++
	John M. Anderson, Hope C. Dawson & Brian D. Joseph. 2010. Historical linguistics.	
	In K. Malmkjaer, ed., The Routledge Linguistics Encyclopedia (3rd edition), Routledge	
	Publishers, pp. 225-51.	
++

Historical linguistics[1]

Introduction

Historical linguistics examines the nature of linguistic change, looking at how and why languages change, and what the underlying forces and processes are that shape, mould and direct modifications in language. Engaging in this enterprise, historical linguists also map the world's languages, reconstruct their earlier states, determine their relationships to one another and, with the use of written documentation, fit extinct languages of the past into the jigsaw puzzle of the world's complex pattern of linguistic distribution. The historian of language must also identify the various influences that are at work in language change relating to both internal conditions in the linguistic system itself and external forces at play, such as language contact, adherence to social norms and the like.

Historical linguistic studies are important for our understanding of human language in general. Study of language change can reveal or test **language universals**, with data from differences between stages of languages being analogous to the typologist's cross-linguistic surveys [*see* LINGUISTIC TYPOLOGY]. Furthermore, the structural, social and biological complexity of language, and its relationships to other forms of communication, can be fully understood only when we know how it responds to internal and external stimuli. Language is always embedded in a social and historical context.

Historical background

We start with a brief overview of the development of historical linguistics. Discussing the history of the field is not just an antiquarian's exercise, but reveals the course of scholarly investigations that led to dramatic and still highly relevant findings in the eighteenth and nineteenth centuries.

Pre-modern era

The works of early Greek and Roman philosophers and grammarians include musings about **etymology** (in the ancient Greek sense, 'the true meaning of a word'), the **origin of language** and the role of **pattern** (**analogy**) in shaping language, issues that have concerned historical linguists ever since.

But it was with the advent of the European Renaissance that historical linguistics began to come into its own as an independent field of inquiry. Both local (typically Indo-European) and farther flung (typically non-Indo-European) languages came under scholarly scrutiny. As trade routes opened up to the East and explorers ranged the lands of the New World, data on exotic languages began to accumulate and stimulate the imagination. Vernacular languages came to be deemed worthy of study, and diversity in the world's linguistic structures was recognised.

An important trend in the seventeenth century was the effort to compare and classify languages in accordance with their resemblances. The study of etymology also gained momentum, but word derivations were still posited by scholars somewhat haphazardly, for instance, by rearranging the letters of some putative source language, especially Hebrew (thought by many to have been the original language).

Early in the eighteenth century, comparative and historical linguistics gained more consistency.

For instance, Job Ludolf in 1702 stated that affinities between languages must be based on grammatical resemblances rather than vocabulary, and among vocabulary correspondences the emphasis should be on simple words such as those that describe parts of the body. In a paper published in 1710, Gottfried Leibniz maintained that no known historical language is the source of the world's languages since they must be derived from a **proto-speech**. He also attempted to establish language classifications and toyed with the idea of a universal alphabet for all languages.

Despite continued interest in the origin of language, especially in the works of Hobbes, Rousseau, Burnett (Lord Monboddo), Condillac and Herder, the fundamental historical study of language can be said to have begun in earnest at this time through efforts to compare and classify languages in accordance with their origins, hypothetical or otherwise. The crowning achievement in the latter part of the eighteenth century came with the discovery that the Sanskrit language of ancient India was related to the languages of Europe and to Latin and Greek.

