3 Consider how each of the following sounds might be specified in terms of features.

   a. [b]
   b. [n]
   c. [s]
   d. [x]
   e. [l]
   f. [u]
   g. [e]
   h. [a]

4 Consider how each of the following classes of sounds might be specified in terms of features.

   a. voiced obstruents
   b. fricatives
   c. lateral consonants
   d. high vowels
   e. back rounded vowels

5 Is there such a thing as a natural class of sounds?
6 Can you find any justification for abstract features which are not definable in acoustic, articulatory or perceptual terms?
7 Should features be true universals or is it legitimate for them to take on (slightly) different meanings when applied to different languages?
8 What criteria should be used to evaluate rival feature systems?

11 The Progress of Phonology

Introduction

This chapter draws the book together, by surveying the theoretical development of the subject. The introductory section (11.1) calls attention to the theoretical underpinnings of phonetics and phonology. Subsequent sections proceed more or less historically through perspectives and schools.

The classic perspectives and terminology were developed in the late nineteenth and first half of the twentieth centuries (11.3–11.8). The second half of the chapter deals mainly with generative phonology (11.9) and elaborations of it or reactions to it (especially 11.10–11.14). The emphasis here is on currents of theory: other chapters give more details of phonemics (chapter 4) and generative phonology (chapter 5).

The topics of the sections are:

- phonetics and phonology before the twentieth century (11.2)
- phonemic phonology (11.3)
- the traditions of phonetics (11.4)
- early North American phonology (11.5)
- the Prague School (11.6)
- glossematics and stratificational phonology (11.7)
- Firthian prosodic phonology (11.8)
- generative phonology (11.9)
- natural generative phonology (11.10)
- natural phonology (11.11)
- autosegmental and CV phonology (11.12)
- metrical phonology (11.13)
- lexical phonology (11.14)
- dependency phonology (11.15)
- optimality theory (11.16)
- prosodic phonology (11.17)
- phonology in the laboratory (11.18).

The conclusion emphasizes the fundamental nature of theoretical discussion in any scientific approach to reality (11.19).

11.1 Currents of theory

We began this book on a functional footing, declaring that language has the ultimate function of conveying meaning and that the task of analysis is to investigate how that function is achieved through subsidiary functions, such as articulation and perception (section 1.2 above).

Functional linguists commonly emphasize the systemic and structural organization of language: language functions by virtue of the choices available to speakers, whether choice of words, selection of options within the grammatical system, or exploitation of phonological distinctions. The term 'system' indicates that we operate with the finite options available to us within the language we are using, and the significance of any particular selection within a system rests in the contrast between what is selected and what could have been selected. In a phonological system, for example, the choices are limited and make sense only by reference to the system itself, a point which has long been recognized in discussion of the 'phonemic principle' or 'phonological distinctiveness' (section 4.2 above). The term 'structure' is less precise, being used sometimes in much the same way as 'system', to indicate that choices are made within a 'structured' scheme or framework, but sometimes to refer to the linear organization of language. In this second sense, structure can be contrasted with system, reflecting the two dimensions of linguistic organization that are often referred to as 'syntagmatic' and 'paradigmatic'. Syntagmatic relations are linear or sequential, operative for example in the coarticulation or assimilation of adjacent sounds or in the organization of alliteration or rhyme across longer stretches of language. Paradigmatic relations are those that exist among the options in a system, for example between a word in a text and other words that might have been used in its place or between a phoneme and the other phonemes to which it is opposed. (For an influential functional view of language, see Halliday 1978, 1985a; see also Sampson 1980, ch. 5, for a general account of functional linguistics. Bybee 2001 sets the scene for recent alternative functional approaches to phonology.)

The essential role of theory in such description of language is underlined by the frequent use of '-isms', as in functionalism, structuralism and systemicism. These words are often employed to declare or reject a theoretical standpoint. Loosely, one may acknowledge the importance of linguistic structure by aspiring to 'structuralism' or describing oneself as a 'structuralist'. More strictly, '-isms' may define theoretically limiting positions, so that 'structuralism' may profess a belief that structures are the only true reality and that all meaning and value are achieved within structures. A structuralist in this sense may believe, for example, that human society – as well as language – is a structure, and that concepts or values such as justice, beauty and truth are functional or meaningful only by virtue of their expression within such a structure. Ultimately, when fundamental beliefs of this kind are explored and exposed, they must indeed be revealed as beliefs, resting on human faith or commitment of one kind or another. And this does not mean that we can somehow purify

ourselves of any such belief and stick to solid facts and obvious truths. Rather, all human undertakings ultimately rest on commitment, and even our notions of what counts as science, and of how facts can be shown to be solid, and of which truths are obvious, are themselves reducible to beliefs – to a belief, for example, in the power and primacy of human reason, or in the self-contained and self-justifying nature of the world, or in a reality that is ultimately beyond human understanding and explanation.

That modern science often fails to probe these underlying beliefs is itself a matter of theoretical commitment. Scientific inquiry, as practised in industrialized countries, has been strongly influenced by a belief that concrete reality is the only reality of life, or, less arrogantly, by a belief that it is only material reality that is amenable to scientific investigation or that human experience provides the only valid data of science. The most famous versions of such beliefs are characterized as materialism, empiricism and positivism. Those who explicitly defend these beliefs may set them against what they negatively describe as metaphysics, mentalism or transcendent beliefs – speculation going beyond the evidence of the human senses, for example, or a willingness to accept the existence of that which cannot be directly observed, such as mental concepts or the human mind itself. Positivism is particularly associated with the name of Auguste Comte (1798–1857), a French philosopher who began a modern obsession with the 'positive' data of human experience, but positivism draws on a long tradition of empiricist thinking that makes human experience the determinant of reality. Via such philosophers as Bertrand Russell (1872–1970), a version of positivism (usually known as 'logical positivism') remains pervasive and influential. It bears at least some responsibility for the widely held view that esthetics, ethics and religious beliefs are entirely a matter of individual self-satisfaction or social convention, and that national and political ambitions cannot go beyond an ill-defined pursuit of material well-being for a reasonable number of citizens. For logical positivism allows only logic and mathematics as legitimate undertakings that transcend sensory experience: truth is what can be shown to be true by universal logic (and mathematics) or by observation. Moral and esthetic judgements are not so much wrong as meaningless, for they are neither logically true nor derived from observation.

While positivism can be blamed for many of the woes of modern industrial society – for example for that kind of materialist medicine which treats human beings as purely physical organisms and struggles to find any moral ground on which to base concepts of caring and service – it has undoubtedly had some beneficial effects, particularly in pointing to the limits of human knowledge and the need to study language carefully. Paradoxically, the very search for certain knowledge, based firmly on logic and experience, led to the realization that statements about causation, scientific law and generalizations from experience were just as vulnerable to positivist puritanism as were moral and esthetic judgements. As a consequence, it is now recognized in orthodox science, though not always in popular presentations of it, that one cannot prove a generalization by scientific methods of observation and experiment. One cannot prove, for instance, that a cricket ball thrown in the air will always return to earth: repeated experiments will presumably confirm our experience and demonstrate

that the ball keeps coming down, but they cannot *prove* that this is what will always happen, indefinitely and without exception. At best, the conclusion must be in terms of probability and expectation rather than certainty about the future (always assuming one remains a positivist, unwilling to admit a measure of faith in the predictability of the universe). What *can* be shown by experimental means, quite decisively, is that a generalization is *not* true. Thus the hypothesis that a cricket ball is lighter than air can, if necessary, be *dis*proved by a single experimental trial. Hence positivist science has introduced a note of caution into our formulation of truth, a reminder of the provisional nature of human knowledge. Partly because of this attention to the phrasing of statements, many twentieth-century philosophers, particularly in the English-speaking world, devoted much of their time to the analysis of language, so much so that for some of them philosophy and linguistic analysis became synonymous.

The important questions are often the ones that nobody will admit into discussion, and linguists, in common with other scientists, have at times taken a theoretical position for granted, as if no alternative were even conceivable. But theoretical debate has at times been vigorous and prominent. Leonard Bloomfield (1887–1949), highly influential in the 1930s in North America, explicitly argued for what he called a 'materialistic' or 'mechanistic' explanation of human behaviour, as opposed to a 'mentalistic' view (1933, pp. 32ff.). Bloomfield's explanation belonged within behaviourism, itself an application of positivist thinking to psychology which sought to describe human behaviour entirely in terms of stimuli and responses. This behaviourism denied the existence of nonphysical factors such as ideas and intentions, except to the extent that they could be reinterpreted as physical states or changes in the human organism.

In another instance of theoretical debate, behaviourism itself came under overt attack when Chomsky (1959) reviewed Skinner's account of what he called 'verbal behaviour' (1957). Chomsky not only criticized the behaviourist explanation of language – some would say demolished it once and for all – but provided a spirited exposition of a new mentalism, unashamedly talking of the organization of the human *mind* and concepts. Chomsky's mentalism is more accurately described as rationalism, for it makes reason, or the rational organization of the mind, the central characteristic of human beings. Here again, an '-ism' can be taken in a restrictive or reductionist sense, for rationalism can be accused of overlooking other, nonrational aspects of human life, especially the extent to which reason is constrained and shaped by social motives and structure.

These brief remarks can hardly be regarded as a history of modern philosophy of science but are meant to suggest that philosophical issues are crucial. Even the assumption that one can survey various perspectives or schools of thought presupposes a certain vantage point. It is not uncommon for university teachers to give courses of precisely the kind that pretend to look fairly but critically at all major points of view; but it has to be asked of such surveys where the viewer is meant to be standing. It may be that a degree of superior objectivity is implied, perhaps unstated because it would be too arrogant to express, or that scrutiny of everyone else's views sits oddly with unwillingness to

adopt any position of one's own. Or reluctance to grapple with incompatible alternatives may be dignified by yet another '-ism' as eclecticism. Having said that, we shall attempt a survey of our own, but without any claim to give equal value to everything that has been written about phonetics and phonology, and without any attempt to hide from the reader our own commitment to a functional view of language in which system and structure are foundational. Readers who wish to delve further into theoretical assumptions are urged to consult at least one or two major overviews of the development of modern linguistics such as Robins (1979) or Sampson (1980). These works and their bibliographies give more than adequate pointers to further reading.

11.2 Phonetics and phonology before the twentieth century

Interest in pronunciation is far older than the pursuit of phonetics and phonology as academic subjects. Several centuries before Christ, Indian scholars were devoting themselves to the description of Sanskrit and achieving remarkable accuracy in articulatory phonetics. Although their primary concern seems to have been to maintain the correct pronunciation of what was already becoming a classical language, their observations about points and manners of articulation and other aspects of pronunciation reveal an interest that qualifies as scientific in the best sense of the term (Allen 1953).

Progress is not inevitable: many who came later remained ignorant of this early work in phonetics and did not equal it, let alone improve on it. Modern European civilization owes many debts to Ancient Greece and Rome, but phonetics is not one of them. The Greek grammarian Dionysius Thrax, for example, bequeathed a curious misunderstanding of the nature of voicing. Writing around 100 years before Christ, he recognized that the spoken Greek of his time had both voiceless aspirated and voiceless unaspirated plosives, i.e. both /p t k/ and /p^h t^h k^h/. But he considered voiced plosives /b d g/ to be 'middle', intermediate between the two voiceless types. The resulting habit of labelling voiced consonants with the misleading Latin term *mediae* persisted well into the nineteenth century.

While Greek and Roman scholars did not match the phonetic and phonological brilliance of ancient India, they were interested in related issues, such as the orthographic representation of spoken forms, and it should not be forgotten that the modern European style of alphabetic writing has its roots in the Greek adaptation of Phoenician symbols. The Greek innovation was to develop separate vowel letters alongside the consonants, thus establishing a convention which is now standard in modern European orthographies. By contrast, many other writing systems still use symbols which stand for entire syllables or morphemes or treat vowels as diacritic or subsidiary features of consonants. The Japanese *Hiragana* syllabary, for example, has in principle a distinct symbol for each syllable of the language; and various Semitic writing systems – including

the one which the Greeks adapted – either omit the vowels or write them above or below the preceding consonant. (If English were to follow the Semitic practice, we might write something like $b^a n^a n^a$ or *b'n'n'* rather than the familiar *banana*.) Whether the Greek alphabetic innovation is entirely beneficial remains an open question: we have noted already (especially in sections 3.1 and 10.10) that many of our worries about segmenting speech may be inappropriately influenced by our familiarity with an alphabetic writing system.

