alphabet with symbols which represent sounds. Such a set of symbols does exist and is called the 'phonetic alphabet'. We will consider how these symbols are used to represent both the consonant and vowel sounds of English words and what physical aspects of the human vocal tract are involved in the production of those sounds.

Phonetics

The general study of the characteristics of speech sounds is called **phonetics**. Our primary interest will be in **articulatory phonetics**, which is the study of how speech sounds are made, or 'articulated'. Other areas of study within phonetics are **acoustic phonetics**, which deals with the physical properties of speech as sound waves 'in the air', and **auditory** (or perceptual) **phonetics**, which deals with the perception, via the ear, of speech sounds. One other area, called **forensic phonetics**, has applications in legal cases involving speaker identification and the analysis of recorded utterances.

Articulation: voiced and voiceless

In articulatory phonetics, we investigate how speech sounds are produced using the fairly complex oral equipment we have. We start with the air pushed out by the lungs up through the trachea (the 'windpipe') to the larynx. Inside the larynx are your vocal cords which take two basic positions.

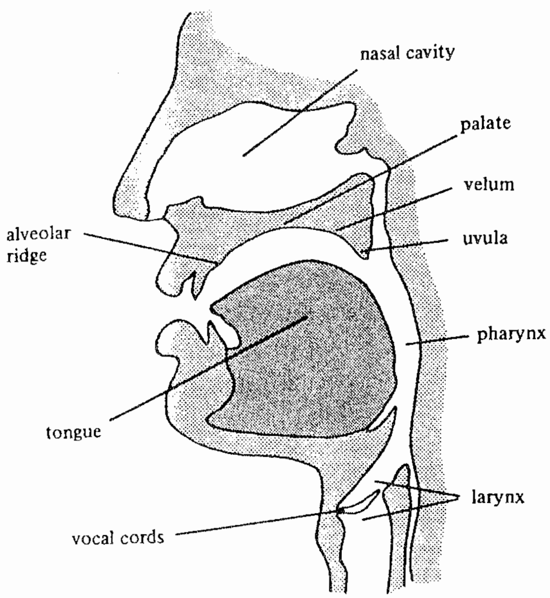
(1) When the vocal cords are spread apart, the air from the lungs passes between them unimpeded. Sounds produced in this way are described as **voiceless**.
(2) When the vocal cords are drawn together, the air from the lungs repeatedly pushes them apart as it passes through, creating a vibration effect. Sounds produced in this way are described as **voiced**.

The distinction can be felt physically if you place a fingertip gently on the top of your 'Adam's apple' (i.e. part of your larynx) and produce sounds like Z-Z-Z-Z or V-V-V-V. Because these are voiced sounds, you should be able to feel some vibration. Keeping your fingertip in the same position, make the sounds S-S-S-S or F-F-F-F. Because these are voiceless sounds, there should be no vibration. Another trick is to put a finger in each ear, not too far, and produce the voiced sounds (e.g. Z-Z-Z-Z) to hear some vibration, whereas no vibration will be heard if the voiceless sounds (e.g. S-S-S-S) are produced in the same manner.

Place of articulation

Once the air has passed through the larynx, it comes up and out through the mouth and/or the nose. Most consonant sounds are produced by using the ongue and other parts of the mouth to constrict, in some way, the shape of he oral cavity through which the air is passing. The terms used to describe nany sounds are those which denote the place of articulation of the sound: :hat is, the location, inside the mouth, at which the constriction takes place.

What we need is a slice of head. If you crack a head right down the middle, you will be able to see which parts of the oral cavity are crucially involved in speech production. To describe the place of articulation of most consonant sounds, we can start at the front of the mouth and work back. We can also keep the voiced–voiceless distinction in mind and begin using the symbols of the phonetic alphabet to denote specific sounds. These symbols will be enclosed within square brackets [ ].



Bilabials. These are sounds formed using both (= bi) upper and lower lips (= labia). The initial sounds in the words *pat*, *bat* and *mat* are all bilabials. They are represented by the symbols [p], which is voiceless, and [b] and [m], which are voiced. The [w] sound found at the beginning of *way*, *walk* and *world* is also a bilabial.

Labiodentals. These are sounds formed with the upper teeth and the lower lip. The initial sounds of the words *fat* and *vat* and the final sounds in the words *safe* and *save* are labiodentals. They are represented by the symbols [f], which is voiceless, and [v], which is voiced. Notice that the final sounds of *laugh* and *cough*, and the initial sound of *photo*, despite the spelling differences, are all pronounced as [f].

Dentals. These sounds are formed with the tongue tip behind the upper front teeth. The term **interdental** is sometimes used to describe a manner of pronunciation with the tongue tip between (=inter) the upper and lower teeth. The initial sound of *thin* and the final sound of *bath* are both voiceless dentals. The symbol used for this sound is [θ], usually referred to as 'theta'. It's the symbol you would use for the first and last sounds in the phrase *three teeth*.

The voiced dental is represented by the symbol [ð], usually called 'eth'. This sound is found in the pronunciation of the initial sound of common words like *the*, *there*, *then* and *thus*. It's also the middle sound in *feather* and the final sound of *bathe*.

Alveolars. These are sounds formed with the front part of the tongue on the alveolar ridge, which is the rough, bony ridge immediately behind the upper teeth. The initial sounds in *top*, *dip*, *sit*, *zoo* and *nut* are all alveolars. The symbols for these sounds are quite easily remembered - [t], [d], [s], [z], [n]. Of these, [t] and [s] are voiceless, whereas [d], [z] and [n] are voiced.

It may be clear that the final sounds of the words *bus* and *buzz* have to be [s] and [z] respectively, but what about the final sound of the word *raise*? The spelling is misleading because the final sound in this word is voiced, and so must be represented by [z]. Notice also that despite the different spelling of *knot* and *not*, both of these words are pronounced with [n] as the initial sound.

Other alveolars are the [l] sound found at the beginning of words such as *lap* and *lit*, and the [r] sound at the beginning of *right*, *write* and *rip*.

Alveo-palatals. If you feel back behind the alveolar ridge, you should find a hard part in the roof of your mouth. This is called the palate. Sounds which are produced with the tongue at the very front of the palate, near the alveolar ridge, are called alveo-palatals. Examples are the initial sounds in the words *shout* and *child*, which are voiceless.

Although there are two letters in the spelling of 'sh' and 'ch', the sounds are represented by the single phonetic symbols [š] and [č] respectively. The small mark above the symbols is called 'wedge'. So, the word *shoe-brush* begins and ends with the voiceless alveo-palatal sound [š] and the word *church* begins and ends with the voiceless alveo-palatal sound [č].

One of the voiced alveo-palatal sounds, represented by the symbol [ž], is ot very common in English, but can be found as the middle consonant ound in words like *treasure* and *pleasure*, or the final sound in *rouge*. The ther voiced alveo-palatal sound is represented as [ǰ] and is the initial sound n words like *joke* and *gem*. The word *judge* and the name *George* both begin nd end with the sound [ǰ], despite the obvious differences in spelling.

One sound which is produced with the tongue in the middle of the palate s the [y] sound to be found at the beginning of words like *you* and *yet*. This ound is usually described as a **palatal**.

Velars. Even further back in the roof of the mouth, beyond the hard palate, ou will find a soft area which is called the soft palate, or the velum. Sounds roduced with the back of the tongue against the velum are called velars. There is a voiceless velar sound, represented by the symbol [k], which occurs ot only in *kid* and *kill*, but is also the initial sound in *car* and *cold*. Despite he variety in spelling, this [k] sound is both the initial and final sound in the words *cook*, *kick* and *coke*. The voiced velar sound to be heard at the beginning of words like *go*, *gun* and *give* is represented by [g]. This is also the final sound in words like *bag*, *mug* and, despite the spelling, *plague*.

The velum can be lowered to allow air to flow through the nasal cavity and thereby produce another voiced velar which is represented by the symbol [ŋ], typically referred to as 'angma'. In written English, this sound is normally spelled as the two letters 'ng'. So, the [ŋ] sound is at the end of *sing*, *sang* and, despite the spelling, *tongue*. It would occur twice in the form *ringing*. Be careful not to be misled by the spelling – the word *bang* ends with the [ŋ] sound only. There is no [g] sound in this word.

Glottals. There is one other sound that is produced without the active use of the tongue and other parts of the mouth. It is the sound [h] which occurs at the beginning of *have* and *house*, and, for most speakers, as the first sound in *who* and *whose*. This sound is usually described as a voiceless glottal. The 'glottis' is the space between the vocal cords in the larynx. When the glottis is open, as in the production of other voiceless sounds, but there is no manipulation of the air passing out through the mouth, the sound produced is that represented by [h].

Charting consonant sounds

Having described in some detail the place of articulation of English consonant sounds, we can summarize the basic information in the following chart. Along the top of the chart are the different labels for places of articulation and, under each, the labels –V (= voiceless) and +V (= voiced). Also included in this chart, on the left-hand side, is a set of terms used to describe 'manner of articulation' which we will discuss in a later section.

	Bilabial		Labio-dental		Dental		Alveolar		Alveo-palatal		Velar		Glottal	
	–V	+V	–V	+V	–V	+V	–V	+V	–V	+V	–V	+V	–V	+V
Stops	p	b					t	d			k	g		
Fricatives			f	v	θ	ð	s	z	š	ž				
Affricates									č	ǰ				
Nasals		m						n				ŋ		
Approximants		w						l,r		y			h	

Notes on the chart. This chart is far from complete. It does contain the majority of consonant sounds used in the basic description of English pronunciation. There are, however, several differences between this basic set of symbols and the much more comprehensive chart produced by the International Phonetic Association (IPA). The most obvious difference is in the range of sounds covered.

The IPA aims to describe the sounds of all languages and includes, for example, symbols for the velar fricative sound you may have heard in the German pronunciation of the "ch" part of *Bach* or *Achtung*. It also includes sounds made with the back of the tongue and the uvula (below the velum) which represents the "r" parts of the French pronunciation of *rouge* and *lettre*. Uvular sounds also occur in many American Indian languages. Other, non-English sounds such as pharyngeals (produced in the pharynx) occur in Semitic languages such as Arabic. There are many more.

Another shortcoming of the chart above is the single entry covering *r* sounds in English. There can be a lot of variation among speakers in the pronunciation of the initial sound in *raw* and *red*, the medial sound in *very*, and the final sound in *hour* and *air*. Different symbols (e.g. [ɹ], [R]) may be encountered in transcriptions where the different *r* sounds are distinguished.

Finally, the IPA uses different symbols for a few of the sounds represented here. These alternatives are [ʃ] = [š]; [ʒ] = [ž]; [tʃ] = [č]; [dʒ] = [ǰ] and [j] = [y]. For a fuller discussion of the use of IPA symbols, see Ladefoged (1992).

Manner of articulation

So far, we have concentrated on describing consonant sounds in terms of where they are articulated. We can, of course, describe the same sounds in terms of how they are articulated. Such a description is necessary if we wish to be able to differentiate between some sounds which, in the preceding discussion, we have placed in the same category. For example, we can say that [t] and [s] are both voiceless alveolar sounds. How do they differ? They differ in their manner of articulation, that is, in the way they are pronounced. The [t] sound is one of a set of sounds called 'stops' and the [s] sound is one of a set called 'fricatives'.

Stops. Of the sounds we have already mentioned, the set [p], [b], [t], [d], [k], [g] are all produced by some form of complete 'stopping' of the airstream (very briefly) and then letting it go abruptly. This type of consonant sound resulting from a blocking or stopping effect on the airstream is called a stop (or a 'plosive'). A full description of the [t] sound at the beginning of a word like *ten* is as a 'voiceless alveolar stop'. On occasion, only the manner of articulation is mentioned, as when it is said that the word *bed*, for example, begins and ends with 'voiced stops'.

Fricatives. The manner of articulation used in producing the set of sounds [f], [v], [θ], [ð], [s], [z], [š], [ž] involves almost blocking the airstream, and having the air push through the narrow opening. As the air is pushed through, a type of friction is produced and the resulting sounds are called fricatives. If you put your open hand in front of your mouth when making these sounds, [f] and [s] in particular, you should be able to feel the stream of air being pushed out. A word like *fish* will begin and end with 'voiceless fricatives'. The word *those* will begin and end with the 'voiced fricatives' [ð] and [z].

Affricates. If you combine a brief stopping of the airstream with an obstructed release which causes some friction, you will be able to produce the sounds [č] and [ǰ] These are called affricates and occur at the beginning of the words *cheap* and *jeep*. In the first of these, there is a 'voiceless affricate', and in the second a 'voiced affricate'.

Nasals. Most sounds are produced orally, with the velum raised, preventing airflow from entering the nasal cavity. However, when the velum is lowered and the airstream is allowed to flow out through the nose to produce [m], [n] and [ŋ], the sounds are described as nasals. These three sounds are all voiced. Words like *morning*, *knitting* and *name* begin and end with nasals.

Approximants. In the set of sounds called approximants, the articulation of each is strongly influenced by the following vowel sound. Indeed, the sounds [w] and [y] are sometimes called 'semi-vowels' or 'glides', because they are typically produced with the tongue moving, or 'gliding', to or from the position of a nearby vowel. Both [w] and [y] are voiced, occurring at the beginning of *we*, *wet*, *you* and *yes*.

Also voiced are the two initial approximants in *led* and *red*. The [l] sound is formed by letting the airstream flow around the sides of the tongue as it makes contact with the alveolar ridge. The type of sound for which we are using the [r] symbol is formed with the tongue tip raised and curled back behind the alveolar ridge. The [l] and [r] sounds are also sometimes called 'liquids'.

The sound [h] is a voiceless approximant which, in common words like *Hi* or *hello*, simply begins the pronunciation of the following vowel as if it was voiceless.

The glottal stop and the flap. Two common terms used to describe ways of pronouncing consonants are not included in the chart presented earlier.

The **glottal stop**, represented by the symbol [ʔ], occurs when the space between the vocal cords (the glottis) is closed completely, very briefly, and then released. Try saying the expression *Oh oh!*. Between the first *Oh* and the second *oh*, people typically produce a glottal stop. Some people do it in the middle of *Uh-uh* (meaning 'no'), and others put one in place of 't' in pronouncing *Batman*. You can also produce a glottal stop if you try to say the words *butter* or *bottle* without pronouncing the *-tt-* part in the middle. This sound is considered to be characteristic of Cockney (London) speech, but it is also used by Scottish speakers and New Yorkers.

If, however, you are an American English speaker who pronounces the word *butter* in a way that is close to 'budder', then you are making a **flap**. It is represented by [D] or sometimes [ɾ]. This flap is produced by the tongue tip being thrown against the alveolar ridge for an instant. Many Americans tend to flap the [t] and [d] consonants between vowels so that, in casual speech, the pairs *latter* and *ladder*, *writer* and *rider*, *metal* and *medal* do not have distinct middle consonants. They all have flaps. The student who was told about the importance of 'Plato' in class and reported it as 'play-dough' was clearly a victim of a misinterpreted flap.

This rather lengthy list of the phonetic features of English consonant sounds is not presented as a challenge to your ability to memorize a lot of terminology and symbols. It is presented as an illustration of how a thorough description of the physical aspects of speech production will allow us to characterize the sounds of spoken English, independently of the vagaries of spelling found in written English. There are, however, some sounds which we have not yet investigated. These are the types of sounds known as vowels and diphthongs.

Vowels

While the consonant sounds are mostly articulated via closure or obstruction in the vocal tract, vowel sounds are produced with a relatively free flow of air. They are all typically voiced. To describe vowel sounds, we consider the way in which the tongue influences the 'shape' through which the airflow must pass. To talk about place of articulation, we think of the space inside the mouth as having a front versus a back and a high versus a low area. Thus, in the pronunciation of *heat* and *hit*, we talk about 'high, front' vowels, because the sound is made with the front part of the tongue in a raised position.

In contrast, the vowel sounds in *hot* and *hat* are produced with the tongue in a relatively lower position and are described as 'low, back' vowels. The next time you're facing the bathroom mirror, try saying *heat, hit, hot, hat*. For the first two, your mouth will stay fairly closed, but for the last two, your tongue will move lower and cause your mouth to open wider. (You may also notice that the sounds of relaxation and pleasure, if you're getting any, typically contain back vowels.)

The terminology for describing vowels is usually presented in the form of a chart, as shown below, which provides a means of classifying the most common vowel sounds of English.