Sanskrit and its impact on the West

The first known reference in the West to Sanskrit occurred at the end of the sixteenth century when Filippo Sassetti wrote home to his native Italy about the *lingua Sanscruta* and some of its resemblances to Italian. Others, too, such as B. Schulze and Père Coerdoux, made similar observations on the resemblance of Sanskrit to Latin and European languages. The importance of these relationships came to the fore in 1786, however, when Sir William Jones, a judge in the English colonial administration, announced to the Royal Asiatic Society in Calcutta that Sanskrit, Greek, Latin, Gothic and Celtic seemed to have the same origin, a language that perhaps no longer existed. In his words (in Lehmann 1967: 15):

> The Sanskrit language, whatever be its antiquity, is of a wonderful structure; more perfect than the Greek, more copious than the Latin, and more exquisitely refined than either, yet bearing to both of them a stronger affinity, both in the roots of verbs and in the forms of grammar, than could possibly have been produced by accident; so strong indeed, that no philologer could examine them all three, without believing them to have sprung from some common source which, perhaps, no longer exists: there is a reason, though not quite so forcible, for supposing that both the Gothic and the Celtic, though blended with a very different idiom, had the same origin with the Sanskrit; and the Old Persian might be added to the same family.

Interest in the discovery mounted and, early in the nineteenth century, Sanskrit was being studied in the West. Sanskrit philological studies were initiated in Germany by W. von Schlegel about the time the first Sanskrit grammar in English was published. The linguistic study of this language set in motion the comparison of Sanskrit with languages of Europe, forming the first period in the growth of historical linguistics and setting **comparative linguistics** on a firm footing. Meanwhile, systematic etymological studies helped clarify and cement the family ties of the Indo-European languages. The modern era of historical linguistic studies can be said to have been launched at this point.

The introduction of Sanskrit and its subsequent study in Europe was a prime inducement to **comparative-historical linguistics** (which came to be known also as **comparative philology**). It came at an auspicious moment: the time was right for more cohesive approaches than the sporadic attempts of earlier scholars. It is generally accepted that the nineteenth century is the era *par excellence* of comparative-historical linguistics – a century in which most of the linguistic efforts were devoted to this subject, led (in the main) by German scholarship.

The nineteenth century

A few of the best-known historical linguists of the early nineteenth century are the Dane Rasmus Rask and the Germans Franz Bopp and Jacob Grimm. Bopp (1791–1867) published a work in 1816 comparing the verbal conjugations of Sanskrit, Persian, Latin, Greek and German. After adding Celtic and Albanian, he called these

the **Indo-European family of languages**. Bopp has often been considered the father of Indo-European linguistics.

Rask (1787–1832) wrote the first systematic grammars of Old Norse and Old English and, in 1818, he published a comparative grammar outlining the **Scandinavian languages**, noting their relationships to one another. Through comparisons of word forms, he brought order into historical relationships, matching a letter of one language to a letter in another, so that regularity of change could be observed.

Jacob Grimm (1785–1863) restricted his studies to the **Germanic family**, paying special attention to **Gothic** due to its historical value (having been committed to writing in the fourth century). This endeavour allowed him to see more clearly than anyone before him the systematic nature of **sound change**. Within the framework of comparative Germanic, he made the first statements on the nature of umlaut (see below) and ablaut, or, as it is sometimes called, vowel gradation (as found, for example, in German *sprechen*, *sprach*, *gesprochen* 'speak, spoke, spoken'), and developed, more fully than Rask, the notion of *Lautverschiebung*, or sound shift.

One specific case he examined is referred to as **Grimm's Law** ('law' in the sense of a statement of regular behaviour), or the **First Germanic Sound Shift**. Grimm's *Deutsche Grammatik*, published in 1822, contained general statements about similarities between Germanic obstruents – i.e. plosives, affricates and fricatives – and their equivalents in other languages. Using the old terms of Greek grammar where T = *tenuis* (p, t, k), M = *media* (b, d, g) and A = *aspirate* (f, θ, x), he noted:

Proto Indo-European	=	*Germanic*
T		A
M		T
A		M

A modern tabulation of his conclusions would appear as:

Indo-European	>	*Germanic*
p		f
t		θ
k		x
Indo-European	>	*Germanic*
b		p
d		t
g		k
Indo-European	>	*Germanic*
bh		b
dh		d
gh		g

Interest also began to develop in the causes of language change. Jacob H. Bredsdorff (1790–1841), a disciple of Rask, proposed in 1821 such factors as mishearing, misunderstanding, misrecollection, imperfection of speech organs, indolence, the tendency towards analogy, the desire to be distinct, the need for expressing new ideas and influences from foreign languages.