Most societies which have developed or adopted a writing system have shown some degree of interest – even if meagre or misguided – in pronunciation or phonological analysis. While spoken language is typically unconscious, writing is far less so, for the product remains before us for inspection and reconsideration (Halliday 1985a, pp. xxiii–xxv; 1985b). The existence of a written form of expression not only invites reflection on the relationship between speech and writing but also creates a distance between speakers and their language that encourages them to treat language as an object of analysis. In China, a system of written characters was in use by 2000 BC, and by the time Chinese scholarship became known to Europeans there was a long Chinese tradition of linguistic studies. Even though the use of characters can scarcely have encouraged segmentation, the Chinese developed an analysis of syllables into 'initials' and 'finals', where the 'final' corresponds to what we might describe as the rhyming portion of a syllable. (Under this kind of analysis, English *sea*, *flee*, *suit*, *flute* might be considered to consist of initials /s-/ and /fl-/ and finals /-i:/ and /-u:t/. It is worth noting that some phonologists are now recognizing units of this kind in languages other than Chinese; see section 11.13 below.) In Korea, Chinese characters were long used to write Korean but an indigenous alphabet, said to have been commissioned by King Sejong, came into use in the middle of the fifteenth century. This alphabet, apparently a genuine local invention and not an adaptation of an existing alphabet, represented a break with character writing, as its 28 letters included separate symbols for vowels as well as consonants. Sensitivity to pronunciation is revealed in the relationship among the symbols – for example, the symbols for fortis voiceless obstruents are essentially doubled versions of the symbols for the corresponding lenis obstruents.

It is of course important not to confuse phonology and spelling. All human languages are spoken languages and can be analysed and described phonologically; but many of them have no written form or have only recently begun to be written. And in any case, some writing systems do not neatly match phonological organization. As we have already had cause to note, English spelling often obscures the patterns of phonological organization. The written form of words such as *psalm* and *psychic*, for instance, suggests that English words can begin with the consonant cluster /ps/, whereas in fact these words begin, in spoken English, with a single consonant /s/, and indeed it is a systematic feature of the phonological structure of English that words cannot begin with clusters of consonant plus /s/. On the other hand, English structure does tolerate words that end with sequences of voiceless plosive plus /s/, i.e. /ps/ /ts/ and /ks/. But this regularity is again obscured in written English, by orthographic devices such as the 'silent *e*' on *apse* and *copse*, or the use of a single letter *x* to represent /ks/ in *fox* and *six*. Nevertheless, written and spoken language are not entirely unrelated to each other, and discussion of the written may sometimes – though certainly not always – reflect insight into the spoken.

In many cases, little survives to testify to the insights and achievements of previous generations. We are fortunate to have any record at all of the work of an Icelandic grammarian of the twelfth century. His main aim was to reform the spelling of Icelandic, which was already being written in an adaptation of the Roman alphabet, but his discussion does indicate some thinking about the phonological organization of the language, and suggests a clear grasp of what we would nowadays call phonemic contrasts, minimal pairs and allophonic variants. The name of this scholar is no longer known and his treatise was not published until the nineteenth century (see Haugen 1972). In quite a different part of the world, Sequoyah (1760–1843), who never learned to speak or read English, succeeded in designing a syllabary for the Cherokee language. He experimented with pictographs before finally adopting various letters from English, Greek and Hebrew (without knowing what these symbols stood for in the source languages) to represent Cherokee syllables. His syllabary was widely used for some time, and seems to be based on a sensible phonological analysis of Cherokee syllables, but we know next to nothing of Sequoyah's thinking in devising the system.

11.3 The phoneme

By the latter part of the nineteenth century, phonetics had been established as part of the modern European scientific enterprise. Interests in spelling and pronunciation were now benefiting from technological advances that made it possible to investigate speech by instrumental methods. At the same time, horizons widened. Where scholars had previously tended to focus on their own languages, the nineteenth century brought, particularly in Germany, a flowering of historical phonology that tried to encompass all the sound changes that had taken place in the development of Indo-European languages. And accompanying this expansiveness was a growing interest in the various spoken dialects of Europe and in hitherto unwritten languages outside Europe, many of which were spoken in areas now under the control of the European colonial powers.

The concept of the phoneme became important not only for its relevance to practical problems such as how to represent the pronunciation of dialects and languages that had never been transcribed before, but also as a keystone of modern phonological theory. In a sense, the word 'phoneme' merely provided a technical term for a concept that was already known – for example to Sanskrit scholars and the Icelandic grammarian. Yet the origin of the term is somewhat obscure, and its meaning continues to be controversial.

The term is usually ascribed to Baudouin de Courtenay (1845–1929), a Polish linguist who taught in Russian universities from 1870. He actually seems to have taken the term over from Kruszewski, a fellow Pole who studied under him from 1878 at the University of Kazan. But the Swiss linguist Ferdinand

de Saussure (1857–1913) had already used the French version of the term 'phoneme' in an article published in 1878, and he in turn had adopted the word from French predecessors who had almost certainly used it as a convenient translation of the German *Sprachlaut*, 'language sound'. It would be wrong to suppose that all of these early users of the term meant exactly the same by it. The common thread is the need to treat discernibly different sounds as a single sound for functional or descriptive purposes; but it is evident that early usage, foreshadowing continuing disagreement, was not uniform (Anderson 1985, pp. 65–82).

Baudouin de Courtenay, whose own use of the term 'phoneme' seems to have shifted during his lifetime, fell back on what is now commonly described as a 'psychological' or 'intentional' definition of the phoneme. This definition proposes that the phoneme represents a mental image or intention and that variants or alternate realizations of the phoneme are to be regarded as different actualizations of a single underlying 'ideal' or 'intended' sound. This mentalism – or versions of it – achieved some popularity among European scholars and was kept alive in North America by Sapir (section 11.5 below), who wrote unapologetically of the 'psychological reality' of phonemes. It was nevertheless overshadowed by alternative conceptions before being revived by modern generative phonology in the 1960s. Chapter 3 of Anderson (1985) provides a detailed account of the work of Baudouin de Courtenay and Kruszewski.

11.4 The traditions of phonetics

In Britain in particular, phonetics was already a creditable pursuit in the sixteenth and seventeenth centuries. No doubt encouraged by – and contributing to – the strong empiricist flavour of the British scientific tradition, phoneticians such as Henry Sweet (1845–1912) and Daniel Jones (1881–1967) were more interested in the description and transcription of speech than in the concept of the phoneme as a matter of theory. Sweet was aware of German linguistic scholarship and reportedly somewhat hostile to it. He did not use the term 'phoneme' at all, but did distinguish between 'broad' and 'narrow' phonetic transcription: broad transcription recorded speech in symbols that were sufficient to convey the relevant distinctive differences, whereas a narrow transcription included phonetic information of the kind which was not contrastive within the system but which might be of importance to the dialectologist noting precise details. Thus a broad transcription of English Received Pronunciation (RP) might show simple voiceless plosives; a narrow transcription might show that these plosives are also markedly aspirated. Sweet's broad transcription is at least roughly equivalent to a phonemic transcription, and Sweet's notion of the phoneme (if it can be called that) is of a functionally distinctive unit rather than a psychological entity.

Jones maintained Sweet's use of 'broad' and 'narrow' and was familiar with the development of phonology in Europe. Under his leadership, University

College London became a centre of practical phonetics, and Jones himself was renowned for his documentation of English pronunciation and his attention to training in articulatory and auditory skills. He tended to regard phonology as subsidiary to phonetics and is considered to be the author of the 'phonetic' view of the phoneme. In this he retreats even from Sweet's recognition of distinctiveness and describes the phoneme as a set of similar sounds that are in complementary distribution. The extreme empiricist flavour is not surprising but it must be said that Jones was not entirely consistent and that he continued, for example, to recognize the practical importance of minimal pairs, which clearly reflect phonemic contrasts (section 4.2 above). Jones's successors have continued his high standards of phonetics, often still oriented to 'ear training' and the teaching of English pronunciation to foreign learners, but have been generally less suspicious of phonological theorizing. Gimson's work on English phonetics (1980, first published in 1962), essentially a successor to Jones's *An outline of English phonetics* (first published in 1918), straightforwardly acknowledges the phoneme as a contrastive unit.

Jones's own views are set out in a paper on the history and meaning of the term 'phoneme' (Jones 1957) as well as in a book-length treatment of the phoneme (1962).

11.5 Phonology in North America

Franz Boas (1858–1942) was born and educated in Germany but settled in the United States after he had begun to study American Indian culture. An anthropologist rather than a linguist, he stressed the need to respect the diversity of culture and to study a cultural system (including language) on its own terms. He laid the foundation for phonetic and grammatical studies of American Indian languages, and influenced men like Edward Sapir (1884–1939) and Leonard Bloomfield, who combined high standards of scholarship with an enthusiastic interest in recording and analysing unwritten languages. Sapir's phonology was explicitly 'mentalist' (section 11.3 above), while Bloomfield allied himself with the new behaviourist psychology and began a tradition of linguistic description which, taken at its worst, can be accused of studying linguistic forms without proper regard for meaning.

Sapir's understanding of phonology is set out in two influential papers. The first, on 'Sound patterns in language' (1925), promotes the psychological reality of sounds within a linguistic system and contends that there are ways of determining the 'place' of a sound in a system that go beyond the articulatory and acoustic nature of the sound (cf. section 10.7 above). The second paper (1933) is explicitly entitled 'The psychological reality of phonemes' and appeals to evidence from field work on North American Indian languages. Sapir's examples are well worth study and reflection. In one account he describes how a speaker of Sarcee (Alberta, Canada) felt that two words in his own language differed in pronunciation even though he could not substantiate this from the

pronunciation itself; Sapir shows how he later came to understand that this was because the two words differed morphophonemically and compares this with the way in which even English speakers who pronounce *soared* and *sawed* identically might still 'feel' a difference between the two words because of their awareness of related forms such as *soaring* and *sawing*. In effect, Sapir is suggesting that we can hear what is not there in the phonetic record, by what he calls 'collective illusion'.

Bloomfield's views, as set out in his major work *Language* (1933, esp. ch. 5), have proved more influential than Sapir's but are in some ways contradictory. He professes a materialist concern with 'actual speech' but none the less refers to the distinctiveness of phonemes in ways that would be acceptable to Prague School functionalists (section 11.6 below). Chapter 6 of *Language* is a survey of articulatory processes entitled 'types of phonemes', which risks some confusion between the phoneme as a distinctive unit within a system and the speech sound as a convenient descriptive device of general phonetics, although he does speak of some sounds as 'variants' and others as 'separate phonemes'; and he concludes the chapter by noting the possibility that the 'same phoneme' may be produced by quite different articulatory mechanisms. Bloomfield's rather programmatic view of phonology and the ways in which it was taken up and elaborated are further discussed in Fischer-Jørgensen (1975) and Anderson (1985, chs 10 and 11).

In the 1940s and 1950s followers of Sapir and Bloomfield vigorously debated and applied the principles of phonology. Some of the exchanges – for example about whether phonemes were 'physical' or 'fictitious' or whether one could analyse a language phonologically without knowing any of the grammar – may strike the modern reader as pedantic. Nevertheless, much of our modern terminology, such as 'allophone' and 'complementary distribution', was elaborated in this period, and the American experience of analysing and discussing American Indian languages has proved normative for much comparable work done elsewhere, for example on indigenous languages of Australia and Papua New Guinea.

The continuing interest in developing analytical techniques is reflected in Kenneth Pike's *Phonemics* (1947), significantly subtitled *A technique for reducing languages to writing*. The book remains unusually thoughtful and comprehensive, and has been studied by hundreds of field linguists, many of them missionary linguists working with Wycliffe Bible Translators, an organization which Pike helped to found.

Though it was certainly not Pike's intention to restrict phonology to a matter of analytical technique and orthographic design, questions of transcription have often been dominant in modern phonology, especially in the English-speaking world. Sometimes neatness seems to become an end in itself. A classic North American example, originally proposed by Trager and Bloch in 1941 but later modified, is an analysis of English vowels into six, namely /i e a o ə u/, representing the vowels heard in *pit*, *pet*, *pat*, *pot*, *cut* and *put*. Additional vowels are accounted for by postulating that each vowel may be followed by /j/ /w/ or /h/, so that, for example, *beat* can be represented as /bijt/, *boat* as /bowt/ or /bəwt/ and *law* as /loh/. Thus we have an array of 24 vowels, as shown in table 11.5.1.

Table 11.5.1 Trager and Bloch's vowel transcription system (originally due to Trager and Bloch 1941)

Simple vowel	i	e	a	o	ə	u
Vowel + j	ij	ej	aj	oj	əj	uj
Vowel + w	iw	ew	aw	ow	əw	uw
Vowel + h	ih	eh	ah	oh	əh	uh

While there may be a certain appeal in economizing on symbols and making a symmetrical table, it has to be conceded that not all of the possibilities occur in any one variety of English. Trager and Bloch are reduced to noting that /əj/ occurs in a New York City pronunciation of *bird*, but not in General American, that /əh/ occurs in (some varieties of) British English (*burr*, *furred*), and so on. Thus the transcription becomes a general notational scheme for varieties of English rather than an analysis or description of a phonological system, and its attractiveness rests in the neatness and potential of the notation rather than in systemic validity.

