CAMBRIDGE TEXTBOOKS IN LINGUISTICS

Psycholinguistics

Michael Garman

PSYCHOLINGUISTICS

MICHAEL GARMAN

DEPARTMENT OF LINGUISTIC SCIENCE
UNIVERSITY OF READING

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 1990
Reprinted 1991, 1994, 1996, 2000

Printed in the United Kingdom at the University Press, Cambridge

British Library Cataloguing in Publication Data
Garman, Michael
Psycholinguistics – (Cambridge textbooks in linguistics)
I. Psycholinguistics
I. Title
401′.9

Library of Congress Cataloguing in Publication data
Garman, Michael.
Psycholinguistics / Michael Garman.
p. cm. – (Cambridge textbooks in linguistics)
Includes bibliographical references.
ISBN 0 521 25675 5 (hardback) – ISBN 0 521 27641 1 (paperback)
1. Psycholinguistics. I. Title. II. Series
[DNLM: 1. Psycholinguistics. BF 455 G233p]
P37.G33 1990
401′.9 – dc20 89–18645 CIP

ISBN 0 521 25675 5 hardback
ISBN 0 521 27641 1 paperback

ETA

1
Characteristics of the language signal

1.1 Introduction

1.1.1 *Preview*

The first part of this book (chs. 1 to 3) surveys the elements of psycholinguistics. For this purpose it is convenient to think of most types of observable language behaviour as comprising three levels: (a) the *language signal*, which we shall take to cover all the forms of language expression which are generated and perceived by language users, including writing as well as speech; (b) the *neurophysiological activity* involved both in the first and the next level; (c) the *language system*. While the first two levels relate to physical entities, the third is abstract, and may be implemented even when we are not using palpable language signals at all, as in silent verbal reasoning, contemplation of our language, and general language knowledge.

These three levels define the first three chapters. In this first chapter we shall review the properties of the language signal. Since these derive, at least in part, from the operations of our neurophysiological systems, they can help to determine the limits of functioning of those systems. In the second chapter we shall consider aspects of the neurophysiological systems themselves: since they are involved, in some way, in constituting whatever mechanisms subserve language behaviour, a knowledge of them must be a necessary, if not a sufficient, basis for theories of language processing, which we shall deal with in the second part of the book. Finally, in the third chapter we shall address the issue of the language system: not directly, since this is too controversial a task, but indirectly, by reviewing the sorts of evidence that are to hand for an empirical approach.

We start by reviewing the more important characteristics of the speech signal (section 1.2), both as it is generated by the human articulatory apparatus (1.2.1) and carried through the air (1.2.2) for processing by the auditory system (1.2.3). Then, because a very large part of our everyday contact with language is based on signs which we perceive and produce as writing, we review some properties of writing systems (section 1.3), looking at some of the major types of writing system (1.3.1.–1.3.3) before considering the English system in some detail (1.3.4). Finally, we consider briefly the adaptation of a writing system to tackle the difficult task of representing the characteristics of the speech signal, in however limited a fashion (section 1.4). This helps to remind us that the representation of speech in such a written medium as this book is necessarily indirect and tends to be selective of certain characteristics over others.

1.1.2 ***Language processing***

First, to set the scene for our discussion, let us consider two very simple diagrams of language processing, as set out in figures 1.1 and 1.2. The first is of a quite familiar sort, reproduced here from Denes and Pinson (1963), who refer to it as 'the speech chain'. It shows schematically two heads, one talking to the other, The talking head is listening to (monitoring) its own speech output via a feedback link which is important for maintaining good control of articulatory targets. But the main channel of information flow is from the talking head to the listening head. In this flow three stages, or levels, are distinguished: the *linguistic*, concerned with the formulation of the message; the *physiological*, concerned with the expression/reception of the signal carrying the message; and the *acoustic*, which is distinguished by being the only level which is outside of, and common to, both individuals – the air gap which has to be bridged for the speech chain to be completed between the heads. It is the acoustic level which is placed at the centre of this picture, the central link in the chain. Now consider the second diagram (fig. 1.2). It has just one (schematic) head, which is busy acting as a central language processor, both receiving and sending signals, through two channels, the articulatory–auditory (speech), and the manual–visual (writing). We might refer to

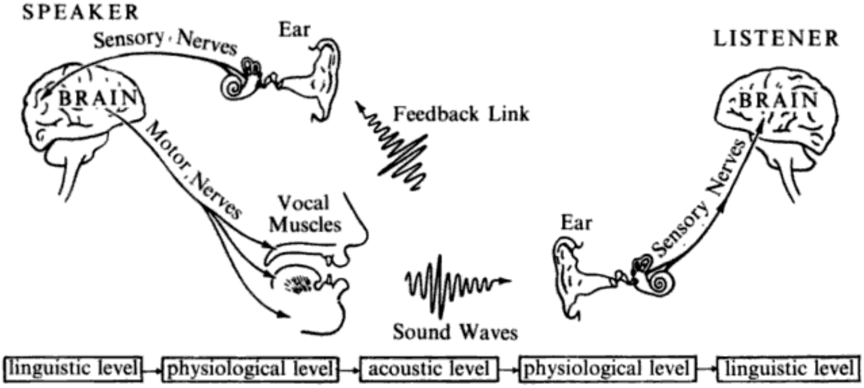


Figure 1.1 The 'speech chain'. (From Denes and Pinson 1963: fig. 1.1, p. 4.)

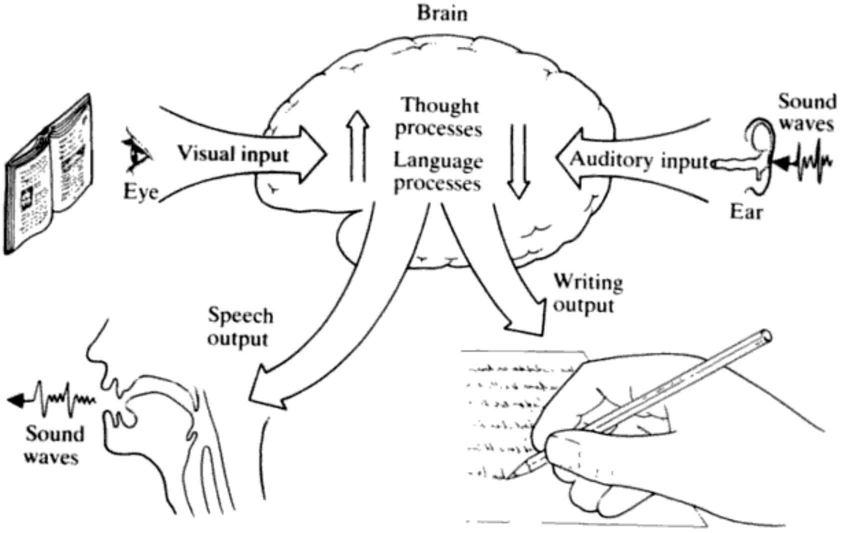


Figure 1.2 The 'language switchboard'

this picture as 'the language switchboard'. At the centre of these input–output events is the brain, capable of reconciling the considerable physical and physiological differences between these events, so that it can recognise and generate the 'same' message in different forms. We tend to take this ability so much for granted that we may not fully appreciate what a task this is that the brain manages so well. Consider, for example, what is going on when someone says something like the following:

(1) I've got the *j, e, l, l, y* (/'dʒeɪ 'i: 'el 'el 'waɪ/) in the fridge

(perhaps in the presence of a child who is about to be given a surprise treat). Notice, first, that such an utterance would not be possible except in a language community that has a way of speaking the written forms of words. Your auditory system perceives these speech signals, and delivers the results of this processing to your brain; at this point, the events are interpreted, not as a sequence of words (thus, for instance /'waɪ/ is not interpreted as the word *why*), but as spoken versions of letters. The appropriate letter-sequence is somehow assembled in such a way that it can be used to address the dictionary of stored written forms of words that you carry in your head, and you find the item *jelly*.

This is just one, albeit quite complex, example of input–output relations in language processing. There are numerous other 'special' examples, including

Table 1.1 *Some input–output relations in language use*

Input	Output
Visual–linguistic reading the word *jelly* on a box in the fridge	Manual–non-linguistic picking the box out
Visual–non-linguistic seeing a jelly in the fridge	Articulatory 'can I have some of this jelly?'
Auditory–linguistic hearing a request to say what you would like for tea	Articulatory 'can I have some jelly?'
Auditory–linguistic hearing a request to find the jelly in the fridge	Manual–non-linguistic picking out the right item
Intentional wishing to label the box containing jelly	Manual–linguistic writing on the lid

the ability to recognise words from letter shapes traced in the palm of the hand; and a host of others that we carry out as part of our everyday activities, such as those set out in table 1.1. Both diagrams in figures 1.1 and 1.2 are attempting to sketch the same thing, which is an impressionistic picture of the framework within which these and other sorts of language processing are carried on. One could easily add a written language dimension to the first diagram, for example, to bring it more into line with the second. But even if we did this, their emphases would be rather different, with the first putting the physical signal (whether acoustic or graphic) at the centre, while the second puts the brain at the centre, and asks us to consider how all the input–output events in language use might be mediated. It is this second picture that this book attempts to deal with.

In so doing, we encounter a problem. If we attend first to the auditory/articulatory systems, say, and then shift our attention to the brain, we are not thereby making the transition from the 'physiological' to the 'linguistic' level, as figure 1.1 implies. The brain is a physiological entity but language is not, so looking for language in the brain is a problematic matter. It is possible, however, to approach this problem from a number of complementary angles:

1. to examine the physiological foundations of language in the brain, as well as in the auditory/articulatory systems, and the visual/manual systems (ch. 2);
2. to consider the basic elements of the language system, as one

source of evidence concerning what sorts of constructs the human language faculty manipulates, together with aspects of human language performance, as revealed in experimental and naturalistic situations (ch. 3);

3. to construct processing models (the second part of the book, chs 4–7);
4. to examine how language processing breaks down in relation to specific brain injury (ch. 8);
5. to approach the issue from the standpoint of formal theories of language.

In this latter way, we can ask how formal properties of language must constrain our theories of the *mind*, and we may then consider the relationship of this construct to what we know of the brain. This approach is the most ambitious one, at least in respect of the role it envisages for linguistic theory in this enterprise, and is beyond the scope of this book. (Aspects of it can be found in Fodor 1983, Bever, Carroll and Miller 1984, and in Chomsky 1986.)

1.1.3 *Language signals*

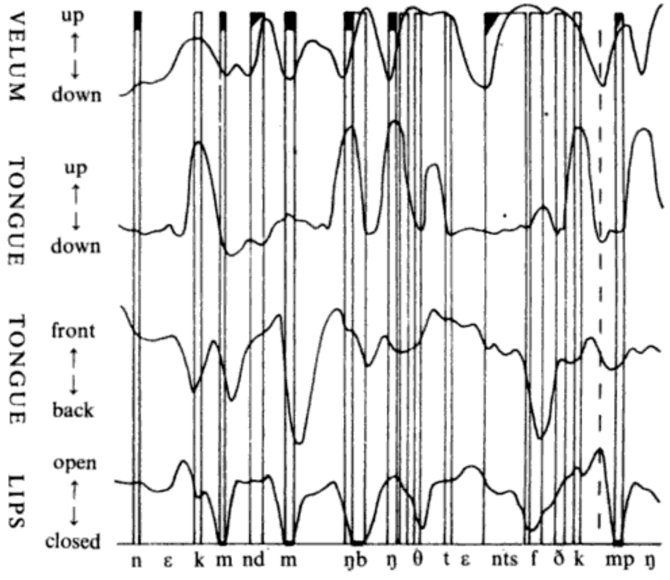
Returning to the main point of this chapter, we should consider how it is that we normally make contact with the signals of language. Apart from the written forms of English, for example, there are other forms such as sign language, semaphore, morse code, braille and so on. Some of these systems are direct encodings of particular forms of a language, as in the case of semaphore, morse code and braille. Among sign languages direct encodings are found in so-called 'finger-spelling', and in aspects of the Paget-Gorman sign system, which is based on English and has signs that signal particular morphemes (e.g. word endings) as well as others that more abstractly represent the vocabulary of English words. In contrast to this, British Sign Language (BSL) and American Sign Language (ASL) are independent of English or any other language (see Woll and Kyle 1983 for detailed discussions of sign language characteristics). In this book we shall be concerned with just the spoken and written forms of language, since speech as a form of language is universal and the written forms are also basic in all the language communities that have them.

Another point which may not be immediately obvious, perhaps, is that within both of these forms of language, spoken and written, distinctions may exist from the point of view of the producer vs the receiver; speaker vs listener, and writer vs reader. For the speaker 'speech' involves the control of the

The object level of description

We are not concerned immediately here with the description of the organs of speech, or how they are controlled; these topics will be addressed in the next chapter. Our purpose at this stage is to consider the signal as it arises from the movements of the articulators.

A good place to start from is the sort of record that is produced from cine-fluorographic film (Kent 1983), as illustrated in figure 1.3. In this technique, very small radio-opaque metallic pieces are attached to crucial articulators in such a way as not to impede their natural performance, and a radiographic film is then taken (usually from one side of the head) on which dynamics of speech show up as traces of the interactive movements of the metallic pieces on the articulatory organs. These can then be plotted as separate traces, all co-ordinated with reference to a single time scale.



Movements of the velum, tongue, and lips recorded by lateral cinefluorography during the sentence 'Next Monday morning bring three tents for the camping trip.' Movements of velum and tongue were recorded as displacements of radiopaque markers attached to these articulators. The segments blacked in at the top are nasal consonants. A partial phonetic transcription at the bottom of the illustration identifies major articulatory events. The dashed vertical line marks one example of nearly simultaneous movements of two articulators, in this case, velum and lips.