Some of his ideas are still viable today. For instance, it is recognised that the tendency towards **analogy**, i.e. speakers' desire for uniformity and for regular patterns, causes language to become more rather than less regular in syntax, morphology and phonology. Colloquial speech – which popular, though rarely expert, opinion often classifies as **indolent** – can also eventually result in changes in pronunciation, spelling, grammatical patterning and semantics. And the speech organs certainly are involved in sound changes as well, though we would now speak in terms of physiological constraints on the vocal tract rather than imperfections. The influence from foreign languages is clearly observable when words are borrowed from another language, as when *pizza* entered English from Italian or when *weekend* entered Danish from English. This is often motivated by the need of speakers of a language to express a new idea or name a new thing – pizzas were at one time unfamiliar in the USA and Britain, and at one time Danish did not have a word that could express the conceptualisation of the weekend as a whole. Similarly, new inventions often result in the need for new terminology, as when the advent of computers led to the coinage of the term *software* by analogy with *hardware*, which was itself borrowed from another sphere, namely that of the traditional metal fittings used in strengthening things made of wood.

In the mid-nineteenth century, one of the most influential linguists, August Schleicher

(1821–68), set about reconstructing the hypothetical parent language from which most European languages were derived – the **proto-language** (see below). He also devised the *Stammbaumtheorie* or **genealogical family-tree model** of the **Indo-European languages** (see below). He worked out a typological classification of languages based on the work of his predecessors in which he viewed languages as isolating, agglutinating or inflectional [*see* LINGUISTIC TYPOLOGY]. On a more philosophical level, he brought to linguistics three important concepts mostly rejected today but which at the time stimulated much discussion and work in the discipline; namely: that language is a natural organism, that it evolves naturally in the Darwinian sense, and that language depends on the physiology and minds of people (that is, it has racial connotations). In short, he stimulated a new and different approach to language study – a **biological approach**.

The work of Schleicher represents a culmination of the first phase of historical linguistics in the nineteenth century. In the second half of the century the discipline of linguistics became more cosmopolitan as scholars in countries other than Germany began seriously to investigate linguistic problems. Germany, however, remained the centre of linguistic attention throughout the century.

In 1863, Hermann Grassmann, a pioneer in internal reconstruction (see below), devised a phonetic law based on observations of the Indo-European languages, showing why correspondences established by Grimm did not always work. His **Law of the Aspirates** demonstrated that, when an Indo-European word had two aspirated sounds [*see* ARTICULATORY PHONETICS] in the same syllable, one (usually the first) underwent deaspiration. For example, Sanskrit *da-dhā-mi* 'I put' < **dha-dhā-mi* shows the reduplicated syllable of the root reduced through loss of aspiration (the asterisk indicates that the form is reconstructed). This exception to Grimm's Law, where Sanskrit [d] corresponds to Germanic [d] (compare English *do*) and not to [t], then, proved to be a law itself.

In 1875, still another phonetic law was proposed by Karl Verner (1846–96). This succeeded in accounting for other exceptions to Grimm's statements by showing that the position of the Indo-European accent was a factor in the regularity of the correspondences. For example, Indo-European [t] in [*pətḗr] became [ð] in Germanic [faðar], not [θ], as might be expected. The accent later shifted in Germanic to the first syllable.

In his 1870 *Corsi di Glottologia*, Graziadio Ascoli (1829–1907) demonstrated by comparative methods that certain [k]s elsewhere in Indo-European correspond to Sanskrit [ʃ] (transliterated as *ś*). Compare the word for 'one hundred':

Latin	***c**entum*
Greek	*(he)**k**aton*
Old Irish	***c**et*
Sanskrit	***ś**ata*
English	***h**undred*

By the principles of comparative reconstruction (see below), such correspondences allowed for the positing of an original stop that became a fricative in Sanskrit, thereby ending the belief that Sanskrit was the oldest and closest language to the proto-form or parent language.