Further details of Trager and Bloch's analysis, which has remained influential in the USA, can be found in Trager and Bloch's 1941 paper (Makkai 1972, pp. 72–89, with notes on pp. 4, 72). See also Gleason's adaptation of the system and commentary (1961, pp. 27–39, 320–5).

11.6 The Prague School

By the 1920s, the terms 'phoneme' and 'phonology' were well known to European linguists. More importantly, de Saussure (section 11.3 above) had left a legacy of modern structuralism which greatly influenced linguistics in general. Working within this structuralist tradition were, among others, a group of scholars known from 1926 as the Linguistic Circle of Prague. In phonology, two members of the Circle stand out: Roman Jakobson (1896–1982), who began his career in Moscow but moved to Czechoslovakia (now the Czech Republic) and worked there in the 1930s before fleeing via Scandinavia to the USA; and Nikolai Trubetzkoy (1890–1938), also of Russian origin, who was a professor in Vienna from 1923 until his death.

Following de Saussure's emphasis on the differential function of linguistic elements, both Jakobson and Trubetzkoy attached great importance to the OPPOSITIONS among phonemes rather than to the phonemes themselves. Thus to say that English has phonemes /s/ and /z/ is a statement about a distinction which English speakers make and recognize rather than a claim about phonemes as mental images or phonetic entities. This was a significant insight, which seemed to accord with linguistic experience. By the very nature of spoken language, a speaker is aware of differences and reacts to mispronunciation or interference with the system of oppositions ('Was the name *Buss* or *Buzz*?',

'Did you say *sip it* or *zip it*?', and so on). But the isolation of individual phonemes from their spoken context is neither a typical nor an easy task. Most speakers seem incapable of doing it in any systematic way, and, in literate societies, usually resort to naming letters and spelling out a word rather than attempting to articulate separate phonemes.

Jakobson (and others of the Prague School) published actively during the 1920s and 1930s, but it was Trubetzkoy who provided the School's most comprehensive and widely consulted work on phonology, *Grundzüge der Phonologie* (Principles of phonology), which first appeared in 1939, the year after his death. Besides discussing the nature of distinctive oppositions in theoretical terms, Trubetzkoy also surveys analytical procedures ('rules' for determining the phonemic system of a language) and gives extensive examples of the different oppositions of various languages. He follows through the implications of the structural approach in a number of ways, particularly in the classification of oppositions. For example, some oppositions within a language are 'proportional', i.e. distinguish more than one pair of phonemes, while others are 'isolated', i.e. are restricted to just one pair. In English, for instance, the voicing distinction is proportional (relevant for p/b, t/d, k/g, f/v, etc.) whereas the l/r opposition is isolated (no parallel cases). Trubetzkoy is also responsible for the concepts of 'neutralization' and 'archiphoneme' (section 4.9 above), which are consistent with a functional view of the phoneme. For, if the phoneme is characterized by its opposition to other phonemes, then it follows that the /p/ in words such as *spin* and *spa* (where there is no potential opposition to /b/ in *sbin* or *sba*) is of different functional status from the /p/ in words such as *pin* and *par* (where there *is* opposition to /b/ in *bin* or *bar*).

Jakobson and Trubetzkoy also initiated modern distinctive feature theory. The notion of component features is already implicit in the idea of opposition: /n/ is nasal by opposition to /d/, alveolar by opposition to /m/, and so on. The notion was made explicit by Jakobson's and Trubetzkoy's recognition of such features as 'differential qualities' or 'relevant properties' (section 10.5 above). This further strengthened their point that phonemes represented points in a system rather than physical or mental entities. It was now possible to conceive of the phoneme as a 'bundle' of distinctive features, a simultaneous set of oppositions. (For further details of the Prague School phonologists and their concerns, see Fischer-Jørgensen 1975, ch. 3; Anderson 1985, ch. 4.)

11.7 Glossematics and stratificational phonology

Glossematics is much more than an approach to phonology. It is a general theory of language, elaborated by two Danish linguists, Louis Hjelmslev (1899–1965) and Hans Jorgen Uldall (1907–57). Glossematics is neither popular nor widely understood, but has exercised some influence on the development of phonology (which within glossematics is termed 'phonematics'). Hjelmslev's presentation of this theory at the Congress of Phonetic Sciences in London in

1935, which drew approval from Jakobson, affirmed that a phoneme must be defined by means of its function in language, not by physical or psychological criteria. For Hjelmslev, linguistic function included more than distinctive opposition, and he was not averse to classifying and interpreting sounds on the basis of their distribution and alternation. Accordingly, he entertained such possibilities as analysing French /ɛ:/ as /ɛə/ and Danish /ŋ/ as /ng/ (Fischer-Jørgensen 1975, p. 134). His tolerance of a high degree of abstraction is also evident in the positing of a phoneme /h/ in French (Anderson 1985, p. 158): the /h/ is entirely abstract in that it is never pronounced, but it serves to account for lack of elision. Thus words on the left below begin with a vowel (despite their orthographic *h*) and the preceding article *le* is reduced to *l'*; those on the right also begin with a vowel but show no such elision and are therefore credited with an initial /h/ which is unpronounced but blocks the elision:

l'habit ('the clothes')	le havre ('the harbour')
l'harnais ('the armour')	le haricot ('the bean')
l'homme ('the man')	le homard ('the lobster').

Stratificational phonology is, again, part of a wider theory of language. Developed in the USA in the 1960s, it falls within the broad tradition of Saussurean structuralism and shows particular influence from glossematics, notably the emphasis on language as a network of relationships rather than a set of elements. The stratificational view is that language is organized on distinct levels or 'strata', the one of most relevance to phonology being the 'phonemic stratum'. The units of this stratum, phonemes, are represented as points in a network which links each phoneme in three directions. Oversimplifying somewhat, phonemes are

1 realizations of morphemic elements;
2 subject to the phonotactics (i.e. the pattern specifying how phonemes can be sequentially combined);
3 realized as (combinations of) features.

In a full display of relationships, the English phoneme /k/ would therefore be linked to

1 each element which it realizes (the first segment of the morpheme *cat*, the second of *sky*, the first and last of *critic*, and so on);
2 the tactic pattern determining that /k/ can follow initial /s/, can precede /r/, and so on;
3 the various features by which it is realized, i.e. [voiceless], [dorsal], etc.

In fact the network is even more complex than this suggests, for realizations are mediated via alternation patterns. Viewed from the morphemic stratum, the first and last elements of the morpheme *critic* are not identical: the first is always realized as /k/ but the last may be either /k/ (as in *critic*) or /s/ (as in *critic-ism*). Such alternation is handled by saying that the segmental components of morphemes

are 'morphons' (not phonemes) and that morphons are realized as phonemes. In English there will be one morphon which is realized only by the phoneme /k/, another which is realized by /k/ and /s/ in alternation with each other.

Elaborations of this kind make stratificational phonology intricate both in its terminology and in its diagrammatic displays (see Lamb's foundational work, 1966a, b). Whether for this reason or because it was overshadowed by the greater popularity of generative phonology in the USA in the 1960s and 1970s, it has relatively few champions today. Nevertheless, stratificational phonology is an impressive outline of a structuralist perspective which incorporates many important concepts such as levels of organization, phonemes and phonetic features. Its bold attempt to formalize the entire network of relationships (including phonotactics and alternations, not just phonemic oppositions and allophonic realizations) is an improvement upon the simpler varieties of North American phonemics, and it deserves better than to be submerged in the shifting of the phonemic concept itself towards a more morphophonemic notion.

Both glossematics and stratificational phonology receive detailed attention in Makkai (1972), under the heading of 'The Copenhagen School and stratificational phonology', and in Fischer-Jørgensen (1975). In addition, glossematic phonology is evaluated in considerable detail by Anderson (1985, ch. 6) and stratificational phonology is reviewed by Sommerstein (1977, ch. 4).

11.8 Firthian prosodic phonology

Like stratificational phonology, Firthian phonology offered thought-provoking insights into phonology but never achieved a wide following. Unlike stratificational phonology, it questions the centrality of the phoneme as a segmental unit. Traditionally it was known as 'prosodic phonology', but this now usually refers to another theory that deals with prosodic or phonological domains, and prosodic structure at several levels (see section 11.17 below). In this brief overview, we will use the more current terms 'Firthian prosodic phonology' or 'Firthian prosodic analysis' (FPA).

The founder of Firthian prosodic phonology was J. R. Firth (1890–1960), who held a chair at the School of Oriental and African Studies in the University of London from 1944 to 1956. Firth himself wrote only a few papers on phonology but his ideas were developed and applied by pupils and successors, who practised 'Firthian linguistics' or belonged to the 'the Firth School'.

Firth broke with the English tradition of Sweet and Jones (section 11.4 above) and tried to take English phonology away from its preoccupation with phonetic description and segmental transcription. His starting point was solidly structuralist. He recognized systems as reflections of paradigmatic oppositions (a set of phonemes in opposition to each other is a system) and structures as reflections of syntagmatic relations (a syllable, for example, is a sequential structure). But he disputed the traditional concern with systems (especially phonemes) at the expense of structures. He drew attention to the subphonemic

components of speech and to the extent to which such features may spread across successive segments. In this respect he and his followers have something in common with others who, from the 1930s on, were increasingly interested in features as the ultimate elements of phonological description.

The thesis of the Firthian school is that various components of the flow of speech do not lend themselves to analysis into discrete segments. This point has always had force regarding intonation and other such 'suprasegmental' phenomena (chapter 9 above), but the Firthian school extended it to other features such as nasality and lip rounding. A favourite example to demonstrate the point is the pattern of vowel harmony found in languages such as Turkish and Hungarian. In general, vowel harmony means that successive vowels agree in certain features. Under Turkish vowel harmony, a vowel other than the first in a word may be low unrounded or high: other features of these noninitial vowels are simply taken from the first vowel. For example, 'my house' is *evim*, 'my nation' *ulusum*, 'my arm' *kolum* and 'my rose' *gülüm*. In all four examples, the high vowels of noninitial syllables copy their backness and roundedness from the first syllable. All four also end in what is grammatically the same suffix meaning 'my'. Other examples can be found in table 4.7.1 above, where the plural suffix is seen as *-ler* after front vowels and *-lar* after other vowels, and the genitive suffix (like the suffix 'my') has different high vowels agreeing with the backness and roundedness of the preceding vowel. Table 4.7.2 (repeated below as table 11.8.1 for convenience) shows the vowels as two systems, reflecting the way in which the phonemic options are constrained by the operation of harmony. The notion that the vowels constitute two (sub)systems is itself in keeping with Firthian thinking.

Vowel harmony is of course not just a matter of vowel articulation, but of pervasive tongue or lip settings that must affect intervening consonants as well, even if less audibly. Hence a Firthian prosodist would argue that the fronting and rounding of the vowels is not located in the vowels but extends throughout the relevant stretch of speech. (Compare remarks on coarticulation in section 4.1 above.) To capture this in the notation, we can extract the pervasive features and show them as 'prosodies' of the word (or other appropriate unit). Notice that this implies that the vowels themselves do not carry any marking

Table 11.8.1 Turkish vowels

(a) FULL SYSTEM (in first syllable of a root)

	Front		*Central/back*	
	Unrounded	*Rounded*	*Unrounded*	*Rounded*
High	i	ü	ı	u
Low	e	ö	a	o

(b) SUBSYSTEM (in noninitial syllables, including suffixes)

High	I
Low	A

I is realized as /i/, /y/, /ɨ/ or /u/ according to harmony.
A is realized as /e/ or /a/ according to harmony.

as (non)front or (un)rounded but simply take on these values from the prevailing prosody. Using /V/ for a high vowel (unspecified for frontness or rounding) and superscript /y/ for frontness and /w/ for lip rounding, we might then transcribe *ulusum* as /ʷVlVsVm/ and *gülüm* as /ʷʸgVlVm/. Under such an analysis there is only one other segmental vowel besides /V/ – let us call it /A/. This vowel is low (or nonhigh), again taking its specification as (non)front from the prevailing prosody. Thus the analysis differs sharply from one that recognizes eight vowel phonemes subject to certain constraints on their sequential cooccurrence. Here there are two phoneme-like units and two prosodies. The phoneme-like units are in fact termed PHONEMATIC UNITS within prosodic theory, underlining their difference from the phonemes of, say, North American phonemics or the Prague School.