	Front	Central	Back
High	i ɪ		u ʊ
Mid	e ɛ	ə	o ɔ
Low	æ	ʌ	a

The easiest way to become familiar with the distinctions within the set of vowel sounds is to have some examples of familiar words which for a lot of American English speakers, most of the time, contain those sounds. The following list goes from the high front vowels through to the low back vowel and ends with three diphthongs:

[i]	*see, eat, key*	[ʊ]	*put, could, foot*
[ɪ]	*hit, myth*	[o]	*no, know, though*
[e]	*tail, great, weight*	[ɔ]	*raw, fall, caught*
[ɛ]	*pet, said, dead*	[a]	*cot, father, body*
[æ]	*sat, ban*	[ay]	*my, buy, eye*
[ə]	*above, sofa*	[aw]	*cow, loud*
[ʌ]	*putt, blood, tough*	[ɔy]	*boy, void*
[u]	*move, two, glue*		

Diphthongs. The last three symbols in the list above contain two sounds. These 'combined' vowel sounds are called diphthongs. Note that in each case they begin with a vowel sound and end with a glide. With the majority of single vowel sounds, the vocal organs remain relatively steady, but in pronouncing diphthongs, we move from one vocalic position to another.

This process of **diphthongization** can actually happen with a wide range of vowel sounds and is more common in some varieties of English (e.g. Southern British) than in others. Most Americans pronounce the word *say* as [sey], with a diphthong rather than a single vowel. You will also hear the pronouns *they* as [ðey], *you* as [yuw] and *we* as [wiy], all diphthongized. If you try to pronounce the consonants and diphthongs in the following transcription, you should recognize a traditional speech training exercise: [haw naw brawn kaw].

Notes on the vowel chart. Vowel sounds are notorious for varying between one variety of English and the next, often being a key element in what we recognize as different accents. So, you may find that some of the words offered here as examples are not normally encountered with the vowel sounds as listed. Also, some of the sounds shown here may not be commonly used in your dialect. It may be, for example, that you make no distinction between the vowels in the words *caught* and *cot*. Some transcriptions only use [a] for this back vowel sound.

Or, you may not make a significant distinction between the central vowels [ə] and [ʌ]. If not, then just use the symbol [ə], called 'schwa'. In fact, in casual speech, we all use schwa more than any other single sound. It is the unstressed vowel in the everyday use of words like *afford*, *collapse*, *oven*, *photograph*, *wanted*, and in the common words *a* and *the*.

There are many other variations in the actual physical articulation of the sounds we have considered here. The more we focus on the subtle differences of the actual articulation of each sound, the more likely we are to find ourselves describing the pronunciation of small groups or even individual peakers. Such subtle differences allow us to identify individual voices. But hose differences don't help us understand how we know what total trangers with unfamiliar voices are saying. To make sense of that, we need o look at the more general sound patterns of a language, also known as phonology'.

Study questions

Try pronouncing the initial sounds of the following words and then determine the place of articulation (e.g. bilabial, alveolar, etc.) of each:

(a) hand ________ (b) foot ________ (c) toe ________
(d) belly ________ (e) chin ________ (f) thigh ________
(g) calf ________ (h) knee ________

Which of the following words end with voiceless (–V) sounds and which end with voiced (+V) sounds?

(a) crash ________ (b) bang ________ (c) smack ________
(d) thud ________ (e) wham ________ (f) splat ________

Identify the manner of articulation (e.g. stop, fricative, etc.) of the initial sounds in the following words:

(a) silly ________ (b) crazy ________ (c) jolly ________
(d) merry ________ (e) dizzy ________ (f) happy ________
(g) loony ________ (h) funny ________

Which written English words are usually pronounced as transcribed here?

(a) klɪŋ ________ (b) fes ________ (c) šip ________
(d) ðə ________ (e) hu ________ (f) bæk ________
(g) bɔt ________ (h) haw ________

Produce a phonetic transcription of the most common pronunciation you hear of the following words:

(a) she ________ (b) tape ________ (c) dope ________
(d) walk ________ (e) sigh ________ (f) fell ________
(g) these ________ (h) thought ________

Discussion topics/projects

(ii) Below is a set of English words with different written forms representing the same sounds in a number of ways. Can you identify the alternative spellings of the sounds [i], [f] and [e]?

elephant, rare, marines, pear, hay, feet, quay, air, suite, weigh, giraffe, pier, tough, keys, meat, Sikh.

(ii) How many different ways of spelling the sounds [s], [k], [š] and [ɛ] can you discover in English words?

B (i) Using the first two examples as a guide, can you provide a description, in terms of manner of articulation and voicing (if necessary), of your pronunciation of the initial consonants of the following English words?

(a) mist (NASAL) (g) thin
(b) bat (VOICED STOP) (h) near
(c) far (i) tall
(d) wall (j) joke
(e) rope (k) shop
(f) zoo (l) gun

(ii) What criteria did you use to decide if the voicing aspect was 'necessary' or not?

C Two other distinctions used in describing the articulation of vowels are called 'rounded versus unrounded' and 'tense versus lax'.
(i) The description 'rounded' applies to the shape of the lips. With the help of a mirror or a co-operative friend, try to decide which English vowels are normally rounded and which are not. Is there any other term that you would use to describe a common recognizable shape of the lips for some vowels?

(ii) The term 'tense' means that the vowels are produced with extra muscular effort whereas 'lax' vowels do not require this effort. Can you identify which of the English vowels are easily identifiable as 'tense' or 'lax'? Are there any contexts (e.g. presence or absence of other sounds) that make this identification easier?
(You can consult any of the Phonetics texts in the Further reading section for help with this one.)

D Consider the following set of transcribed 'words'. Can you divide the set into those forms which are English words, those which could not possibly be English words, and those which are not English words at this time, but might possibly become English words? How do you make the decision regarding what goes in the second or third group?

(1) θlem	(5) ksln	(9) črls
(2) θrlnz	(6) šlop	(10) blʌnk
(3) θiətər	(7) kwlk	(11) fɛrt əm
(4) sɔng	(8) zun	(12) bɔyllŋ

E English has a number of expressions such as chit-chat and flip-flop which never seem to occur in the reverse order (i.e. chat-chit, flop-flip). Perhaps you can add examples to the following list?

criss-cross	hip-hop	riff-raff
dilly-dally	knick-knacks	see-saw
ding-dong	mish-mash	sing-song
fiddle-faddle	ping-pong	tick-tock
flim-flam	pitter-patter	zig-zag

(i) Can you think of a phonetic description of the regular pattern in these expressions?
(ii) What kind of phonetic description might account for these other common pairings?

fuddy-duddy	hocus-pocus	namby-pamby
fuzzy-wuzzy	hurly-burly	razzle-dazzle
hanky-panky	lovey-dovey	roly-poly
helter-skelter	mumbo-jumbo	super-duper

Further reading

Introductory chapters on phonetics can be found in all linguistics textbooks, such as Chapter 3 of Akmajian *et al.* (1990), Chapter 2 of Finegan & Besnier (1989), Chapter 5 of Fromkin & Rodman (1993), or Chapter 2 of O'Grady *et al.* (1989). The standard textbook is Ladefoged (1992). Other texts are Calvert (1993). The standard textbook is Ladefoged (1992). Other texts are Calvert (1993), Catford (1988), Clark & Yallop (1995), Edwards (1992), MacKay (1987) and Roach (1991). More specialized discussion of the issues can be found in Abercrombie (1967), Fromkin (1985), Hardcastle & Laver (1995), Ladefoged & Maddieson (1995), Laver (1994), Levelt (1989) and Ramsaran (1990). Pullum & Ladusaw (1986) is a useful guide to phonetic symbols and Crystal (1991) is a dictionary. Section 17 of Crystal (1995) provides information on regional variants of pronunciation. Texts with a greater emphasis on describing pronunciation are Gimson (1994) and Kreidler (1989), and on teaching pronunciation (mainly to learners of English), see Avery & Ehrlich (1992), Celce-Murcia *et al.* (1996), Prator & Robinett (1985) or Dalton & Seidlhofer (1994). On auditory and acoustic phonetics, see Denes & Pinson (1993), Fry (1979) or Lieberman & Blumstein (1988). On forensic phonetics, try Baldwin & French (1990) or Hollien (1990). The closely related area of speech science is described in Borden *et al.* (1994).

6 The sound patterns of language

Ever on the search for legal jokes not necessarily connected with the death penalty, I consulted a friend who is still practising. She said a member of her chambers was in court one Monday morning when the judge said, "I'm afraid we'll have to adjourn this case, I have written my judgment out, but I left it in my cottage in Devon and I can't get it sent here until tomorrow." "Fax it up, my Lord," the helpful barrister suggested, to which his Lordship replied, "Yes, it does rather."

Quoted in Crystal (1995)

In the preceding chapter, we investigated the physical production of speech sounds in terms of the articulatory mechanisms of the human vocal tract. That investigation was possible because of some rather amazing facts about the nature of language. When we considered the human vocal tract, we did not have to specify whether we were talking about a fairly large male, over six feet tall, weighing over 200 pounds, or about a rather small female, about five feet tall, weighing 100 pounds. Yet those two physically different individuals would inevitably have physically different vocal tracts, in terms of size and shape. In a sense, every individual has a physically different vocal tract. Consequently, in purely physical terms, every individual will pronounce sounds differently. There are, then, potentially thousands of physically different ways of saying the simple word *me*.

In addition, each individual will not pronounce the word *me* in a physically identical manner on every occasion. Obvious differences occur when the individual is shouting, is asking for a sixth martini, or is suffering from a cold. Given this vast range of potential differences in the actual physical production of a speech sound, how do we manage consistently to recognize all those versions of *me* as the form [mi], and not [ni], or [si], or [ma], or [mo], or something else entirely? The answer to that question is provided to a large extent by the study of phonology.

Phonology

Phonology is essentially the description of the systems and patterns of speech sounds in a language. It is, in effect, based on a theory of what every speaker of a language unconsciously knows about the sound patterns of that language. Because of this theoretical status, phonology is concerned with the abstract or mental aspect of the sounds in language rather than with the actual physical articulation of speech sounds. Phonology is about the underlying design, the blueprint of the sound type, that serves as the constant basis of all the variations in different physical articulations of that sound type in different contexts.

Thus, when we think of the [t] sound in the words *tar*, *star*, *writer* and *eighth* as being 'the same', we actually mean that, in the phonology of English, they would be represented in the same way. In actual speech, these [t] sounds are all very different.

However, all those articulation differences in [t] sounds are less important than the distinction between the [t] sounds in general and the [k] sounds, or the [f] sounds, or the [b] sounds, because there are meaningful consequences related to the use of one rather than the others. These sounds must be distinct meaningful sounds, regardless of which individual vocal tract is being used to pronounce them, because they are what make the words *tar*, *car*, *far* and *bar* meaningfully distinct. Considered from this point of view, we can see that phonology is concerned with the abstract set of sounds in a language which allows us to distinguish meaning in the actual physical sounds we say and hear.

Phonemes

Each one of these meaning-distinguishing sounds in a language is described as a **phoneme**. When we considered the basis of alphabetic writing in Chapter 2, we were actually working with the concept of the phoneme as the single sound type which came to be represented by a single symbol. It is in this sense that the phoneme /t/ is described as a sound type, of which all the different spoken versions of [t] are tokens. Note that slash marks are conventionally used to indicate a phoneme, /t/, an abstract segment, as opposed to the square brackets, as in [t], used for each phonetic, or physically produced, segment.

An essential property of a phoneme is that it functions contrastively. We know that there are two phonemes /f/ and /v/ in English because they are the only basis of the contrast in meaning between the forms *fat* and *vat*, or *fine* and *vine*. This contrastive property is the basic operational test for determining the phonemes which exist in a language. If we substitute one sound for another in a word and there is a change of meaning, then the two sounds represent different phonemes. The consonant and vowel charts presented in Chapter 5 can now be seen as essentially a mapping out of the phonemes of English.

The terms which were used in creating that chart can be considered 'features' which distinguish each phoneme from the next. If the feature is present, we mark it with a plus (+) sign; if it's not present, we use a minus (–) sign. Thus, /p/ can be characterized as [–voice, +bilabial, +stop] and /k/ as [–voice, +velar, +stop]. Because these two sounds share some features, they are sometimes described as members of a natural class of sounds. The prediction would be that sounds which have features in common would behave phonologically in some similar ways. A sound which does not share those features would be expected to behave differently.

For example, /v/ has the features [+voice, +labiodental, +fricative] and so cannot be in the same 'natural class' as /p/ and /k/. Although other factors will be involved, this feature-analysis could lead us to suspect that there may be a good phonological reason why words beginning with /pl-/ and /kl-/ are common in English, but words beginning /vl-/ are not. Could it be that there are some definite sets of features required in a sound in order for it to occur word-initially before /l/? If so, then we will be on our way to producing a phonological account of permissible sound sequences in the language.

Phones and allophones

While the phoneme is the abstract unit or sound-type ('in the mind'), there are many different versions of that sound-type regularly produced in actual speech ('in the mouth'). We can describe those different versions as **phones**. Phones are phonetic units and will appear in square brackets. When we have a set of phones, all of which are versions of one phoneme, we refer to them as the **allophones** of that phoneme.

For example, the [t] sound in the word *tar* is normally pronounced with a stronger puff of air than is present in the [t] sound of the word *star*. If you put the back of your hand in front of your mouth as you say *tar*, then *star*, you should have some physical evidence of the **aspiration** (the puff of air) accompanying the [t] sound in the initial position of *tar* (but not in *star*). This aspirated version is represented more precisely as [t^h]. That's one phone. In the last chapter, we noted that the [t] sound between vowels in a word like *writer* often becomes a 'flap', which we represented as [D]. That's another phone. In the pronunciation of a word like *eighth*, the influence of the final dental [θ] sound causes a dental articulation of the [t] sound. This would be represented more precisely as [t̪]. That's yet another phone. There are other variations of this sound which, like [tʰ], [D] and [t̪], can be represented differently in a detailed, or narrow, phonetic transcription. Because these variations form a set of phones, they would typically be referred to as allophones of the phoneme /t/.

The crucial distinction between phonemes and allophones is that substituting one phoneme for another will result in a word with a different meaning (as well as a different pronunciation), but substituting allophones only results in a different (and perhaps odd) pronunciation of the same word.

Let's take another brief example. In English, there is a difference in pronunciation of the /i/ sound in words like *seed* and *seen*. In the second word, the effect of the nasal consonant [n] makes the [i] sound nasalized. This nasalization can be represented by a diacritic [˜], called 'tilde', over the symbol [ĩ] in narrow phonetic transcription. So, there are at least two phones, [i] and [ĩ], used in English to realize a single phoneme. They are allophones of /i/ in English.

It is possible, of course, for two languages to have the same pair of phonetic segments, but to treat them differently. In English, the effect of nasalization on a vowel is treated as allophonic variation because the nasalized version is not meaningfully contrastive. In French, however, the pronunciation [mɛ] is used for one word *mets*, meaning 'dish', and [mɛ̃] for a different word *main*, meaning 'hand'. Also, [so] for *seau*, meaning 'pail', contrasts with [sõ] for *son*, meaning 'sound'. Clearly, in these cases, the distinction is phonemic.

Minimal pairs and sets

Phonemic distinctions in a language can be tested via pairs and sets of words. When two words such as *pat* and *bat* are identical in form except for a contrast in one phoneme, occurring in the same position, the two words are described as a **minimal pair**. More accurately, they would be classified as a minimal pair in the phonology of English. (Arabic does not have this contrast between the two sounds.) Other examples of English minimal pairs are *fan – van*, *bet – bat*, *site – side*. Such pairs have been used frequently in tests of English as a second language to determine non-native speakers' ability to understand the contrast in meaning resulting from the minimal sound contrast.

When a group of words can be differentiated, each one from the others, by changing one phoneme (always in the same position), then we have a **minimal set**. Thus, a minimal set based on the vowel phonemes of English would include *feat*, *fit*, *fat*, *fate*, *fought*, *foot*, and one based on consonants could have *big*, *pig*, *rig*, *fig*, *dig*, *wig*.

Phonotactics

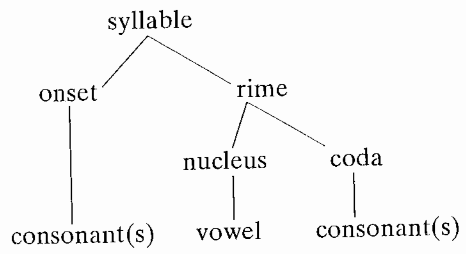
This type of exercise involving minimal sets also allows us to see that there are indeed definite patterns to the types of sound combinations permitted in a language. In English, the minimal set we have just listed does not include forms such as *lig* or *vig*. As far as I know, these are not English words, but they can be viewed as possible English words. That is, your phonological knowledge of the pattern of sounds in English words would allow you to treat these forms as acceptable if, at some future time, they came into use. They represent 'accidental' gaps in the vocabulary of English.

It is, however, no accident that forms such as [fsIg] or [rnIg] do not exist or are unlikely ever to exist. They have been formed without obeying some constraints on the sequence or position of English phonemes. Such constraints are called the **phonotactics** of a language and are obviously part of every speaker's phonological knowledge. Because these constraints operate on units larger than the single segment, or phoneme, we have to consider the basic structure of that larger phonological unit called the syllable.