The formulation of such sound laws, which appeared to be systematic and regular to the extent that exceptions were laws themselves, gave rise to one of the most important and controversial theories in historical linguistics, promulgated in the doctrine of the Neogrammarians or *Junggrammatiker*.

The Neogrammarian era

Inspired in 1868 by the ideas of Wilhelm Scherer (1841–86) who, in his 1868 book on the history of the German language (Scherer 1868), advocated fixed laws in sound change, the Neogrammarian movement soon dominated linguistic enquiry. To account for situations where phonetic laws were not upheld by the data, Scherer looked to **analogy** as the explanation for change. The chief representatives of the movement – Karl Brugmann, Hermann Osthoff, Berthold Delbrück, Jacob Wackernagel, Hermann Paul and August Leskien – held that phonetic laws were similar to laws of nature in the physical sciences in their consistency of operation. In 1878, in the first volume of a journal edited by Brugmann (1849–1919) and Osthoff (1847–1909), *Morphologische Untersuchungen*,

they delineated the Neogrammarian doctrine and the special designation *junggrammatische Richtung* ('Neogrammarian School of Thought'). The crux of their doctrine was, as Osthoff (1878: 326) put it: 'sound-laws work with a blind necessity' and all discrepancies to these laws were the workings of analogy. Centred around the University of Leipzig, the Neogrammarians saw in sound change the application of laws of a mechanical nature opposed by the psychological propensity of speakers towards regularisation of forms.

The Neogrammarian doctrine did not go unopposed. For example, the psychologist Wilhelm Wundt (1832–1920) found fault with their views relating to psychological aspects of language. In addition, Hugo Schuchardt (1842–1927) of the University of Graz published an article in 1885 on sound laws in which he considered language change to be due to a mixing process both within and outside language, leading to the formulation of a **Substratum Theory**, in which languages are influenced by a mixture of populations (see below).

One further key conceptual innovation of the era came with the work of Ferdinand de Saussure (1857–1913) of the University of Geneva. His view of language as a system of arbitrary signs in opposition to one another and his separation of **synchronic** (descriptive) linguistics and **diachronic** (historical) linguistics into two distinct spheres of investigation earned him the reputation as one of the founders of structural linguistics [*see* INTRODUCTION].

The twentieth century and the modern era

After Saussure and the rise of generative linguistics in the middle of the twentieth century, the field of **descriptive linguistics** developed rapidly while historical linguistics and comparative studies lost their pre-eminence.

Today, among the disciplines that make up the broad field of linguistics (descriptive, historical, sociological, psychological, etc.), historical linguistics, from once being the embodiment of the discipline, has become another branch of the multivaried area of investigation. Contemporary advancements in historical-comparative language studies have been on the practical side, with the collection of data and reformulation of previous work. On the theoretical side, much has come from advancements in descriptive linguistics and other branches of the discipline – for example, from structural concepts such as the phoneme, and refinements in phonetics, to more stringent application of ordered rules and underlying structures, statistical methods and their relationship to language change and language universals, and increased understanding of the social factors relevant to the spread of change.

Principles, methods, objectives and data of historical linguistics

Certain principles in the field of historical linguistic enquiry are taken as axiomatic; for example:

- All languages are in a continual process of change.
- Language change is regular and systematic, allowing for unhindered communication among speakers.
- Linguistic and social factors are interrelated in language change.
- All languages are subject to the same kinds of modifying influences, including the constraints and restrictions associated with the notion of 'possible human language'.

To elaborate on this last point, a linguistic change or state not attested in known languages would be suspect if posited for an earlier stage through reconstruction. A sound change like [b] → [k] between vowels would be considered unlikely on phonetic grounds. Similarly, no system of consonants in any known language consists entirely of voiced fricatives, so that any reconstruction that ignored this observation and posited only voiced fricatives would be highly questionable. [*See* ARTICULATORY PHONETICS.]