Prosodies in the Firthian style are not limited to extensive components but also include demarcative phenomena of various kinds, such as the English 'intrusive' /r/ in e.g. *draw(r)ing* or *banana(r)oil*, or the German glottal stop in a word such as *Be*[ʔ]*amte* ('official'), where the consonantal element may be thought of as a boundary marker between two successive vowels or syllables, rather than as a phoneme of the same status as other consonants. And a further difference from most versions of phonemics is that Firthian prosodic phonology is explicitly polysystemic, in the sense that the set of phonematic units need not be considered a single undifferentiated set. In Firthian prosodic analysis, the vowels of Turkish are clearly not a single system of eight phonemes (although they do turn out, arguably, to be a single system of two phonematic units). More generally, wherever there are restrictions on the distribution of phonemes within units such as syllables or words, sets of phonemes occurring in specific positions may constitute distinct subsystems. Thus in English the system of consonants standing syllable-initial before /r/, namely

/ p t k b d g f θ ʃ /,

is not the same as the system that applies syllable-initial before /l/, namely

/ p k b g f s /.

Sommerstein gives a helpful outline of Firthian prosodic phonology (1977, ch. 3) and Lyons contrasts the prosodic approach with North American phonemics (1962). Both authors explain the relevance of Turkish in further detail and give references to works by Firth and other prosodists. Firthian prosodic analysis has never gone away and was championed from the 1980s by a group of researchers with links to York University in England (e.g. Kelly and Local 1989, Ogden and Local 1994). See Ogden and Kelly (2003) and Coleman (2004) for their timely reaffirmation of Firth's contribution to phonological theory.

11.9 Generative phonology

After moving to the USA, Roman Jakobson (section 11.6 above) continued a distinguished career as a linguist, and in the 1950s joined forces with Morris Halle

(also in the USA) and Gunnar Fant (working in Sweden) in the development of distinctive feature theory (section 10.5 above). In the 1960s Halle and Chomsky, at the Massachusetts Institute of Technology, elaborated a new approach to phonology which came to be known as generative phonology (chapter 5 above).

Generative phonology belonged to a new school of linguistics, transformational–generative theory. Those who embraced this theory were critical of prevalent interests, particularly in North America, and Chomsky himself accused his 'structuralist' predecessors of undue concern with inventories of elements and a classificatory or 'taxonomic' approach to linguistic analysis. Instead, linguistic description ought to aim to construct a grammar that would 'generate' linguistic forms. The phonological component of such a grammar would be a set of phonological rules applying to the underlying forms of the language and yielding surface phonetic representations. Since both underlying and surface forms were represented in features, the rules essentially changed feature specifications (section 5.3 above), and the shape of a phonological description was indeed radically different from a typical inventory of phonemes and allophones. Moreover, the attention to the formal conventions governing rules and their operation (sections 5.4–5.6 above) went hand in hand with a new interest in what was phonologically possible and what impossible in language. The formalism, often offputting to the newcomer, did bring explicitness to hypotheses about phonological organization and supported a new emphasis on claims about the nature of human language.

Orthodox generative phonology is part of a model of language (more strictly a model of 'linguistic competence') which proposes that underlying representations are converted into surface representations by the application of rules. The model went through several modifications in the 1960s, and one version of it is presented in figure 11.9.1. The model shows phonology as a component 'fed' by a syntactic component that generates grammatical sequences of the language.

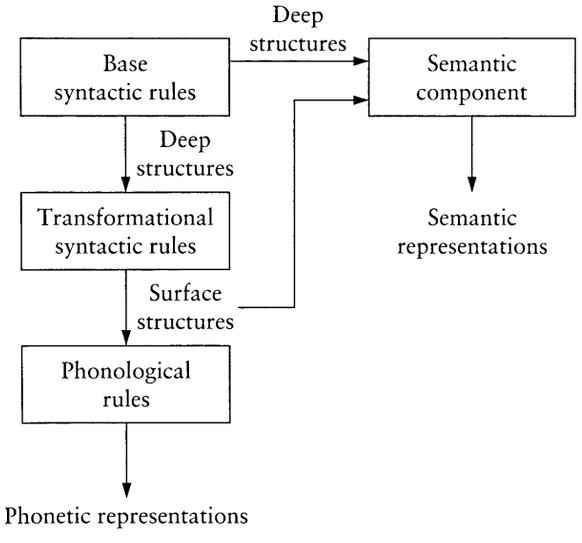


Figure 11.9.1 A generative model of grammar

These grammatical structures – so-called 'surface structures' – are complete with lexical items and reflect the grammatical rules of the language. The lexical items in surface structures bring with them their underlying phonological representations in the form of feature matrices. The surface structures serve as input to the phonological rules, which, responding both to underlying phonological representations and to their syntactic and phonological contexts, generate a phonetic representation.

The model is an idealization in that it portrays the competence of an 'ideal speaker–hearer'. Indeed, generative scholars explicitly contrasted competence and performance, excluding 'performance factors' from consideration (Chomsky and Halle 1968, pp. 1–3). Competence is viewed as knowledge, and the generative model is meant to have psychological import. Thus a grammar (in one sense of the word) is competence represented as rules: the grammar is 'internalized' by speakers, constructed from data in the process of acquisition, that is, and used in linguistic performance (Chomsky 1964, pp. 8–10). Chomsky and Halle specifically propose that phonological representations 'are mentally constructed by the speaker and the hearer and underlie their actual performance in speaking and "understanding"' (Chomsky and Halle 1968, p. 14).

Details of the generative approach have been given in chapter 5 above, and it is sufficient here to note that the abstraction and mentalism of orthodox phonology were contentious. Particularly in the USA, generative phonology was followed by a number of reactive movements. None of these could afford to ignore the pre-eminence which generative linguistics achieved in the 1970s: some of them stressed their disagreement with aspects of the generative orthodoxy, others claimed rather to be modifying or refining the generative model. But whatever their stance, these phonological programmes brought fragmentation into the generative tradition, and, in one way or another, undermined the generative thrust against the more traditional concerns with distinctiveness and structure. Several of these postgenerative movements are reviewed in the following sections.

11.10 Natural generative phonology

Natural generative phonology (NGP) emerged from a number of papers by Vennemann in the early 1970s and is most comprehensively expounded by Hooper in a 1976 book (Vennemann 1972, 1974a, b; Hooper 1976). As the title of this 'school' suggests, its proponents do not claim to depart radically from the mainstream of generative phonology. They describe their school as 'based in part on transformational generative theory as developed since the mid-1950s' but point to a major difference concerning the 'abstractness of phonological representations and rules' (Hooper 1976, p. xi).

In fact, NGP is quite radical in its attack on abstractness, though less now than in its earliest formulations. At one stage, Vennemann had proposed to rule out any underlying form that was not identical to a surface form: if a morpheme showed no alternation, then its underlying form must be identical

to its surface form; if there was alternation, then the underlying form must be identical to one of the surface allomorphs. Hooper herself assesses this proposal and states that it goes too far (1976, pp. 117 ff.). Consider, for example, pairs of words showing different vowels reduced to [ə], depending on where the stress falls, such as

melody ['mɛlədi]	melodic [mə'lɒdɪk]
heretic ['hɛɹətɪk]	heretical [hə'ɹɛtɪkəl]
demon ['di:mən]	demonic [də'mɒnɪk]
telephone [tɛlə'foʊn]	telephonist [tə'lɛfənɪst].

A strict constraint on abstractness would mean that one of the surface forms would have to be chosen as underlying. But, of each pair of forms given above, neither seems genuinely underlying in the context of a generative description: if the term 'underlying form' has any value at all, the root should not contain any occurrence of [ə], as this vowel is derived by reduction from other vowels.

Hooper's solution is to abandon abstract underlying forms altogether – although she does revive the concept of archiphonemic representation of the kind entertained by Trubetzkoy and the Prague School phonologists (sections 4.9 and 11.6 above). Rules are now to be regarded as generalizations across surface forms rather than as the means of *generating* surface forms. Hence Hooper is able to say that within NGP, rules and representations are directly related to surface forms and that phonological analysis is more concrete and more realistic than in Chomsky and Halle's (1968) *The sound pattern of English* (SPE) (Hooper 1976, pp. xi–xii, 1–11, 119ff.; 1979, pp. 106–7). As she puts it: 'The major claim of natural generative phonology is that speakers construct only generalizations that are surface-true and transparent . . . An important property of surface-true generalizations is that they are all falsifiable in a way that the more abstract generalizations of generative phonology are not' (1979, p. 106). The formal apparatus of NGP offers no prospect of highly abstract underlying forms, undercutting much of the discussion engendered by SPE about the ordering and interaction of rules. In a sense, NGP directs phonology back towards the more concrete concerns of phonemics. This point is underlined by Hooper's recognition of a distinction among rules that virtually revives the traditional categorization into phonetic (allophonic) and morphophonemic rules. Hooper distinguishes between rules that refer only to phonetic information and reflect the 'automatic' pronunciation habits of a speaker (which she terms P-rules), and rules that refer to grammatical or lexical contexts and often do admit exceptions (MP-rules) (Hooper 1976, p. 15; 1979, pp. 107–8). A comparable distinction is made by many of those who have reacted against orthodox generative phonology, and will emerge again in different guises below.

11.11 Natural phonology

Though similar in name to natural generative phonology, natural phonology represents a more dramatic departure from the mainstream of generative

phonology. It has its origins in David Stampe's dissertation on natural phonology (submitted to the University of Chicago in 1973 and published in 1979). Stampe begins his dissertation in the context of children's acquisition of phonology and draws attention to what he calls 'phonological processes'. A phonological process is 'a mental operation that applies in speech to substitute, for a class of sounds or sound sequences presenting a specific common difficulty to the speech capacity of the individual, an alternative class identical but lacking the difficult property' (1979, p. 1). These processes are not rules of the language, acquired as the child masters language, but reflections of what we might call the child's inbuilt tendencies. Thus, by the very nature of the human articulatory and perceptual organism, a child will prefer to articulate plosives as voiceless rather than voiced (because of the relative difficulty of maintaining voicing while the supraglottal tract is closed off) or will prefer to nasalize vowels next to nasal consonants (again for reasons of articulatory ease). Processes are revealed in the consequent substitutions which children make in the early stages of acquisition – for example, when they neutralize the voicing distinction of English by substituting voiceless plosives for voiced.

The application of phonological processes is not as straightforward as simple examples might suggest. Not only are there many such processes, but some of them are contrary to others. For example, a process of vowel denasalization reflects the goals of articulatory ease and auditory distinctness in vowel production; but this is to some extent countered by the process of nasalizing vowels next to nasal consonants (Stampe 1979, pp. 17–23). In early stages of language acquisition, the unconstrained operation of natural processes will tend to reduce every potential utterance to something like a monosyllabic [pa] (Stampe 1979, pp. xvii, 2). As the child comes closer to an adult competence, processes will be suppressed or limited in response to the demands of the phonological system. Hence, if acquiring a language in which nasalized vowels are distinctive, a child will have to suppress the relevant natural processes and thus achieve control of vowel nasalization; but in a language in which vowel nasalization is not distinctive, the natural process of nasalizing vowels next to nasal consonants may persist as an 'allophonic rule' of adult speech (Stampe 1979, pp. 27–8).

Stampe appears to turn generative phonology on its head. What we thought of as rules constituting a phonological system are now seen as processes motivated by the nature of production and perception. Phonological acquisition is a matter of suppressing or constraining innate tendencies rather than of learning rules. Stampe does leave room for phonological rules, however. These are indeed acquired, but they differ sharply from processes. The English alternation of /g/ and /dʒ/ (as in, for example, *analogous* with [g] but *analogy* with [dʒ] is governed by an acquired rule. Unlike a natural process, such a rule is open to exceptions and easily suspended: many speakers fail to apply the rule consistently and pronounce *analogous* with [dʒ] or *pedagogy* with [g], for example, and even those who follow the rule could easily produce the 'wrong' pronunciations if they wished. By contrast, the phonetic consequences of natural processes, such as aspiration of voiceless plosives and lengthening of vowels before voiced obstruents, are much harder for native speakers to discern and overcome (Stampe 1979, pp. 45–7).

Among the various North American 'schools' of phonology which represent reactions to orthodox generative phonology, natural phonology is the least inclined to proclaim its faithfulness to generative principles. In their useful outline of natural phonology, Donegan and Stampe (1979) appeal to phonological traditions that are much older than SPE: 'Natural phonology is a modern development of the oldest explanatory theory of phonology . . . Its basic thesis is that the living sound patterns of languages, in their development in each individual as well as in their evolution over the centuries, are governed by forces implicit in human vocalization and perception' (1979, p. 126).

Donegan and Stampe claim that their theory is natural because it seeks to explain why language is the way it is. The theory offers genuine explanation by presenting language not as merely conventional but as a 'natural reflection' of the needs, capacities, and world of its users (Donegan and Stampe 1979, p. 127). Donegan and Stampe are critical of positivism, which gives priority to exhaustive scientific description (p. 127; cf. section 11.1 above). They reject (underlying) morphophonemic representation in favour of a more traditional phonemic representation (which, following Sapir, they interpret as reflecting the phonological intention of speech; pp. 158ff., esp. pp. 163–7). And they conclude that although both structuralist and generative phonology have a well-developed methodology, neither of them is a theory in the true sense, since neither is genuinely open to falsification by data (pp. 167–8). Threads of these arguments are developed more fully by linguists like Bybee (2001), who strongly attest the primacy of the link between language usage and phonological structure.