Syllables and clusters

A **syllable** must contain a vowel (or vowel-like) sound. The most common type of syllable in language also has a consonant before the vowel, often represented as CV. Technically, the basic elements of the syllable are the **onset** (one or more consonants) and the **rime**. The rime (also written as 'rhyme') consists of the vowel, which is treated as the **nucleus**, plus any following consonant(s), treated as the **coda**.

Thus, syllables like *me*, *to* or *no* have an onset and a nucleus, but no coda. They are known as 'open' syllables. When a coda is present, as in the syllables *up*, *cup*, *at* or *hat*, they are called 'closed' syllables. The basic structure of the kind of syllable found in English words like *green* (CCVC), *eggs* (VCC), *and* (VCC), *ham* (CVC), *I* (V), *do* (CV), *not* (CVC), *like* (CVC), *them* (CVC), *Sam* (CVC), *I* (V), *am* (VC), is shown in the accompanying diagram.



Both the onset and the coda can consist of more than one consonant, also known as a **consonant cluster**. The combination *st* is a consonant cluster (CC) as onset in the word *stop*, and as coda in the word *post*. There are many CC onset combinations permitted in English phonotactics, as in *black*, *bread*, *trick*, *twin*, *flat* and *throw*, with approximants (/w/, /r/, /l/) frequently appearing in second position. (Note that *throw* begins with only two consonants, /θ r/, once again showing that spelling is not a good guide in phonology.)

English actually can have larger onset clusters, as in *stress* and *splat*, consisting of three consonants (CCC). The phonotactics of these larger onset consonant clusters in English is not difficult to describe. The first consonant must always be /s/, followed by one of the voiceless stops (/p/, /t/, /k/) and then one of the approximants (/r/, /l/, /w/). You can check if this description is adequate for the combinations in *splash*, *spring*, *strong*, *scream* and *square*. Does the description also cover the second syllable in the pronunciation of *exclaim*? How about /ɛk-skleym/? Remember that it is the onset of the syllable that is being described, not the beginning of the word.

It is quite unusual for languages to have consonant clusters of this type. Indeed, the syllable structure of many languages (e.g. Hawaiian or Japanese) is predominantly CV. It is also noticeable in English that large consonant clusters are frequently reduced in casual conversational speech, particularly if they occur in the middle of a word. This is just one example of what is often discussed in terms of 'co-articulation effects'.

Co-articulation effects

In much of the preceding discussion, we have been describing the speech sounds as if they are always pronounced carefully and deliberately, almost in slow motion. Speech isn't like that very often. Mostly our talk is fast and spontaneous, and it requires our articulators to move from one sound to the next without stopping. The process of making one sound almost at the same time as the next is called **co-articulation**. There are two well-known co-articulation effects, called 'assimilation' and 'elision'.

Assimilation

When two phonemes occur in sequence and some aspect of one phoneme is taken or 'copied' by the other, the process is known as **assimilation**. In terms of the physical production of speech, one might assume that this regular process is occasioned by ease of articulation in everyday talk. In isolation, you would probably pronounce /I/ and /æ/ without any nasal quality at all. However, in saying words like *pin* and *pan*, the anticipation of forming the final nasal consonant will make it 'easier' to go into the nasalized articulation in advance and consequently the vowel sounds in these words will be, in precise transcription, [Ĩ] and [æ̃]. This is a very regular feature of English speakers' pronunciation. So regular, in fact, that a phonological rule can be stated in the following way: 'Any vowel becomes nasal whenever it immediately precedes a nasal.'

This type of assimilation process occurs in a variety of different contexts. It is particularly noticeable in ordinary conversational speech. By itself, you may pronounce the word *can* as [kæn], but, if you tell someone *I can go*, the influence of the following velar [g] will almost certainly make the preceding nasal sound come out as [ŋ] (a velar) rather than [n] (an alveolar). The most commonly observed 'conversational' version of the phrase is [aykəŋgo]. Notice that the vowel in *can* has also changed to schwa [ə] from the isolated-word version [æ]. In many words spoken carefully, the vowel receives stress, but in the course of ordinary talk, that vowel may no longer receive any stress and reduce to schwa. For example, you may pronounce *and* as [ænd] in isolation, but in the casual use of the phrase *you and me*, you almost certainly say [ən], as in [yuənmi].

Elision

Note that in the last example, in the environment of preceding and following nasals, the [d] sound of *and* has simply disappeared. The [d] sound is also commonly 'omitted' in the pronunciation of a word like *friendship*, [frɛnšIp]. This 'omission' of a sound segment which would be present in the deliberate pronunciation of a word in isolation is technically described as **elision**. In consonant clusters, especially in coda position, /t/ is a common casualty in this process, as in the typical pronunciation [æspɛks] for *aspects*, or in [himəsbi] for *he must be*. You can, of course, slowly and deliberately pronounce the phrase *we asked him*, but the process of elision in casual speech is likely to produce [wiæstIm]. Vowels also disappear, as in [ɛvri] for *every*, [IntrIst] for *interest*, [kæbnIt] for *cabinet*, and [spowz] for *suppose*.

These two processes of assimilation and elision occur in everyone's speech and should not be treated as a form of sloppiness or laziness in speaking. In fact, consistently avoiding the regular patterns of assimilation and elision used in a language would result in extremely artificial-sounding talk. The point of investigating phonological processes (only a very small number of which have been explored here) is not to arrive at a set of rules about how a language should be pronounced, but to try to come to an understanding of the regularities and patterns which underlie the actual use of sounds in language.

Study questions

1 What is the test used for determining phonemes in a language?
2 Which of the following words would be treated as minimal pairs?
pat, pen, more, heat, tape, bun, fat, ban, chain, tale, bell, far, meal, vote, bet, pit, heel
3 How does an allophone differ from a phoneme?
4 What's the difference between an open and a closed syllable?
5 Which segments are most likely to be affected by elision in the pronunciation of the following words?
(i) *postman* (ii) *government* (iii) *sandwich* (iv) *pumpkin*

Discussion topics/projects

A (i) In all of the following English words there is an onset consonant cluster with /r/ in second position. Is there any way (using voice, place and manner features) to describe the kind of consonant that can appear before /r/ in these clusters, and to exclude any other consonants?
brave, crash, freak, growl, pray, shriek, three, trick

(ii) When the second part of an onset cluster is /l/, as in *black* and other words, which features are required in the first consonant?

B English words like *audible* and *edible* can be made negative by adding *in-* to produce *inaudible* and *inedible*.
(i) What types of assimilation processes appear to have been involved in the pronunciation of the following negatives?
impossible, illegal, irresponsible, immature, indecent, incomplete, ingratitude, insane

(ii) Can you think of any reason why these assimilation processes do not happen with *un-* in words like *unpleasant* or *unreal*? (You can consult chapter 1 of Harris, 1994, for some suggestions.)

C In this chapter, aspiration of a stop consonant in word-initial position was illustrated.
(i) Can you work out, by testing how /p/ is pronounced in the following words, whether "word-initial" is in fact the most accurate description of that phenomenon?
personal, rapid, desperate, computer, competition, special, description, empathy, peculiar

(ii) If you can identify the stressed syllable in each of these words, does that help you arrive at a different phonological description?

D The use of plural *-s* in English has three different, but very regular, phonological alternatives. You add:
/s/ to words like *ship, bat, book* and *cough*
/z/ to words like *cab, lad, cave, rag* and *thing*
/əz/ to words like *bus, bush, judge, church* and *maze*
(i) Can you work out the set of sounds which regularly precedes each of these alternatives?
(ii) What features do each of these sets have in common?

E There seems to be some phonological rule involved in the different pronunciations of the past tense *-ed* form in English. It is /t/ at the end of *walked* and *passed*; it is /d/ at the end of *jogged* and *played*; it is /əd/ at the end of *bounded* and *vaulted*.
(i) What is the phonological pattern in these words (and any others you can add) that determines which *-ed* version is used?

(ii) Do you think that one of these phonological forms for *-ed* is more basic, with the others being derived from it in a regular way? Which, and how?

Further reading

All basic texts in linguistics have a chapter on phonology, as in Chapter 4 of Akmajian *et al.* (1990), Chapter 3 of Finegan & Besnier (1989), Chapter 6 of Fromkin & Rodman (1993) and Chapter 3 of O'Grady *et al.* (1993). Many of the other texts listed at the end of the previous chapter, on phonetics, also deal with phonology, particularly Clark & Yallop (1995), Ladefoged (1992) and Roach (1991). For a useful set of basic exercises in phonology, see Halle &

Clements (1983). Well-established reference works are Chomsky & Halle (1991), Hyman (1975) and Lass (1984). A different perspective is offered in Hogg & McCully (1987). Other relatively accessible treatments are in Carr (1993), Halle (1990), Hawkins (1984), Katamba (1989) and Maddieson (1984). Edwards & Shriberg (1983) and Lowe (1994) address phonological issues in speech therapy. Specifically focusing on the phonology of English are Giegerich (1992), Jensen (1993), Lodge (1984) and Wolfram & Johnson (1982). Brown (1990) presents an extended treatment of assimilation and elision. Historical perspectives are provided by Anderson (1985) and Jones (1989). Other theoretical treatments are available in Goldsmith (1994), Harris (1994) and Kaye (1988).

7 Words and word-formation processes

When, in about 1820, a congressman named Felix Walker was accused of speaking drivel - which, evidently, he was – he replied that he was speaking to the people of Buncombe County, North Carolina, his district. Almost immediately his congressional colleagues began referring to any political claptrap or bombast as *speaking to Buncombe*. Soon the phrase had spread beyond Washington and was being abbreviated to *buncombe*, often respelled *bunkum*, and eventually further contracted to *bunk*. *Debunk* did not come until 1927. *Bunkum* in turn begat *hokum* – a blend of *hocus* and *bunkum*. Thus with a single fatuous utterance, the forgotten Felix Walker managed to inspire half a page of dictionary entries.

Bill Bryson (1994)

Around 1900, in New Berlin, Ohio, a department-store worker named J. Murray Spangler invented a device which he called an *electric suction sweeper*. This device eventually became very popular and could have been known as a *spangler*. People could have been *spanglering* their floors or they might even have *spanglered* their rugs and curtains. The use could have extended to a type of person who droned on and on (and really sucked), described as *spanglerish*, or to a whole style of behavior called *spanglerism*. However, none of that happened. Instead, Mr Spangler sold his new invention to a local businessman called William H. Hoover whose Hoover Suction Sweeper Company produced the first 'Hoover'. Not only did the word *hoover* (without a capital letter) become familiar all over the world, but in Britain, people still talk about *hoovering* (and not *spanglering*) their carpets.

The point of this small tale is that, although you had never heard of Mr. Spangler before, you really had no difficulty coping with the new words: *spangler*, *spanglering*, *spanglered*, *spanglerish* or *spanglerism*. That is, you can very quickly understand a new word in your language (a **neologism**) and accept the use of different forms of that new word. This ability must derive in part from the fact that there is a lot of regularity in the word-formation processes in your language. In this chapter, we shall explore some of those basic processes by which new terms are created.

Word-formation processes

In some respects, the study of the processes whereby new words come into being in a language like English seems relatively straightforward. This apparent simplicity, however, masks a number of controversial issues, some of which we shall consider in the following chapter. Despite the disagreements among scholars in this area, there do seem to be some regular processes involved, and in the following sections we shall cover the technical terms used to describe those processes and identify examples currently in use which are the result of those processes.

It should be remembered that these processes have been at work in the language for some time and many words in daily use today were, at one time, considered barbaric misuses of the language. It is difficult now to understand the views expressed in the early nineteenth century over the "tasteless innovation" of a word like *handbook*, or the horror expressed by a London newspaper in 1909 over the use of the newly coined word *aviation*. Yet many terms of recent currency cause similar outcries. Rather than act as if the language is being debased, we might prefer to view the constant evolution of new terms and new uses of old terms as a reassuring sign of vitality and creativeness in the way a language is shaped by the needs of its users. Let us consider the ways.

Coinage

One of the least common processes of word-formation in English is **coinage**, that is, the invention of totally new terms. The most typical sources are invented trade names for one company's product which become general terms (without initial capital letters) for any version of that product. Older examples are *aspirin*, *nylon* and *zipper*; more recent examples are *kleenex*, *teflon* and *xerox*. It may be that there is an obscure technical origin (e.g. *te*(tra)-*fl*(uor)-*on*) for such invented terms, but after their first coinage, they tend to become everyday words in the language.

Borrowing

One of the most common sources of new words in English is the process simply labeled **borrowing**, that is, the taking over of words from other languages. Throughout its history, the English language has adopted a vast number of loan-words from other languages, including *alcohol* (Arabic), *boss* (Dutch), *croissant* (French), *lilac* (Persian), *piano* (Italian), *pretzel* (German), *robot* (Czech), *tycoon* (Japanese), *yogurt* (Turkish) and *zebra* (Bantu). Other languages, of course, borrow terms from English, as can be observed in the Japanese use of *suupaamaaketto* ('supermarket') and *rajio* ('radio'), or Hungarians talking about *sport*, *klub* and *futbal*, or the French discussing problems of *le stress*, over a glass of *le whisky*, during *le weekend*.

A special type of borrowing is described as **loan-translation**, or **calque**. In this process, there is a direct translation of the elements of a word into the borrowing language. An interesting example is the French term *un gratte-ciel*, which literally translates as 'a scrape-sky', or the German *Wolkenkratzer* ('cloud scraper'), both of which were used for what, in English, is normally referred to as a *skyscraper*. The English word *superman* is thought to be a loan-translation of the German *Übermensch*, and the term *loan-word* itself is believed to have come from the German *Lehnwort*. Nowadays, some Spanish speakers eat *perros calientes* (literally 'dogs hot'), or *hot dogs*. The American concept of 'boyfriend' was a borrowing, with sound modification, into Japanese as *boyifurendo*, but as a calque into Chinese as 'male friend' or *nan pengyu*.

Compounding

In some of those examples we have just considered, there is a joining of two separate words to produce a single form. Thus, *Lehn* and *Wort* are combined to produce *Lehnwort* in German. This combining process, technically known as **compounding**, is very common in languages like German and English, but much less common in languages like French and Spanish. Obvious English examples would be *bookcase*, *fingerprint*, *sunburn*, *wallpaper*, *doorknob*, *textbook*, *wastebasket* and *waterbed*.

This very productive source of new terms has been well documented in English and German, but can also be found in totally unrelated languages, such as Hmong, in South East Asia, which combines *hwj* ('pot') and *kais* ('spout') to produce *hwjkais* ('kettle'). The forms *pajkws* ('flower' + 'corn' = 'popcorn') and *hnabloojtes* ('bag' + 'cover' + 'hand' = 'glove') are recent creations.

Blending

This combining of two separate forms to produce a single new term is also present in the process called **blending**. However, blending is typically accomplished by taking only the beginning of one word and joining it to the end of the other word. In some parts of the United States, there's a product which is used like *gasoline*, but is made from *alcohol*, so the 'blended' term for referring to this product is *gasohol*. If you wish to refer to the combined effects of *smoke* and *fog*, there's the term *smog*. In places where they have a lot of this stuff, they can jokingly make a distinction between *smog*, *smaze* (smoke + haze) and *smurk* (smoke + murk). Some other commonly used examples of blending are *bit* (binary/digit), *brunch* (breakfast/lunch), *motel* (motor/hotel), *telecast* (television/broadcast) and the *Chunnel* (Channel/tunnel), connecting England and France.

The recent phenomenon of fund-raising on television that feels like a marathon is typically called a *telethon* and if you are excessively crazy about video, you may be called a *videot*. *Infotainment* (information/entertainment) and *simulcast* (simultaneous/broadcast) are also new blends from life with television. To describe the mixing of languages, people refer to *Franglais* (French/English) and *Spanglish* (Spanish/English). In order to send information fast, you may use a *telex* (teleprinter/exchange) or, via computer, a *modem* (modulator/demodulator), or you may decide to send a *fax*. But that's not a blend. It's an example of our next category.

Clipping

The element of reduction which is noticeable in blending is even more apparent in the process described as **clipping**. This occurs when a word of more than one syllable (*facsimile*) is reduced to a shorter form (*fax*), often in casual speech. The term *gasoline* is still in use, but occurs much less frequently than *gas*, the clipped form. Common examples are *ad* ('advertisement'), *bra* ('brassiere'), *cab* ('cabriolet'), *condo* ('condominium'), *fan* ('fanatic'), *flu*, *perm*, *phone*, *plane*, *pram*, *pub* and *sitcom* ('situation comedy'). English speakers also like to clip each other's names, as in *Al*, *Ed*, *Liz*, *Mike*, *Ron*, *Sam*, *Sue* and *Tom*.