The **diachronic study** of language may be approached by comparing one or more languages at different stages in their histories. **Synchronic** studies underlie historical investigations inasmuch as an analysis of a language or a part thereof at period *A* can then be compared to a descriptive study at period *B*. For example, an investigation of English at the time of Chaucer, and another of Modern English, would reveal a number of differences. Similarly, a descriptive statement of Latin and one of Modern French would disclose very different systems in phonology and

morphosyntax. The **historical linguist** attempts to classify these differences and to explicate the manner and means by which they came about.

When the various historical facts of a language are discovered, the investigator might then establish general rules based on the data. These rules will demonstrate in more succinct form the manner in which the language changed and how it differs from other related languages.

Rules of change may be written in several ways: [t] → [d]/V_V states that the sound [t] becomes [d] in the environment between vowels. Such rules can also be stated in **feature specification**:

$$\begin{bmatrix} +\text{consonantal} \\ +\text{plosive} \\ +\text{coronal} \\ +\text{anterior} \\ -\text{voiced} \end{bmatrix} \rightarrow [+\text{voiced}] / [+\text{vocalic}] \_\_ [+\text{vocalic}]$$

When, as is often the case, an entire class of sounds – for example, [p t k] – behaves in an identical manner, instead of different rules for each sound, one rule suffices:

$$\begin{bmatrix} +\text{consonantal} \\ +\text{plosive} \\ -\text{voiced} \end{bmatrix} \rightarrow [+\text{voiced}] / [+\text{vocalic}] \_\_ [+\text{vocalic}]$$

If we were to compare Latin and Italian, we would find such words as:

Latin	*Italian*	
noctem	*notte*	'night'
octo	*otto*	'eight'
lactem	*latte*	'milk'
factum	*fatto*	'fact'
lectum	*letto*	'bed'

In these examples, and others that could be added, we discover that Latin [k] (e.g., in [noktem]) became Italian [t] in the environment before [t]. This assimilatory change (see below) is a general process in Italian and can be stated in rule-like fashion as: [k] → [t]/__[t], or it can be stated in feature specifications. The rule helps account for the differences between Latin and Italian, and between Italian and other Romance languages, where a different set of changes apply to give, say, Spanish *noche* [nótʃe] and French *nuit* [nyɪ].

Objectives of the practitioners of historical linguistics vary. Excluding here language changes resulting from evolutionary or maturation processes of developing neuro-anatomical structures of *Homo sapiens*, some historical linguists are concerned with phonological, morphological, syntactic and semantic changes that occur in languages over a given period of time, to acquire an understanding of the mechanisms underlying the modifications and to seek explanations for them. Answers to these questions also bear on the nature of the species and may be sought within cognitive and physiological parameters that govern the behaviour of the species.

Other historical linguists may be more concerned with reconstruction and comparison of languages to arrive at historical relationships indicating common origins of languages, which allow them to be grouped into families. The geographical distribution of families is of paramount importance in our understanding of migrations and settlement patterns over the surface of the earth.

Sociological aspects of language change encompassing questions of dialect, style, prestige, taboos, changes in social behaviour, technology and even individual needs to be different are also important considerations in the understanding of cultural associations and ultimately human behaviour.

The changes that languages undergo make up the data for historical linguists and are themselves generally transmitted by and derived from written documentation or reconstructed from the languages in question if such records are not available.

In cases where the underlying language of the documentation is known, such as Old English, Latin and Sanskrit, the investigator must try to determine the orthoepic features of the language through knowledge of the writing system employed, through commentary on the language by contemporary authors, by rhyme and by the pronunciation of the descendent languages.