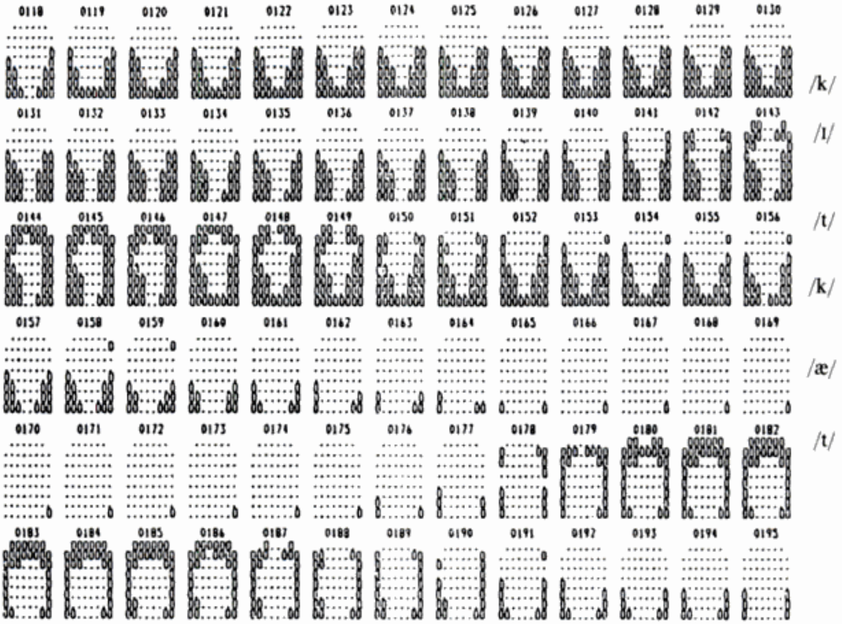
Figure 1.3 Lateral cine-fluorographic record of velum, tongue and lip movements in continuous speech. (From Kent 1983: fig. 4.7, p. 69.)

Most outstandingly, perhaps, in this articulatory display, several *channels* (each representing a particular articulatory organ) are in simultaneous activity. This immediately leads to very great difficulties of *segmentation*, since the articulatory cycles of one channel are more or less independent of those of another. Thus, what is a segment boundary for one channel may not be one for another. And we have no way of deciding, in advance, which channels, if any, are more 'important' for our purpose than others. All we can do, it seems, is to look for points of coincidence between channels; the more channels that agree on any particular 'segment', the better established that segment will be. But for the most part, we find it more difficult to segment the articulatory record than to ascertain sound segments in the acoustic record. This is simply a fact of articulatory phonetics, springing from the many-to-one relationship that we have observed to exist between the speech organs and the complex acoustic signal. The segmentation provided in figure 1.3 is only partly derivable from the information contained in the recorded traces; for the most part it has been imposed on them, derived independently from a simultaneous audio-recording. Nevertheless, there is much to learn from representations like these. Consider first the lips trace, at the bottom of the chart in figure 1.3. There are four points where the lips close: for *Monday*, *morning*, *bring* and *camping*. There is also a further point where they approximate: for the labiodental fricative in *for*. Apart from these points, however, it is emphatically *not* the case that the lips simply remain passively in the 'open' position. Instead, the upper areas of the trace show continuous movement. This is fundamentally for two reasons: first, because the lips are, fairly passively, reflecting other movements, e.g. of the lower jaw, which carries the lower lip up and down with it; and second, because the lips may be involved in *secondary articulations*, as shown in this example at the point in *three*. Here the lips approximate to the same degree as in *for*, reflecting the secondary lip-protrusion on the /r/ segment. Thus, the upper part of the trace both carries general information about other aspects of speech dynamics and indicates specific types of (non-primary) articulation. Notice, finally, that the lips trace indicates that considerable velocities can be achieved by these articulators, in the relatively vertical traces shown leading into, and out of, full closures – particularly, from the open vowel in *camping*.

There is something more like an identifiable base-line in the case of the tongue trace, although here too passive movement effects are found as a result of jaw movements (e.g. the jaw partially closes for the initial consonant in *morning*). Notice, however, that this is over-ridden in the case of *Monday*, where the following vowel requires a lower tongue position. Finally, consider the velum trace (up indicating an oral articulation, down allowing for degrees of opening into the nasal cavity). Here, it may be surprising to note that the points of relatively complete closure are fairly few and far between, occurring in the sequences of oral closures that immediately follow nasal consonants in *Monday*, *morningbring*, *tentsfor* and *camping*, and also in *three*, and *forthe*. For the rest, about 60 per cent of the time, the velum is lowered by some appreciable amount (around one third). But probably the most significant pattern to spot in the velum trace is that, in nasal-plus-oral consonant sequences like those just identified, the velum is actually moving towards closure during the oral closure for the nasal consonant, and in some cases prior to that. In other words, the consonant, as defined by tongue/lip closure, ends less nasally, as defined by velic closure, than it began. This can be observed in the case of *Monday*, *morning*, *bring*, *tents*, *camping*. And, in the case of *camping*, the velic closure starts to occur even before the vowel is terminated by oral closure. Against these instances, compare the velic trace for *next*, *Monday* and *morning*, where the velum is consistently lowered through the oral closure for the nasal consonant. We therefore have to reckon with *variability* in speech production, as in perception. One sort of variability is the product of *articulatory context*: thus, the tongue may be readied for the next vowel even while maintaining position for a preceding consonant; or it may move with greater or less precision depending upon some external factor, such as *speech rate*. Another sort of variability is found as the product of *articulatory timing*: essentially the same gesture (e.g. velic closure) may be traced to coincide with one or other aspect of articulatory movement in other channels. In the case of *camping* above, the coordination is such that, simultaneously, the tongue is maintaining positions for /a/, the lips are moving to close for /m/, and the velum is rising to close for /p/. Both these types of variability result in what is referred to as *coarticulation*.

What this means is that, at any given point in time, the articulators may be observed to be carrying out aspects of both preceding and following speech targets, in an overlapping fashion. If we take a cross-sectional slice through the articulatory continuum, therefore (like the vertical lines in figure 1.3), we find a complex signal, in which anticipatory movements for upcoming targets are interleaved with the execution of current targets. This means, in turn, that coarticulation is related to the concept of articulatory variance, since it follows that the articulatory movements associated with any one target will vary from one context to another.

Another instrumental approach, which is particularly revealing of coarticulatory aspects of tongue-contact patterns in speech production, is found in electropalatography (Hardcastle 1984; Hardcastle, Morgan-Barry and Clark 1987). The sorts of phenomena that can be recorded by this technique are



Computer printout of lingual–palatal contacts during production of the word *kitkat* by a normal speaker. Sample interval is 10 msec., and frames are numbered from left to right. In each palate diagram the electrode positions are arranged schematically, the top representing the anterior part of the oral region and the bottom, the posterior part. For descriptive purposes in the text, the palate is divided into zones based on traditional phonetic labels: the top three rows are referred to as the 'alveolar' zone, rows 4 and 5 as the 'palatal' zone, and rows 6, 7 and 8 as the 'velar' zone.

Figure 1.4 Computer printout of electropalatographic record of lingual–palatal contacts during production of the word *kitkat*. (From Hardcastle, Morgan-Barry and Clark 1987: fig. 1, p. 174.)

illustrated in figure 1.4. The layout consists of six rows of thirteen 'frames', each frame representing a picture of tongue–palate contact, sampled at intervals of 10ms (starting from the top left and running to bottom right). Each frame represents schematically an artificial palate worn by the speaker and having eight rows of contact points, each row except the first (at the front of the palate, at the top of the frames in the figure) having eight contact points (sixty-two contact points in all). Where the tongue makes contact with any of these points, an electrical circuit is closed for the duration of the contact, and the contact is recorded; the frames in the figure show dots (.) for points that are not currently contacted by the tongue, and empty circles (◦) for those that are.

The top three rows represent the alveolar position, the middle two the palatal and the lower three the velar. In these terms, we can see the initiation of the word *kitkat* with a lingual contact bilaterally in the palatal to velar areas, closing to a full velar contact, maintaining bilateral palatal contact (row 1). This bilateral palatal contact is characteristic of the following /ɪ/ vowel, as can be seen in the next row; toward the end of row 2 we observe the beginnings of alveolar closure (for the /t/) being overlaid on the bilateral palatal pattern, and this is effected at the start of row 3. Also in this row, however, we can see the overlap between alveolar and velar closures (for the medial –/tk/– sequence). Row 4 shows the minimal contact pattern characteristic of the open vowel following, and this continues also into the fifth row. Towards the end of the fifth row, we have the beginnings of the final /t/ segment, showing first as a narrow bilateral palatal–velar contact (distinct from the thicker contact observed in the corresponding phase of the first /t/ closing from the high front vowel /ɪ/), and subsequently as full alveolar closure, extending into the final row. It may be appreciated that, while the rows are organised here in terms of the six main articulatory segments, they show considerable overlaps and contextual specialisations, reflective of the coarticulatory processes of connected speech.

Linguistic segmentation and articulatory variance

Given this complex phenomenon of articulatory variance, we should consider its implications for relationships between observed speech production and linguistic descriptions. One, fairly standard, viewpoint has been to adhere to the notion of linguistic–phonetic input (comprising elements of the systematic phonetic level, in the terms of Chomsky 1964 – see also Postal 1968). Several models derive from this viewpoint, one of which (Perkell 1980) we shall be considering in some detail in chapter 4. All such approaches need to account in some way for the mismatch between discrete input segments and coarticulated output.

A well-established alternative view (MacNeilage 1970) is that the input to speech production processing is essentially non-linguistic. Instead of encoding abstract linguistic segments into articulatory movements, speech production is viewed as starting from target articulatory goals (such as 'velic closure', 'labial opening', etc.) which are essentially invariant in their sensory-motor implications. By definition, however, such targets are teleological in nature: that is, they lie 'out there' as end-states to be worked towards, so that articulatory variance tends to enter the picture as the inevitable consequence of degrees of success in achieving the targets. The inevitability arises because of the probabilistic nature of human behaviour and the fact that successive targets may make conflicting demands on the articulators (e.g. 'velic opening' immediately followed by 'velic closure'). Such targets may actually be *perceptually* based:

motion in the locality of the articulatory gestures. They then disturb neighbouring air particles, which in turn transmit the motion to others, while other articulatory gestures modify and control the nature of the particle movement that results. Eventually, the acoustic signal that is so generated acts directly on the tympanic membrane of the listener's ear. At this stage, the airborne phase of the signal ends.

Fry points out that there are two sorts of particle motion that can be described as waves. One is that of a wave across the surface of water, where the particles (of water) move up and down, at right angles to the direction of the wave (it is the wave that moves transversely, not the particles of water) – this is the *transverse* wave. The other is the *longitudinal* wave, in which particles of the transmitting medium move to and fro in the same line of movement as the wave itself. This is the case with wave propagation in air. The air particles may be likened to people in a chain carrying buckets of water to a fire, each individual moving backwards and forwards within a restricted space, alternatively receiving and delivering the next bucket. As particles of air crowd together, local air pressure increases, and as they pull apart, it decreases: these represent the peaks and troughs of the air pressure wave.

Amplitude, frequency and intensity

There are three dimensions to such waves which we must recognise: they have *amplitude, frequency* and *intensity.*

Amplitude is the displacement of particles from their position of rest, either in one direction or the other. Elasticity is the property that tends to hold particles of any medium in one position, hence resisting displacement and leading to a pendulum-like reaction against displacement in one direction, resulting in an overshoot-displacement in the other direction. Gradually, with no further external force being applied, the particle will come to rest in its original position.

Frequency is the number of times that the pattern of displacement either side of the position of rest occurs in a unit of time. It is conventional to refer to cycles of displacement (one cycle being described by the motion of a particle as it goes through each pattern of displacement and overshoot) per second: originally abbreviated as cps (cycles per second), this unit is now more generally referred to as Hertz (after the German physicist, Heinrich Hertz), and abbreviated as Hz (kiloHertz, kHz, 1,000 cycles per second, are also convenient large units for speech).

Intensity is the hidden member of these three dimensions of the sound signal, since it relates to the energy in the sound, and this may be expressed in terms of greater amplitude (displacement of particles at a given frequency) or frequency (more rapid oscillation of particles at a given amplitude). The intensity of a sound must be increased by n squared for an increase of n in the frequency of a sound of given amplitude; and by n squared also for an increase of n in the amplitude of a sound of given frequency. Intensity is measured in decibels (dB), units on a logarithmic comparison scale ranging from 0dB, on the threshold of audibility, to 130dB, equivalent to the sound of a jet aircraft at 120 feet. Conversation at normal levels and distances is rated as 60dB. The ear is used to handling speech arriving as whispers (30dB) to the loudest shouting that speakers can manage (75dB) (Fry 1979).

Simple versus complex waves

Any departure from a perfectly regular, symmetrical wave form (a *sinusoidal* form, or a *sine wave*) results in what is referred to as a complex wave form. Simple wave forms arise from such instruments as a tuning fork, where the to-and-fro motion of the prongs is evenly distributed either side of the position of rest. Most of the sounds we find in everyday experience, and speech sounds are among these, are produced from instruments that are more complex in their vibrating properties than tuning-fork prongs; they move more easily in one direction than another, for instance, or they are constrained (damped) by some other instrument contacting them. These situations give rise to complex wave forms, which can be analysed out in terms of constituent simple wave forms in particular time (or phase) relationship to each other. The constituency is expressed in Fourier's Theorem, and the process by which complex wave forms are recognised by calculating their constituent simple wave forms is referred to as Fourier analysis.

Sound quality and harmonics

For complex wave forms, the constituent sinusoidal wave forms – or components – consist of (a) the lowest-frequency component, called the fundamental frequency, and (b) other components that are at whole-number multiples of this frequency (i.e. twice that frequency, three times that frequency, four times that frequency, and so on). These latter are called the harmonics of the fundamental (second harmonic, third harmonic, fourth harmonic, etc.). Depending on the characteristics of the sound-producing system, some of these harmonic frequencies may be of greater amplitude than others, and it is this harmonic structure of a sound that gives it its characteristic quality. The shape of the vocal tract at any given moment is related to the harmonic–amplitude structure of the sound that results.

Periodic versus aperiodic sounds

Certain sounds do not have this orderly relationship between fundamental and harmonic frequencies, however, and these are referred to as aperiodic sounds. Sounds generated by taking frequencies at random and adding them together are of this type. Fry (1979) notes that these sounds are 'those which the ear and brain class as *noises*; the sound of escaping steam and the sound of something being fried in the pan are good examples of natural noises' (1979: 83). In speech, sounds that have a supraglottal source, e.g. the turbulence of air escaping through an oral constriction, as for [s], are aperiodic; by contrast, sounds that arise through pulse waves deriving from the opening and closing of the glottis in the larynx are periodic (and hence have harmonics on which differential amplitude characteristics can establish a structure).