There must be something about educational environments that encourages clipping because just about every word gets reduced, as in *chem*, *exam*, *gym*, *lab*, *math*, *phys-ed*, *poly-sci*, *prof* and *typo*.

Backformation

A very specialized type of reduction process is known as **backformation**. Typically, a word of one type (usually a noun) is reduced to form another word of a different type (usually a verb). A good example of backformation is the process whereby the noun *television* first came into use and then the verb *televise* was created from it. Other examples of words created by this process are: *donate* (from 'donation'), *opt* (from 'option'), *emote* (from 'emotion'), *enthuse* (from 'enthusiasm'), *liaise* (from 'liaison') and *babysit* (from 'babysitter'). Indeed, if you *backform* anything, you have used a backformation.

One very regular source of backformed verbs in English is based on the pattern: *worker – work*. The assumption seems to have been that if there is a noun ending in *-er* (or something close in sound), then we can create a verb for what that noun-*er* does. Hence, an *editor* must *edit*, a *sculptor* must *sculpt* and *burglars*, *peddlers* and *swindlers* must *burgle*, *peddle* and *swindle*.

A particular type of backformation, favored in Australian and British English, produces forms technically known as **hypocorisms**. First, a longer word is reduced to a single syllable, then *-y* or *-ie* is added to the end. Perhaps the most familiar versions of this process are the words *movie* ('moving pictures') and *telly* ('television'). It has also produced *Aussie* ('Australian'), *barbie* ('barbecue'), *bookie* ('bookmaker'), *brekky* ('breakfast') and *hankie* ('handkerchief'). You can probably guess what *Chrissy pressies* are.

Conversion

A change in the function of a word, as, for example, when a noun comes to be used as a verb (without any reduction), is generally known as **conversion**. Other labels for this very common process are 'category change' and 'functional shift'. A number of nouns, such as *paper*, *butter*, *bottle*, *vacation*, can, via the process of conversion, come to be used as verbs, as in the following sentences: *He's papering the bedroom walls*; *Have you buttered the toast?*; *We bottled the home-brew last night*; *They're vacationing in France*. These conversions are readily accepted, but some examples, such as the noun *impact* being used as a verb, seem to *impact* some people's sensibilities rather negatively.

The conversion process is particularly productive in modern English, with new uses occurring frequently. The conversion can involve verbs becoming nouns, with *guess*, *must* and *spy* as the sources of *a guess*, *a must* and *a spy*. Phrasal verbs (*to print out*, *to take over*) also become nouns (*a printout*, *a takeover*) One complex verb combination (*want to be*) has become a very useful noun as in *He isn't in the group, he's just a wannabe.*

Verbs (*see through*, *stand up*) also become adjectives, as in *see-through material* or *a stand-up comedian*. Or adjectives, such as *dirty*, *empty*, *total*, *crazy* and *nasty*, can become the verbs *to dirty*, *to empty*, *to total*, or the nouns *a crazy* and *a nasty*. You may even hear of people *doing the nasty*.

Some compound nouns have assumed adjectival or verbal functions, exemplified by *the ball park* appearing in a *ball-park figure* or asking some-one to *ball-park an estimate of the cost*. Other nouns of this type are *carpool*, *mastermind*, *microwave* and *quarterback*, which are all regularly used as verbs. Other forms, such as *up* and *down*, can also become verbs, as in *They up the prices* or *We down a few brews*.

It is worth noting that some converted forms shift substantially in mean-ing when they change category. The verb *to doctor* often has a negative sense, not normally associated with the source noun *a doctor*. A similar kind of reanalysis of meaning is taking place with respect to the noun *total* and the verb *run around*, which do not have negative meanings. However, after conversion, if you *total* your car (= verb), and your insurance company gives you *the runaround* (= noun), then you will have a double sense of the negative.

Acronyms

Some new words, known as **acronyms**, are formed from the initial letters of a set of other words. These can remain essentially 'alphabetisms' such as *CD* ('compact disk') or *VCR* ('video cassette recorder') where the pronuncia-tion consists of the set of letters. More typically, acronyms are pronounced as single words, as in *NATO*, *NASA* or *UNESCO*. These examples have kept their capital letters, but many acronyms lose their capitals to become every-day terms such as *laser* ('light amplification by stimulated emission of radiation'), *radar* ('radio detecting and ranging'), *scuba* ('self contained underwater breathing apparatus') and *zip* ('zone improvement plan') code. You might even hear talk of a *snafu* which is reputed to have its origins in 'situation normal, all fouled up', though there is some dispute about the f-word in there.

Names for organizations are often designed to have their acronym repre-sent an appropriate term, as in 'mothers against drunk driving' (*MADD*) and 'women against rape' (*WAR*). Some new acronyms come into general use so quickly that many speakers do not think of their component mean-ings. Recent innovations in banking such as the *ATM* ('automatic teller machine') and the required *PIN* ('personal identification number') are regularly heard with one of their elements repeated, as in *I sometimes forget my PIN number when I go to the ATM machine.*

Derivation

In our list so far, we have not dealt with what is by far the most common word-formation process to be found in the production of new English words. This process is called **derivation**, and it is accomplished by means of a large number of small 'bits' of the English language which are not usually given separate listings in dictionaries. These small 'bits' are called **affixes** and a few examples are the elements *un-*, *mis-*, *pre-*, *-ful*, *-less*, *-ish*, *-ism*, *-ness* which appear in words like *unhappy*, *misrepresent*, *prejudge*, *joyful*, *careless*, *boyish*, *terrorism* and *sadness*.

Prefixes and suffixes

In the preceding group of words, it should be obvious that some affixes have to be added to the beginning of a word (e.g. *un-*). These are called **prefixes**. The other affix forms are added to the end of the word (e.g. *-ish*) and are called **suffixes**. All English words formed by this derivational process use either prefixes or suffixes, or both. Thus, *mislead* has a prefix, *disrespectful* has both a prefix and a suffix, and *foolishness* has two suffixes.

Infixes

There is a third type of affix, not normally to be found in English, but fairly common in some other languages. This is called an **infix** and, as the term sug-gests, it is an affix which is incorporated inside another word. It is possible to see the general principle at work in certain expressions, occasionally used in fortuitous or aggravating circumstances by emotionally aroused English speakers: *Hallebloodylujah!*, *Absogoddamlutely!* and *Unfuckingbeliev-able!*. In the movie *Wish You Were Here*, the main character expresses her aggravation (at another character's trying to contact her) by screaming *Tell him I've gone to Singabloodypore!* The expletive may even have an infixed element, as in *godtripledammit!*. We could view these 'inserted' forms as a special version of infixing. However, a much better set of examples can be provided from Kamhmu, a language spoken in South East Asia. These examples are taken from Merrifield *et al.* (1962):

('to drill')	*see-srnee*	('a drill')
('to chisel')	*toh-trnoh*	('a chisel')
('to eat with a spoon')	*hiip-hrniip*	('a spoon')
('to tie')	*hoom-hrnoom*	('a thing with which to tie')

It can be seen that there is a regular pattern whereby the infix *-rn* is added to verbs to form corresponding nouns. If this pattern is generally found in the language and you know that the form *krnap* is the Kamhmu word for 'tongs', then you should be able to work out what the corresponding verb 'to grasp with tongs' would be. It is *kap*.

Multiple processes

Although we have concentrated on each of these word-formation processes in isolation, it is possible to trace the operation of more than one process at work in the creation of a particular word. For example, the term *deli* seems to have become a common American English expression via a process of first 'borrowing' *delicatessen* (from German) and then 'clipping' that borrowed form. If you hear someone complain that *problems with the project have snowballed*, the final term can be noted as an example of 'compounding', whereby *snow* and *ball* have been combined to form the noun *snowball*, which has then undergone 'conversion' to be used as a verb. Forms which begin as 'acronyms' can also undergo other processes, as in the use of *lase* as a verb, the result of 'backformation' from *laser*. In the expression, *waspish attitudes*, the form *WASP* ('white Anglo-Saxon Protestant') has lost its capital letters and gained a suffix (*-ish*) in the 'derivation' process.

An acronym that never seems to have had capital letters comes from 'young urban professional', plus the *-ie* suffix, as in hypocorism, to produce the word *yuppie* (first recorded in 1984). The formation of this new word, however, was helped by a quite different process, known simply as **analogy**, whereby words are formed to be similar in some way to existing words. *Yuppie* was made possible as a new word by the earlier existence of *hippie* and the other short-lived analogy *yippie*. The term *yippie* also had an acronym basis ('youth international party'), but was generally used for students protesting the Vietnam war in the United States. One joke has it that *yippies* just grew up to be *yuppies*. And the process continues. Another analogy, with the word *yap* ('making shrill noises'), has recently helped label some of those noisy young professionals as *yappies*.

Many such forms can, of course, have a very brief life-span. Perhaps the generally accepted test of the 'arrival' of recently formed words in a language is their published appearance in a dictionary. However, even this may not occur without protests from some, as Noah Webster found when his first dictionary, published in 1806, was criticized for citing words like *advocate* and *test* as verbs, and for including such 'vulgar' words as *advisory* and *presidential*. It would seem that Noah had a keener sense than his critics of which new word-forms in the language were going to last.

Study questions

1 Which of the following expressions is an example of 'calque'? How would you describe the others?

(a) *luna de miel* (Spanish) – *honeymoon* (English)
(b) *mishin* (Japanese) – *machine* (English)
(c) *tréning* (Hungarian) – *training* (English)

2 The term *vaseline* was originally created as a trade name for a product, but has become an ordinary English word. What is the technical term used to describe this process?

3 Identify the affixes used in the words *unfaithful*, *carelessness*, *refillable* and *disagree*, and decide whether they are prefixes or suffixes.

4 Can you identify the word-formation processes involved in producing the italicized forms in these sentences?

(a) Laura *parties* every Saturday night.
(b) Tom was worried that he might have *AIDS*.
(c) Zee described the new toy as *fantabulous*.
(d) Eliza exclaimed, "*Absobloominglutely!*"

5 More than one process was involved in the creation of each of the forms indicated below. Can you identify them?

(a) I just got a new *car-phone*.
(b) Shiel wants to be a *footballer*.
(c) The negotiators *blueprinted* a new peace proposal.
(d) Another *carjacking* has just been reported.

Discussion topics/projects

A The compound word *birdcage* is formed from a noun *bird* plus another noun *cage*, while the word *widespread* is formed from an adjective *wide* and a verb *spread*. So, compounds differ in terms of the types of elements which are combined. Can you identify the different elements involved in each of the following compounds?
bedroom, *blackbird*, *brainwash*, *catfish*, *clean-shaven*, *crybaby*, *haircut*, *heartbeat*, *hothouse*, *house-sit*, *hovercraft*, *leadfree*, *madman*, *ready-made*, *seasick*, *sunflower*, *sunrise*, *telltale*, *threadbare*, *watchdog*, *well-dressed*, *wetsuit*

B The work of Bruce Downing and Judy Fuller (at the University of Minnesota) in a study of the language of Hmong refugees now living in the United States has produced some interesting examples of new word-formations designed to cope with new objects and experiences. If you are given the translation equivalents of some Hmong terms, can you work out the English equivalents of the Hmong compounds which follow?

kws ('artisan'); *kev* ('way'); *ntaus* ('hit', 'mark'); *ntoo* ('tree'); *nqaj* ('rail'); *ntawv* ('paper'); *niam* ('mother'); *hlau* ('iron'); *tshuaj* ('medicine'); *tsheb* ('vehicle'); *kho* ('fix'); *hniav* ('teeth'); *mob* ('sick'); *cai* ('right', 'law'); *dav* ('bird', 'hawk'); *daim* ('flat'); *muas* ('buy').

niam hlau ('mother iron' = a magnet)		*kev kho mob*
kws ntawv	*kws ntaus ntawv*	*kws kho tsheb*
kws tshuaj	*kws hlau*	*kev cai*
kev nqaj hlau	*tsheb nqaj hlau*	*kws kho hniav*
kws ntoo	*dav hlau*	*daim ntawv muas tshuaj*

C A number of interesting word-formation processes can be discerned in some of the following examples. Can you identify what is going on in these, and have you come across any comparable examples?

When I'm ill, I want to see a doc, not a vet.
I was a deejay before, but now I emcee in a nightclub.
That's a-whole-nother problem.
The deceased's cremains were scattered over the hill.
He's always taking pills, either uppers or downers

D Only a handful of the English words borrowed from other languages are presented in this chapter

(i) Can you find out, by consulting a dictionary (an etymological dictionary if possible), which of the following words are borrowings and from which languages they came?
advantage, assassin, caravan, cash, child, clinic, cobalt, cockroach, crime, have, laundry, measles, physics, pony, ranch, scatter, slogan, violent, wagon, yacht, zero

(ii) While you have that dictionary, try to discover the source of the following 'eponyms' (words derived from names of people or places):
biro, blurb, boycott, cardigan, denim, diesel, fahrenheit, nicotine, sandwich, saxophone, watt

E In deriving new words via a suffix such as *-able*, there seems to be some constraint on what is permitted. The words in the first column below are 'acceptable' (that's one!) formations, but the forms in the other columns are not. Can you work out what the rule(s) might be for making new adjectives with the suffix *-able*? (Hint: if X is *fixable*, then someone can fix X.)

breakable	*?carable*	*?dieable*
doable	*?chairable*	*?downable*
inflatable	*?deskable*	*?oldable*
movable	*?hairable*	*?redable*
understandable	*?housable*	*?runable*
wearable	*?pencilable*	*?sleepable*

Further reading

There are a number of general treatments of word-formation in English, such as the textbooks by Adams (1973) and Bauer (1983). Espy (1978) and Levi (1978) are other traditional presentations. More technical treatments are offered by Aronoff (1976) and Di Sciullo & Williams (1987). Comprehensive reference works are Marchand (1969) or Quirk *et al.* (1985), Appendix 1, which is mainly based on British English. For American English, the journal *American Speech* regularly carries articles on new word-formations and Algeo (1991) presents a collection from that source. Other collections are Green (1991), Le May *et al.* (1988), Barnhart *et al.* (1990) and Mish (1986). Books with a more social (and entertaining) perspective are Bryson (1994) for American English and Howard (1990) or Hughes (1988) for British English. Carver (1991) presents a historical perspective, Randall (1991) explains the distinctions between many confused terms and Allan & Burridge (1991) explore why some words are considered better than others. Aitchison (1994), Allan (1986), Lipka (1990) and Miller (1991) all include discussions of word-formation, with Allan (1986) illustrating many hypocorisms. McArthur (1992) is a good resource, as are Barnhart (1988) on etymology and Mossman (1933) for acronyms. Volume 3 of Shopen (1985) contains several papers on word-formation processes in different languages. For an exhaustive survey of contemporary examples of conversion, see Clark & Clark (1979) and on infixing in English, see McMillan (1980). A survey of English words in other languages is presented in Viereck & Bald (1986).

8 Morphology

BAMBIFICATION: The mental conversion of flesh and blood living creatures into cartoon characters possessing bourgeois Judeo-Christian attitudes and morals.

Douglas Coupland (1991)

Throughout the preceding chapter, we approached the description of processes involved in word-formation as if the unit called the 'word' was a regular and easily identifiable form. This doesn't seem unreasonable when we look at a text of written English, since the 'words' in the text are, quite obviously, those sets of things marked in black with the bigger spaces separating them. Unfortunately, there are a number of problems with using this observation as the basis of an attempt to describe language in general, and individual linguistic forms in particular.

Morphology

In many languages, what appear to be single forms actually turn out to contain a large number of 'word-like' elements. For example, in Swahili (spoken throughout East Africa), the form *nitakupenda* conveys what, in English, would have to be represented as something like *I will love you*. Now, is the Swahili form a single word? If it is a 'word', then it seems to consist of a number of elements which, in English, turn up as separate 'words'. A very rough correspondence can be presented in the following way:

ni	*-ta*	*-ku*	*-penda*
'I	will	you	love'

It seems as if this Swahili 'word' is rather different from what we think of as an English 'word'. Yet, there clearly is some similarity between the

languages, in that similar elements of the whole message can be found in both. Perhaps a better way of looking at linguistic forms in different languages would be to use this notion of 'elements' in the message, rather than to depend on identifying 'words'. The type of exercise we have just performed is an example of investigating forms in language, generally known as **morphology**. This term, which literally means 'the study of forms', was originally used in biology, but, since the middle of the nineteenth century, has also been used to describe that type of investigation which analyzes all those basic 'elements' which are used in a language. What we have been describing as 'elements' in the form of a linguistic message are more technically known as **morphemes**.