In dealing with primary written sources inscribed in an unknown language, the investigator must decipher the texts in order to gain a clear view of the underlying linguistic structure. The performance of this task must take into

account the kind of writing system used, the direction of writing and the phonetic basis underlying the orthographic signs. Morphemes and morpheme boundaries must be determined, syntactic features assessed and semantic properties determined.

Phonological change

Regularity of sound change

[For explanation of the phonetic terms in this and the following sections, *see* ARTICULATORY PHONETICS.]

In talking about pronunciation changes, we draw a technical but crucial distinction between **changes in sound** and **sound change proper**, for there can be changes in the phonetic realisation of words that have nothing to do with sound change in its strictest sense (that is, sound change proper). When we speak of sound change proper, we mean modifications in the sounds of a language that are regular and systematic, applying in the same manner in all instances of a specified phonetic environment. The reflexes of the Latin vowel [a], for example, demonstrate this principle.

Latin [a] regularly became French [ɛ] when [a] was accented and free, that is, in an open syllable, as in [má-rem] and the following examples:

Latin	*French*	
marem	*mer* [mɛʁ]	'sea'
fabam	*fève* [fɛv]	'bean'
patrem	*père* [pɛʁ]	'father'
labram	*lèvre* [lɛvʁ]	'lip'

The accented Latin vowel [a] in an open syllable, but followed by a nasal, resulted in [ɛ̃]:

Latin	*French*	
manum	*main* [mɛ̃]	'hand'
panem	*pain* [pɛ̃]	'bread'
planum	*plain* [plɛ̃]	'plane'
famen	*faim* [fɛ̃]	'hunger'

But there are also cases where Latin [a] became French [a], and while these may at first glance appear to have been exceptions to the above rule, they were in fact the result of another regular sound change in which accented [a] behaved predictably in a closed environment, that is, in a closed syllable or one blocked by a consonant, as in [pár-te], [vák-ká], etc. Compare:

Latin	*French*	
partem	*part* [paʁ]	'part'
vaccam	*vache* [vaʃ]	'cow'
carrum	*char* [ʃaʁ]	'cart'
cattum	*chat* [ʃa]	'cat'

And when Latin [a] was closed by a nasal consonant, the result was a nasal [ã] as in:

Latin	*French*	
campum	*champ* [ʃã]	'field'
grande	*grand* [grã]	'large'
annum	*an* [ã]	'year'
manicam (mancam)	*manche* [mãʃ]	'sleeve'

Since the environment dictated the sound change, the conditions of the modifications can be established along the following lines (where . = syllable boundary, C = oral consonant, N = nasal consonant):

[a]> [ɛ]/_ . C
[ɛ̃]/_ . N
[a]/_ C .
[ã]/_ N .

This general rule requires clarification based on further environmental factors that regularly affect the vowel [a]. For example:

Latin	*French*	
alterum	*autre* [otʁ]	'other'
valet	*vaut* [vo]	'is valued'

where [a] plus [l] becomes [au] and subsequently monophthongises to [o].

Beginning in the period of Late Old French, the vowel [ɛ] (from [a]) underwent a further change to become [e] when the syllable became open through the loss of a final consonant, cf.:

Latin	*French*	
clavem	*clé* [kle]	'key'
pratum	*pré* [pre]	'meadow'

When [a] was unaccented, it underwent another set of changes, which resulted in [ə] or [a] as in:

Latin	*French*	
camisam	*chemise* [ʃəmiːz]	'shirt'
amicum	*ami* [ami]	'friend'

The treatment of [a] in the above examples is intended to be indicative of the kind of regularity found in sound change and shows the value of looking to finely grained phonetic environments in determining the correct formulation of sound changes (proper).

Processes of sound change

The mechanisms by which sound change occurs involve changes in the features of a sound (e.g., voiceless, voiced, plosive, fricative) or the addition, loss or movement of sound segments. Many such changes are of an anticipatory nature in that a modification takes place due to the influence of a following sound; for example, the **assimilation** of [k] → [t]/_[t] in Latin *octo* [okto] to Italian *otto* 'eight' is of this type, in which the feature velar is changed to dental before a following dental sound. Compare:

[k]	[t]
voiceless	voiceless
plosive	plosive
velar	dental

Other processes of this type include **nasalisation**, as in Latin *bonum* to Portuguese *bom* [bõ] 'good', where a non-nasal vowel acquires the nasality of a following nasal consonant.