11.12 Autosegmental and CV phonology

The phrase 'autosegmental phonology' is the title of Goldsmith's dissertation submitted to the Massachusetts Institute of Technology in 1976 and published in the same year. Goldsmith's initial concern is with what may seem to be a limited and particular problem, that of segmental organization, or more particularly, that of phenomena which have 'evaded segmental classification' (Goldsmith 1976, p. 6). The longest chapter in the thesis is devoted to the 'tonology' of Igbo, a west African tonal language, and Goldsmith includes substantial attention both to other tonal languages and to stress and intonation in English.

Goldsmith's work nevertheless goes beyond tone and intonation, and the implications of his thesis have been increasingly extended and elaborated; many of the concepts underpinning his earlier work are still evident in current phonological practice. His thesis announces a claim about the 'geometry' of phonetic representations (p. 6) in the context of what he calls 'the absolute slicing hypothesis' (the hypothesis that speech can be phonologically represented as successive discrete segments, pp. 16–17). His fundamental point is that speech, observed as articulatory activity, consists of gestures – such as tongue movement, lip movement and laryngeal activity – which are coordinated, but which by no means start and finish all at the same instant. The point is a familiar

one in modern phonetics (sections 4.1 and 7.17 above), and Goldsmith's reiteration of it leads him to what he calls a 'multilinear phonological analysis in which different features may be placed on separate tiers' (1979, p. 202). The tiers are connected to each other by 'association lines', which allow for the fact that there may not always be a neat one-to-one mapping between tiers. Thus an autosegmental notation can show tonal features on a different tier, represented below segmental features, e.g.

Disyllabic word with high tone on each syllable:

```
baka
 | |
H H
```

Disyllabic word with high tone then low tone:

```
baka
 | |
H L
```

The vertical lines are the normal association lines mapping tones on to syllables. In many tonal languages, however, a high tone becomes, by anticipatory assimilation, a falling tone when followed by a low tone. If so, this can be shown as the consequence of both the high tone and the low being mapped on to a single syllable:

```
baka
 \|
H L
```

This provides a flexible way of associating tones with segmental features, such that tones are not, as it were, swallowed up within a strictly segmental notation (cf. section 9.4 above).

The approach can be extended to other features. Nasality, for instance, may also be represented on a separate tier, allowing for similar spreading across segmental boundaries. Where a consonant is prenasalized and the preceding vowel nasalized, we may represent

[dāmba] as

```
d  a  b  a
   \  |
    N
```

Goldsmith himself remarks that the system of analysis was 'originally suited to fit the intricacies of African tone languages' (1979, p. 212), but various papers developing the autosegmental perspective (Clements 1977, 1981, 1984, 1985; Goldsmith 1979, 1985) show how similar notation can be applied to other phenomena, especially vowel harmony, where the spread of vowel features across segmental boundaries can be treated analogously to the examples of tone and nasality shown above.

If separate tiers are indeed a reflection of the parallel articulatory activities of speech, then we may expect features to be generally autosegmental – independent of each other, that is – in the early stages of a child's phonological acquisition. In acquiring their mother tongue, children must 'deautosegmentalize' (Goldsmith 1976, pp. 160ff.; 1979, pp. 214ff.). In other words, it is necessary for a child to learn the appropriate language-specific restructuring of the phonetics into segments (1979, p. 214) or to learn 'which sets of feature-specifications on separate tiers may be merged together to form an acceptable segment in that language' (1976, p. 164).

The notion of tiers is reminiscent of Firthian prosodic phonology (section 11.8 above), and Kenstowicz – who gives a useful general assessment of autosegmental phonology – comments on the way in which generative phonology is 'recapitulating in part some of the insights of the Firthian School' (1994, p. 311). Lass (1984, p. 269) doubts whether autosegmental phonology, 'despite its formal sophistication', is 'much more than a notational variant of prosodic analysis'. Goldsmith certainly acknowledges an affinity and seems to take Firth to be an ally against American phonologists who have too rigidly adhered to segmental discreteness (1979, pp. 203–4). He argues nevertheless that Firthian prosodies are a more limited set than the features of an autosegmental analysis, and places autosegmental phonology within the generative tradition of formalism, maintaining a discourse of 'rules' and 'well-formedness conditions', even though his theme is (auto)segmental organization rather than feature-changing processes.

The concept of tiers is also found in CV phonology, which arose from work by Kahn on syllabic organization and by McCarthy on Semitic languages (Kahn 1980; McCarthy 1981; Clements and Keyser 1983; Kenstowicz 1994, ch. 8). The original contribution of CV phonology is the postulation of a CV tier, a tier of C and V 'slots' which are filled by segments. Often segments (or the set of features represented by a segment) can be mapped straightforwardly on to these CV positions. Charting the relationships in similar fashion to autosegmental notation, we might represent English *map* and *landed* as

```
C  V  C        C  V  C  C  V  C
|  |  |        |  |  |  |  |  |
m  æ  p        l  æ  n  d  ə  d
```

But, just as autosegmental phonology allows other kinds of mapping, so CV phonology offers the possibility of capturing the special nature of complex segments that traditionally require structural interpretation (section 3.14 above). Thus a diphthong can be shown as two vowel qualities functioning as, or filling the position of, a single vowel; or a lengthened or geminate consonant can be represented as a single segment spreading over two C positions:

```
  V          C  C
 / \         \ /
a   i  = [ai]  t  = [t:]
```

In keeping with their generative antecedents, both autosegmental and CV phonology incorporate derivational processes into their phonological modelling.

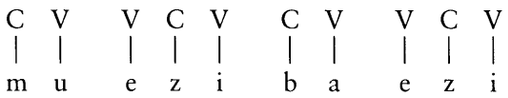
Thus phenomena such as lengthening of underlying short segments and secondary articulations derived from underlying adjacent segments are captured by reallocation of association lines. In Luganda, for instance, some nouns take the prefix /mu-/ in the singular and /ba-/ in the plural (Goldsmith 1989, ch. 2, based on work by Clements). Examples are

mukazi ('woman')	bakazi ('women')
mulimi ('cultivator')	balimi ('cultivators')

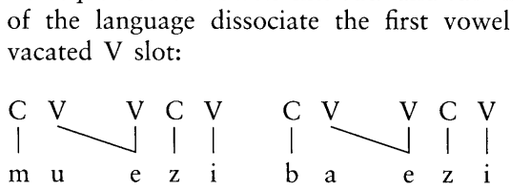
But in some forms we find a slightly different pattern:

mweezi ('sweeper')	beezi ('sweepers')
mwaana ('child')	baana ('children').

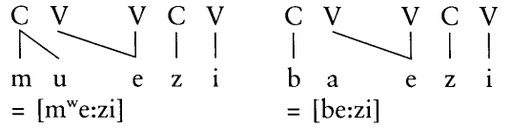
Here the prefixes appear as /mw-/ and /b-/, and the vowel following the prefix is always long. These forms can be derived from 'regular' underlying forms. For 'sweeper(s)', the underlying forms are



As sequences of vowels like /ue/ and /ae/ are not tolerated in Luganda, rules of the language dissociate the first vowel and associate the second to the vacated V slot:

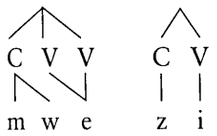


If the dissociated vowel is high, it can combine with the preceding consonant as secondary labialization; otherwise, if it remains dissociated, it will not be realized phonetically. And in either case, the following vowel, having now been associated to two V positions, will be realized as a long vowel.

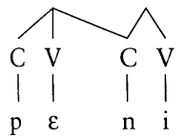


The CV tier, also referred to as the 'skeletal' tier or 'timing' tier, now forms part of what Clements and Keyser call 'a universal theory of the syllable' (Clements and Keyser 1983, p. 25; cf. section 11.13 below). The tier not only defines the timing of segmental organization (by, for example, determining that a vowel occupies two V slots) but also takes over the role of the feature [syllabic] (Clements and Keyser 1983, pp. 10–11): syllabic organization is shown by a tree structure dominating the CV tier.

For instance, /mweezi/ might be represented as two syllables as follows



where the C and V units constituting a syllable are dominated by a single node. This model allows for interesting and useful possibilities, such as assigning a consonant to two successive syllables, as in English *penny*, where it is phonologically appropriate to take the /n/ to be both final in the first syllable and initial in the second syllable:

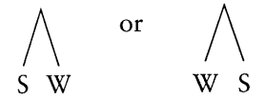


The model has affinities with metrical phonology, to which we turn in the following section, and Goldsmith (1989) suggests that the convergence of autosegmental, CV and metrical phonology is yielding 'a new synthesis'. Many of the seminal papers in these areas of phonology are brought together in a very useful volume edited by Goldsmith (1999).

11.13 Metrical phonology

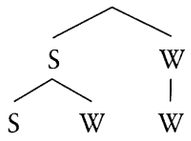
Yet again, metrical phonology has its origins in a doctoral dissertation (Liberman 1985). Just as autosegmental phonology began with tone and was then extended to other phenomena, metrical phonology began as a theory of stress (see section 9.7 above) and later widened its horizons. As noted by van der Hulst and Smith (1982b, p. 30), metrical theory has now 'invaded' the territory of autosegmental phonology. As clear evidence of this, the prevailing framework of intonational analysis today is known as the autosegmental-metrical framework (Ladd 1996).

The starting point of metrical phonology is an assumption about the nature of stress and its representation, namely that stress patterns reflect an underlying structure in which stronger and weaker constituents are juxtaposed. To say that a certain syllable is stressed is to make a judgement about its strength relative to adjacent syllables (cf. sections 9.3 and 9.6 above). Using the kind of tree structure noted in the preceding section, we can display the stress patterns of disyllabic words as either

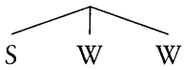


where S and W simply indicate stronger and weaker constituents. Much of metrical theory is then devoted to explaining how more complex patterns are

derived from these basic patterns within certain postulated constraints. It is assumed, in some versions of metrical theory, that the relationship between S and W is binary, so that polysyllabic patterns entail subsidiary branching, e.g.

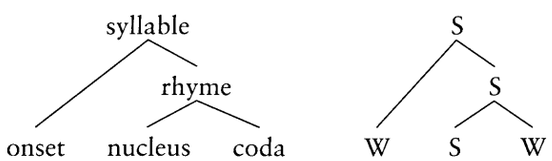


and not



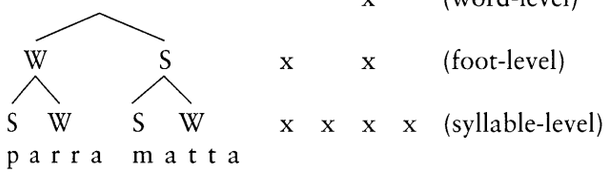
Attempts to draw up procedures for the assignment of English stress under such a model (usefully surveyed by van der Hulst and Smith 1982b, pp. 30ff.) confronted various criteria. These were related both to the formal nature of the process (whether stress assignment proceeds from right to left throughout all words, and how subsidiary branching is organized, for instance) and to the properties of a word which may be said to affect stress assignment (such as morphological structure, syllabic structure and the presence of specific segments such as 'tense' vowels). This discussion was part of a revival of interest in the concepts of feet and syllables, an interest evident also within autosegmental and CV phonology (section 11.12 above). In the new formalism, the foot, traditionally recognized in English poetry and used also by writers such as Halliday (sections 9.3 and 9.6 above), could also be identified as a tree structure. Thus the word 'catastrophic' has two feet revealed as

By the mid-1980s, the syllable – having been totally ignored within standard generative phonology – was attracting considerable attention in North America. It was argued that the syllable was a significant unit which must be recognized within phonological theory, and, in keeping with the spirit of generative phonology, efforts were made to formalize the structure of the syllable. Using fairly traditional terms (reminiscent of those used in Chinese linguistics; section 11.2 above), we can take a syllable to consist of a RHYME preceded (usually) by an ONSET. The rhyme may in turn consist of a PEAK or NUCLEUS, sometimes followed by a CODA. Interestingly, this structure can be handled by the general formula originally proposed for stress patterns. Compare the two patterns below:



Metrical phonology offers an alternative way of expressing such structures, in the form of a so-called METRICAL GRID. Suppose we take a tree of the sort

shown above, and convert it into a grid by making entries at levels corresponding to the levels of the tree. The tree on the left below reflects the stress pattern of the word *Parramatta*, with greatest stress on the third syllable, and minimal stress on the second and fourth syllables. The tree can be mapped on to a grid, as shown on the right, in which the x-entries correspond to nodes on the tree: the grid thus provides an alternative visual display, with the greatest degree of stress represented by the column having the greatest number of entries.