Morphemes

We do not actually have to go to other languages such as Swahili to discover that 'word-forms' may consist of a number of elements. We can recognize that English word-forms such as *talks*, *talker*, *talked* and *talking* must consist of one element *talk*, and a number of other elements such as *-s*, *-er*, *-ed*, *-ing*. All these elements are described as morphemes. The definition of a morpheme is "a minimal unit of meaning or grammatical function". Let's clarify this definition with some examples. We would say that the word *reopened* in the sentence *The police reopened the investigation* consists of three morphemes. One minimal unit of meaning is *open*, another minimal unit of meaning is *re-* (meaning 'again'), and a minimal unit of grammatical function is *-ed* (indicating past tense). The word *tourists* also contains three morphemes. There is one minimal unit of meaning, *tour*, another minimal unit of meaning *-ist* (meaning 'person who does something'), and a minimal unit of grammatical function *-s* (indicating plural).

Free and bound morphemes

From these two examples, we can make a broad distinction between two types of morphemes. There are **free morphemes**, that is, morphemes which can stand by themselves as single words, e.g. *open* and *tour*. There are also **bound morphemes**, that is, those which cannot normally stand alone, but which are typically attached to another form, e.g. *re-*, *-ist*, *-ed*, *-s*. You will recognize this last set as a group of what we have already described in Chapter 7 as affixes. So, all affixes in English are bound morphemes. The free morphemes can be generally considered as the set of separate English word-forms. When they are used with bound morphemes, the basic word-form involved is technically known as the **stem**. For example:

undressed			**carelessness**		
un-	*dress*	*-ed*	*care*	*-less*	*-ness*
prefix	stem	suffix	stem	suffix	suffix
(bound)	(free)	(bound)	(free)	(bound)	(bound)

It should be noted that this type of description is a partial simplification of the morphological facts of English. There are a number of English words in which the element which seems to be the 'stem' is not, in fact, a free morpheme. In words like *receive*, *reduce*, *repeat* we can recognize the bound morpheme *re-*, but the elements *-ceive*, *-duce* and *-peat* are clearly not free morphemes. There is still some disagreement over the proper characterization of these elements and you may encounter a variety of technical terms used to describe them. It may help to work with a simple distinction between those forms like *-ceive* and *-duce* as 'bound stems' and other forms like *dress* and *care* as 'free stems'.

Free morphemes

What we have described as free morphemes fall into two categories. The first category is that set of ordinary nouns, adjectives and verbs which we think of as the words which carry the 'content' of messages we convey. These free morphemes are called **lexical morphemes** and some examples are: *boy*, *man*, *house*, *tiger*, *sad*, *long*, *yellow*, *sincere*, *open*, *look*, *follow*, *break*. We can add new lexical morphemes to the language rather easily, so they are treated as an 'open' class of words.

The other group of free morphemes are called **functional morphemes**. Examples are: *and*, *but*, *when*, *because*, *on*, *near*, *above*, *in*, *the*, *that*, *it*. This set consists largely of the functional words in the language such as conjunctions, prepositions, articles and pronouns. Because we almost never add new functional morphemes to the language, they are described as a 'closed' class of words.

Bound morphemes

The set of affixes which fall into the 'bound' category can also be divided into two types. One type we have already considered in Chapter 7, the **derivational morphemes**. These are used to make new words in the language and are often used to make words of a different grammatical category from the stem. Thus, the addition of the derivational morpheme *-ness* changes the adjective *good* to the noun *goodness*. The noun *care* can become the adjectives *careful* or *careless* via the derivational morphemes *-ful* or *-less*. A list of derivational morphemes will include suffixes such as the *-ish* in *foolish*, the *-ly* in *badly* and the *-ment* in *payment*. It will also include prefixes such as *re-*, *pre-*, *ex-*, *dis-*, *co-*, *un-* and many more.

The second set of bound morphemes contains what are called **inflectional morphemes**. These are not used to produce new words in the English language, but rather to indicate aspects of the grammatical function of a word. Inflectional morphemes are used to show if a word is plural or singular, if it is past tense or not, and if it is a comparative or possessive form. English has only eight inflectional morphemes, illustrated in the following:

*Let me tell you about Jim**'s** two sister**s**.*
*One like**s** to have fun and is always laugh**ing**.*
*The other lik**ed** to study and has always tak**en** things seriously.*
*One is the loud**est** person in the house and the other is quiet**er** than a mouse.*

From these examples, we can see that two of the inflections, *-'s* (possessive) and *-s* (plural) are attached to nouns. There are four attached to verbs, *-s* (3rd person present singular), *-ing* (present participle), *-ed* (past tense) and *-en* (past participle). There are two inflections, *-est* (superlative) and *-er* (comparative) attached to adjectives. Note that, in English, all inflectional morphemes listed here are suffixes.

Noun +	*-'s, -s*
Verb +	*-s, -ing, -ed, -en*
Adjective +	*-est, -er*

There is some variation in the form of these inflectional morphemes, with, for example, the possessive sometimes occurring as *-s'* (*those boys' bags*) and the past participle as *-ed* (*they have finished*).

Derivational versus inflectional

The difference between derivational and inflectional morphemes is worth emphasizing. An inflectional morpheme never changes the grammatical category of a word. For example, both *old* and *older* are adjectives. The *-er* inflection (from Old English *-ra*) simply creates a different version of the adjective. However, a derivational morpheme can change the grammatical category of a word. The verb *teach* becomes the noun *teacher* if we add the derivational morpheme *-er* (from Old English *-ere*). So, the suffix form *-er* can be an inflectional morpheme as part of an adjective and also a distinct derivational morpheme as part of a noun. Just because they (*-er*) look the same doesn't mean they do the same kind of work. In both cases, they are bound morphemes.

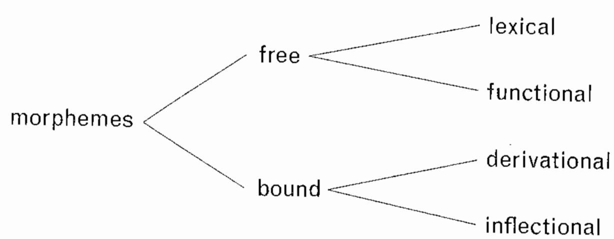
Whenever there is a derivational suffix and an inflectional suffix attached to the same word, they always appear in that order. First the derivational *-er* attaches to *teach*, then the inflectional *-s* is added to yield *teachers*.

Morphological description

Armed with all these terms for the different types of morphemes, you can now take most sentences of English apart and list the 'elements'. As an example, the English sentence *The girl's wildness shocked the teachers* contains the following elements:

The	*girl*	*-'s*	*wild*	*-ness*
(functional)	(lexical)	(inflectional)	(lexical)	(derivational)
shock	*-ed*	*the*		
(lexical)	(inflectional)	(functional)		
teach	*er*	*s*		
(lexical)	(derivational)	(inflectional)		

As a useful way to remember the different categories of morphemes, the following chart can be used:



Problems in morphological description

The rather neat chart presented above conceals a number of outstanding problems in the analysis of English morphology. So far, we have only considered examples of English words in which the different morphemes are easily identifiable as separate elements. Thus, the inflectional morpheme *-s* is added to *cat* and we get the plural *cats*. What is the inflectional morpheme which makes *sheep* the plural of *sheep*, or *men* the plural of *man*? A related question concerns the inflection which makes *went* the past tense of *go*. And yet another question concerns the derivation of an adjective like *legal*. If *-al* is the derivational suffix, as it is in forms like *institutional*, then what is the stem? No, it isn't *leg*.

These problematic issues, and many others which arise in the analysis of different languages, have not been fully resolved by linguists. The solutions to these problems are clearer in some cases than in others. The relationship between *law* and *legal* is a reflection of the historical influence of other languages on English word-forms. The modern form *law* is a result of a borrowing into Old English from Old Norse, over 1,000 years ago. The modern form *legal* is a borrowing from the Latin form *legalis* ('of the law'). Consequently, there is no derivational relationship between the two forms in English, nor between the noun *mouth* (an Old English form) and the adjective *oral* (a Latin borrowing). It has been pointed out that an extremely large number of English forms owe their morphological patterning to languages like Latin and Greek. Consequently, a full description of English morphology will have to take account of both historical influences and the effect of borrowed elements.

Morphs and allomorphs

The solution to other problems remains controversial. One way to treat differences in inflectional morphemes is by proposing variation in morphological realization rules. In order to do this, we draw an analogy with some processes already noted in phonology (Chapter 6). If we consider 'phones' as the actual phonetic realization of 'phonemes', then we can propose **morphs** as the actual forms used to realize morphemes. Thus, the form *cat* is a single morph realizing a lexical morpheme. The form *cats* consists of two morphs, realizing a lexical morpheme and an inflectional morpheme ('plural'). Just as we noted that there were 'allophones' of a particular phoneme, then we can recognize **allomorphs** of a particular morpheme.

Take the morpheme 'plural'. Note that it can be attached to a number of lexical morphemes to produce structures like 'cat + plural', 'sheep + plural', and 'man + plural'. Now, the actual forms of the morphs which result from the single morpheme 'plural' turn out to be different. Yet they are all allomorphs of the one morpheme. It has been suggested, for example, that one allomorph of 'plural' is a zero-morph, and the plural form of *sheep* is actually 'sheep + ø'. Otherwise, those so-called 'irregular' forms of plurals and past tenses in English are described as having individual morphological realization rules. Thus, 'man + plural' or 'go + past', as analyses at the morpheme-level, are realized as *men* and *went* at the morph-level.

Other languages

This type of analytic approach is not without its critics, particularly when applied to other languages. Yet, the absence of a comprehensive analytic system should not discourage us from exploring and describing some of the morphological features of other languages. Some patterns appear to be describable in terms of the basic categories we listed earlier. The first example below is from English and the second is from Aztec:

Stem	Derivational	Inflectional	
DARK	+ *-EN* ('make')	+*-ED* ('past')	= *DARKENED*
MIC ('die')	+ *TIA* ('cause to')	+ *-S* ('future')	= *MICTIAS* ('will kill')

Different patterns occur in other languages. Let's look at some sample data, adapted from examples originally presented in Gleason (1955), and try to work out which morphological features can be identified. The first is from Kanuri, a language spoken in Nigeria.

Kanuri

('excellent')	*karite*	- *nəmkarite*	('excellence')
('big')	*kura*	- *nəmkura*	('bigness')
('small')	*gana*	- *nəmgana*	('smallness')
('bad')	*dibi*	- *nəmdibi*	('badness')

From this set, we can propose that the prefix *nəm-* is a derivational morpheme which can be used to derive nouns from adjectives. Discovering a regular morphological feature of this type will enable us to make certain predictions when we encounter other forms in the language. For example, if the Kanuri word for 'length' is *nəmkurugu*, then we can be reasonably sure that 'long' is *kurugu*.

Different languages also employ different means to produce inflectional marking on forms. Here are some examples from Ganda, a language spoken in Uganda:

Ganda

('doctor')	*omusawo*	- *abasawo*	('doctors')
('woman')	*omukazi*	- *abakazi*	('women')
('girl')	*omuwala*	- *abawala*	('girls')
('heir')	*omusika*	- *abasika*	('heirs')

From this small sample, we can observe that there is an inflectional prefix *omu-*, used with singular nouns, and a different inflectional prefix *aba-*, used with the plural of those nouns. If you are told that *abalenzi* is a Ganda plural, meaning 'boys', you should be able to determine the singular form, meaning 'boy'. It is, of course, *omulenzi*.

The following data from Ilocano, a language of the Philippines, will serve to illustrate a quite different method for marking plurals:

Ilocano

('head')	úlo	- ulúlo	('heads')
('road')	dálan	- daldálan	('roads')
('life')	bíag	- bibíag	('lives')
('plant')	múla	- mulmúla	('plants')

In these examples, there seems to be repetition of the first part of the singular form. When the first part is *bi-* in the singular, the plural begins with this form repeated, *bibi-*. The process involved here is technically known as **reduplication** and several languages use this repetition device as a means of inflectional marking. Having seen how plurals differ from singular forms in Ilocano, you should be able to take this plural form *taltálon* ('fields') and work out what the singular ('field') would be. If you follow the pattern observed, you should get *tálon*.

Finally, here are some intriguing data provided by Lisa Miguel, who speaks Tagalog, another language of the Philippines:

Tagalog

basa ('read')	*tawag* ('call')	*sulat* ('write')
bumasa ('Read!')	*tumawag* ('Call!')	*sumulat* ('Write!')
babasa ('will read')	*tatawag* ('will call')	*susulat* ('will write')

If we assume that the first form in each set is some type of stem, then it appears that in the second member of each set an element *-um-* has been inserted after the first consonant. It must be an example of an infix. In the third member of each set, note that the change in form involves, in each case, a repetition of the first syllable. So, the marking of future reference in Tagalog appears to be accomplished via reduplication. If you know that *lapit* is the verb meaning 'come here' in Tagalog, how would you expect the expressions 'Come here!' and 'will come here' to be realized? How about *lumapit* and *lalapit*? And if you hear *lalakad* ('will walk'), you can guess that *lakad* will translate as 'walk'.

It may have occurred to you as we were exploring all these features of morphology that the discussion often seemed to be connected to what was traditionally called 'grammar'. That would have been an accurate observation and we shall continue the exploration in the following chapter.

Study questions

1 (a) List the 'bound' morphemes to be found in these words: *misleads, previewer, shortened, unhappier, fearlessly.*

(b) In which of the following examples should the 'a' be treated as a bound morpheme: *a boy, apple, atypical, AWOL?*

2 What are the functional morphemes in the following sentence: *The old man sat on a chair and told them tales of woe.*

3 What are the inflectional morphemes in the following phrases:

(a) *the singer's songs* (c) *the newest style*

(b) *it's raining* (d) *the cow jumped over the moon*

4 What would we list as allomorphs of the morpheme 'plural' from this set of English words: *dogs, oxen, deer, judges, curricula?*

5 Provide the equivalent forms, in the languages listed, for the English translations shown on the right below.

Tagalog	'buy'	*bili*	'will buy'
Kanuri	'sweetness'	*nəmkəǰi*	'sweet'
Kamhmu	'an ear ornament'	*srnal*	'to place in earlobe'
Ganda	'twin'	*omuloŋgo*	'twins'
Ilocano	'windows'	*tawtáwa*	'window'
Kamhmu	'a small package'	*trniap*	'to fold a small 'package'
Tagalog	'eat'	*kain*	'Eat!'

Discussion topics/projects

A In the following examples from Turkish (thanks to Feride Erkü), there is some variation in the form of the inflectional morpheme for marking plural.

(i) Can you provide the missing forms in the table?

('man')	*adam*	- *adamlar*	('men')
('gun')	______	- *toplar*	('guns')
('lesson')	*ders*	- ______	('lessons')
('place')	*yer*	- *yerler*	('places')
('road')	______	- *yollar*	('roads')
('lock')	______	- *kilitler*	('locks')
('arrow')	*ok*	- ______	('arrows')
('hand')	*el*	- ______	('hands')
('arm')	*kol*	- ______	('arms')
('bell')	______	- *ziller*	('bells')
('friend')	______	- *dostlar*	('friends')
('apple')	*elma*	- ______	('apples')

(ii) What are the two plural morphs?

(iii) Next, consider *a* and *o* as representing back vowels, while *e* and *i* represent front vowels. Under what conditions are the two different plural morphs used?

(iv) How would you describe the translation equivalents of *your*, and the conditions for their use, on the basis of the following Turkish expressions:

dishin	('your tooth')	*topun*	('your gun')
okun	('your arrow')	*dersin*	('your lesson')
kushun	('your bird')	*kibritlerin*	('your matches')

(v) While English usually marks location with prepositions (i.e. ***in** a house* or ***at** a place*), Turkish has postpositions (i.e. *house-**in*** or *place-**at***). After studying the following examples, you should be able to identify the three versions of the 'location' suffix and the conditions for their use.

('book')	*kitap*	- *kitapta*	('in a book')
('chair')	*koltuk*	- *koltukta*	('in a chair')
('room')	*oda*	- *odada*	('in a room')
('restaurant')	*lokanta*	- *lokantada*	('in a restaurant')
('house')	*ev*	- *evde*	('in a house')
('place')	*yer*	- *yerlerde*	('in places')
('hand')	*el*	- *ellerimde*	('in my hands')
('road')	*yol*	- *yollarta*	('in roads')

For your final task: when Turkish borrowed (from French) the word *randevu*, meaning 'an appointment', how do you think they expressed 'in an appointment'?