Often a velar consonant becomes a palatal consonant under the influence of a following front vowel that pulls the highest point of the tongue from the velar forward into the palatal zone; such a **palatalisation** is exemplified by Old English *kin* [kɪn] becoming Modern English *chin* [ʧɪn], or Latin *centum* [kentum] becoming Italian *cento* [ʧɛnto] 'one hundred'.

A specific kind of assimilation, referred to as **sonorisation**, involves the voicing of voiceless consonants and appears to be motivated primarily by voiced surroundings. For example, voiceless [p], [t] and [k] became [b], [d] and [g] in the environment between vowels in an earlier stage of Spanish, as in the following examples:

Latin	*Spanish*		
cupa	*cuba* [ˈkúba]	'vat'	[p] → [b]
vita	*vida* [ˈbida]	'life'	[t] → [d]
amica	*amiga* [aˈmiga]	'friend'	[k] → [g]

Assimilation may take place over syllable boundaries, as occurs in the process affecting vowels commonly called **umlaut**. For example, the Proto-Germanic form *[musiz] gave Old English [miːs] (Modern English *mice*) when the tongue position for the vowel in the first syllable was drawn forward through the influence of the front articulation of the vowel in the second syllable. Similarly, Latin *feci* 'I made' gave rise to Spanish *hice* when the influence of the Latin vowel [i] raised [e] to [i] through assimilation. Final [i] subsequently lowered to [e]. Compare also Latin *veni* 'I came' and Spanish *vine*.

The opposite of assimilation, **dissimilation**, modifies a segment so that it becomes less like another, often neighbouring, segment in the word. Dissimilation is less frequent than assimilation in the known histories of the world's languages. The conditioning sound may be adjacent to the sound that undergoes change, or dissimilation may operate at a distance. The first case is illustrated by Latin *luminosum* 'luminous', which became Spanish *lumbroso* when, after the loss of unaccented [i], the resultant nasal + nasal cluster [mn] dissimilated to [mr] and subsequently became [mbr]. The nasal [n], in losing its nasal quality and changing to [r], became less like the adjacent [m]. The second case is illustrated by Latin *arbor* 'tree', which became Spanish *arbol* when [r] changed to [l] under the influence of the preceding [r].

The addition of a segment into a particular environment of the word, **epenthesis**, is essentially a form of anticipation of a following sound and may involve either consonants or vowels. The Middle English verb *glymsen* gave rise to Modern English *glimpse* through the insertion of an epenthetic [p] in the environment [m_s]. The inserted sound develops in the transition between the bilabial [m] and the voiceless and oral [s]. Compare Old English *þunrian*, Modern English *thunder*.

We see epenthesis also at work in the adaptation of foreign loan words to native phonological patterns. For example, Basque speakers borrowed a number of words from late Latin with

certain consonant clusters not found in Basque. Vowels were inserted in the borrowed words to make them more compatible with the Basque phonological system, which, for example, tended to avoid sequences of plosive plus [r]; compare:

Latin	*Basque*	
[krus]	[guruts]	'cross'
[libru]	[libiru]	'book'

The addition of a word-initial segment applied generally to facilitate the pronunciation of an initial consonant cluster is a process referred to as **prothesis**; for example,

Latin	*Spanish*	
schola [skola]	*escuela* [eskwela]	'school'
stella [stela]	*estrella* [estreʎa]	'star'

Sounds are also subject to deletion. The two most common processes of vowel deletion are **apocope** and **syncope**, which are especially common in environments after accented syllables. In word-final position, apocope has been common in the history of many languages including French. Compare:

Latin	*French*	
cane [kane]	*chien* [ʃjɛ̃]	'dog'
caru [karu]	*cher* [ʃɛʁ]	'dear'

The loss of a word-medial vowel, or syncope, occurs in English in words such as *vegetable* [ˈvɛʤtəbl̩], where the unaccented second syllable lost the vocalic segment. The process does not commonly occur in English, however, but appears much more readily in the Romance languages.