1.2.3 *Perceptual factors*

The acoustic signal properties

A good way to start considering the nature of the acoustic signal is to look at a spectrogram. Figure 1.5 shows a spectrogram of a phrase, *rapid writing*, with the broad phonetic identification of segments indicated. One of the features you will notice is that the relative timing of these segments as represented in the acoustic signal is quite varied: some are of relatively long duration, while others are very brief. Another feature is that the segments do not have sharp boundaries, by and large: the centres of the segments are more clearly identifiable than are their boundaries. Indeed, the phonetic identification provided in this example has been carried out in this way; the positioning of the phonetic symbols is in line with their spectrographic centres, and the boundaries have been left unmarked. These two features, then, belong to the *time* dimension of the signal. The next feature to notice is that different parts of the signal, as identified along the time line, have distinct energy patterns spread vertically. The vertical axis represents the *frequency* dimension of the display, and it is possible to see at once that it would be too simple to say that some sounds are of higher or lower frequency than others. Rather, each identified sound has a *pattern* of frequencies – a frequency profile – and these profiles are distinct from each other. But, because speech is basically a continuous phenomenon, these frequency profiles do not show *sharp* discontinuities from those on either side of them. They merge and flow, one into the other, along the duration dimension. This is why the boundaries are harder to locate than the centres. There is also a third dimension, of *intensity*, which is indicated by relative darkness in figure 1.5 but can be seen more clearly in figure 1.6. The

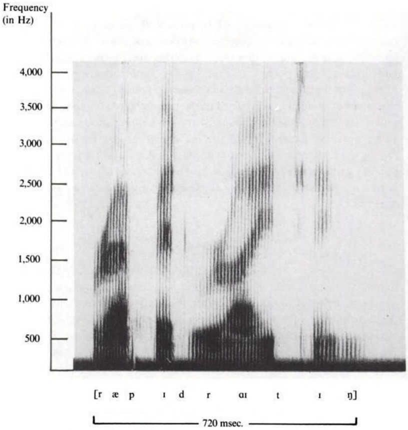


Figure 1.5 Spectrographic representation of the phrase *rapid writing*

peaks of intensity appear on the spectrogram as dark bands of energy, known as *formants*. These reflect the changing resonance characteristics of the vocal tract as it assumes different configurations. These three dimensions, of time, frequency and intensity (amplitude may be thought of as the overall intensity of a section of the signal, across all frequencies), represent a good deal of the acoustic aspect of the speech signal. But we must bear in mind that the acoustic form of the signal derives from the aerodynamic effects, which in turn result from the consequences of articulatory movements on surrounding air particles – and there is certainly some loss of information through these transduction stages. First, there are quite palpable articulatory distinctions which appear to get lost in aerodynamic encoding, and hence cannot be identified in the acoustic signal: e.g. the distinction between alveolar [n] in *ten* and dental

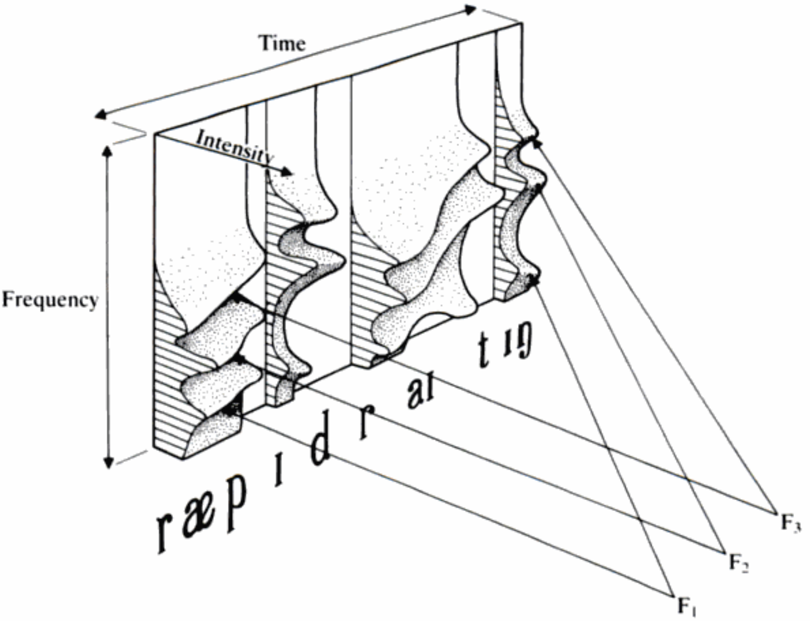


Figure 1.6 Schematic three-dimensional diagram of the spectrographic form of *figure 1.5*

[n̪] on *tenth* (Fudge 1967); or the velar closure coarticulated with alveolar closure in the case we noted above in figure 1.4; or the momentary drop in air pressure in the forward part of the oral tract which results from a sudden enlarging of that cavity, obliterating a continued increase in pressure below the larynx; and so on.

Secondly, the normal possibilities for contamination of the aerodynamic/acoustic relationship are endless, as will be appreciated by anyone who has made a spectrogram of speech recorded in a less-than-soundproof environment. Typically, what comes in at the ear is not just what is being said by the person you are listening to at the moment, but a host of other sounds, speech as well as non-speech. Consequently, a major issue in the understanding of speech perception in normal, noisy environments is the ability to 'lock on' to a subset of the acoustic information arriving at the ear, and to define it as a distinct channel, which may be kept open, or abandoned in favour of another, in spite of the fluctuating real-time values both within and around it. In this connection, we should bear in mind that phase relations, which can prove important in discriminating speech from non-speech sound, are not found in the spectrographic record.

It is not necessary or possible to provide an exhaustive review of different writing systems in what follows (aspects of this study may be found in Edgerton 1952; Gelb 1963; Pei 1965; Haas 1976a; and Sampson 1985). Instead, we shall concentrate on some of the major differences of organisation, and then discuss the system that English uses in this context. We may also note that it is not appropriate for us to go into the physical properties of light, and the ways in which the photoreceptors of the eye are activated by the wavelengths involved in processing written language symbols. There is no real parallel here to the role of air as the transmitting medium of speech. Light does not serve as the expressing medium of written language (hand and finger movements do not manipulate light in the act of writing). Thus writing can be accomplished in the dark, whereas no speech is possible in a vacuum (apart from the indirect skill of lip-reading); and switching out the light does not erase the symbols from the page. Accordingly, we shall not organise our discussion in terms of production, transmission and perception here, as we did in discussing the auditory signal (section 1.2), since these aspects do not relate in a straightforward way to the conventional differences that we are concerned with. These are represented in

1. ideographic systems
2. syllable-based systems (syllabaries)
3. alphabetic systems

which we shall consider in sections 1.3.2–1.3.4 below. The processing implications of such distinctions will be reserved for discussion in chapters 4 and 5. However, we shall briefly review some general issues regarding production and perception in 1.3.5.

The status of writing systems

From a linguistic point of view, it is usual, and appropriate, to stress the derivative status of written forms of language, in spite of the degree of independence that some written language forms may attain (Lyons 1968). The argument is based on numbers (relatively few of the world's languages are possessed of a native script), on history (writing emerged comparatively late on in the evolution of languages) and on logic (a spoken language is not affected by the rise, or the loss of, a written representation for it). There are many difficult points of detail with this last argument, since certain distinct features of style, involving all levels of language, may rest upon the existence of a written representation for a language, but these hardly affect the main issue. From a psycholinguistic point of view, however, the visual modality of language processing gives rise to questions that go to the heart of our understanding of how language is instantiated in the individual. Certainly, reading and writing develop later in the individual's language development than do speaking and listening abilities; but how they develop, and what subsequently happens to the individual's total language capacities, are fundamental issues.

One logical possibility is the situation where a new means of expression/perception is developed for one and the same language ability; another is where a totally new language ability is developed in relation to the written means of expression and perception. It may be suspected that most people's experience lies somewhere between these extremes of compound and coordinate relationship, but exactly where is dependent on a number of factors, including the way in which the writing system is taught and acquired, the degree of literacy achieved and the formal properties of the system. We are mainly concerned with the latter here.

1.3.1 ***Possible writing systems***

General linguistic accounts of written language (conveniently reviewed in Henderson 1982) frequently start by establishing the boundaries of possible scripts and continue by elaborating three main types within these. The continuum of graphic representation is bounded at one extreme by a variety of *pictographic* possibilities. These may be relatively unstereotyped, as in the case of a representational picture, telling a story; or symbolically constrained, as in the case of internationally agreed road signs, where a red border signifies a warning or prohibition. Even in these latter, conventionalised pictographs, however, the relationship to linguistic forms of the message tends to be approximate and variable. For intance, the sign in figure 1.7 might be verbalised as 'Warning: stone chippings are liable to be thrown up from vehicles' wheels along this surface', but this particular linguistic version is only as good as all the other possible paraphrases. In short, the sign is directly representing a message, rather than a particular linguistic form of a message, and hence falls outside the accepted scope of 'written language'. At the other extreme, we shall wish to exclude machine-based recordings of the acoustic signal (spectrograms, mingograms, and the like), since they selectively present formal features of language for the purpose of refining the analysis of these features,

Figure 1.7 A standard warning sign from the British Highway Code. (From *The highway code*. London: HMSO.)

rather than for the more ordinary purposes of communication (Marshall 1976).

In between these extremes, the sort of picture set out in figure 1.8 is frequently recognised. This sort of taxonomy is usually accompanied by reference to the diachronic interpretation of these types, whereby meaning-based systems tend to become sound-symbolic over time; and all known alphabetic systems are said to derive from some relatively recent (conventionally, Phoenician) development representing the high-water mark of orthographic development. Thus, it is argued, for example, that no further refinement of alphabetic writing systems into feature-based notations is possible, since this would actually worsen their practicability and efficiency; the ratio of graphic information (marks on the page) to perceived speech distinctions would become unacceptably high.

Problematically, this sort of typology can create the misleading impression that actual writing systems are purely of one type or another. Possibly the most striking example of a mixed system is Japanese, where logographic and syllabary symbols coexist in the written forms of utterances; a more mundane example is to be found in English, which has a number of non-alphabetic elements, as we shall see. But even in the less obviously mixed types, we shall note that Chinese logographs contain sound-representative components; and that among the alphabetic elements of English, words may be distinguished by their visual spelling patterns, within certain limits of sound-representativeness. Thus, for example, being able to read and understand forms such as *boy* and *buoy* is a matter of relating letter structure to word meaning as well as to word sounds.

First, however, let us sketch out more neutrally the different sorts of representation of language that may exist – and coexist – with an orthographic system. As figure 1.9 shows, there are three relevant orders of unit, which may be illustrated in the threefold ambiguity of the word *word*: we may mean a word in the sense of its sound shape (phonological word), or in the sense of its grammatical properties (grammatical word), or in the sense of its lexical iden-

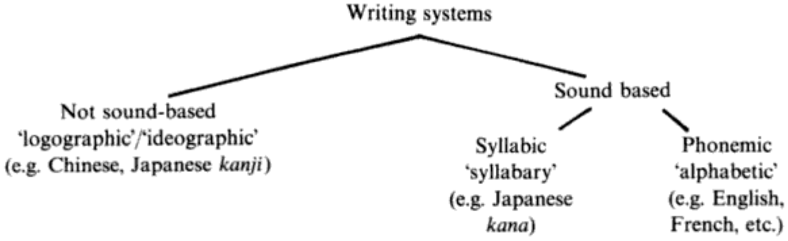
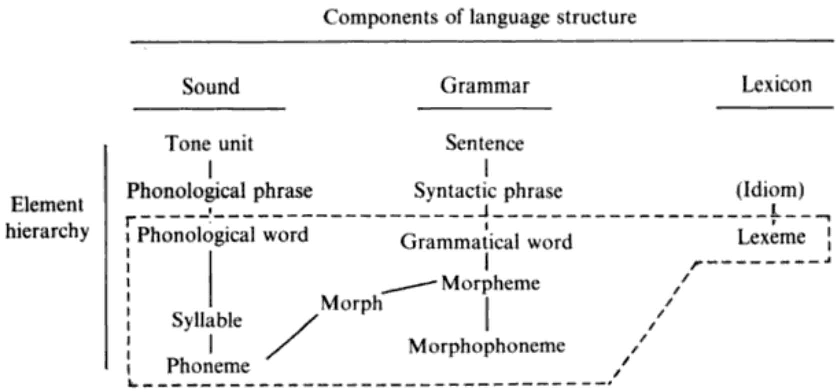


Figure 1.8 A generalised typology of writing systems



(Elements within the dotted lines are available for orthographic mapping)

Figure 1.9 Possible points of contact between language elements and units of writing systems

tity (the word as a lexeme). As an example, consider the word *foolish.* Phonologically, it consists of two syllables, organising five phonemes, /fu:-lɪʃ/, with stress on the first syllable. Grammatically, it consists of two morphemes, {fool} and -{ish}, the first being a noun, the second an adjectivalising suffix, and the whole form being an adjective. Lexemically, it is a member of the set *fool, foolish, foolishly, foolishness, folly* (which, in this sense of the term, are all forms of the same 'word').

It is no accident that we have chosen to illustrate these three orders of language structure from the level of the word. In the phonological hierarchy, word forms constitute phonological phrases, which build into tone units; in the grammatical hierarchy, words are constituents of syntactic phrases and thence of sentences; and lexically, there are no other units to select (apart from idioms such as *hit it off*). Presumably, no orthographic system builds on phonological or syntactic phrases, or larger units still, for the same reason that these do not constitute the basic units of form–meaning correspondence in language structure: this would entail having different symbols to represent each of *the boy, a boy, the girl, a girl* and so on. So, given that orthographic representations map from units of word size (or less), we can see at once that any particular system might represent *either* aspects of the *form* of words (whether phonological, morphological or lexemic) *or* the semantic aspect. And if the orthographic system maps from units of less than word size then these will either be the semantically empty phonological units of the syllable and the phoneme, or, in the grammatical hierarchy, morphemes which again can be captured either as form-based or semantic entities. For our discussion here, the item morph in figure 1.9 may be considered to be a grammatically relevant phonological entity, as in the alternative forms of the stem in *knife*, *knive-s*.

Given this framework, we shall now briefly characterise some of the salient features of Chinese orthography, followed by Japanese as an example of a 'mixed' script, then illustrate from a typical syllabary, as used for Kannada, a Dravidian language of southern India, and finally turn to a consideration of English.

1.3.2 ***Chinese (Mandarin)***

The spoken languages (often referred to as 'dialects') that are Chinese include Mandarin, Cantonese, Hokkien and other mutually unintelligible languages. They can all be written in the same traditional script; the examples given here are illustrated with phonological forms from Mandarin. There are basically three sorts of character to consider.

Simple characters

These include:

(2)	人	木	其
	rén	mù	qí /tɕʰi/
	person	tree	his, her, its, etc.

The first two of these illustrate the simplest sort of meaning-based symbol, even preserving something of their representational (pictographic) basis: we might use the term *ideographic* for them. The third is representationally much more abstract, as well as being more complex in terms of its component strokes; for this we might use the term *logographic*, since there is no hint of an attempt here to portray the idea of the word, only to provide a distinct symbol for the word itself. Of all these characters we may say that there is no connection between their stroke structure and the sound structure of the words they represent: nothing marks the tone or the phonemic sequence in any way; we shall say that there is nothing *phono*-graphic about them.