B Here are some further examples of Swahili sentences. Can you work out the forms which correspond to the elements in the English translations?

alipita	('she passed by')
alikupiga	('she beat you')
waliondoka	('they left')
nilimlipa	('I paid him')
niliwapika	('I cooked them')
nitakupenda	('I will love you')
utawauza	('you will sell them')
utanipiga	('you will beat me')

nitaondoka	('I will leave')
tuliwapenda	('we loved them')
tutapita	('we will pass by')
watamlipa	('they will pay him')

C Remembering the morphological processes identified in Tagalog, can you extend the analysis to describe the elements and processes involved in the following examples, also from Tagalog?

hanap	('look for')	*sulat*	('write')
hinanap	('was looked for')	*sinulat*	('was written')
humahanap	('is looking for')	*sumusulat*	('is writing')
hinahanap	('is being looked for')	*sinusulat*	('is being written')
basag	('break')	*tawag*	('call')
binasag	('was broken')	*tinawag*	('was called')
bumabasag	('is breaking')	*tumatawag*	('is calling')
binabasag	('is being broken')	*tinatawag*	('is being called')

D In English, the idea of possession can be marked by an inflectional suffix (*-'s*) on the noun representing the 'possessor', placed before the noun that is 'possessed' (as in the English examples below). The other examples below are from a West African language called Basari, spoken in Ghana (data adapted from Jackson, 1985).

(i) Can you describe how 'possession' is expressed in these examples?

('chief') *uboti* ('wife') *unimpu* ('farm') *kusaau*

('man's chief')	*uninja botiu*
('a man's wife')	*uninja nimpuu*
('one wife')	*unimpu ubo*
('this man's one wife')	*uninja-nee nimpuu ubo*
('one farm')	*kusaau kubo*
('a man's farm')	*uninja saaku*
('a man's one farm')	*uninja saaku kubo*
('one man's farm')	*uninja ubo saaku*

(ii) If the Basari word for 'mortar' is *kukuntuu*, then how would you translate *uninja-nee nimpuu kuntuuku*?

E In English, plural forms such as *mice* appear to be treated in a different way from plurals such as *rats*. If you tell people that a place is infested with mice or rats, they will accept the compounds *mice-infested* and *rat-infested*, but not **rats-infested*. This suggests that the forms with the regular plural affix (*-s*) follow a different rule in compounding than irregular plural forms such as *mice*. Can you think of a way to state the rule (or sequence of rules) that will accommodate all the examples given here? (You can get some help from Gordon, 1985, or the summary in chapter 5 of Pinker, 1994). An asterisk (*) before a word means that it is an unacceptable form in the language.

teethmarks	*the feet-cruncher*
clawmarks	*the finger-cruncher*
**clawsmarks*	**the fingers-cruncher*
lice-infested	*a people-mover*
roach-infested	*a dog-mover*
**roaches-infested*	**a dogs-mover*

Further reading

Most introductory linguistics texts have a section on morphology, for example, Akmajian *et al.* (1990), Chapter 2, Fromkin & Rodman (1993), Chapter 2, or O'Grady *et al.* (1993), Chapter 4. An array of interesting exercises involving a wide variety of different languages can be found in Gleason (1955). Other good resources are Cowan & Rakusan (1985) and Jackson (1985). A particularly clear presentation of the relationship between morphemes and morphs is in Brown & Miller (1991) and why the distinction is necessary is covered in Chapter 5 of Lyons (1968). A comprehensive textbook on the subject is Matthews (1991). Other texts (some quite complex) are Bauer (1988), Bybee (1985), Katamba (1993), Jensen (1990) and Spencer (1981). Interest in morphology was much greater in earlier works on language and you might like to go back to Bloomfield (1933) for one approach and then try Hockett (1954; 1958) for another.

9 Phrases and sentences: grammar

A memo at my school warns against the sins most often committed by students against the English language. Watch out for the "gross reference": "Bill hit Fred on the head with the bat. It was hard. It made him mad." Well, I don't know how "gross" this is. If I were Fred I'd be mad too, and I'd jump a few logical steps to get back at that hooligan Bill. Then there is, of course, the ever-threatening "dangling modifier"; "While thinking about Sue, the Honda hit the fence." "John ran to the door and yelled at the dog in his underwear." Being somewhat familiar with how students live nowadays, I can see the dog wearing John's underwear. In fact, I saw it late last night. The memo also warns of the "double negative": "I can't hardly believe it." "I can't get no satisfaction." Come now. Are the Rolling Stones wrong? **Andrei Codrescu (1989)**

We have already considered two levels of description used in the study of language. We have described linguistic expressions as sequences of sounds which can be represented phonetically. For example:

ð̲əlʌk̲ibɔ̲y̲z

voiced fricative — voiceless stop — diphthong

We can take the same linguistic expression and describe it as a sequence of morphemes. For example:

the	luck	y	boy	s
functional	lexical	derivational	lexical	inflectional

With these descriptions, we could characterize all the words of a language in terms of their phonetic and morphological make-up.

Grammar

However, we have not yet accounted for the fact that these words can only be combined in a limited number of patterns. We recognize that the phrase *the lucky boys* is a well-formed piece of English, but that the following two 'phrases' are not at all well-formed:

**boys the lucky* **lucky boys the*

(Beside each of these ill-formed structures there is an asterisk (*), which is a conventional way of indicating that a structure is ill-formed, or ungrammatical.)

So, we need a way of describing the structure of phrases and sentences which will account for all of the grammatical sequences and rule out all the ungrammatical sequences. Providing such an account involves us in the study of **grammar**. We should note that this term is frequently used to cover a number of different phenomena.

Types of grammar

Each adult speaker of a language clearly has some type of 'mental grammar', that is, a form of internal linguistic knowledge which operates in the production and recognition of appropriately structured expressions in that language. This 'grammar' is subconscious and is not the result of any teaching. A second, and quite different, concept of 'grammar' involves what might be considered 'linguistic etiquette', that is, the identification of the 'proper' or 'best' structures to be used in a language. A third view of 'grammar' involves the study and analysis of the structures found in a language, usually with the aim of establishing a description of the grammar of English, for example, as distinct from the grammar of Russian or French or any other language.

Given these three concepts, we can say that, in general, the first may be of most interest to a psychologist, since it deals with what goes on in people's minds, and the second may be of interest to a sociologist, since it has to do with people's social attitudes and values. The third is what occupies many linguists, since the concern is with the nature of language, often independently of the users of the language. The study of grammar, in this narrow sense of the study of the structure of expressions in a language, has a very long tradition.

The parts of speech

You may already be familiar with many of the terms used in a grammatical description, particularly the terms for the parts of speech, as illustrated in this sentence:

The	*lucky*	*boys*	*saw*	*the*	*clowns*	*at*
article	adjective	noun	verb	article	noun	preposition

the	*circus*	*and*	*they*	*cheered*	*loudly*
article	noun	conjunction	pronoun	verb	adverb

Simple definitions of these terms can be presented in the following way:

Nouns are words used to refer to people, objects, creatures, places, qualities, phenomena and abstract ideas as if they were all 'things'.
Adjectives are words used, typically with nouns, to provide more information about the 'things' referred to (*happy* people, *large* objects, *cute* creatures, *stupid* ideas).
Verbs are words used to refer to various kinds of actions (*run*, *jump*) and states (*be*, *seem*) involving the 'things' in events.
Adverbs are words used to provide more information about the actions and events (*slowly*, *suddenly*). Some adverbs (*really*, *very*) are also used with adjectives to modify the information about 'things' (*really* large objects, *very* stupid ideas).
Prepositions are words (*at*, *in*, *on*, *near*, *with*, *without*) used with nouns in phrases providing information about time (*at* five, *in* the morning), place (*on* the table, *near* the window) and other connections (*with* a knife, *without* a thought) involving actions and things.
Pronouns are words (*me*, *they*, *he*, *himself*, *this*, *it*) used in place of noun phrases, typically referring to things already known (*he* likes *himself*, *this* is *it*!)
Conjunctions are words (*and*, *but*, *although*, *if*) used to connect, and indicate relationships between, events and things (we swam *although* it was very cold).

Simple definitions of this type are useful for identifying most forms in a language like English, but they are never completely accurate. A different approach might focus on some other properties of the parts of speech. For example, a noun can be defined as a form that comes after an article (*the*, *a*) and can take inflections for possessive (*-'s*) and plural (*-s*). Of course, not all nouns (e.g. *information*, *mud*) have all these characteristics. Procedures for the structural analysis of the parts of speech are presented later.

Traditional grammar

These terms, used to label the grammatical categories of words in sentences, come from traditional grammar, which has its origins in the description of languages like Classical Latin and Greek. Since there were well-established grammatical descriptions of these older languages, it seemed appropriate to adopt the existing categories from these descriptions and apply them in the analysis of languages like English. After all, Latin and Greek were the languages of scholarship, religion, philosophy and 'knowledge', so the grammar of these languages was taken to be the best grammar.

Traditional categories

In addition to the terms used for the parts of speech, traditional grammatical analysis also gave us a number of other categories, including 'number', 'person', 'tense', 'voice' and 'gender'. These categories can be discussed in isolation, but their role in describing language structure becomes clearer when we consider them in terms of **agreement**. For example, we say that the verb *likes* 'agrees with' the noun *boy* in the sentence *The boy likes his dog*. This agreement is partially based on the category of **number**, that is, whether the noun is singular or plural. It is also based on the category of **person**, which covers the distinctions of first person (involving the speaker), second person (involving the hearer) and third person (involving any others). The different forms of English pronouns are usually described in terms of person and number, in that we have first person singular (*I*), second person singular (*you*), third person singular (*he*, *she*, *it*), first person plural (*we*), and so on. So, in the sentence *The boy likes his dog*, we have a noun *boy*, which is third person singular, and the verb *likes* 'agrees with' the noun.

In addition, the form of the verb must also be described in terms of another category, that of **tense**. In this case, the verb (*likes*) is in the present tense, which is distinguished from the past tense (*liked*). The sentence is also in the **active voice**, with *the boy* doing the liking. An alternative is the **passive voice**, in which the liking is done to *the boy*, as in *The boy is liked by his dog*, or just *The boy is liked*.

Our final category is that of **gender**, which helps us describe the agreement between *boy* and *his* in our example sentence. In English, we have to describe this relationship in terms of **natural gender**, mainly derived from a biological distinction between male and female. The agreement between *boy* and *his* is based on a distinction English makes between reference to male entities (*he*, *his*), female entities (*she*, *her*), and sexless entities, or animals, when the sex of the animal is irrelevant (*it*, *its*).

This type of biological distinction is quite different from the more common distinction found in languages which use **grammatical gender**. In this latter sense, nouns are classified according to their gender class and, typically, articles and adjectives take different forms to 'agree with' the gender of the noun. Spanish, for example, has two grammatical genders, masculine and feminine, illustrated by the expressions *el sol* ('the sun') and *la luna* ('the moon') respectively. German uses three genders, masculine *der Mond* ('the moon'), feminine *die Sonne* ('the sun') and neuter *das Feuer* ('the fire'). Note the different forms of the articles in both the Spanish and German examples, corresponding to differences in the gender class of the nouns. Also note that the gender distinction is not based on a distinction in sex. A young girl is biologically 'female', but the German noun *das Mädchen* is grammatically 'neuter'. The French word *le livre* ('the book') is grammatically masculine, but we would not consider books to be biologically male. So, the grammatical category of gender is very usefully applied in describing a number of languages (including Latin), but may not be as appropriate in describing English.

Traditional analysis

The notion of 'appropriateness' of analytic categories has not always been a consideration. In traditional grammar books, tables such as the following were often presented for English, constructed by analogy with similar tables of forms in Latin grammars. The forms for the Latin verb amare ('to love') are listed on the right.

Present tense, active voice	First person, singular	*I love*	*amo*
	Second person, singular	*you love*	*amas*
	Third person, singular	*he loves*	*amat*
	First person, plural	*we love*	*amamus*
	Second person, plural	*you love*	*amatis*
	Third person, plural	*they love*	*amant*

Note that each of the Latin verb forms is different, according to the categories of person and number, yet the English forms are, with one exception, the same. Thus it makes some sense, in describing a language like Latin, to have all those descriptive categories to characterize verb forms, yet it seems a rather extravagant descriptive system for English. The influence of Latin, however, goes beyond the descriptive labels.

The prescriptive approach

It is one thing to adopt the grammatical labels (e.g. 'noun', 'verb') to categorize words in English sentences; it is quite another thing to go on to claim that the structure of English sentences should be like the structure of sentences in Latin. Yet this was an approach taken by some grammarians, mainly in eighteenth-century England, who set out rules for the correct or 'proper' use of English. This view of grammar as a set of rules for the 'proper' use of a language is still to be found today and may be best characterized as the **prescriptive approach**. Some familiar examples of prescriptive rules for English sentences are as follows:

(1) You must not split an infinitive.
(2) You must not end a sentence with a preposition.

There are, of course, many such rules which generations of English teachers have attempted to instill in their pupils via corrections, as when the sentence *Mary runs faster than me* is 'corrected' to read *Mary runs faster than I*. And *Who did you see?* is 'corrected' to *Whom did you see?* And never begin a sentence with *and*.

It may, in fact, be a valuable part of one's education to be made aware of this 'linguistic etiquette', or the 'proper' use of the language. If it is a social expectation that someone who writes well should obey these prescriptive rules, then social judgments such as "poorly educated" may be made about someone who does not follow these rules.

However, it is worth considering the probable origins of these rules and asking whether they are appropriately applied to the English language. Let us take one example: "You must not split an infinitive."

Captain Kirk's infinitive

The infinitive in English has the form *to* + the verb, as in *to go*, and can be used with an adverb such as *boldly*. So, at the beginning of each televised 'Star Trek' episode, Captain Kirk used the expression *To boldly go* .... This is an example of a split infinitive. Captain Kirk's English teacher should have taught him to say *To go boldly* or *Boldly to go*. If Captain Kirk had been a Roman space traveler, speaking Latin, he would have used the expressions *ire* ('to go') and *audacter* ('boldly'). Now, in saying *Ire audacter* ... in Latin, Captain Kirkus would not even have the opportunity to split his infinitive (*ire*), because Latin infinitives are single words and just do not split.

So, it would be very appropriate in Latin grammar to say that you cannot split an infinitive. But is it appropriate to carry this idea over into English, where the infinitive does not consist of a single word, but of two words, *to* and *go*? If it is a typical feature of the use of English that speakers and writers do produce forms such as *to boldly go* or *to solemnly swear*, then we may wish to say that there are structures in English which differ from those found in Latin, rather than to say that the English forms are 'bad" because they are breaking a supposed rule of Latin grammar.

The descriptive approach

It may be that using a well-established grammatical description of Latin is a useful guide for studying some languages (e.g. Italian or Spanish), is less useful for others (e.g. English), and may be absolutely misleading if you want to describe some non-European languages. This last point became clear to those linguists who wanted to describe the structure of North American Indian languages at the end of the nineteenth century. The categories and rules which were appropriate for Latin grammar just did not seem to fit the Indian languages encountered. As a consequence, throughout the present century, a rather different approach has been taken. Analysts collect samples of the language they are interested in and attempt to describe the regular structures of the language as it is used, not according to some view of how it should be used. This is called the **descriptive approach** and it is the basis of most modern attempts to characterize the structure of different languages.

Structural analysis

One type of descriptive approach is called **structural analysis** and its main concern is to investigate the distribution of forms (e.g. morphemes) in a language. The method employed involves the use of 'test-frames' which can be sentences with empty slots in them. For example:

The __________ makes a lot of noise.
I heard a __________ yesterday.

There are a lot of forms which can fit into these slots to produce good grammatical sentences of English (e.g. *donkey*, *car*, *dog*, *radio*, *child*, etc.). Consequently, we can suggest that because all of these forms fit in the same test-frame, they are likely to be examples of the same grammatical category. The label we give to this grammatical category is, of course, 'noun'. However, there are many forms which do not fit the test-frames above. Examples would be *Cathy*, *it*, *the dog*, *a car*, and so on. For these forms, we require different test-frames, which could be like this:

__________ makes a lot of noise.
I heard __________ yesterday.

Among the forms which fit these test-frames are *Cathy*, *Anna Banana*, *it*, *the dog*, *an old car*, *the professor with the Scottish accent*, and many more. Once again, we can suggest that these forms are likely to be examples of the same grammatical category. The common label for this category is 'noun phrase'. By developing a set of test-frames of this type and discovering what forms fit the slots in the test-frames, you can produce a description of (at least some) aspects of the sentence structures of a language.

Immediate constituent analysis

An approach with the same descriptive aims is called **immediate constituent analysis**. The technique employed in this approach is designed to show how small constituents (or components) in sentences go together to form larger constituents. In the following sentence, we can identify eight constituents (at the word level): *Her father brought a shotgun to the wedding*.

How do those eight constituents go together to form constituents at the phrase level? Does it seem appropriate to put the words together as follows?

brought a *father brought* *shotgun to* *to the*

We don't normally think of these combinations as phrases in English. We are more likely to say that the phrase-like constituents here are combinations of the following types: *Her father*, *a shotgun*, *the wedding*, which are noun phrases; *to the wedding*, which is a prepositional phrase; *brought a shotgun*, which is a verb phrase.