The illustration here is of the simplest possible kind. A detailed exposition of metrical theory, in coursebook style, can be found in Hogg and McCully (1987). Van der Hulst and Smith (1982b) offer a thorough evaluation and comment on the 'competition' caused by the expansion of both autosegmental and metrical theory to include the linear organization of speech in general (1982b, pp. 42–5). They refer to a number of possibilities – including Halle and Vergnaud's suggestion (1981) that there are two kinds of harmony, 'metrical' and 'autosegmental' – but they admit they are unable to offer a unified theory. Anderson et al. (1985, p. 203) are slightly more optimistic that the various models of suprasegmental representation, including autosegmental and metrical phonology, are less different than appears at first sight and that a single model may 'perhaps be developed from the various frameworks'. And Goldsmith, in a detailed assessment of the different schools (1989), takes a highly positive view of a new integrated perspective. One of the classic metrical analyses of English stress is presented in Halle and Vergnaud (1987), and Hayes (1995) presents an excellent survey of stress analyses in the metrical tradition. Kenstowicz (1994) also sets work on syllabic organization and stress in its wider context: chapter 6 of his book is a survey of the major results of generative research on the syllable, and chapter 10 discusses the phenomenon of stress and a series of analyses inspired by metrical theory. Likewise, Roca and Johnson (1999, chs 11–13) cover the major developments in metrical theory until the late 1990s, and Gussenhoven and Jacobs (2005, ch. 14) and Kager (1999, ch. 4) provide a reassessment of metrical stress rules in terms of a constraints-based optimality theory framework (see section 11.16 below).

11.14 Lexical phonology

Among all the attempts to modify and extend orthodox generative phonology in North America, lexical phonology reflects most clearly the concerns of pre-generative phonemics. Originally developed by Strauss, Kiparsky and Mohanan,

it shows a revived interest in morphology and asserts a level of representation which is comparable to that of taxonomic phonemics (Strauss 1982; Kiparsky 1985; Mohanan 1985, 1987; Goldsmith 1989, ch. 5).

In a useful overview, Kaisse and Shaw (1985) point out that despite the willingness to recognize value in traditional phonemics, lexical phonology is not as concrete as, say, natural generative phonology or natural phonology (sections 11.10 and 11.11 above). Lexical phonology does allow for abstract underlying forms and in that light is 'a standard generative phonology' (Kaisse and Shaw 1985, p. 3). What the title of the school reflects is a distinction between 'lexical' and 'postlexical' components of description. Lexical rules are fed by the morphology (itself a subject of considerable debate in the postgenerative era): the morphological component supplies the various affixed and compounded forms of the language, and lexical rules then apply, to modify these forms in accordance with the phonological requirements of the language. In English, a lexical rule might ensure that the final consonant of stems such as *logic*, *critic* and *electric* is 'softened' to /s/ before the suffixes *-ism* and *-ity*; or another lexical rule might apply to the suffix *-ed* to devoice the /d/ in forms like *tapped* and *licked*, in conformity with the patterning of English consonant clusters. At this stage of derivation, only distinctive features are relevant (in the classic sense of 'distinctive'), and lexical representations and lexical rules make no reference to redundant or 'allophonic' features (such as, in English, the voicing of nasal consonants or the aspiration of voiceless plosives). The postlexical rules, applying to the output of lexical rules, include those that apply to larger domains than words – rules, for instance, that need to refer to phrasal structure or that apply across word boundaries. In English, the assimilation of /s/ and /z/ to /ʃ/ and /ʒ/ before /j/ must be postlexical, since it applies not only within words (as in *tension* and *usual*) but also across word boundaries (as in *I miss you* or *as you wish*). Rules of the postlexical component also fill in the redundant features that have been unspecified in the lexical component.

It is noteworthy that lexical rules are by and large 'morphophonemic' in traditional terms, including the rules familiar from SPE which apply to tense and lax vowels (*sane*, *sanity*, etc.). Postlexical rules are similar to Stampe's natural processes (section 11.11 above) or the allophonic processes of traditional phonemics (section 4.3 above). Thus postlexical rules do not tolerate exceptions, can apply across word boundaries and may yield phonetic values such as 'heavily aspirated' or 'partially devoiced'. The consequence is that the output of lexical rules – termed 'lexical representation' – is in some respects quite similar to a traditional phonemic transcription. It is recognized by lexical phonologists as a significant level within phonology, one which is likely to be real to native speakers in the sense that, for example, they are conscious of the different vowels in *sane* and *sanity* determined by lexical rules, but unaware of the extent to which they voice the plosive or nasalize the vowels in *sanity* (Kaisse and Shaw 1985, pp. 4–8).

It is tempting but unfair merely to dismiss lexical phonology as the generativists' rediscovery of phonemics. Lexical phonology is clearly generative in its style of theoretical modelling and its commitment to rule-based description (including even the principle of cyclic rule application; section 5.6 above). Early

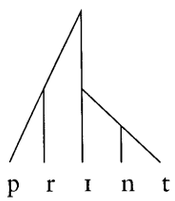
proponents of generative phonology who made a point of being scornful of taxonomic phonemics might have some cause to be embarrassed but there have always been those within generative phonology who remained open to phonemic insights (for example, Schane 1971 and Hyman 1975). Moreover, lexical phonology continues to grapple with the problems of describing English morphology and morphophonemics. These problems are real, given the extent of morphophonemic alternation in English and the difficulty of determining what is truly patterned or rule-governed (by genuine processes such as assimilation) and what is odd irregularity (such as the forms of 'to be').

Goldsmith (1989) includes a chapter on lexical phonology which again holds out some promise of a synthesis of postgenerative trends in phonology. Kenstowicz (1994, ch. 5) provides a thorough outline of lexical phonology, concluding with a detailed review of some of the 'unresolved problems' that confront this model (pp. 227 ff.). Gussenhoven and Jacobs (2005, ch. 8) also present a clear and concise overview of lexical phonology, as do Roca and Johnson (1999, ch. 16).

11.15 Dependency phonology

Dependency phonology (Anderson et al. 1985, Anderson and Ewen 1987) shares much of the modern interest in structures such as feet and syllables and in the organization of features below the level of the segment. We have already noted the way in which features are treated in dependency phonology (section 10.11 above) and we review here the wider concept of dependency that underlies this work.

It is possible to model the structural organization of speech in a way that is reminiscent of metrical tree structures (section 11.13 above) but different in important respects. A monosyllabic word like English *print* might be displayed as follows:



As in other kinds of tree diagram, the single node at the top can be said to dominate the structure, defining the unit – here a syllable in which the vowel serves as head or nucleus. But in dependency phonology there are no category symbols (such as the S and W used in metrical models), and structural relations are shown by 'dependency' alone, reflected in the tree diagram. Thus the vowel in our example is most prominent, and the consonants are subordinate or dependent. But dependency extends further than this, for the diagram shows

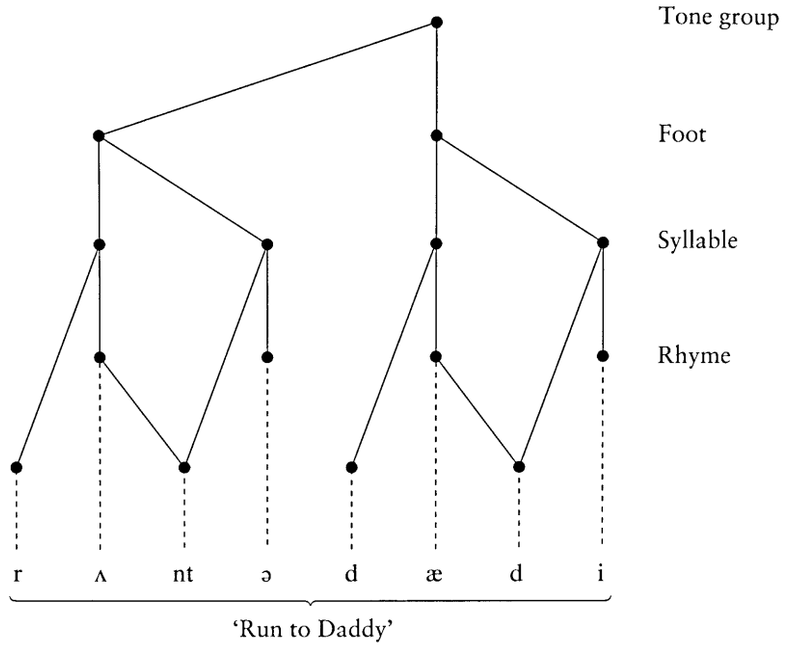


Figure 11.15.1 Dependency structure of an English utterance *Run to Daddy*
Adapted from: Anderson and Ewen 1987, p. 101.

the vowel both as head of the syllable and as head of the rhyme /ɪnt/. Moreover, /r/ is shown to be head of the initial consonant cluster, and /n/ head of the final cluster; conversely /p/ is dependent on /r/, /t/ on /n/ and both clusters are dependent on the nuclear vowel.

Anderson and Ewen (1987, pp. 96ff.) use this kind of notation for much larger structures. One of their examples is given in figure 11.15.1, where the structural levels reflected in the diagram are labelled on the right. The utterance 'run to Daddy' is shown to be a single tone group, two feet, and four syllables. The diagram models prominence relations in direct fashion, incorporating points that are now standard in modern phonology, such as the recognition that a consonant can belong to two syllables (as the second /d/ in 'Daddy' does).

We saw earlier (section 10.11) that features or components can also 'preponderate' in dependency phonology. Where a language distinguishes /e/ from /ɛ/, the higher vowel may be represented with frontness preponderant over lowness, as {|i⇉a|}, the lower with lowness preponderant over frontness, as {|a⇉i|}. But Anderson and Ewen (1987, pp. 127ff.) point out that this relationship of preponderance can also be portrayed in a similar way to prominence relations within linear structure. Hence |i| dominating |a| can be shown as

i
|
a.

A diphthong such as [ai] can be shown with the first target predominant:

a i.

As Anderson and Ewen put it, preponderance relationships among features, and prominence relationships among linear elements, are parallel: 'preponderance is the intrasegmental analogue of prominence' (1987, p. 128).

11.16 Optimality theory

Optimality theory (OT) emerged in the 1990s with the publication of technical reports by Prince and Smolensky in 1991 and 1993 (now published as Prince and Smolensky 2004) and McCarthy and Prince (1993). OT is widely used in phonology and morphophonology research, as well as in other areas of generative grammar. Only key elements of OT are explained here and the reader is directed to Prince and Smolensky (2004) and Kager's textbook on OT (Kager 1999) for full elucidation of the framework (see also Roca and Johnson 1999, ch. 19, and Gussenhoven and Jacobs 2005, ch. 3, for excellent summaries). There are a number of components of OT that make it appear different from earlier theoretical models. Phonological (or for that matter linguistic) generalizations, rather than being expressed as generative rules that produce a particular output, can be thought of in terms of resolution of competing demands or constraints on the form of the output or surface form. The approach to phonology outlined in chapter 5 embodies the derivational approach, whereby the output or surface form is the result of a series of rules that operate on an underlying form of a word or morpheme, generating at each stage of the derivation a specific output, which is in turn operated on by any following rules in the derivation process. OT, by contrast, is a constraint-based approach. Constraints are universal, are governed by markedness principles and are violable. Constraints are ranked relative to each other and according to their strength.

An OT framework, considered to be part of Universal Grammar, has several additional essential components. The GEN function produces a number of potential outputs, and the EVAL function acts as the policeman of the model, ruling which ranking of constraints produces the 'optimal' output candidate. For an output to be optimal, it must have incurred the smallest number of violations of the highest-ranking constraints. Every constraint can be violated. There are different families of constraints. For example, 'faithfulness' ensures that the surface forms are identical to the input or underlying form, and by contrast 'markedness' is an additional force that makes sure that language-specific differences come into play either to create contrast, or to give priority to cross-linguistically preferred forms (see section 11.6 above). 'Structural' constraints are also a crucial component of the model; in other words, candidates

are admitted to an analysis by virtue of their structural well-formedness. So we can think of each language as having a particular hierarchy of universal constraints. Even if two of the constraints appear to be contradictory, as long as the 'right' one is ranked higher than the other, the two can coexist in the same constraint hiearchy for a particular language. All of this information is captured in a representative device called a 'tableau'.

One of the most famous illustrations of this framework is presented by McCarthy and Prince (1993) in their technical report on generalized alignment, which sought to show how OT can provide a principled treatment of the ways in which the edges of prosodic or morphological constituents align with one another. The authors give what is now considered a classic analysis of Tagalog *um*-infixation (McCarthy and Prince 1993, pp. 22–5). If we take the root /aral/ and add the /um/ affix, we get the inflected form /umaral/ meaning 'teach', but with the root /gradwet/, we get the inflected form /grumadwet/. McCarthy and Prince showed that the optimal output can be achieved for each variant by invoking two constraints, first a syllable structure constraint, and then an 'edge-most' or 'align' constraint, which deals with how far a linguistic entity like the *um*-infix can be from the edge of some unit (like the prosodic word or stem). The relevant constraints are shown in (a) and (b).

a. NO-CODA (i.e. syllables must be open)

b. ALIGN-um, L, Stem, L (align left edge of affix with left edge of stem).