However, some simple forms may represent more than one word, such as that for *gōng* /kūŋ/, 'public' (also 'male of species'), and the character for *gàn* /kàn/, 'trunk' (also 'work'). The fact that the same character may be used for two (or more) words that happen to share the same phonological form is important: it means that, in such cases, the character is taking on a phonographic function. It may be regarded as logographic in respect of one of the words it stands for (usually one is seen as basic to it), but when we say that it 'stands for' this word, we cannot rigorously exclude its potential for representing the sound as well as the meaning; and it is the sound aspect that is have rather literally interpreted the sort of information that these characters convey. In practice, only a learner of the Chinese script would interpret 松 in (7) as 'sort of tree, and pronounced like *gōng*'. For a fluent reader, the character as a whole would map onto the lexical item {sōng: 'pine tree'}; and indeed onto its homophone {sōng: 'loose'}. What this means is that the force of the radical component is frequently lost, and the result of this may be seen in the existence of certain sequences such as:

(8) 木 板
mù bǎn
tree board = wooden plank or board

Here, no new character is formed, but rather a (syntactic) sequence of characters. But notice that the first character, the full form of *mù*, is required to distinguish *wooden* from other sorts of *board.*

In all, there are now 214 radicals, which serve as the organising principle of Chinese dictionaries. There are 1,585 compound characters listed in the largest Chinese dictionary as being built on the radical 木 *mù*. Most, but not all, of them have something to do with wood; but the full symbol 木 *mù* also functions in many other characters as a 'phonetic' component.

Finally, we should stress the point that it is too much of a simplification to say that the Chinese script is not phonographic. We have tried to show that it does indeed have phonographic aspects, and that the control of it by native writers and readers requires skilful interplay of both sound and meaning dimensions in the script. Historically, this can be seen in the way new characters have been devised for the language: Ong Tee Wah (1980) relates the story of how the second-century lexicographer Xu Shen provided a character for the word *xī* /ʃī/, 'west', by using the form 西 which also appears in 栖 *qī* /jī/, 'nesting bird', with the gloss that the bird went to its nest at the time of the setting of the sun in the west (possibly the forms involved were phonologically more similar in his day). The same sort of interplay can be found today, for instance in the devising of suitable characters for newspaper reports to use in referring to visiting foreign dignitaries (S. E. Martin 1972 notes the possibilities that exist for intentional sound–meaning interplay in this connection).

There are many details that have been left out of this account (see French 1976, Sampson 1985), but it should suffice to make the point that the Chinese writing system does embody certain phonographic principles. We have to observe also that the phonographic aspects tend to be related to whole word-forms, rather than to phonological segments, and that individual component strokes of the written characters do not themselves represent aspects of the phonological pattern; that is to say, the phonographic representation is fundamentally non-compositional. Or, putting it another way, it is sound *similarity* rather than *components* of sound structure, that the characters capture.

1.3.3 ***Syllabaries***

Among those writing systems that are fundamentally phonographic, we shall expect to find the ability to represent components of sound structure. One very important type is that of *syllabaries*, which are based on the syllable unit of the spoken language; and in this connection a question immediately arises as to how far they are radically distinct from those, *alphabetic*, systems that are based on individual consonant and vowel segments.

We shall look first at the writing system of Japanese, which uses a syllabary alongside a Chinese type of system. Having briefly noted the main characteristics of the Japanese syllabary, we shall then consider a rather different type, as found in Kannada, a Dravidian language spoken and written in Karnataka state in south India. This is an example of a highly compositional syllabary, within which the consonant and vowel segments can easily be discerned; as such, it will take us towards our consideration of the alphabetic system used in English (section 1.3.4).

Japanese

The Japanese writing system is a hybrid of (a) *kanji* characters, deriving from Chinese ideo-/logographs, and (b) *kana*, syllabary elements that belong to either of the subtypes *hiragana* (cursive kana) or *katakana* (square kana) (Morton and Sasanuma 1984). The overall characteristics of this mixed system, particularly the interplay of meaning-symbolic and sound-symbolic elements, are well described in Martin (1972). The kanji characters are used for major lexical items (nouns, verbs, adjectives), while, of the kana elements, hiragana is used for grammatical morphemes (particles, auxiliary verbs, etc.), and katakana is used for representing loan words and foreign names. Each of these syllabary systems has seventy-one characters, using basic forms and diacritics (see Morton and Sasanuma 1984). Both systems appear to have arisen from simplifications of kanji characters, and hence share some general similarities; but there are nonetheless some striking differences of form between corresponding symbols in the two systems.

Each system is typical of syllabaries generally in having two fundamental types of sound syllable directly represented in terms of characters: the vowel-type (V), without preceding or following consonant; and the open-syllable type (CV), consisting of a consonant followed by a vowel. In Japanese kana, as traditionally organised, there are five vowel characters, corresponding to the alphabetic symbols *a*, *i*, *u*, *e* and *o*. In addition, there are separate symbols for each of the CV combinations arising from the following consonants with these vowels: *k*, *s*, *t*, *n*, *h*, *m*, *r*, *g*, *z*, *d*, *b*, *p*. Some other symbols represent CV sequences with a restricted range of vowels: *ya*, *yu*, *yo*, *wa*, *wo*. There is also a symbol for the nasal consonant without included vowel. The total inventory is seventy-one symbols.

In such a system, it is clearly crucial to understand the nature of the relationship, in character composition, between those shapes that represent the individual V sounds and those that represent the CV patterns. In both forms of kana, this relationship is fundamentally obscure: while it is possible to spot similarities of form here and there among V and CV characters that represent the same vowel sound, this phonographic correspondence goes largely unrepresented. It is also important to observe how consistently individual consonant sounds are represented in CV symbols; and here the kana system again provides very little overt correspondence. The result is that kana symbols by and large are highly syllabic, possessing very little internal structure which would permit their decomposition into constituent consonant and vowel shapes. See figure 1.10, where the top row consists of the *k*-series of CV symbols in hiragana; it will be quite apparent that it is impossible to detect a consistent *k*-feature across these forms. By contrast, note that the corresponding *g*-series (row 4 in fig. 1.10) is systematically related to the *k*-series by using the same basic symbol, with the addition of a voicing diacritic (two short strokes in the upper right corner). This feature is found on certain other voiced consonant series also. But it should be noted that this type of compositionality

Japanese (Hiragana)	か	き	く	け	こ
Kannada	ಕ	ಕಿ	ಕು	ಕೆ	ಕೊ
English	ka	ki	ku	ke	ko
Japanese (Hiragana)	が	ぎ	ぐ	げ	ご
Kannada	ಗ	ಗಿ	ಗು	ಗೆ	ಗೊ
English	ga	gi	gu	ge	go

Figure 1.10 An illustrative selection of Japanese (Hiragana) and Kannada symbols with English equivalents

operates from syllable level to *features*, rather than to the *segments* of alphabetic notation.

Kannada

By contrast, the Kannada language is possessed of a syllabary that has distinctly alphabetic implications. It may be thought of, as a first approximation, as having fifty basic symbols, made up of twelve vowel types (V), thirty-four CV types, and four other positionally restricted elements that need not detain us here. The count of thirty-four CV types, however, is based only on the included vowel *a*: i.e., there are thirty-four $C + a$ symbols. In addition, these may be supplied with other vowel diacritics, generating separate symbols for each of the CV patterns recognised in the system – a total of more than 400 symbols in all. Perhaps because of the size of this system, it is systematically compositional in nature, with relatively consistent shapes for both consonantal and vocalic identities. And these identities are essentially segmental (alphabetic) in size. See figure 1.10, rows 2 and 5: these clearly show that there are distinct consonant symbols for the *k*- and *g*-series, and that there is consistency of both consonant and vowel representation in the CV symbols illustrated (not all the vowel possibilities that Kannada recognises are shown). A further aspect of this system that underlines its alphabetic potential is that it is possible to represent consonants alone (i.e. without an included vowel diacritic) by providing the basic consonant symbol with a 'vowel-deletion' diacritic: this is used, for example, in pedagogic contexts, in formulae such as '$k^{\o}$ (the symbol with the vowel deletion diacritic) $+ a = k^{a}$ (the symbol with the *a* diacritic)'. The other contexts in which such consonant-alone symbols might be used is in consonant clusters: thus the word /yatna/ 'effort', might be represented as $y^{a}t^{\o}n^{a}$. In practice, this solution is much less used nowadays, in favour of a range of consonantal diacritics below the base symbol; in these terms, /yatna/ would be represented as $y^{a}t_{n}{}^{a}$. This solution reflects the fundamentally syllabic nature of the symbols, as does the practice of ignoring word boundaries in continuous text, in favour of syllabic structure: thus /ond(u)/ 'one' + /iṭṭige/ 'brick', which in speech usually shows elision of the parenthesised vowel, is representable as $ond^{i}ṭ_{ṭ}{}^{i}g^{e}$. (The symbol 'ṭ' represents a retroflex consonant.)

The Kannada writing system, then, is fundamentally syllabic, but has, as we have seen, a number of alphabetic aspects. The existence of more than one organising principle within a single writing system should not surprise us: after all, the speech signal contains syllables as well as smaller and larger units, and it is the function of a phonographic script to provide useful clues to the sound structure of the language forms that it represents. But, beyond this, we have also seen that phonographic clues may even sit alongside logographic ones. At this point, we may turn to the case of the English writing system.

1.3.4 English

Examining the nature of our own writing system (see also Haas 1969; Albrow 1972) is in some ways a harder task than looking at other systems, since we have to step back and examine our assumptions (e.g. that ours is basically an alphabetic system, that it is therefore phonographic, and at the level of the phoneme, and so on).

We may start by considering the inventory of alphabetic elements that English makes use of, and for this purpose it is convenient to use the keys of an ordinary typewriter, as set out in figure 1.11.

Non-alphabetic elements

Perhaps what is immediately striking about this array is the fact that there are so many non-alphabetic elements, such as £, &, %, etc. (Edgerton 1941). Furthermore, this list could easily be extended: . in *9.186* ('nine *point* one eight six'); × in *5 × 4* ('five *times* four'; in computer applications, * has this function); *X* in *Xmas* ('*Christ*mas', also '*Ex*mas') and in common abbreviations such as *Xian* ('*Christ*ian') and *KingsX*, *QueensX* (abbreviating the major railway stations on the London–Aberdeen route), + similarly in *Kings +* and *Charing +* (the usual form for 'Kings *Cross*', 'Charing *Cross*' on London buses), and so on. Further extensions still can be effected by compiling basically alphabetic elements into such conventional units as *ms*, *dB*, *Hz*,

1	2	3	4	5	6	7	8	9	0	¾	=
*	"	/	$	£	_	&	'	(	)	¼	+
q	w	e	r	t	y	u	i	o	p	-	[
Q	W	E	R	T	Y	U	I	O	P	?	!
a	s	d	f	g	h	j	k	l	;	]	⅔
A	S	D	F	G	H	J	K	L	:	@	⅓
z	x	c	v	b	n	m	,	.	½		
Z	X	C	V	B	N	M	,	.	%		

Figure 1.11 The character inventory of a standard English typewriter

a phoneme: *sh* for /ʃ/, *ng* for /ŋ/, *ee* for /iː/, and so on. In these cases, we have to recognise a set of digraphs, or complex symbols, in which the constituent elements are not compositionally phonographic (though, as in the cases of *sh*, *ng*, etc., they are clearly chosen on a 'near neighbour' principle). Essentially, then, these violations are more apparent than real, and we should extend our list of basic alphabetic symbols to include *sh*, *ng*, *ee*, *th*, and a host of other digraphs. However, this does not eliminate all problems: *th*, for instance, is used for both /θ/ and /ð/; and *ti* in *nation*, *ration*, maps on to /ʃ/. Furthermore, in *cushion*, *fashion*, we may wish to recognise a trigraph *shi* which is related to both *sh* and *ti*. We should, however, consider these problems to derive not from the existence of polygraphic letter sequences as such, but from a distinct characteristic of English orthography, which affects single letter symbols as well; this is the phenomenon of many-to-many mappings between sounds and letters. Figure 1.12 illustrates the situation. It is perhaps not surprising that, traditionally, English spelling has been seen as essentially unsystematic: cf. G. B. Shaw's deliberately perverse spelling of *fish* as *ghoti* (*gh* as in *rough*, *o* as in *women*, *ti* as in *nation*). We should mention here also the well-known problem of 'silent letters', as in *sword*, *scissors*, *psychology*, *mnemonic* and the like. Note that in other cases, such as *e* in *sale*, the so-called 'silent *e*' is usually described as having the function of determining the phonological correspondence of the preceding vowel symbol: a sort of discontinuous digraph, perhaps, *a–e*. But this is complicated further in words like *face*, where *e* also is said to function in 'softening' *c*: a convergence here of *a–e* and *ce*. We may conclude, then, that there is a considerable body of evidence that tends towards the view that English spelling is compositionally phonographic only *postlexically*; in other words, the grapheme–phoneme mappings become apparent *after* the word in question has been identified, and are not an adequate basis on which to achieve its identification.

Spelling patterns in English

Contrary to possible first impressions, however, English spelling is not simply perverse. Three types of example will suffice to make the point.

First, consider again Shaw's outrageous *ghoti*. This example partly makes its point by selecting from the exceptional use of *o* in *women*, but also by ignoring positional factors. Thus, *gh* = /f/ is a feature of word-final position only, as in *cough*, *rough*; and so also in *coughing*, *roughly* where the forms are built up by suffixation onto a word. Similarly, *ti* = /ʃ/ is only found word-medially. So we learn from Shaw's example not that English spelling is simply perverse but that it is complex, and involves positionally restricted grapheme–phoneme mappings.