This analysis of the constituent structure of the sentence can be represented in different types of diagrams. One type of diagram simply shows the distribution of the constituents at different levels.

Her	father	brought	a	shotgun	to	the	wedding

This type of diagram can be used to show the types of forms which can substitute for each other at different levels of constituent structure.

Her	father	brought	a	shotgun	to	the	wedding
The	man	saw	the	thief	in	a	car
Fred		took	Jean		to	Honolulu	
He		came			here		

Labeled and bracketed sentences

An alternative type of diagram is designed to show how the constituents in sentence structure can be marked off via labeled brackets. The first step is to put brackets (one on each side) around each constituent, and then more brackets around each combination of constituents. For example:

$$[\,[\,[\text{The}]\,[\text{dog}]\,]\;[\,[\text{followed}]\;[\,[\text{the}]\,[\text{boy}]\,]\,]\,]$$

With this procedure, the different constituents of the sentence are shown at the word level [*the*], at the phrase level [*the boy*], and at the sentence level [*The dog followed the boy*].

We can, of course, label each constituent with grammatical terms such as 'Art' (= article), 'N' (= noun), 'NP' (= noun phrase), 'V' (= verb), 'VP' (= verb phrase) and 'S' (= sentence). In the following diagram, these labels are placed beside each bracket which marks the beginning of a constituent. The result is a labeled and bracketed analysis of the constituent structure of the sentence.

$$[_{S}\ [_{NP}\ [_{Art}\ \text{The}]\ [_{N}\ \text{dog}]\ ]\ [_{VP}\ [_{V}\ \text{followed}]\ [_{NP}\ [_{Art}\ \text{the}]\ [_{N}\ \text{boy}]\ ]\ ]\ ]$$

In performing this type of analysis, we have not only labeled all the constituents, we have exposed the **hierarchical organization** of those constituents. In this hierarchy, the sentence is higher than, and contains, the noun phrase. The noun phrase is higher than, and contains, the noun. We shall return to this concept of the hierarchical organization of grammatical structure in the next chapter.

Before moving on, however, we should note that constituent analysis is not only for the description of English sentences. We can take a sample sentence from a language with a structure quite different from English and apply the same type of analysis.

A Gaelic sentence

Here is a sentence from Scottish Gaelic which would be translated as *The boy saw the black dog*:

Chunnaic	*an*	*gille*	*an*	*cu*	*dubh*
'saw'	'the'	'boy'	'the'	'dog'	'black'

One very obvious difference between the structure of this Gaelic sentence and its English counterpart is the fact that the verb comes first in the sentence. Another noticeable feature is that, when an adjective is used, it follows rather than precedes the noun. We can represent these structural observations in our diagram.

$$[_{S}\ [_{V}\ \text{Chunnaic}]\ [_{NP}\ [_{Art}\ \text{an}]\ [_{N}\ \text{gille}]\ ]\ [_{NP}\ [_{Art}\ \text{an}]\ [_{N}\ \text{cu}]\ [_{Adj}\ \text{dubh}]\ ]\ ]$$

It is not, of course, the aim of this type of analysis that we should be able to draw complicated-looking diagrams in order to impress our friends. The aim is to make explicit, via the diagram, what we believe to be the structure of grammatical sentences in a language. It also enables us to describe clearly how English sentences are put together as combinations of phrases which, in turn, are combinations of words. We can then look at similar descriptions of sentences in other languages, Gaelic, Japanese, Spanish, Arabic, or whatever, and see clearly what structural differences exist. At a very practical level, it may help us understand why a Spanish learner of English produces phrases like **the wine white* (instead of *the white wine*), using a structural organization of constituents which is possible in Spanish, but not in English.

Study questions

1 Give the traditional terms for the grammatical categories of words used in the following sentence (e.g. *boy* = noun): *The boy rubbed the magic lamp and suddenly a genie appeared beside him.*

2 What prescriptive rules for the 'proper' use of English are not obeyed in the following sentences?

(a) *That's the girl I gave my roller skates to.*
(b) *He wanted to simply borrow your car for an hour.*

3 Most modern attempts to characterize the structure of sentences are based on a particular approach. What is this approach called, and what general principle is adhered to in such an approach?

4 Present a labeled and bracketed analysis of this sentence:
The policeman chased a robber.

5 Given the following English translations of some other Gaelic words, can you translate the sentences which follow: *mor* ('big'), *beag* ('small'), *bhuail* ('hit'), *duine* ('man').

(a) *Bhuail an gille beag an cu dubh.*
(b) *Chunnaic an cu an duine mor.*

Discussion topics/projects

A The grammatical category of 'tense' was mentioned briefly in this chapter and a distinction between present and past tense in English was noted. It has been claimed (Palmer, 1983: 193) that English does not have a future tense form of the verb, although it does have many ways of referring to future time. Consider the following sentences and decide what kind of time-reference is involved. Then, consider whether the labels 'past', 'present' and 'future' are appropriate for describing the verb forms used.

Water will freeze at zero degrees Centigrade.
I'll leave if you want.
If Bucky phones, tell him I am asleep.
I wish I had a million dollars.
Your plane leaves at noon tomorrow.
You always listen to the same songs.
We're going to visit Cairo next year.
She said Jim was leaving next week.
Shall we dance?
They were about to leave when I arrived.

B The types of grammatical descriptions we have considered would simply treat the following examples as English sentences and present a description of their form and structural organization. Is this what everyone considers as 'grammar'? Might there be more to say about sentences like these?

(i) *I don't know nothing about that*
(ii) *You wasn't here when he come looking for you*
(iii) *There's hundreds of students in there*
(iv) *Do you wanna go? Are you gonna go?*
(v) *Are y'all coming to see us soon?*
(vi) *That chair's broke, you shouldn't ought to sit on it*
(vii) *I never seen them when they was doing that*
(viii) *If you would have come with, we would have had more fun*

C Can you produce a single diagram, following the format of an immediate constituent analysis, which would incorporate all the constituents of the following sentences? What problems have to be resolved in an exercise like this?

A friend borrowed my car in June. They arrived yesterday.
My parents bought two tickets at Christmas. Suzy left.
We saw that film during the summer. The thief stole it last year.

D Here are some sample sentences from two different languages. The first set is from Latin and the second set is from Amuzgo, a language of Mexico. (The examples used are adapted from data in Merrifield *et al.*, 1962.) Work out the basic constituent structure of the sentences from each language, and then describe them in terms of the phrase level constituents.

(1)	*puellae aquilas portant*	'The girls carry the eagles'
	feminae columbas amant	'The women love the doves'
	puella aquilam salvat	'The girl saves the eagle'
	aquila columbam pugnat	'The eagle fights the dove'
	femina aquilam liberat	'The woman frees the eagle'
(2)	*macei'na tyocho kwi com*	'The boy is reading a book'
	kwil'a yonom kwi w'aa	'The men are building a house'
	nnceihnda yusku kwi com we	'The woman will buy a red book'
	kwil'a yonom ndee meisa	'The men are making three tables'
	macei'na kwi tyocho com t'ma	'A boy is reading the big book'

E The structural analysis of a basic English sentence (NP V NP) is often described as Subject Verb Object or *SVO*. The basic sentence order in Gaelic (V NP NP) is described as Verb Subject Object or *VSO*. After looking at the following examples, how would you describe the basic sentence order in Japanese? Are there any other differences you can note between the structures of English and Japanese?

(1) *Jakku-ga* *gokkoo-e* *ikimasu*
Jack school to go
('Jack goes to school')

(2) *Jakku-ga* *gakkoo-de* *eigo-o* *naratte* *imasu*
Jack school at English learn be
('Jack is learning English at school')

(3) *Jakku-ga* *tegami-o* *kakimasu*
Jack letter write
('Jack writes a letter')

(4) *Jakku-ga* *shinbun-o* *yomimasu*
Jack newspaper read
('Jack reads a newspaper')

(5) *kore-ga* *Jakku-ga* *tateta* *uchi* *desu*
this Jack built house is
('This is the house that Jack built')

Further reading

There are many reference grammars for contemporary English, notably Quirk *et al.* (1985), or in a shorter version, Leech & Svartvik (1994). Others are Alexander (1988), Frank (1993) and Sinclair (1990). Coursebooks on English grammar include Downing & Locke (1992), Givon (1993), and Jacobs (1995). Other frequently consulted texts on grammatical categories are Atkinson *et al.* (1988), Huddleston (1984; 1988), Jespersen (1924), Lyons (1968), Palmer (1983) and Robins (1980). Hurford (1994) is a useful resource on grammatical terminology. A really clear treatment of constituent structure is available in Brown & Miller (1991), which can also be consulted on Gaelic sentence structure. Fuller discussion of the parts of speech can be found in Jackson (1985). For a comprehensive work on grammatical gender, see Corbett (1991) and on grammatical voice, see Klaiman (1991). On prescriptive grammar, see Bolinger (1980), Bryson (1990) and Chapter 15 of Lakoff (1990). Pedagogical grammar, particularly in connection with teaching English as a second language, is represented in Celce-Murcia & Larsen-Freeman (1983) and Rutherford (1987). On the need for a different approach to the grammatical description of North American Indian languages, go back to the introduction in Boas (1911).

10 Syntax

After a lecture on cosmology and the structure of the solar system, William James was accosted by a little old lady who told him that his view of the earth rotating round the sun was wrong.

"I've got a better theory," said the little old lady.

"And what is that, madam?" inquired James politely.

"That we live on a crust of earth which is on the back of a giant turtle."

"If your theory is correct, madam,' he asked, "what does this turtle stand on?"

"You're a very clever man, Mr. James, and that's a very good question," replied the little old lady, "but I have an answer to it. And it's this: the first turtle stands on the back of a second, far larger, turtle, who stands directly under him."

"But what does this second turtle stand on?" persisted James patiently.

To this, the little old lady crowed triumphantly, "It's no use, Mr. James, it's turtles all the way down."

Adapted from J. R. Ross (1967)

In the course of the preceding chapter, we moved from a consideration of general grammatical categories and relations to specific methods of describing the structure of phrases and sentences. If we concentrate on the structure and ordering of components within a sentence, we are studying what is technically known as the **syntax** of a language. The word 'syntax' came originally from Greek and literally meant 'a setting out together' or 'arrangement'. In earlier approaches to the description of syntax, as we saw in Chapter 9, there was an attempt to produce an accurate analysis of the sequence or the ordering 'arrangement' of elements in the linear structure of the sentence. While this remains a major goal of syntactic description,

more recent work in syntax has taken a rather different approach in accounting for the 'arrangements' we observe in the structure of sentences.

Generative grammar

Since the 1950s, particularly developing from the work of the American linguist Noam Chomsky, there have been attempts to produce a particular type of grammar which would have a very explicit system of rules specifying what combinations of basic elements would result in well-formed sentences. (Let us emphasize the word "attempts" here, since no fully worked-out grammar of this or any other type yet exists.) This explicit system of rules, it was proposed, would have much in common with the types of rules found in mathematics. Indeed, a definitive early statement in Chomsky's first major work betrays this essentially mathematical view of language: "I will consider a language to be a set (finite or infinite) of sentences" (Chomsky, 1957: 13).

This mathematical point of view helps to explain the meaning of the term **generative**, which is used to describe this type of grammar. If you have an algebraic expression like $3x + 2y$, and you can give x and y the value of any whole number, then that simple algebraic expression can **generate** an endless set of values, by following the simple rules of arithmetic. When $x = 5$ and $y = 10$, the result is 35. When $x = 2$ and $y = 1$, the result is 8. These results will follow directly from applying the explicit rules. The endless set of such results is 'generated' by the operation of the explicitly formalized rules. If the sentences of a language can be seen as a comparable set, then there must be a set of explicit rules which yield those sentences. Such a set of explicit rules is a **generative grammar**.

Some properties of the grammar

A grammar of this type must have a number of properties, which can be described in the following terms. The grammar will generate all the well-formed syntactic structures (e.g. sentences) of the language and fail to generate any ill-formed structures. This is the 'all and only' criterion (i.e. *all* the grammatical sentences and *only* the grammatical sentences).

The grammar will have a finite (i.e. limited) number of rules, but will be capable of generating an infinite number of well-formed structures. In this way, the productivity of language (i.e. the creation of totally novel, yet grammatical, sentences) would be captured within the grammar.

The rules of this grammar will also need the crucial property of **recursion**, that is, the capacity to be applied more than once in generating a structure. For example, whatever rule yields the component *that chased the cat* in the sentence *This is the dog that chased the cat*, will have to be applied again to get *that killed the rat* and any other similar structure which could continue the sentence *This is the dog that chased the cat that killed the rat* ... .

You can do the same recursive thing with phrases specifying a location, beginning with *The book was on the table*. This sentence tells us where the book was. Where was the table? *Near the window*? Okay, where was the window? *In the hallway*? Okay. Putting this type of recursive effect into a single sentence will lead us to: *The book was on the the table near the window in the hallway beside the* ... . There is, in principle, no end to the recursion which would yield ever-longer versions of this sentence, and the grammar must provide for this fact.

Basically, the grammar will have to capture the fact that a sentence can have another sentence inside it, or a phrase can have another phrase of the same type inside it. (Recursion is not only to be found in descriptions of sentence structure. It is an essential part of the little old lady's view of the role of turtles in cosmic structure, as quoted at the beginning of this chapter.)

This grammar should also be capable of revealing the basis of two other phenomena: first, how some superficially distinct sentences are closely related, and second, how some superficially similar sentences are in fact distinct. We need some exemplification for these points.

Deep and surface structure

Two superficially distinct sentence structures would be, for example, *Charlie broke the window* and *The window was broken by Charlie*. In traditional terminology, the first is an active sentence and the second is passive. The distinction between them, it can be claimed, is a difference in their **surface structure**, that is, the syntactic form they take as actual English sentences. However, this difference in superficial form disguises the fact that the two sentences are very closely related, even identical, at some less 'superficial' level. This other 'underlying' level, where the basic components shared by the two sentences would be represented, has been called their **deep structure**. The deep structure is an abstract level of structural organization in which all the elements determining structural interpretation are represented. So, the grammar must be capable of showing how a single underlying abstract representation can become different surface structures.

Structural ambiguity

On the second point noted above, let us say that we had two distinct deep structures expressing, on the one hand, the fact that 'Annie had an umbrella and she whacked a man with it'; and, on the other hand, that 'Annie whacked a man and the man happened to be carrying an umbrella.' Now, these two different concepts can, in fact, be expressed in the same surface structure form: *Annie whacked a man with an umbrella*. This sentence is structurally ambiguous. It has two different underlying interpretations which would be represented differently in the deep structure.

Groucho Marx knew how to have fun with structural ambiguity. In the film 'Animal Crackers', he first says *One morning I shot an elephant in my pyjamas*, then follows it with *How he got into my pyjamas I'll never know*. In the non-funny interpretation, the structural unit *in my pyjamas* is an addition, attached to the end of the structural unit *I once shot an elephant*. In the alternative (ho, ho) interpretation, the structural unit *an elephant in my pyjamas* is a necessary internal part of a structure that would otherwise be incomplete, *I once shot* ... .

Phrases can also be structurally ambiguous, as when you come across an expression like *old men and women*. The underlying interpretation can be either *old men* plus *old women* or *old men* plus *women* (no age specified). The grammar will have to be capable of showing the structural distinction between these underlying representations.

Different approaches

We have considered some of the requirements which would have to be met by a complete syntactic description of a language. However, this area of linguistic investigation is notorious for giving rise to very different approaches to producing that description. For some, the only relevant issues are syntactic ones, that is, how to describe structure, independently of 'meaning' considerations. For others, the 'meaning component' is primary. In some later versions of generative grammar, the level of deep structure is essentially taken over by a 'meaning' or semantic interpretation which is assigned a structural or syntactic form in its surface realization. (We shall explore 'meaning' issues in Chapters 11 and 12.)

Unfortunately, almost everything involved in the analysis of generative grammar remains controversial. There continue to be many different approaches among those who claim to analyze language in terms of generative grammar, and many more among those who are critical of the whole system. Rather than explore controversies, let us look at some of the really basic features of the original analytic approach and see how it is all supposed to work. First, we need to get the symbols straightened out.

Symbols used in syntactic description

We have already introduced some symbols (in Chapter 9) which are quite easily understood as abbreviations for the grammatical categories involved. Examples are 'S' (= sentence), 'N' (= noun), 'Art' (= article) and so on. We need to introduce three more symbols which are commonly used.

The first of these is in the form of an arrow →, and it can be interpreted as 'consists of'. It will typically occur in the following format:

NP → Art N

This is simply a shorthand way of saying that a noun phrase (e.g. *the book*) consists of an article (*the*) and a noun (*book*).