The following two tableaux show the operations of these two constraints (McCarthy and Prince, pp. 24–5) that achieve the optimal output in each case: *grumadwet* and *umaral*. The number of '*' shows the number of times the generated candidate violates the constraint. Recall that the candidate with the smallest number of higher-ranking constraint violations is the optimal one. The arrow shows the favoured candidate. When candidates are ruled out as optimal, an exclamation mark '!' is placed in the cell showing the number of violations. In the tableau shown in (1), we see that the first two candidates incur three violations of the NO-CODA constraint, because this output contains three closed syllables. The first candidate does not violate ALIGN-*um*, but the second candidate does. The important thing about the second constraint is that it is measured in terms of *distance* from the left edge of the stem, so rather than putting a row of '*'s for each violation, McCarthy and Prince list the linguistic material (i.e. phonemes in this case) that separate the affix from the left edge of the stem (indicated by a '['). So in the case of the third candidate, the violation is 'gr'; in other words, the affix is two segments away from the left edge of the stem. However, because this constraint is ranked as less important than NO-CODA, these violations are less serious and it is the fatal violations of NO-CODA that rule out the first two candidates. The third candidate is the optimal one because it incurs only two violations of NO-CODA, and it only violates ALIGN-*um* minimally. The case shown in (2) is very straightforward. It has only one violation of NO-CODA, and as McCarthy and Prince (1993, p. 25) point out, '*um* is optimally positioned as a classical prefix', and does not violate ALIGN-*um* at all.

(1)

Candidates	NO-CODA	Align-*um*
[-um.grad.wet.	***!	
[g-um-rad.wet.	***!	g
→[gr-u.mad.wet.	**	gr

(2)

Candidates	NO-CODA	Align-*um*
→ [-u.ma.ral.	*	Ø
[a.-um.ral.	**	a
[a.r-u.mal.	*	ar!

This brief illustration of the framework shows many of its noteworthy features. For example, the NO-CODA constraint reflects the widely accepted premise that open syllables are unmarked. Some knowledge of other elements of preferred syllabification principles is also built in, particularly with respect to the syllable onsets of the listed forms. This presumably also cuts down on the number of generated candidates. It also suggests that *-um-* is best treated as a prefix, and not some 'exotic and idiosyncratic morphological entity' (McCarthy and Prince 1993, p. 25), and the optimal candidate reflects the preference for avoiding overviolation of the highest-ranked phonological NO-CODA constraint.

As is clear from this example, OT is widely accepted today as a framework for reinterpreting or repackaging phonological and morphophonological (or indeed morphological) analyses from the past decades. It is an extension of the generative tradition, and still uses many of the phonological representation devices and principles that came out of earlier stages in the evolution of generative theory. In OT treatments of stress, for example, constraints like FT-BIN (all feet must be binary) and so forth are assumed in most analyses. Many questions remain to be answered. For instance, it is still theoretically possible for many ill-formed candidate outputs to be generated by a given set of constraints, which seems somewhat counter-intuitive for a model that purports to have cognitive underpinnings. It is also questionable whether all of the constraints posited in some OT treatments of various phonological phenomena are truly universal. The fact that all constraints are violable also leads to accusations of the model

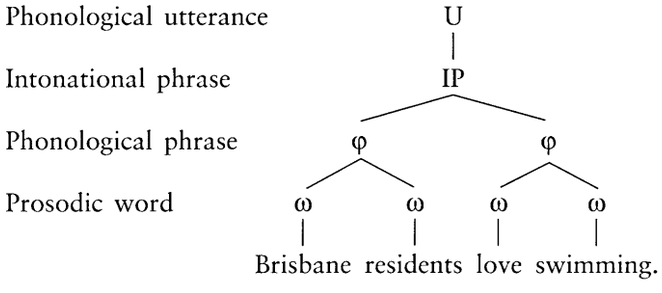
or theory being 'unfalsifiable'. Nevertheless, it constitutes the major milestone in the evolution of generative linguistic theory in recent decades.

OT is by no means the only constraints-based phonological theory to have emerged from the early 1990s. Declarative approaches to phonology, drawing on the architecture and language of nontransformational grammatical theories like HPSG (Head-driven Phrase Structure Grammar), are also built around well-formedness constraints (e.g. Bird 1995, Scobbie et al. 1996). Like OT, declarative theories of phonology are nonderivational. Coleman (1995) provides an excellent elucidation of the main components of declarative approaches.

11.17 Prosodic phonology

Prosodic phonology (not to be confused with Firthian prosodic analysis, described in section 11.8 above) concerns itself largely with presenting a view of phonological structure that incorporates levels of constituency above the syllable and the foot. In chapter 9 we made the point that prosody in this sense is not synonymous with suprasegmental features, but rather refers to the hierarchical organizational structure of an utterance. In chapter 1 of their celebrated monograph on prosodic phonology, Nespor and Vogel (1986, p. 1) state that 'the phonological component cannot be considered a homogeneous system but rather must be seen as a set of interacting subsystems, each governed by its own prinicples such as the metrical grid, lexical phonology, autosegmental phonology and prosodic phonology'. This is a sentiment that continues to be shared by phoneticians and phonologists alike, although a range of models of the prosodic hierarchy has been proposed over the years (e.g. Selkirk 1984, Hayes 1989). Prosodic phonology is built on the premise that prosodic constituency 'draws on different types of phonological and nonphonological information' that contribute to the definition of each domain (Nespor and Vogel 1986, p. 2). We have already encountered this notion of prosodic constituency in this chapter and in chapter 9.

A typical hierarchical representation of some of the relevant domains referred to in prosodic phonology is presented below:



We could add in the metrical foot and syllable (section 11.13 above) below the level of the word. Nespor and Vogel (1986) also include the 'clitic group',

which is between the phonological phrase and the prosodic word. Gussenhoven and Jacobs (2005, ch. 15.2) present a good summary of some of the heuristics that are used to determine prosodic constituency of prosodic domains, including boundary strength, edge tones, postlexical phonological processes and phonology-sensitive syntax. Shattuck-Hufnagel and Turk (1996) show how different approaches to prosodic phonology have evolved.

One of the key elements of the hierarchical structure shown above is the strict layer hypothesis (e.g. Selkirk 1984, Nespor and Vogel 1986). This has had a few different permutations over the years, but essentially the idea is that at each level of the hierarchy, the relevant constituent (e.g. intonational phrase) will consist of one or more constituents at the next level down (e.g. the phonological phrase or intermediate phrase, after Beckman and Pierrehumbert 1986). Prosodic words will be composed of feet, and feet will be composed of syllables. It is perfectly possible for there to be violations of the strict layering hypothesis, and a constituent may dominate another constituent that is a couple of levels down. One of the best examples of this is the concept of extrametricality (e.g. Hayes 1995), whereby a prosodic word can dominate a syllable directly, if that syllable has not been parsed into a foot.

The utterance-level constituent is meant to represent the widest domain span of phonological rules, and it usually corresponds to syntactic sentence, although according to Selkirk (1984) it may include more than one sentence. Gussenhoven and Jakobs (2005, pp. 222–3) cite the Nespor and Vogel (1986) example of /r/ insertion or linking /r/ in non-rhotic varieties of English. It is relatively common to insert an /r/ between a word that ends in a nonhigh vowel and a following morpheme that starts with a vowel (e.g. as when 'india ink' is pronounced 'india-r-ink', or 'Spa is in Belgium' as 'Spa-r-is in Belgium'). This 'r'-insertion process is blocked at the utterance level. The intonational phrase is an intonationally defined constituent with (optional) left and right edge tones (see section 9.8 above). In English, the right edge of this constituent is usually also associated with final lengthening of the final syllable or foot. The phonological phrase tends to correspond to the syntactic phrase, and in some languages this is the domain for certain kinds of phonological processes, including the English rhythm rule (Hayes 1989). Some models also do not distinguish between the intermediate intonational phrase, the accentual phrase and the phonological phrase. There is a range of definitions in the literature for the prosodic word, but most would agree that this constituent does not always coincide with a morphological or grammatical word. Certain phonological processes may not 'cross the boundary' in morphologically complex words or compounds in some languages. Nespor and Vogel (1986, pp. 120–1) use Turkish vowel harmony (section 11.8 above) as an illustrative example. Compound words in Turkish effectively behave like two separate prosodic words because harmony is restricted within each component of the compound and does not spread across them.

Nespor and Vogel (1986), Selkirk (1984) and the instigators of the metrical phonology revolution (Liberman and Prince 1977) initiated a major transformation in modern phonology in the latter part of the twentieth century. Many of the insights brought to the forefront of phonological research by these scholars continue to fuel wide-ranging debate about the nature of phonological organization,

as well as providing a framework for experimental work in the areas of stress and intonation.

11.18 Phonology in the laboratory

While many phonologists in the 1970s and 1980s may seem to have been reacting to SPE – Anderson refers to 'a variety of knights-errant' (1979, p. 2) – some maintained or developed quite different perspectives, including that of Nespor and Vogel (1986; see section 11.17 above). Another striking example is the emergence of what is generally known as 'experimental phonology'. Since the early 1990s, this has evolved into 'laboratory phonology' (e.g. Kingston and Beckman 1990, Docherty and Ladd 1992, Broe and Pierrehumbert 2000, Gussenhoven and Warner 2002, Local et al. 2004). Both experimental phonology and laboratory phonology represent an attempt to draw together at least three research styles: experimental phonetics, experimental psychology and phonological theory. Neither experimental nor laboratory phonology is a theory as such. Rather, the intention of experimental phonology as expounded by Ohala and colleagues (e.g. Ohala and Jaeger 1986) was to submit hypotheses about phonological organization to testing and validation of the kind which is standard in the experimental sciences, and which had been taken over, to some extent at least, by researchers in fields such as psychology, psycholinguistics and instrumental phonetics. Laboratory phonology is the more recent manifestation of this trend (Pierrehumbert et al. 2000).

The original approach to experimental phonology was not always free of the implication that phonology is speculative and that evidence obtained experimentally is superior to any other kind of evidence. Thus Ohala begins his 'Consumer's guide to evidence in phonology' with the words: 'For the past 30 years phonologists have speculated on how sound patterns in language are represented in the human mind (Chomsky and Halle 1968: viii). The claims made, of course, are only as good as the evidence they are based on' (Ohala 1986, p. 3). In a sense, then, experimental phonology was after all a reaction against generative phonology, or if not a direct reaction, then a reassertion of pregenerative interests. Ohala stresses the importance of evidence in evaluating theories and appeals to the example of physics, in which, he argues, evidence has enabled modern physicists to discard inadequate theories (such as the Ancient Greek hypothesis that all matter consists of only four elements; Ohala 1986, p. 5). In fact, he maintains that 'physics, chemistry and biology first became mature disciplines (with an accompanying marked increase in the rate of successful applications of their theories) when they started relying on and insisting on experimental evidence for claims' (p. 11). Similarly, Ohala and Jaeger express the hope that phonology is developing into 'an experimental discipline' (1986, p. 1) and again refer to the importance of the experimental method as it has been defined in modern Western science (pp. 1–6).

Proponents of experimental phonology take a rather generous view of what constitutes an experiment, and emphasize observation and careful refinement

of one's beliefs rather than reliance on instrumental investigation or statistical processing of results (Ohala 1986, p. 10; Ohala and Jaeger 1986, pp. 2–3). Several authors give some prominence to word games (Ohala 1986, pp. 9–10, Campbell 1986, Hombert 1986). Both Campbell and Hombert suggest that evidence from word games provides a test of the 'psychological reality' of phonological rules and structures. Hombert describes a word game in Bakwiri, a Bantu language which has no consonant clusters other than sequences of nasal followed by homorganic voiced stop. The word game involves reversing the two syllables of two-syllable words, so that e.g. [moko] would become [komo] and [lowa] would become [walo]. Now words such as [komba] and [kondi] are reversed as [mbako] and [ndiko], suggesting that the syllabic structure is [ko.mba] and [ko.ndi] rather than [kom.ba] and [kon.di]. Evidence of this kind has long been used in phonology (as noted by Hombert himself) and can hardly be described as experimental, unless the analysis extends to inventing word games and asking speakers to perform new operations on words (Hombert 1986, pp. 175–6 and 180ff.).