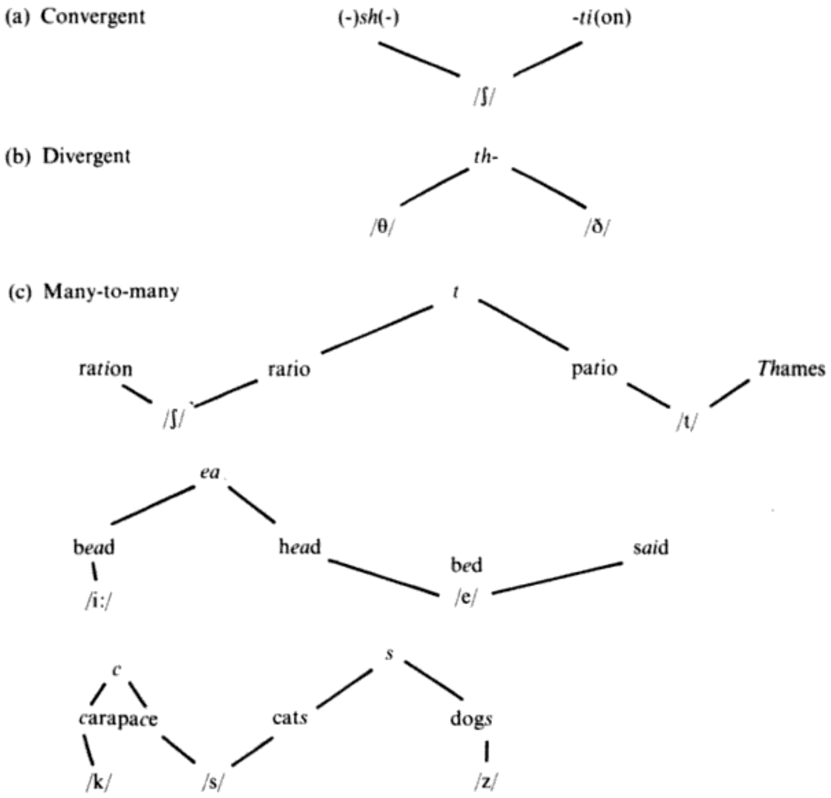


Figure 1.12 Examples of convergent, divergent and many-to-many letter-to-sound relationships in English

Secondly, consider *cats* vs *dogs*: the use of *-s* captures the predictable phonetic relationship between -[s] ~ -[z] in this environment. And in *lint*, *link*, the use of *-n-* for the homorganic nasal consonant shows the same principle at work. There is a limit to the adherence to this principle, however, as shown in *foxes* (where the syllabic form of the affix is distinctly represented with a vowel symbol – but notice that *s* is preserved) and in *limp* (apparently marking the quantal divide between labial and lingual articulations). From these examples it seems apparent that English spelling actually implements the phonemic principle, up to some (imprecise) limit of phonetic difference.

Thirdly, it has more recently been suggested, within the framework of generative phonology, that much of the systematicity of English spelling has been missed by failing to look for it in the right place, i.e. at an abstract level where lexical relationships are to be stated. Thus, in

op*aque* – op*ac*ity
prof*a*ne– prof*a*nity
electri*c* – electri*c*ity

the stability of orthographic *-a-* and *-c/-c-* help to mark word identities in a language where these are obscured by phonological changes to the base forms. Again, the orthography only departs from this principle, in these examples, where a limit on its operation is encountered: in the case of *-qu-/-c-*, **opaquity*, with *-qu-* = /s/, would look odd against *electricity*; and **opace*, with *-c-* = /k/, would look odd against *pace*.

These three examples, taken together, suggest that English spelling may indeed be seen as violating a very simple spelling-to-sound correspondence, but that it actually embodies a set of relationships that are considerably more sophisticated, and which carry a number of advantages. The fact that these advantages – e.g. the capturing of positional effects, environmental constraints, and abstract representations – are open only to those who already know the language well is undeniable but, clearly, not a criticism of the orthography as such.

Residual problems, and conclusions

Contrary to these justifiably emphatic second impressions, however, it is still the case that there are many inconsistencies in the way that English words are spelled. Many of these derive from ancient scribal practice (e.g. writing *o* for *u* in letter environments where *u* would have been easily confused, hence *woman*), from regional differences of both pronunciation and spelling tradition, from attempts at reform that were more or less thorough, or misguided, and so on. There is no 'reason' for *head* to exist alongside *bead*, in the language as it is today; the cause is essentially external, located in the historical accretions that culminated in the present-day script.

What, then, is the nature of the English alphabetic system of orthography? It tends to be phonographic, but is compositionally so only in comparatively few cases (it is no accident that *cat*, *sat*, *mat* are routinely used for beginning readers). Differences in spelling between homophones such as *two* vs *to* vs *too*, *for* vs *four*, *bee* vs *be*, *oar* vs *or*, etc. mark distinctions of grammatical-word identity, in a way which over-rides, yet does not conflict with, their identical status as phonological words. The essence of such differences is that they are visual rather than phonological. As a result, the semantic distinction between, say, *flare* and *flair* is marked, as far as production is concerned, in terms of strokes of the pen, or sequences of typewriter key pressings, and, in perception, in partially distinct visual arrays. The similarity between this state of affairs and what was noted for Chinese will be apparent; and it is somewhat reinforced by the consideration that, in terms of pen-strokes, roughly the same number of elements, on average, go to make up a written word in each of these languages.

Words – real, possible and impossible

It will have become apparent, but is worth stating explicitly, that the basic unit of English orthography is the word. Alphabetic characters exist to spell words (rather than, say, syllables), and orthographic word-forms reveal themselves to be highly resistant to variability (as witness *elevate–elevation–elevator*). Higher-level units than the word, such as the syntactic phrase or sentence, are spelled in a way which is a strict function of the spelling of their constituent words (with no contextual modifications such as are found, for instance, in Kannada). Furthermore, English orthography consistently marks the word status of alphabetic sequences by use of spaces (which thus function as a type of word-level character in the system) and, more-or-less consistently, the hyphen. If the way that words are spelled represents the heart of the system, it will be appropriate for us to concentrate attention on what constraints exist on the form of letter strings between two word spaces.

Constraints on letter sequences

We have stated already that the English alphabet is not phonologically compositional in a strict or thorough-going manner. However, precisely because it is an alphabet, and because individual letters have nominal sound correspondences, the resulting system (a) allows for *literal* compositionality – i.e. we can take individual letters and put them into any sequence, and (b) recognises that these strings have some nominal pronounceability. It is easy to spot **gjmwbk* as an impossible literal string and *gippit* as possible but not having word status. But it is less easy to disentangle these observations from the fact that the nominal pronounceability of **gjmwbk* (as something like [gəˈʤɨmwəbk], let us say) violates phonological sequencing constraints in English, while *gippit* (either as /ˈgɪpɪt/ or /ˈʤɪpɪˈt/) does not. So the question arises: are there any purely literal sequencing constraints? Consider first the case of *qu-* in *quash*, *queen*, *quiz*, etc. where the occurrence of *u* is completely determined by the preceding *q*. At first sight this looks as if it must represent a literal sequence constraint, since of course there is no phonological constraint that /w/ should follow /k/. But the issue is complicated by the fact that *u* in these sequences does represent the phonological element /w/. The grapheme–phoneme correspondence here is something like 'If /k/ is followed by /w/, it is represented by *q*, and /w/ by *u*'. So again, the literal and phonological factors are difficult to disentangle. The search for purely literal constraints is more profitably conducted at points where literal sequences have contextual variants which do not change their nominal pronunciation value, e.g. the letter clusters *tt*, *ck* in *butt*, *butter*, *back*, *backing*, compared to **ttub*, **ckub*, etc. However, even here it is possible to provide a sound-based observation for the letter-cluster forms: they occur only after phonologically short vowels, hence **cacke* (for *cake*), etc.

Grapheme–phoneme correspondences

One of the points to emerge from this discussion is the issue of grapheme–phoneme correspondences. We have seen that as far as real words of the language are concerned, their pronunciation is, by and large, *not* a function of segment-by-segment mappings between letters and sounds. Instead, the letter *sequence* maps on to a *word*, and the word then guides the interpretation of the component letters: thus *-gh* in *cough*, *rough* is /f/, *-ou-* in *cough* is /ɒ/, and so on. It is in this sense that the phonographic compositionality of English spelling can be said to be *postlexical* as opposed to *prelexical*.

On the other hand, we have also recognised the existence of nominal pronunciation values of individual letters, and the fact that these elements are fully compositional. This is clearly crucial to the pronunciation of *non-words* (both possible and impossible types). It is also crucial to the pronunciation of those words which we come across in print for the first time. These are actually of two types, which we may refer to as the *blick* type, and the *misled* type. The *blick* type consists of words having dictionary status, but existing in the personal lexicons of relatively few people (e.g. architects use *blick* to refer to a glimpse of a building between surrounding buildings). The *misled* type is a convenient label for the situation where we are misled (/mis'led/)) into pronouncing a commonly used word in the wrong way (/'maɪzəld/) when we encounter it for the first time written down. In each case, a first scanning of the graphic form relies on (a) our knowledge of the nominal pronunciation values, or grapheme-to-phoneme correspondences, of individual letters, and (b) our knowledge of the phonotactic constraints of the language (since we naturally assume that the item in question is a word which we are simply ignorant of). As it happens, the grapheme-to-phoneme correspondences of *blick* are uniquely constrained, but this is not a defining characteristic of this type. We may actually come up with an incorrect pronunciation for other words of this type, which may persist until some authoritative source (someone who knows, or a dictionary) puts us right. The *misled* type may likewise be either straightforward or complex in terms of grapheme-to-phoneme correspondences. What distinguishes it from the *blick* type is the fact that, at some point, conreason to suppose that Korean readers do not use, for example, real-world knowledge, even though it is not directly represented in their script.

We must conclude that, while it is possible that different types of organising principles in writing systems may affect the nature of the perceptual and production processes that are called into play, it may be very difficult, in practice, to determine what the precise relationship might be between script properties and processing strategies. Certainly, it would appear that such gross distinctions as 'logographic' vs 'syllabary' vs 'alphabetic' provide at best a very haphazard basis for determining the nature of individuals' reading and writing performance.

1.4 Conclusions

1.4.1 *Spoken versus written language signals*

The speech signal is highly complex. It is representable as a dynamic three-dimensional acoustic entity, which, although it has, in some obvious sense, segments within it that correspond to our intuitions regarding speech sounds, is essentially continuous. It is marked by its linear compression, or the interleaving of preceding speech sounds with later ones. This arises as a natural consequence of the parallel and independent operation of the articulators in the vocal tract, which may be thought of as constituting a multichannel transmitting system. The resulting signal is ideally structured for reception by the auditory system, which is able to resolve the frequency and intensity information contained in the signal in a parallel, multi-channel and continuous fashion; it would be incapable of processing the same information in the form of discrete linear events per unit of time. In this respect, and in respect of the frequency and intensity range of their operation, the ear and the mouth work within their mutually compatible ranges (see ch. 2 for further details), and constrain the form of the signal to a high degree.

By contrast, the situation with the written language signal appears to be much less constrained. Many shapes may be generated by the hand–arm system, and many patterns recognised through the visual system: the mutual compatibility of these systems and their characteristic signal properties, are much less at issue than is the case with speech. There is nothing corresponding to the immediacy of the mouth/air/ear transmission chain in written language (though, to be sure, the role of the hand–arm system in signalling to the visual system is much more comparably direct in the case of signing systems such as are found in American or British Sign Languages). Less constrained too is the fashion in which the written language signal may represent the language in question: it may be essentially meaning-based or essentially sound-based, and yet, as we have argued, remain free to mix a range of meaning and sound cues.

It is the product of language awareness within given speech communities; and, in all its various forms, it has been devised as a sufficient and convenient way of representing the language knowledge that has been established first in such communities through the spoken medium.

1.4.2 *Writing the spoken language*

Standard writing systems, even those that are phonographic, do not attempt to represent the properties of the speech signal as such. Instead, they provide clues as to the nature of the linguistic elements being used, and native speakers can use their knowledge of their language to make the connection between written and spoken forms. Thus, for example, when we see a full stop or a comma in written English, we can interpret this in terms of some higher organisation of meaning and grammatical relationship between the elements involved (which may be single words, or longer sequences, representing the parts or wholes of utterances), and can also supply one of the appropriate intonation patterns for the sort of linkage that is conventionally signalled by these devices.

Instrumental representations of the speech signal are also not comprehensive; instead, they focus on certain aspects (some articulatory movements rather than others, or some acoustic properties rather than others), and leave the integration of these with other aspects to the human researcher.

For the scientific study of language, it has been desirable to have a written representation of spoken language, but undesirable to make use of standard writing systems with their reliance on inexplicit native-speaker knowledge. Various systems of phonetic notation have been developed in order to fill this need, and it is convenient to consider them under the headings of *segmental* and *suprasegmental* transcriptions. The fullest sort of phonetic transcription, of course, marks both segmental and supresegmental information, but this is quite rarely used, in practice, outside the field of phonetic analysis. More commonly, and particularly for discussions of grammatical and lexical structure and function, an orthographic transcription is used, supplemented by segmental and suprasegmental phonetic information only where the analyst thinks it necessary.

There are, in general, two problems with traditional segmental phonetic transcriptions for the sort of purpose that we have in this book. The first is that they are too reliant on auditory–impressionistic judgements on the part of the transcriber, and hence may provide a not sufficiently reliable basis on which to build a discussion of, say, speech perception. For this reason, in section 1.2 above we made reference to three types of machine representation of speech.

The second problem is that such transcriptions are loaded with phonetic detail that may be irrelevant for the purpose of representing language units such as words and sentences. Here we have to be careful, however, since it is tempting to 'clean up' a transcription of conversational speech to such a degree that many potentially informative features might be lost. Such features might include hesitations, false starts, filled (e.g. *um* and *er*) and unfilled pauses, repetitions, and so on. In English orthography, a great many of these features can be represented, in terms of their position in the speech stream, and in terms of their nature or identity, in a fairly straightforward fashion. It is also possible to include in such a transcription certain details of intonation (pitch and stress) and certain other details (such as creaky voice, slow onset, etc.). Such details are selected, however, from the full range of speech characteristics that are actually in the signal, so we must bear in mind that the transcription embodies certain decisions about what will be displayed to us for our consideration. We must also recognise that a transcription of this sort does not in any sense give us ready-made language units, such as affix vs word vs phrase vs clause vs sentence, except in so far as (and possibly misleadingly) it provides us with spaces between institutionalised words.

We shall be making use of just such a transcription in chapter 3, where our discussion will focus partly on how we arrive at viable units of analysis out of this sort of representation. For now, it suffices to note that, for all intents and purposes, there is, apparently, no possibility of representing the spoken language on the page in faithful, detailed and comprehensive fashion: it is as well to be aware of this at the outset of a book that will at many points attempt to deal with the spoken language in terms of written symbols.

2
The biological foundations of language

2.1 **Introduction**

2.1.1 ***Preview***

The language signal is generated, and perceived, by the operation of some highly specialised biological systems: auditory and visual pathways from sensory organs to the brain, and motor pathways from the brain to the vocal tract and the hand–arm system. Within the brain itself are ultimately founded not just the representations of the language signal, in its various forms, but also those mediating functions that constitute our general language and cognitive abilities. Before we launch into a consideration of a large and technical research field, we should pause to ask ourselves what we may expect to learn of the nature of language processing from a consideration of what is currently known about these biological systems.