The second symbol used is in the form of parentheses, or round brackets – ( ). Whatever occurs inside these brackets will be treated as an optional constituent. Perhaps an example will make this clear. You can describe an object as *the book*, or as *the green book*. We can say that both *the book* and *the green book* are examples of the category, noun phrase. In order for a noun phrase to occur in English, you may require an article (*the*) and a noun (*book*), but the inclusion of an adjective (*green*) is optional. You can include an adjective, but it isn't obligatory. We can capture this aspect of English syntax in the following way:

NP → Art (Adj) N

This shorthand notation expresses the idea that a noun phrase consists of an obligatory article and an obligatory noun, but may also include an adjective in a specific position. The adjective is optional.

The third symbol used is in the form of braces, or curly brackets – { }. These indicate that only one of the elements enclosed within the brackets must be selected. They are used when there is a choice from two or more constituents. For example, we have already noted, in Chapter 9, that a noun phrase can consist of an expression like *the woman* (Art N), or *she* (pronoun), or *Cathy* (proper noun). We can, of course, write three single rules, as shown on the left below, but it is more succinct to write one rule, as shown on the right below, which incorporates exactly the same information:

NP → Art N
NP → pronoun
NP → proper noun

$$NP \rightarrow \left\{ \begin{array}{l} \text{Art N} \\ \text{pronoun} \\ \text{proper noun} \end{array} \right\}$$

It is important to remember that, although there are three constituents in these curly brackets, only one of them can be selected on any occasion.

We can now present a list of symbols and abbreviations commonly found in syntactic descriptions:

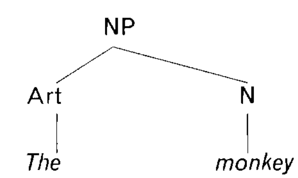
S	sentence	N	noun	Pro	pronoun
PN	proper noun	V	verb	Adj	adjective
Art	article	Adv	adverb	Prep	preposition
NP	noun phrase	VP	verb phrase	PP	prepositional phrase

* = 'ungrammatical sequence'
→ = 'consists of'
( ) = 'optional constituent'
{ } = 'one and only one of these constituents must be selected'

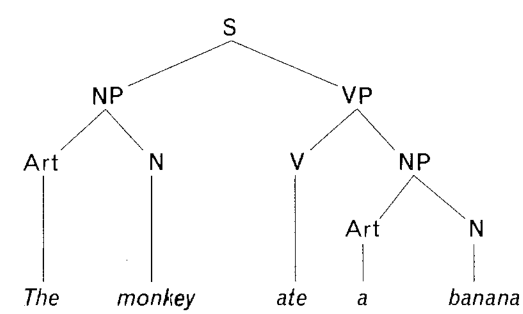
Labeled tree diagrams

In Chapter 9, we considered ways of describing the structure of sentences that (basically) concentrated on the linear sequence of constituents. It is, of course, possible to show the same sequence as, in a more explicit way, 'hierarchically' organized. So, to bring out the hierarchical organization of the labeled and bracketed constituents shown on the left below, we can show the same information in the form of a **tree diagram**, as on the right below:

NP [Art [*The*] N [*monkey*]]



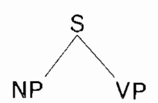
This type of tree-diagram representation contains all the grammatical information found in the other analyses, but also shows more explicitly the fact that there are different levels in the analysis. That is, there is a level of analysis at which a constituent such as NP is represented and a different, lower level at which a constituent such as N is represented. Here's how a whole sentence would look in a tree diagram:



If you start at the top of this tree diagram, you are starting with a sentence (**S**) and then dividing the sentence into two constituents (**NP** and **VP**). In turn, the **NP** constituent is divided into two constituents (**Art** and **N**). Finally, one word is selected which fits the label **Art** (*the*), and another which fits **N** (*monkey*).

Phrase structure rules

We can view this tree-diagram format in two different ways. In one way, we can simply treat it as a static representation of the structure of the sentence at the bottom of the diagram. We could propose that, for every single sentence in English, a tree diagram of this type could be drawn. The alternative view is to treat the diagram as a 'dynamic' format, in the sense that it represents a way of 'generating' not only that one sentence, but a very large number of sentences with similar structures. This alternative view is very appealing since it should enable us to generate a large number of sentences with only a small number of rules. These 'rules' are usually called **phrase structure** rules, and they present the information of the tree diagram in an alternative format. So, instead of the diagram form on the left below, we can use the notation shown on the right below:



$S \rightarrow NP \quad VP$

The rule is then read as: "a sentence consists of a noun phrase followed by a verb phrase". In addition to rules of this type which generate structures, we can also have **lexical** rules which indicate the words to be used for constituents such as N. For example:

N → {*boy, girl, dog,...*}

This means that N is rewritten as *boy*, or *girl*, or *dog*. We can create a set of extremely simple (and necessarily incomplete) phrase structure rules which can be used to generate a large number of English sentences:

$S \rightarrow NP\ VP$

$$NP \rightarrow \left\{ \begin{array}{l} Art\ (Adj)\ N \\ PN \end{array} \right\}$$

$VP \rightarrow V\ NP\ (PP)\ (Adv)$

$PP \rightarrow Prep\ NP$

N → {*boy, girl, dog*}
V → {*saw, followed, helped*}
PN → {*George, Mary*}
Prep → {*with, near*}
Art → {*a, the*}
Adv → {*yesterday, recently*}
Adj → {*small, crazy*}

These rules will generate the grammatical sentences shown below as (1) to (7), but will not yield the ungrammatical sentences shown as (8) to (10):

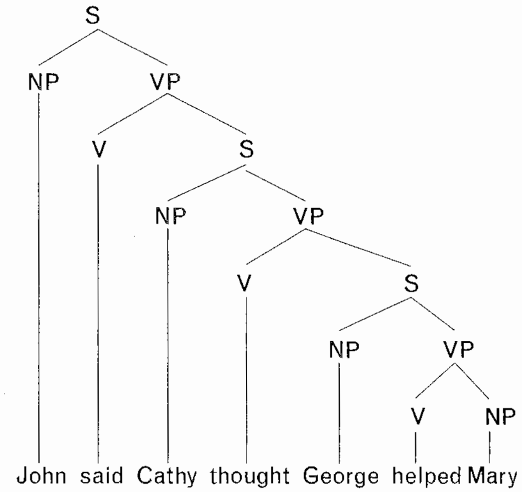
1. The girl followed the boy.
2. A boy helped the dog.
3. The dog saw a girl.
4. Mary helped George recently.
5. George saw a dog yesterday.
6. A small dog followed Mary.
7. The small boy saw George with a crazy dog recently.
8. *Boy the Mary saw.
9. *Helped a girl.
10. *Small dog with girl.

This small set of rules is a good start on creating a phrase structure grammar of English, but we still have not incorporated recursion.

Back to recursion

The phrase structure rules, as presented, have no recursive elements. Each time we rewrote a symbol from the left, we did not include that symbol on the right side of any arrow. We have to be able to repeat some symbols on the right side of the arrow. That is the essence of recursion. We need, for example, to have sentences included within other sentences. We know that *Mary helped George* is a sentence. We also know that *Cathy thought Mary helped George* is a sentence. And, being tediously recursive, we know that *John said Cathy thought Mary helped George* is a sentence.

In order to capture these structures in our rules, we need to add V → {*said, thought*} and PN → {*Cathy, John*} to our lexical rules. We also need to add a crucial recursive rule that says: VP → V S. With these minor additions, we can now represent the structure of a more complex sentence.



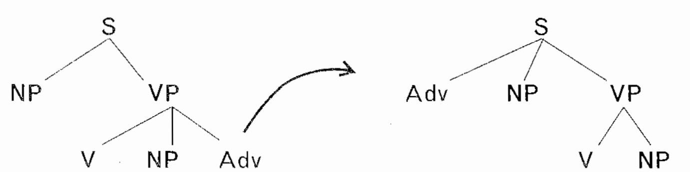
In principle, there is no end to the recursion of sentence structures of this type in the English language and our rule (VP $\rightarrow$ V S) represents that fact.

Transformational rules

One other feature of our phrase structure rules is that they will generate all sentences with fairly fixed word order to the constituents. For example, adverbs will always come at the end of their sentences if we follow the rules we have just illustrated. That is fine for generating the first sentence below, but how would we get the second sentence?

(i) *George helped Mary yesterday.*
(ii) *Yesterday George helped Mary.*

We can think of the *yesterday* element as having been 'moved' to the beginning of the sentence in (ii). In order to do this, we need a set of rules which will change or move constituents in the structures derived from the phrase structure rules. These are called **transformational** rules. Essentially what they do is take a 'branch' of the 'tree' away from one part of the tree diagram, and attach it to a different part. Here is an example of a movement transformation:



(*George helped Mary yesterday*) (*Yesterday George helped Mary*)

We would, of course, specify which constituents can be moved, from where and to where.

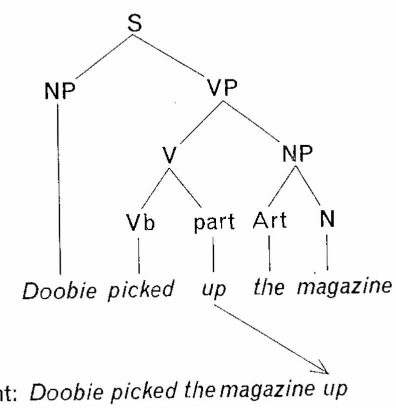
One of the best arguments for having transformational rules involves what seems to be the movement of a very small element in English sentence structure. We recognize that the following two sentences have a great deal in common:

(i) *Doobie picked up the magazine.*
(ii) *Doobie picked the magazine up.*

These sentences contain a verb-particle construction (verb = *pick*; particle = *up*) which can be symbolized as: V $\rightarrow$ Vb part. It is clear that the particle can be separated from the verb and 'moved' to the end of the sentence. A constituent structure analysis, as described in Chapter 9, would have some difficulty accommodating this type of structure. A phrase structure analysis would have to create two distinct tree diagrams. Yet, we intuitively recognize that these two sentences must come from a single underlying source.

Let us propose a single tree diagram source which produces a string of elements like: *NP Verb Particle NP*. Under circumstances like these, let us then propose the optional transformation called 'Particle Movement',

Phrase Structure Tree



Particle Movement:

which takes that structural description and yields the structural change to: *NP Verb NP Particle*.

By using this simple transformational rule, we have provided the means for explicitly relating the two structures in sentences (i) and (ii) above as 'surface' variations of a single underlying structure. It may not seem much, but this type of transformational analysis solved a number of tricky problems for previous syntactic descriptions.

There is, of course, much more involved in generative syntax and other methods of syntactic description. (We have barely scratched the surface structures.) However, having explored some of the basic issues in the syntactic description of language, we must move on, as historically the generative linguists had to do, to come to terms with the place of 'meaning' in linguistic description. This leads us, in the following chapter, to a consideration of the role of semantics.

Study questions

1 In what ways are these expressions 'ambiguous'?

(a) *An American history teacher.*
(b) *Flying planes can be dangerous.*
(c) *The parents of the bride and the groom were waiting.*

2 Can you provide four, 'superficially distinct' sentences which would each have the same 'underlying' structure as one of the following sentences?

(a) *Lara was arrested by the police.*
(b) *She took her coat off.*
(c) *Someone stole my bicycle.*
(d) *I told him to turn down the volume.*

3 Which of the following expressions would be generated by this phrase structure rule: NP → Art (Adj) N?

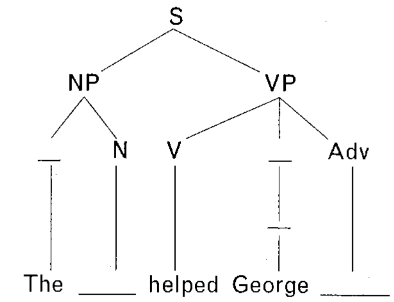
(a) *a radio* (c) *a new student*
(b) *the rusty car* (d) *the screwdriver*

4 Which of the following structures can be changed via the Particle Movement transformation?

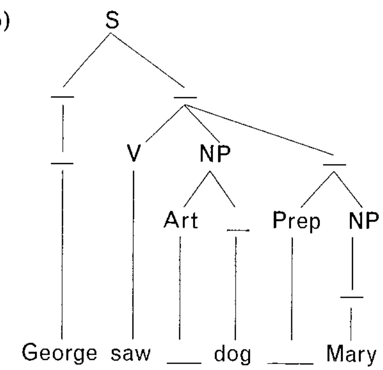
(a) *He put down his glass.*
(b) *She threw away her dress.*
(c) *He pulled off his shirt.*
(d) *They jumped in the pool.*

5 Using the phrase structure rules presented in this chapter, you should be able to complete these labeled tree diagrams.

(a)



(b)



Discussion topics/projects

A Here are some simple phrase structure rules for Scottish Gaelic:

S → V NP NP

NP → { Art N (Adj) / PN }

Art → *an*
N → {*cu*, *gille*}
PN → {*Tearlach*, *Calum*}
Adj → {*beag*, *mor*}
V → {*chunnaic*, *bhuail*}

Only two of the following sentences would be considered well-formed, according to the rules above. First, identify the ill-formed sentences, using the symbol *, then provide labeled tree diagrams for the two well-formed sentences.

1 *Calum chunnaic an gille*
2 *Bhuail an gille mor an cu*
3 *Bhuail an beag cu*
4 *Chunnaic Tearlach an gille*

B Here is a simplified set of phrase structure rules for a language called Ewe, spoken in parts of West Africa. Can you use these rules to write out four different well-formed Ewe sentences? (Note that the syntax of Ewe is different from the syntax of English.)

S → NP VP	N → {*oge, ika, amu*}
NP → N (Art)	Art → *ye*
VP → V NP	V → {*xa, vo*}

C In the chapter we considered one transformational rule which was used for particle movement. Here is a simple version of the passive transformation rule:

Structural Description: NP_1 V NP_2
Structural Change: ⇒ NP_2 *be* V-*ed by* NP_1

(Let us add that the tense, past or present, of the verb (V) in the structural description will determine the tense of the verb (*be*) in the structural change. Also, the symbol ⇒ is used for transformations.) This transformational rule will produce passive versions of a number of the following sentences. First, identify those sentences for which the rule works, and then try to describe what prevents the rule from working on the other sentences.

(1) The cats chased the mouse
(2) Snow White kissed Grumpy
(3) He loves them
(4) Betsy borrowed some money from Jim
(5) The team played badly
(6) The tree fell with a crash
(7) The bank manager laughed
(8) The duckling became a swan

D Each of the following sentences ends with what is called a 'tag question'. For this set of sentences, the process of forming the tag question seems very regular.
(i) Can you produce a simple transformational rule which could be used to add tag questions to basic sentence structures?

(1) She was a dancer, wasn't she?
(2) Zee is a good swimmer, isn't he?
(3) You are ready, aren't you?
(4) They can come, can't they?
(5) Maghna would help, wouldn't she?
(6) You have eaten, haven't you?

(ii) Now, how would you go about making that transformational rule apply for these two sentences?

(7) He smokes a lot, doesn't he?
(8) They arrived early, didn't they?

E There is a principle of syntax called 'structure dependency' that is often used to show that the rules of language depend on hierarchical structure and not on linear position. For example, a young child learning English might be tempted to think that questions of the type in (1) and (2) are formed by moving the second word of the statement to the first position of the question.

(1) a. *Scruffy is tired* (2) a. *You will help him*
b. *Is Scruffy tired?* b. *Will you help him?*

Using the sentences in (1)–(6), try to decide if this 'linear position' rule would be good for forming these English questions and, if not, what would be a better rule?

(3) *Are the exercises in this book too easy?*
(4) *Is the cat that is missing called Blackie?*
(5) *Will the cost of the new book you've ordered be outrageous?*
(6) *Was the guy who scored the winning goal in the final playing for love or money?*

Further reading

All introductory textbooks in linguistics have a section on syntax: try Chapter 3 of Fromkin & Rodman (1993), Chapter 5 of Akmajian *et al.* (1990), or Chapter 5 of O'Grady *et al.* (1993). Lasnik (1990) is a brief review. For more detailed introductory treatments, try Brown & Miller (1991) or Morenberg (1991). Among the other (more advanced) texts on syntax are Borsley (1991), Burton-Roberts (1986), Matthews (1981), McCawley (1988), Sells (1985) and Wekker & Haegeman (1985). A useful reference work is Stockwell *et al.* (1973). For a more functional approach to syntax, see Givon (1990). A good overview of Chomsky's early work is Lyons (1991), or, for selections from Chomsky's early publications, try Allen & van Buren (1971). The basic original works are Chomsky (1957; 1965), with Chomsky (1988) providing a fairly accessible account of his later views. On matters transformational, try Akmajian & Heny (1975), Huddleston (1976) or the widely used textbook by Radford (1988). More recent work in generative syntax is presented in Cook (1988), Freidin (1992), Riemsdijk & Williams (1986) and Haegeman (1991).