Other investigations reported in Ohala and Jaeger (1986) include instrumental studies of articulation and perception, such as Pinkerton's measurements of intraoral air pressure in the production of various kinds of glottalized stop in the Quichean languages of central America, and Kawasaki's study of the degree of vowel nasalization perceived in subjects listening to controlled stimuli. These studies continue the kind of work already being carried out by phoneticians and psycholinguists – and demonstrate very clearly the need to test impressionistic judgements of speech. What is new about such work is not so much its nature as its location in a setting of explicitly phonological questions. In Ohala and Jaeger's own words (1986, p. xi), it is the intersection of traditional phonological questions and experimental means that gives experimental phonology a claim to noteworthiness. Many of the contributors can be said to share this perspective: Lindblom (1986, p. 13) puts his modelling of the vowel space and distance among vowels (sections 2.7 and 7.15 above) in the setting of what he calls a 'functional perspective', namely that 'language form is forged by the sociobiological conditions of its use'; while Nearey and Hogan (1986, p. 141) begin their experimental study of categorical perception with a firm assertion that experimental phonetics is 'inextricably linked with phonology' (cf. section 8.5 above).

It is refreshing that phonetics and phonology began to meet each other more or less formally in this way – that experimental phoneticians continue to be aware of the theoretical assumptions and implications of their work, and that phonologists are equally aware of empirical methods and the need for evidence. Some caution is warranted, however, and the importance of experimental methodology in modern science should not delude us into thinking that knowledge gained through experiments is absolute. Ohala and Jaeger themselves went some way towards this recognition: 'The primary purpose of experimentation is not to create knowledge . . . It is, rather, a way of refining our knowledge. Following Popper (1959), one might even say that in a sense experiments actually destroy knowledge; at least they help to show which of our beliefs about the workings of the world do not agree with observation and hence should be discarded' (1986, p. 2). If taken seriously – as it should be – this view of

scientific method as a procedure of disproving false hypotheses rather than proving true ones means that experiments never yield certainty (cf. section 11.1 above). Equally seriously, an experiment is limited by its very conditions. An experiment in which subjects are asked, for example, to listen to artificial stimuli and record or report what they hear tests precisely what it purports to test, namely the hearing of particular stimuli under particular conditions. It does not test the subjects' perception of the utterances of daily discourse under normal conditions. Now of course if an experiment is well designed it should be possible to make useful inferences from it about normal speaking and hearing. But we should be wary of describing such inferences as 'solid evidence' or 'proof'. Taking relevant examples, we might note that word games do not necessarily reveal more than the rules of the games: it *may* be that the organization of segments and syllables which is exploited in word games is in fact not the organization that is relevant in the actual production of speech. We have referred earlier, for instance, to the observation that speakers of English may regard the vowel complex in *few* both as a diphthong and as a sequence of consonantal /j/ followed by vocalic /u/, depending on which perspective they adopt (section 3.14 above). There is no principled reason why speakers should use the same perspective in word games as in all other aspects of their language use. Likewise, when experimental subjects are tested on their ability to classify sounds according to their auditory similarity, we may reasonably assume that the subjects are drawing on their experience of speaking and hearing under normal conditions and that the experimental results therefore convey something about normal phonological knowledge. But this is not quite the same as saying that experimental results conclusively demonstrate the categories and operations of everyday usage. Nevertheless, subject to such reservations, experimental phonology in its initial manifestation reflected a proper scepticism about speculative pronouncements and a commendable interest in integrating empirical investigation with phonological analysis and theory.

Laboratory phonology as practised today is very sympathetic to many of the original sentiments expressed by experimental phonologists like Ohala. According to Kingston (1999), laboratory phonology tests the predictions of phonological analyses through experimental investigations of phonetic behaviour, rather than relying on the traditional data sources and methodology of phonological analyses (i.e. field-work transcriptions or internal reconstruction, for example). Like experimental phonology, laboratory phonology is not a phonological framework such as OT or dependency phonology, for example. Pierrehumbert et al. (2000, p. 279) describe it as a 'coalition amongst groups of people, with some working in one or another of the various current frameworks, and others working in no phonological framework at all'. It encompasses the use of computational models as well as laboratory techniques to 'evaluate generalizations' about the relationship between phonological structure, the lexicon, and speech production and speech perception (e.g. Pierrehumbert 2003). These techniques range from acoustic phonetic techniques or direct and indirect measurement of articulatory activity to statistical and probabilistic analysis and modelling of the lexicon. In recent years, traditional variationist and language acquisition research and the study of disordered speech have been addressed

by laboratory phonologists as focal areas of activity. While experimental phonetic research has been carried out in these areas for many years, there is increasing acknowledgement that phonological models can be further refined to take into account this kind of variation (see Docherty et al. 1997 and Docherty and Foulkes 2005 for thought-provoking discussion of this). In fact Pierrehumbert and colleagues (2000, pp. 284–5) argue convincingly that the 'modularization of phonetics and phonology' that was still part and parcel of research in experimental and laboratory phonology in the 1980s and 1990s is no longer as widely accepted as it once was.

11.19 Conclusion

Readers may well ponder the ancient wisdom that 'there is no new thing under the sun' but that 'of making many books there is no end'. Certainly some of the controversies of modern phonology seem to lead in circles, and the habit of labelling new trends and emphases as 'schools' exaggerates the impression of proliferation and underplays both the persistence of fundamental issues and the re-emergence of old themes in new dress. Nevertheless, tempting as it is for textbook writers to consolidate and simplify, the truth is that there are genuine differences of theoretical perspective, in phonology as in any field of scholarship.

Seen in this light, the custom of quoting one's antecedents – if done adequately and seriously – is not only a useful indication of historical background but also a declaration of one's place among competing theories. For example, Chomsky appeals to Descartes and seventeenth-century rationalism, Donegan and Stampe to Plato and natural explanation, and Ohala and Jaeger to Popper and the development of modern science (Chomsky 1966, Donegan and Stampe 1979, Ohala and Jaeger 1986). Pierrehumbert et al. (2000, p. 276) similarly call on work in the history and philosophy of science (e.g. Laudan 1996) wherein science is viewed as 'an adaptive human activity'. We cannot simply reconcile these different appeals in an all-embracing review, but should welcome the acknowledgement that theory does matter and cannot be ignored without distorting the nature of research and scholarship.

We have already declared our own preference for a view of phonology which, we believe, draws on a long and healthy tradition of thinking in terms of functions, systems and structures (section 11.1 above). We suggest that those whom we might call the earliest phonologists (section 11.2), even if they did not always expound their theoretical framework, recognized that pronunciation was systematic, that choices ('contrasts' or 'oppositions') mattered, and that phonological choices were finite and constrained (within systems and structures). The contribution of many of the classic figures of twentieth-century phonology has been to develop and refine such insights within a functional perspective. These continue to be refined in the twenty-first century (e.g. Bybee 2001). Once de Saussure had explored the concept of language as system, the way was open for elaboration of phonological description within a broadly structuralist

tradition, and scholars such as Trubetzkoy, Jakobson, Hjelmslev, Pike, Firth and Lamb (sections 11.5–11.8) – all with their own emphases – explored phonological analysis as a means of revealing the organization of language as it is spoken and heard. While this tradition was somewhat overshadowed in the 1960s and 1970s by generative phonology and its aftermath (sections 11.9–11.14), it has remained strong, often forming the backdrop to work that was thought to be outside phonology itself, such as experimental phonetics, psycholinguistics and speech pathology, and often assumed rather than defended in detail by writers whose focus of attention was not phonology itself.

The work in 'functional' or 'systemic' linguistics led by Halliday in the 1970s and 1980s, for example, falls within this tradition: Halliday himself argues for a (social-)functional approach to language in general, and specifically appeals to the general perspective adopted by Hjelmslev, the Prague School and Firth (Halliday 1978, esp. pp. 21, 39).

Advocacy for the orthodox generative view of phonology that emerged in the 1960s is increasingly rare. Even one who is prepared to defend the 'revised standard theory' begins by raising the question of whether he has a position to defend (Anderson 1979, p. 2). Nevertheless, many phonologists still consider themselves to stand in the generative tradition – including some of those contributing to laboratory phonology (section 11.18 above) – but few if any of them accept the classic generative phonology proclaimed in SPE. Kenstowicz (1994) is a remarkably detailed and comprehensive survey of how generative phonology has developed, and it demonstrates not only that generative phonology can be seen as an evolving theoretical enterprise but also that the evolution has been substantial. Trends in autosegmental, CV, metrical and lexical phonology (sections 11.12–11.14 above) are converging towards a new generative model that is largely expressed in the framework of optimality theory (e.g. Prince and Smolensky 2004). Many of the post-SPE generative models are radically different from SPE in a number of respects, but perhaps most fundamentally in their concern with structures such as syllables and feet, and autosegmental structures, and in the shift of focus away from complex derivational processes (applying to relatively abstract underlying forms). The sophisticated representational models (especially multitiered and hierarchically organized structures) that evolved in the last twenty years of the twentieth century are commonplace in any phonology textbook or research article.

Of course the prospect of genuinely blending all currents in phonology – particularly functional approaches with (post)generative approaches – is unrealistic. More is at issue than just notation and terminology, and as long as some of those who work in the generative tradition are still prepared, for example, to see linguistics as cognitive psychology (e.g. McCawley 1986, pp. 37–8) or to assume a model of ordered rules or ranked constraints within various components of a grammar (e.g. Prince and Smolensky 2004), there will remain fundamental differences of theoretical commitment.

There is no room here for an eclecticism which claims to take the best from each approach: the idea that one can pick a few choice fruits while ignoring the trees tends to superficiality rather than omniscience. Neither the investigation of phonetic and phonological questions themselves nor the application

of phonetic and phonological insights to fields such as speech pathology and language teaching can profit from the illusion that there are facts and truths independent of their derivation and expression. Thus if there is scientific maturity in modern phonology (Ohala 1986, pp. 3–5), it is not because there is an agreed unified theory or even a consensus about theoretical issues, and certainly not because there is some body of facts accepted once and for all, but rather because scholars are willing to discuss and explore their theoretical assumptions (see Pierrehumbert et al. 2000 for an excellent discussion of a similar range of issues). The nature of speaking and hearing will continue to be a proper subject of human curiosity, and phonetics and phonology will continue to be relevant wherever speech and hearing need to be explored and understood. What makes phonetics and phonology exciting – perhaps no more than other fields of specialized inquiry, but decidedly no less either – is that we cannot separate the exploration of what lies behind the everyday and the obvious from the confrontation with questions that are fundamental to science in its widest sense.

Exercises

1 Explain briefly what is meant by each of the following.

   a. broad and narrow phonetic transcription
   b. syntagmatic and paradigmatic relations
   c. natural phonology
   d. a postlexical rule

2 Compare as many different definitions of the phoneme as you can. Do these different definitions represent different perspectives on the same reality or do they reflect fundamental differences in theorizing about reality?

3 Which of the following pairs of words could be said to share an underlying phonological form? What criteria do you use to answer this question?

   a. broad, breadth
   b. father, paternal
   c. go, went
   d. moon, month
   e. sink, sank
   f. tempo, time
   g. twice, two
   h. wide, width

4 Is phonology empirical?

5 Briefly describe the importance of each of the following in the development of modern phonology.

   a. Franz Boas
   b. Baudouin de Courtenay

c. Leonard Bloomfield
d. J. R. Firth
e. Louis Hjelmslev
f. Roman Jakobson
g. Daniel Jones
h. Edward Sapir
i. Henry Sweet
j. Nikolai Trubetzkoy

Appendix 1: Phonetic Symbols

1.1 Vowel symbols (see section 2.7)

(a) UNROUNDED

	Front	*Central*	*Back*
High or close	i ɪ	ɨ	ɯ
High-mid or half-close	e	ɘ ɜ	ɤ
Low-mid or half-open	ɛ æ		ʌ
Low or open	a		ɑ

(b) ROUNDED

	Front	*Central*	*Back*
High or close	y ʏ	ʉ ʊ	u
High-mid or half-close	ø		o
Low-mid or half-open	œ		ɔ
Low or open	Œ		ɒ

Diacritics commonly used with vowels

˖ or ˂	fronted, e.g. [u̟], [ụ] or [u˂] for a fronted version of [u]
˗ or ˃	retracted, e.g. [e̠], [ẹ] or [e˃] for a retracted version of [e]
˔ or ˆ	raised, e.g. [e̝] or [eˆ] for a raised version of [e]
˕ or ˇ	lowered, e.g. [e̞] or [ĕ] for a lowered version of [e]
¨ or –	may indicate a centralized vowel (as in [ë], [ɨ] or [a̽])
ː	long, e.g. [uː] or [eː]
ˑ	slightly lengthened, or 'half-long', e.g. [uˑ] or [eˑ]
̥	voiceless or whispered, e.g. [ə̥] or [u̥]
̤	with breathy voice, e.g. [a̤] or [ṳ]
̰	with creaky voice, e.g. [a̰] or [ṵ]
̃	nasalized, e.g. [ã] or [ũ]