In general, the situation may be likened to one or other of the following: in the best case, monitoring the observable performance of some device such as a television set while systematically inspecting and manipulating its circuitry; in the worst case, speculating on the functions of a building by considering its architectural properties. We cannot expect, in even the best case, that biological investigation will explicate concepts such as 'hearing speech', or 'knowing a language', any more than we would expect to get closer to the images on a TV screen by looking in the back of the box. Our expectations must rather be in the direction of gathering evidence that will eventually constrain our understanding of the principles of language processing.

There is a fairly direct relation between what we know of these biological systems and what we know of the auditory and visual signals conveying language and the articulatory and manual generation of such signals. However, there is a much less clear relation between biological concepts and the abstract phenomenon we know as language. This arises partly as a result of uncertainties in the biology of central processing, as well as differences of view among linguists concerning the formal properties of language; and partly because of the difficult problem of relating abstract systems such as language, which belong to the domain of the mind, to specific brain structures and functions (Eccles 1977). Even so, it is necessary for students of psycholinguistics to have some awareness of (a) the general organisation of language-relevant components in the central as well as the peripheral nervous systems, and (b) the characteristic manner of communication within and between these systems. Even such an elementary introduction to these issues as is attempted here may provide a framework within which constraints on theorising might fruitfully be sought. Calvin and Ojemann (1980), Draper (1980), Espir and Rose (1983), Perkins and Kent (1986), Schneiderman (1984), Selnes and Whitaker (1977), Thompson (1967) and Walsh (1978) may be consulted for further details. Helpful illustrations are to be found in standard anatomical references, such as *Gray's Anatomy* (Pick and Howden 1901), or *Cunningham's Manual* (Romanes 1979).

2.1.2 ***Functional relationships***

First, though, we shall introduce the main properties of the human nervous system, to provide the proper setting for our focus on language processing.

Peripheral versus central nervous systems

The human nervous system is conventionally divided into the *peripheral* and the *central* nervous systems. The central nervous system comprises all the neural tissue contained within the skull (the brain) and the spinal vertebrae (the spinal cord) (Perkins and Kent 1986). The peripheral nervous system consists of all the neural tissue outside these bony structures, and connects the central nervous system with the muscles and sensory organs of the body. The brain occupies the superordinate position in this hierarchy, and thus connections which are involved in the brain's control of muscles are said to be *descending* (or *efferent*), while those involved in carrying sensory information to the brain are *ascending* (or *afferent*). For complex motor and sensory activity, however, both afferent and efferent activity is found, since (a) the brain needs to 'feel' the effect of its motor-control output, and (b) sensory perception may require motor-control adjustments to the sensory apparatus for optimal performance.

Descending and ascending connections between the brain and the body may be of three types; the most general one is *contralateral*, whereby, for example, an area of the left side of the brain controls movement in, or registers sensation from, a portion of the right side of the body (and vice versa); in the facial region, however, *bilateral* connections are found, whereby left and right sides of the brain jointly control muscles on both sides of the face; and *ipsilateral* connections are those (usually subsidiary types) where the same side of

These structures include the structures of the brainstem, and also the interbrain, including the *basal ganglia*, the *limbic system* and the *thalamus* (fig. 2.1). Finally, we should mention here the *cerebellum*, a motor coordinating centre with its own two-hemisphere structure, situated above and behind the brainstem, just below the posterior areas of the cerebrum.

2.2 The auditory system

The human auditory system is a sensory-neural complex which has quite general capacities for processing a range of sounds but which also shows some specialisation for the sorts of sounds that are ordinarily used in speech. It has a total frequency range of between 15Hz to 16kHz (Romer 1971a), but is most sensitive in the 1kHz to 4kHz range; stimuli outside this inner range have to be of greatly increased intensity in order to be audible. The linguistically significant parts of most articulated speech sounds are located within about 600Hz to 4kHz, so the mouth and the ear work within compatible limits of comfortable operation.

The auditory system may be thought of in terms of the following organisation:

the first stage: the outer- and middle-ear system (collection and transmission of the airborne signal) (section 2.2.1);
the second stage: the inner-ear system (mechanical analysis of the signal in the cochlea) (section 2.2.2);
the third stage: the mechanical–neural interface (also in the cochlea) (section 2.2.3);
the fourth and fifth stages: the relays of the sensory system, in the brainstem and subcortical nuclei, and finally up to the auditory cortex (section 2.2.4).

2.2.1 *The outer- and middle-ear systems*

The external ear canal (*sound-wave reception*)

Situated in the *outer ear* (with the ear *pinna* as a collecting device), the *ear canal* has certain acoustic properties as a result of its size and shape (fig. 2.2). Much larger ear canals, as in the elephant, favour lower frequencies, while smaller ones, as in cats, help to extend the upper range to around 70kHz. In humans, the 2–6kHz frequency range resonates within the canal (Stillman 1980), and this has the effect of increasing the intensity of sounds within this range. Perkins and Kent (1986) suggest an increase of two to four times for 2½–4kHz sounds. The canal ends in the *tympanic membrane*, which vibrates in sympathy with these amplified sounds (Stillman 1980). Møller (1983) reviews recent work on tympanic membrane dynamics.

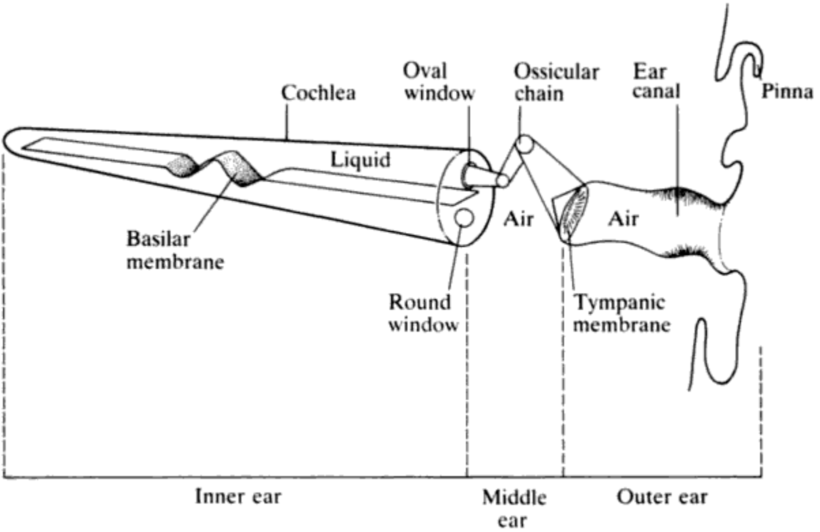


Figure 2.2 Schematic diagram of the outer-, middle- and inner-ear systems. (The basilar membrane is shown displaced in a wave-shape typical of response to a middle-frequency sinusoidal tone.) (Based on Schneiderman 1984: fig. 69, p. 140.)

The ossicular chain (from air waves to mechanical oscillations)

As the tympanic membrane vibrates, the *ossicular chain*, located in the *middle ear*, transfers the frequency, amplitude and temporal characteristics of these movements to the *cochlea* (in the *inner ear*). That is to say, there is a direct correspondence between events arriving at the ossicular chain and those leaving it. What, then, is its purpose?

First, it boosts the incoming signal, preparatory to passing it on to the cochlea. This is necessary because the end of the ossicular chain interfaces with the cochlear liquid, and this is rather resistant to the passage of sound waves. So the ossicular chain transduces the sound waves to mechanical oscillations. These are then boosted to some extent by leverage within the ossicular chain and, more importantly, by the fact that the output area (the *oval window* into the cochlea) is smaller than the input (tympanic membrane) area. In other words, large-amplitude, low-energy compressions and decompressions of the air (the airborne sound) are converted to low-amplitude, high-energy oscillations, or mechanical vibrations, suitable for activating the cochlear mechanisms (see below).

Secondly, the ossicular chain has in-built damping abilities which reduce the potentially harmful effect of very intense sounds.

Thirdly, the ossicular chain has a reflex ability to attenuate low-frequency sounds while preserving (even slightly enhancing) sounds in the higher 1–2kHz range by muscle contraction (Stillman 1980). This is very important for speech, because important acoustic distinctions are carried in the higher-frequency ranges for many sounds. We shall see (below) that certain neuromechanical properties of the auditory system respond in a very gross way to low frequencies, and, because of this, low-frequency sounds have the potential to mask higher-frequency sounds. The ossicular chain thus provides for a first defence against this tendency. Perkins and Kent (1986) estimate that the combined effects of the ossicular chain multiply the mechanical force fourteen times.

2.2.2 *The inner-ear system*

The inner ear is made up of the *cochlea*, which is basically a tube, filled with a liquid, in which is supported a long, flexible structure called the *basilar membrane* (fig. 2.3). The cochlear tube is anatomically a helical structure, like a snail shell. The ossicular chain delivers oscillations to the cochlea via the *oval window*, at the base of the cochlea, to the cochlear liquid, and thence to the basilar membrane. This is narrower, and stiffer, at the oval-window end of the cochlea, and gradually widens, and becomes more flexible, along its length. As a result, it is displaced to its maximum extent at different points, depending on the frequency of the input oscillations: highest-frequency input causes a wave-form displacement of the membrane where it is narrowest (nearest the oval window), and successively lower frequencies result in wave forms which, while propagating from the oval-window end, travel further down the basilar membrane before reaching maximal displacement (fig. 2.3(a)).

Overall, it takes about 5msec. for a travelling wave to reach the farther end of the basilar membrane, and it slows down as it travels (Perkins and Kent 1986). The wave displays a rather abrupt cut-off beyond the point of maximal displacement; by contrast, the build-up to that point is more gentle, particularly for lower-frequency sounds. This means that different frequency sounds have distinct 'wave envelopes' as well as distinct points of maximal displacement (fig. 2.3(b)). It is because lower-frequency sounds yield waves that involve larger areas of the basilar membrane than high-frequency sounds that they tend to have a masking effect. Differences in amplitude of sounds are represented by degrees of displacement of the basilar membrane. It may be appreciated that larger degrees of displacement also involve larger areas of the membrane, so that it is not possible to say, neatly, that frequency is represented purely as distance along the basilar membrane, and amplitude purely as distance of the maximally displaced portion of the membrane from its posi-

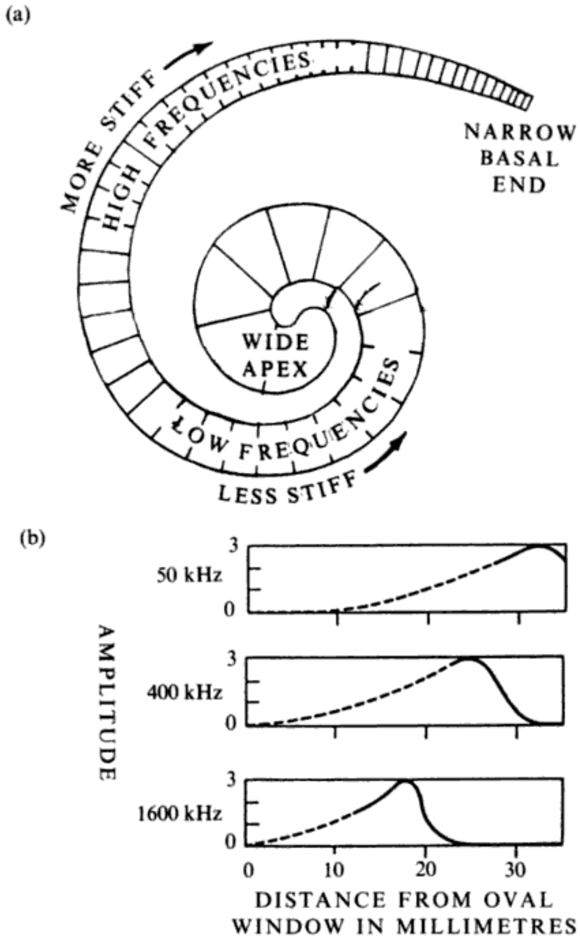


Figure 2.3 Schematic representation of basilar membrane characteristics: (a) the layout of the higher- to lower-frequency response areas; (b) envelopes of displacement for sinusoidal tones of three frequency levels. (From Denes and Pinson 1963: fig. 5, 7, p. 73; Perkins and Kent 1986: fig. 10.1, p. 272.)

tion of rest. The effect of increased amplitude on area of displacement is particularly strong on the higher-frequency side of the wave (the gentler slope of the wave envelope). There is, therefore, an interaction between these two dimensions of sound reception in the mechanical responses of the cochlea. Current understanding of the way the cochlea works builds on the Nobel-prize winning research of von Békésy (see von Békésy 1957).

We may summarise the foregoing by saying that up to this point, the cochlea provides another mechanical analogue of the input signal. But also within the cochlea is another component which has the important task of carrying out the first transduction of the signal from mechanical to electrical response (Stillman 1980): this component is the *organ of Corti.*

2.2.3 *The organ of Corti and the auditory nerve*

The organ of Corti lies along the basilar membrane and effects the link between the movements of that membrane and the nerve impulses which travel along the fibres of the *auditory nerve* (fig. 2.4). (Anatomically, we would strictly have to speak here of the *cochlear nerve*, since the anatomical structure known as the auditory nerve consists of two components; one, the cochlear nerve, involved with hearing; the other, the *vestibular nerve*, part of the system concerned with balance. But we may conveniently refer to the former as the auditory nerve here.)

The organ of Corti consists of a series of *hair-cells* which are attached along the surface of the basilar membrane in such a way that they flex when it is displaced. Each flexing of the hair-cells causes chemical changes which generate small electrical impulses (action potentials) in them that are transmitted along the associated fibres of the auditory nerve. The organ of Corti is responsive to flexing movements in one direction only: that is, on each displacement of the basilar membrane, it responds to the degree and location of movement, and then is passive as the elasticity of the membrane reverses the displacement.

These displacements of the basilar membrane therefore generate patterns of neural impulses which travel along the auditory nerve. The task involved may be appreciated in the light of the number of fibres in the auditory nerve (less than 30,000; by comparison, the optic nerve has about a million), in relation to what they convey, encoding the frequency and amplitude characteristics of the input sound in such an efficient and sensitive fashion that more than 300,000 single tones can be discriminated (Thompson 1967). It is not easy to be clear about how this state of affairs is achieved, as we shall see.

Early attempts to understand this encoding process appealed to (a) the 'frequency' theory (Rutherford 1886), and (b) the 'place' theory (Helmholtz 1863). In our discussion, we shall refer to the first of these as the 'time' theory, to help avoid a confusion between the frequency characteristics of the signal, which are uncontroversial and constitute the datum to be accounted for, and the encoding principles which have been advanced to account for the way in which they might be analysed in the auditory system.

The time theory, then, sees the frequency characteristics of an input sound as being encoded in terms of the timing of the discharge of neural impulses from the organ of Corti along the auditory nerve. Thus, if we assume a rigid

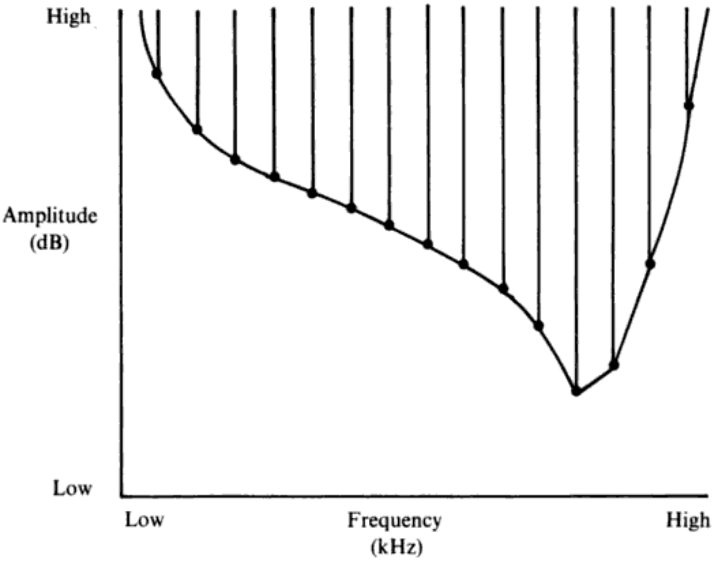


Figure 2.5 Schematic 'tuning curve' of a typical auditory nerve cell. (Lower ends of vertical lines indicate the lowest-amplitude stimuli to which the cell responds at tested frequencies.) (Based on discussion in Thompson 1967: 270–1.)

ing to this theory, adjacent auditory nerve fibres transmit impulses generated on adjacent hair-cells in the organ of Corti, which relate in turn to adjacent regions of the basilar membrane. In this receptotopic arrangement, frequency is encoded as the activation of certain cells and not others. Certain auditory nerve cells exhibit a selective response profile, or 'tuning curve' for particular frequencies, especially those at lower-amplitude levels: see figure 2.5, which shows a fairly typical tuning-curve pattern, with a sharp high-frequency cut-off, and shallower low-frequency slope. Such a response profile acts as a template for frequency analysis of the signal generated in the cochlea. While this interaction of amplitude and frequency in triggering the cell's response is not a problem for the theory, it suggests that the frequency response may be more complex, and variable, than a pure place theory would expect.

More seriously, though, we can see, by referring back to figure 2.3(b), that different wave envelopes along the basilar membrane may overlap to a greater or lesser extent. So, for a given sound at a middle range of amplitude (around 40dB), between 15 and 20 per cent of the basilar membrane and the organ of Corti may be involved. This in turn involves around 10,000 auditory nerve fibres in the transmission of the signal from the cochlea to the higher auditory-processing centres. If the frequency of the sound alters by some small amount, up or down, we may think of a few hundred fibres being lost, or added, at the end points of the activated range, while the central range of activation is maintained. Likewise, small changes in amplitude for a given frequency will show a small proportionate change in the overall pattern of activation. If more than 90 per cent of the action potentials are maintained for slight but perceptible changes of the input sounds, it is a problem to know how the encoding of the input is achieved (Pickles 1982 has further discussion).

For our purposes, it is sufficient to appreciate that frequency and amplitude encoding in the auditory nerve overlap in subtle and complex ways, precisely in the 'speech range' of 1–4kHz. So, to summarise, we may say that:

> the organ of Corti converts the oscillating fluid pressure signal to the form of electrical impulses, for transmission along the auditory nerve and thence to higher auditory-processing centres;
>
> frequency is encoded partly in tonotopic organisation of responses and partly in temporal rate of firing, in ways that are not fully understood;
>
> amplitude is encoded partly in terms of the number and type of auditory nerve cell activation, and partly in terms of the rate of discharge of individual fibres;
>
> within the auditory nerve, there are cells that have 'best frequency' response characteristics, but as amplitude increases a progressive broadening of the response band is observed, especially towards the low frequency end;
>
> frequency and amplitude characteristics of the input signal are observed to interact even at the level of the basilar membrane responses, and it may be that our understanding of how the auditory system encodes these characteristics will improve only with a more precise knowledge of basilar membrane mechanics.

2.2.4 ***The relays of the sensory system***

The cochlear nuclei: the first relay

The auditory nerve conveys neural impulses from the cochlea to the brainstem (fig. 2.6). It enters the medulla high up, near the border with the pons, and connects with groups of nerve cells called the *cochlear nuclei.* These are found each side of the medulla; the left ear connects with the left-side cochlear nuclei, and the right ear with those on the right side. The cochlear nuclei cells effect further refinements in the signal: they effectively extract critical features from the arriving impulse arrays of the auditory nerve. Some respond to the onset of a tone; others discharge continuously while the tone is presented; still others fire to rapid frequency or intensity changes; and so on. A

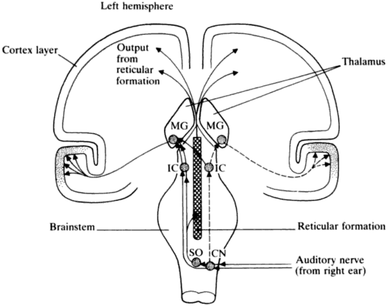


Figure 2.6 Diagram of inputs from right ear, through the higher levels of the auditory system, to the auditory cortex in the left hemisphere; minor input to the right hemisphere is shown by dotted lines. Major cell nuclei in the auditory system are shown as follows: CN, cochlea nucleus; SO, superior olive; IC, inferior colliculus; MG, medial geniculate. (Based on Thompson 1967: fig. 10.28, p. 629; Perkins and Kent 1986: fig. 13.3, p. 350.)

further function is to prevent any spontaneous firing activity (i.e. without stimulus) in the auditory nerve from reaching higher centres and thus reducing the discriminating ability of the system (Thompson 1967).

The superior olivary nuclei: the second relay

Elsewhere in the medulla, in the *superior olive*, extremely important processing of temporal interactions takes place, which requires, crucially, a *bilateral* blending of inputs, from left and right ears. If a sound occurs to the right of your head, there is a small but measurable difference in the time the resulting sound waves will arrive at each of your ears. The right ear will be reached first, then the left. Inter-aural time differences as small as 20µsec. can be processed in the medulla (Stillman 1980). Some medullary neurons respond only to truly simultaneous input to each ear; others only for sounds that arrive at the right ear some critical time intervals before the left, or vice versa. Still other medullary computations involve the 'head shadow' effect, which relates to the drop in intensity at the ear that is further away from the sound source.

This latter is found particularly for high-frequency sounds, and depends in part on the fact that in the medulla, as indeed at every level from the auditory nerve upwards, individual cells are tuned to particular best frequencies, with increasing precision and sharpness of high-frequency cut-off; it also partly results from the fact that the head as a structure tends to filter out lower-frequency components. The medullary coding of sounds in relation to their spatial location, by either method, allows for higher brain processes such as those involved in *attention* to make a selection on the basis of information coming in via other channels.

Fibres from the medullary areas pass through the brainstem bilaterally, with connections being made to the *reticular formation* and the *cerebellum*. The reticular formation is a net-like complex of grey and white matter (nuclei and interconnecting fibres) in the brainstem, and is responsible for integrating as well as relaying sensory inputs, and for readying the cortex as a whole for the arrival of these inputs (French 1957). The cerebellum, while primarily associated with motor inputs and outputs, has a number of sensory inputs including the auditory one, and, like the reticular formation, has rich connections with the cerebral cortex.

The inferior colliculus: the third relay

Further complex intermixing of binaural input takes place in the nuclei of the inferior colliculus, in the midbrain. Neurons of these nuclei include many that are specialised for contralateral (rather than ipsilateral) stimulation (Pickles 1982). Many of the cells exhibit very sharply defined tuning curves, suggestive of further refinement in the processing of the signal characteristics, and cell responses to frequency at this level may be affected by either amplitude or by *frequency modulation* (the phenomenon whereby, in complex wave forms such as are found in speech, certain frequencies of oscillation recur in sufficiently stable cycles to have their own frequency – cycles of cycles, as it were). Pickles (1982) suggests that the inferior colliculus has some of the complex-frequency analysing capacities of the cochlear nucleus, and some of the sound-localising abilities of the superior olive. The inferior colliculus also has a motor output, which appears to be involved in orienting responses (via the superior colliculus, at the top of the midbrain).

Thalamic and cortical auditory systems: the fourth and fifth relays

The major output of the inferior colliculus (apart from that to the superior colliculus) is to an area of the thalamus, represented bilaterally, known as the *medial geniculate body*. While this may be seen as the fourth relay, it has two-way connections with the cells of the auditory cortex, and hence is rather more than simply a relay station (Stillman 1980). Indeed, one of the problems in defining precise cell functions at these highest levels of the auditory system is that their responsiveness depends increasingly on such brain processes as attention, emotion, memory, etc. (in which an important role is played by the reticular formation). In addition, the existence of two-way connections between the cortex and thalamus raises the issue of descending (centrifugal) control in the auditory system, whereby higher centres can 'reach down' to regulate the lower-level processes: about 2 per cent of the auditory nerve fibres carry descending information down to the cochlea and the muscles of the ossicular chain.

Fairly precise locations in the cortex have been demonstrated (e.g. by removing areas of auditory cortex in animals after a period of auditory stimulus training, and observing whether the effects of training are impaired); but interpreting these findings is hard, since the nature of the *training method* used to establish the particular discrimination skill in the animal also has an effect, as well as the actual *portion of cortex* that is removed. It is also not yet possible, from such work with animals, to shed light on the processing of *speech* sounds; but it is possible to show that certain areas of the auditory cortex seem to be involved in *temporal pattern* discriminations. This is important, for the ability to discriminate between a tone sequence ABA vs another tone sequence BAB must involve some form of brief memory (Thompson 1967). This consideration takes us a tiny but welcome step towards the processing of complex temporal sound patterns of speech.

The organisation of the auditory cortex is complex. As for the perception of other sensory modalities, it appears that it is arranged into a 'map' or *projection field*, of the relevant parts of the body (in this case, of the basilar membrane), and that more than one projection field is involved (the cat has six), although one is primary. A special property of the projection field in the auditory cortex may be that, while for *frequency* we can consider the basilar membrane to be laid out on the cortex (as it were, east–west), the dimension of *intensity* is represented at right angles to this (or north–south).

A feature of the organisation of cells in the primary auditory cortex is their columnar arrangement. In certain other areas of the cortex (e.g. the visual cortex) this arrangement is thought to allow for groups of cells of similar response characteristics to be grouped vertically, with sharp differences between them and their columnar neighbours. This in turn may permit rather sharp, discrete, categories of perception. As far as frequency is concerned, however, the columnar organisation of the primary auditory cortex appears to show smooth transitions; but, almost at right-angles to this, the columns mark discrete, step-wise, discriminations of binaural dominance (Pickles 1982).

or up and down. When the eyes are held steady, to maintain a given visual field, they are said to be in *fixation*; normal fixations during the reading process are of the order of 250msec. Moving from one fixation to another is achieved by movements that are called *saccades*. These movements are 'ballistic'; that is, they are executed without an on-line guidance system (like lobbing an object into a particular spot, where everything depends on the accuracy of your preliminary aiming and impelling of the object). They are also very rapid – of the order of 10–20msec., though up to around 50msec. for return sweeps from the end of a line. There seems to be an upper physiological limit of around five saccades per second, so the figure of 250msec. for a fixation is right at that limit. Perception only occurs during fixations, but is nevertheless available, on this basis, during more than 90 per cent of total reading time (at least in principle). Ideally, the procedure for scanning for reading starts with an initial fixation just a few letters in from the start of the first line, covering eight or nine characters; then shifts through 1–4° of visual angle to the right, for the next fixation, which is centred some ten to twelve letter spaces further along the line; and so on (McConkie and Rayner 1975; Bouma 1978; Rayner 1979). In practice, regression movements also occur, particularly if the text is complex, or if the reader is a novice.

The retina of each eye has the form of a hemispherical cup, mounted on a slightly off-centre stalk (the *optic nerve*), which fits snugly into the rear of the eyeball. It is conventionally divided into the *nasal hemiretina* (the area extending from the centre point towards the midline of the head) and the *temporal hemiretina* (the part towards the outer side, or 'temple' of the head). Because the image falling on the retina is passed through the lens, elements in the left visual field impinge on the temporal hemiretina of the right eye and the nasal hemiretina of the left eye; and, conversely, elements in the right visual field are picked up by the left temporal and right nasal retinae. Because of the orientation of the eyes in man, most of our visual field is processed by both eyes (*binocularly*); but there is little or no crossing, or separating, of the responsibilities of the hemiretinae, so that the left and right visual half-fields are normally continuous, without overlap (fig. 2.7). The central area of each retina is particularly sensitive; it has a slight depression, the *fovea*, which has the effect of increasing the retinal surface area in this region, and which is made still more sensitive by virtue of the fact that it has a thinner layer of covering cell-tissue, without blood vessels, and by its composition of photoreceptors (see section 2.3.2). The *foveal field*, at the centre of the visual field (fig. 2.7), is thus particularly well represented on the retinae, and it is those parts of written language that fall within this field that are most available for detailed analysis in the course of reading.

2.3.2 *The retina: the first three stages of signal processing*

From what we have said thus far, it might be thought that the retinae are simply photosensitive membranes which serve the limited function of gathering relevant sense data for higher centres to process. It would, however, be nearer the truth to say that the retinae are actually parts of brain-matter, outfolded from the main mass of the brain (Romer 1971b) in such a way as to be in a position to gather light signals; but their internal cell-structure is sufficiently complex to initiate the processing of these signals too. It may also be that this 'outfolding' concept accounts for the strange orientation of the retinal photoreceptor cells in the eyes – they are so arranged that they point *down* through the retinal tissue, facing away from the incoming light, and towards the outer walls of the eyes. This aspect has been ignored, however, for the sake of convenience, in the discussion and illustration that follows.

The photoreceptors: photo-electrical transduction

Each eye has about 130 million photoreceptors. These are of four types, *rods* and three different types of *cones*. Generally speaking, any area of the retina will be found to have a mixture of these types, but towards the fovea there is an increasingly dense population of cones. Cones require strong illumination, whereas rods are effective in faint light; cones as a group are better for detailed analysis of form (they tend to have one-to-one connections to higher cells), and for colour, while rods give a blurred picture (with many-to-one connections to higher cells) in lighter and darker shades of grey (Hubel 1963; Barrington 1971; Romer 1971b).

These photoreceptors perform the essential first stage in the processing of the stimulus: the sensory transduction into a code of electrochemical impulses. They exhibit a number of horizontal connections, so that different parts of the retinal mosaic (as the banks of receptors are often called) can act together.

Bipolar cells and ganglion cells: the second and third stages

But their chief projections are to the next level of cells (*bipolar cells*) in the retina, where further hierarchical groupings are achieved, and thence to the third and final stage of retinal processing, in the *ganglion cells*. By virtue of their complex connections, these ganglion cells receive input from a number of photoreceptors, and a single photoreceptor may serve more than one ganglion cell (Hubel 1963). In lower animals, ganglion cells are more simply connected to their receptors, and are known to be specialised for certain key visual features, such as *contrast* (i.e. the edge between areas of differential illumination), *movement*, *direction* (of movement), *orientation* (of edge), etc. (see, e.g. Maturana *et al.* 1960, on the frog retina). In mammals, both

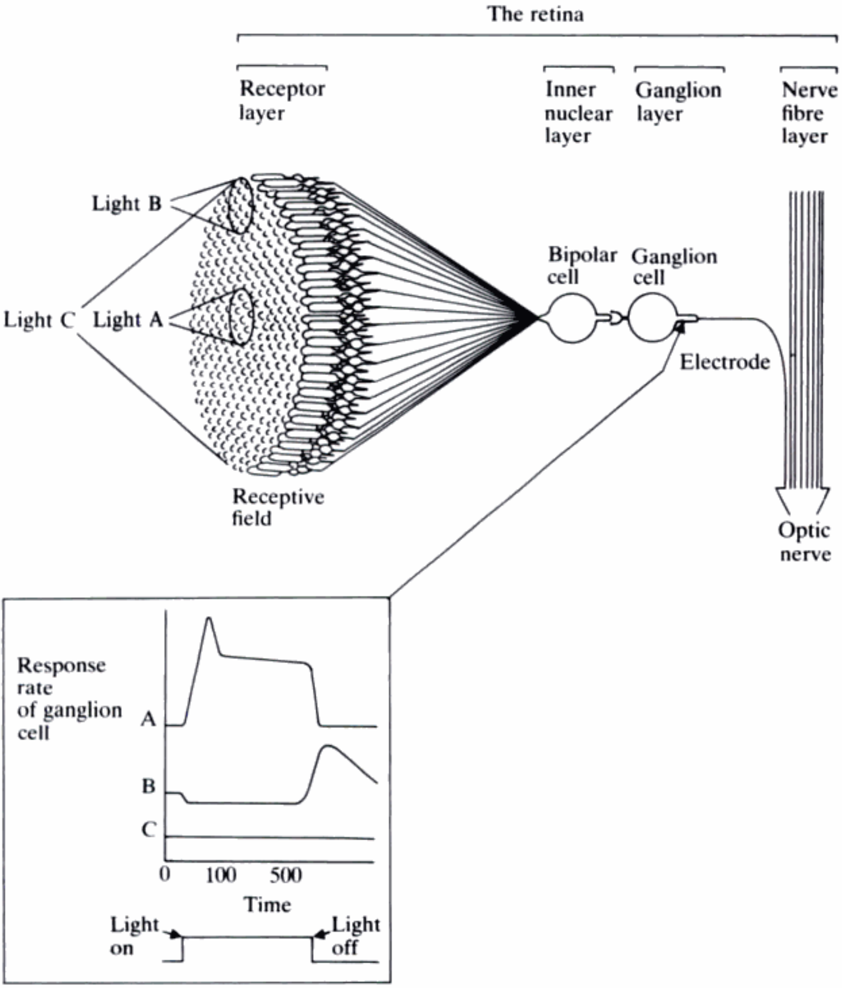


Figure 2.8 Schematic view of a portion of the retina, showing receptive field of a ganglion cell and the corresponding response rate of the cell to three distinct patterns of light stimulation, A, B, C. (Based on Lindsay and Norman: figs. 6.18, 6.19, 6.20, pp. 215–17.)

excitatory and inhibitory connections exist between neighbouring receptors and the ganglion cell served by them. Figure 2.8 (A, B) shows a typical sort of 'on-centre, off-surround' arrangement of a portion of the retina serving a ganglion cell (its receptive field) in the retina of a cat (based on Lindsay and Norman 1977). In situation A, light falling on the central portion of the receptive field causes an increase in the response rate of the ganglion cell (curve A), while, in B, light falling on the surrounding area of the receptive field causes a decrease in ganglion-cell activity, with a brief burst as the light is switched off (curve B). Such a pattern of increased and inhibited cell activity leads to an edge effect in perception. If light falls diffusely over the area C, there may be no increase in the activity of the ganglion cell. In man, it is not certain that such a high degree of specialisation exists at this level as in the case of the lower animals, but it is certain nonetheless that considerable abstraction of stimulus properties does take place. A highly simplified version of this abstraction would posit a limited set of 'trigger' *features*, being processed *in parallel* by specialised cells, with *invariance* of detection as between one feature and another. In practice, invariance is approached rather than achieved, so that, e.g. a cell that fires in response to a particular *edge contrast* (in terms of absolute illumination-levels) may do so only if the edge is in a particular *orientation*, or *location* (in the visual field), or is above some threshold *illumination* value, or shows some combinations of these. Similarly, a simplified view would assume an 'all-or-none' firing response. But this assumption may also require modification, as the response from a particular ganglion cell may be *gated*, or provided with limits of occurrence, in relation to the presence or absence of particular stimulus features. And, within the gated response, differential impulse frequency of discharge may serve as a measure of goodness-of-fit of the stimulus to the characteristics of the detector cell: i.e. a *tuning curve* may exist, as in the case of the auditory system. Any given ganglion cell may thus transmit information simultaneously about, for example, illumination, edge, orientation, etc., thus contributing simultaneously to a number of different higher-level processing systems. It would seem that these stimulus properties are ultimately disentangled by virtue of their parallel transmission by a variety of neurons having different but overlapping sensitivities, along a hierarchical chain. Already, before the signal leaves the retina, something of this hierarchical processing has taken place (Blakemore 1975).

We have already noted that, for lower animals such as frogs, retinal processing goes as far as to distinguish certain environmentally salient feature complexes in a cell-specific way. Most strikingly, certain ganglion cells, highly specialised for convex-edged objects that are moved into and around the visual field, are 'bug-perceivers' (Maturana *et al.* 1960). It would appear that this can only be a partial account of visual perception of objects – otherwise, there would need to be as many uniquely specialised object-detector cells as there are visible objects in the animal's environment (Weisstein 1973). Further, it is not clear that such specialisation exists at this level in man. But it raises the intriguing issue of the link between the optimal modes of operation for ganglion cells (or possibly higher levels of cells) in man and the optimal formal features of man's various writing systems.

2.3.3 ***The optic nerve***

Ganglion cells from all parts of the retina send about a million fibres (Thompson 1967) together from each eye in the form of the optic nerve. This should really be thought of as a brain-internal connecting pathway, rather than as a connection from peripheral to central nervous systems. The fibres are organised in a way that roughly preserves the spatial distribution of the cells over the retina – that is to say, in a *retinotopic* manner. However, cells near the fovea have narrower *receptive fields* than those of the periphery, so the retinotopic bundle of nerve fibres represents the central area of the visual field preferentially over the periphery.

The optic nerve projects to the *lateral geniculate nuclei* (to the rear of the thalamus, very close to the medial geniculate nucleus of the auditory system), and to the *superior colliculi* (also involved in the auditory system) in the mid-brain (fig. 2.7). (Other projections of the optic nerve need not concern us here, as they relate to different phenomena such as the linking of biological rhythms to periods of light and dark, and the mediation of the pupillary response to levels of illumination.)

The projection to the lateral geniculate nuclei is part of the subsystem that eventually projects to the striate cortex area that is concerned with form vision, and which is generally referred to as the *geniculostriate* system.

The projection to the superior colliculi is part of the system which controls eye position and movement (fixation and saccade); because the superior colliculi are in the roof (*tectal* area) of the brainstem, this is generally referred to as the *retinotectal* system (Chamberlain 1983).

The optic chiasm

In the geniculostriate system, the optic nerve fibres innervate the lateral geniculate nuclei in retinotopic fashion; but *en route* from the eyes, a remarkable division of channels has taken place, at the *optic chiasm* (an equivalent term to decussation, from the Greek letter *chi* χ). By virtue of this, the left lateral geniculate nucleus receives input from the right visual field (via the left temporal and right nasal hemiretinae), while the right lateral geniculate nucleus receives that from the left visual field (fig. 2.7). The decussation of fibres in the optic chiasm is roughly 50/50. As a result, each side of the visual system receives an almost identical version of something like the complete half visual-field; *almost* identical, because of the very slight difference in vantage points of the two eyes, and *something like* a half-field because of the slight peripheral area which is not binocular. This 'double exposure' phenomenon, effected at the optic chiasm, lays the basis of stereoscopic vision, which is achieved at higher levels of processing (Romer 1971b).

lobe and the underside of the temporal lobe. Conceivably it is in these regions that integration and synthesis of visual input from the primary visual cortex takes place. However that may be, it is appropriate to conclude this section with the reminder that visual perception is not just the result of extraction processes, but crucially involves active integration, under pressure from *expectations* of *how things should be*. Imposing an active organisation on sensory data thus makes contact with general knowledge as well as specific expectations, and is known to be both fallible and at least partly under voluntary control (witness visual illusions, impossible figures and our ability to transform certain ambiguous stimuli; see Sharpe 1983). And again we return to the point that the role of training and learning may be quite considerable in the sort of visual skill that fluent reading requires, in facilitating synthetic processes of perception that might otherwise be inefficient, prone to error and difficult to transfer from one version of the 'same' stimulus array to another.

For most children, fully stable fixation patterns are observed to be attained by around ten years, considerably after the age at which they are reported to be able to 'read' (Foss and Hakes 1978). The discrepancy here probably relates to developments in central processing, allowing for efficient utilisation of material under fixation. Stein and Fowler (1982) recognise an ocular-motor basis for certain reading difficulties in young children. Since reading small characters at a normal reading distance requires such precision of eye control (down to a quarter of one degree of visual angle), with appropriate convergence of the eyes, they suggest that normally a 'leading (or reference) eye' strategy develops, whereby the signals being perceived from one eye are given priority in determining which foveal image constitutes the relevant input. They further suggest that inducing a 'leading eye' strategy in children who apparently do not use it spontaneously, is helpful in treating early types of reading difficulty. If so, it represents a further illustration of how lower levels (motor control) and higher levels (visual interpretation) are functionally related in the skills required for reading.

2.4 The organisation of language in the brain

Having reached the brain, as it were, by way of the auditory and visual systems, we shall now consider how their outputs to the cortex relate to other areas of language processing in the brain. We shall deal first with the cortex, in its structural and functional aspects (section 2.4.1); then the subcortical pathways and nuclei (section 2.4.2). Finally, we shall consider the roles in language processing of the dominant and non-dominant cerebral hemispheres (section 2.4.3).

2.4.1 ***The cortex***

We shall consider the structural aspects of the cortex first, and then turn to the hypothesised functions.

Structural aspects

Structurally, the cortex presents: (a) a convoluted surface which can be mapped (the *topographical* approach); and (b) a cross-sectional structure which can be described in terms of its tissue composition (the *histological* approach). The latter leads naturally into a consideration of the various cell types and structures (or *cyto-architectonics*).

The topographical features constitute the most obvious starting point, since, as we have noted, the cortex is not smooth but folded into *gyri* standing between small *sulci* or large *fissures* (fig. 2.9). The major fissures are: the *longitudinal* (dividing the two hemispheres); the *central*, or *Rolandic* (demarcating the *frontal* and *parietal lobes*); and the *Sylvian* (separating the *temporal* lobe from the frontal and the parietal lobes). We should also note the *superior temporal* fissure dividing the upper surface of the temporal lobe from the lower. The temporal and parietal lobes share a boundary with the *occipital* lobe which is the posterior portion of the cortex. This is actually not very clearly distinguished in terms of surface features.

Histologically, the cortex is not uniform over the cerebral hemispheres. Overall, it comprises an envelope of roughly six layers of cells, of varying types, depth and organisation, about one-eighth of an inch thick (Perkins and Kent 1986). In detail, it has been mapped into distinct, conventionally numbered, areas (called Brodmann's areas; see Brodmann 1909). In many cases, histological evidence shows gradual, rather than abrupt, changes in cell composition between surface divisions such as lobes, although the occipital lobe does show a sharp histological boundary (Thompson 1967).

Functional aspects

Concerning the fundamental division of the two cerebral hemispheres, we have noted a parallelism of function that may be complementary (in the case of the contralateral connections between body and brain) or mutual (in the case of bilateral connections). But there is also striking asymmetry of function between the hemispheres: one hemisphere tends to be dominant for a range of functions, including handedness and certain aspects of language. Dominance is normally a left-hemisphere characteristic (thus, right-handedness is ensured by left-hemisphere dominance, by virtue of the characteristic contralateral control of the body by the brain). Although there are exceptions, we shall assume the left hemisphere to be the dominant one in

This book is an introduction to psycholinguistics, the study of human language processing. It deals with the central area of this field, the language abilities of the linguistically mature, monolingual adult. It aims to be comprehensive in its coverage, dealing with both written and spoken language, their comprehension and production and the nature of linguistic systems and models of processing.

The book is divided into two parts. Part I identifies and investigates the main contributory areas of study, concerning the nature of the language signal, the biological foundations of language (including auditory and visual systems, the organisation of language in the brain and articulatory and manual systems) and the sources of evidence on the abstract language system. Part II reviews a number of processing models and issues, covering perception and production of speech and writing, lexical storage and retrieval and the comprehension and production of multi-word utterances. The final chapter examines the issues that arise in the context of brain damage and the consequent impairment of language processing (in aphasia and related disorders), an additional important source of evidence and area of process modelling.

Psycholinguistics provides an overview of the major contemporary issues surrounding the psychological foundations of language, most of which have roots in the last decade of research. It assumes a basic grounding in linguistic theory, but it is drawn from the author's considerable experience of teaching this subject, and has thus been designed to be accessible to students of linguistics and psychology, as well as for any reader with an interest in the psychological foundations of human language. It will be an essential work for students and specialists alike.

CAMBRIDGE
UNIVERSITY PRESS

ISBN 0-521-27641-1

9 780521 276412