Latin	*Spanish*	*French*	
viride	*verde*	*vert*	'green'
lepore	*liebre*	*lièvre*	'rabbit'
calidu	*caldo*	*chaud*	'hot'

Consonantal loss in word-final position is also common among many languages. Again, we see in French the deletion of consonants in forms such as Latin *pratu* → French *pré* via **pret*. Other word positions are also vulnerable to deletion of segments. Old and Middle English employed the cluster [kn-] as in *knight*, *knot*, *knee*; the [k] was lost between Middle and Modern English.

A change in the relative position of sounds is referred to as **metathesis**. Adjacent sounds may be affected, as in the Old English *beorht*, yielding Modern English *bright*, where CV*r*C became C*r*VC. Sounds separated by some phonetic distance may also undergo metathesis as, for example, vernacular Latin *mirac(u)lu* 'miracle' became Spanish *milagro* through the transposition of [l] and [r].

A number of other processes are often at work in sound change. Stated briefly, some further changes that affect consonants are:

aspiration	[t]	→	[tʰ]
affrication	[t]	→	[ts]
labialisation	[t]	→	[tʷ]
prenasalisation	[t]	→	[ⁿt]
glottalisation	[t]	→	[tˀ]
velarisation	[t]	→	[ɬ]
rhotacisation	[z]	→	[r]

Or, the opposite changes occur: deaspiration, deaffrication, etc. Further processes observed among vocalic segments are:

raising	[e]	→	[i]
lowering	[i]	→	[e]
fronting	[o]	→	[ø]
backing	[ø]	→	[o]
rounding	[i]	→	[y]
unrounding	[y]	→	[i]
lengthening	[a]	→	[aː]
shortening	[aː]	→	[a]
diphthongisation	[e]	→	[ie]
monophthongisation	[ie]	→	[e]

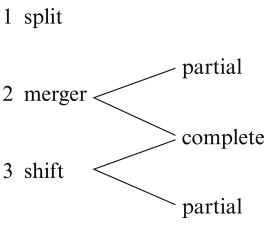
An entire syllable may also undergo loss, a process called **haplology** when a repetitive syllable is involved, cf. Latin **stipipendium* → *stipendium* 'wages'.

Change in phonological systems

As we have seen, phonemes develop variants in accordance with environmental conditions and are the result of influences exercised through phonetic processes such as assimilation. We know, for example, that English vowels have nasalised variants preceding nasal consonants, as in the word *can't*, but not in other environments, compare *cat* – phonetically (US) [kʰæ̃nt] vs. [kʰæt]. These phonetic changes have no impact

on the overall phonological system, since the variation is conditioned and predictable, affecting only the distribution of allophones [*see* PHONEMICS].

Sound changes that result in an increase or reduction in the number of phonemes in a language, or lead to the replacement of phonemes by others, are generally referred to as **splits** or **mergers**. A change in which several phonemes are replaced in a systematic way is called a shift, which also may be partial or complete:



If, in English, nasal consonants were to disappear, the form *can't* would be represented phonetically as [k^hãt] and would, in fact, contrast with *cat* as /kãt/, /kæt/, with the distinguishing feature of nasal versus non-nasal vowel. What was once a phonetic feature of the language, through the loss of the nasal consonant would then become a phonemic feature brought about by phonological split. Something similar to this occurred in French, where nasal and non-nasal vowels distinguish meaning:

Latin	*French*	
bonus	*bon* /bõ/	'good'
bellus	*beau* /bo/	'pretty, handsome'

At some stage in the history of English, allophonic conditioning led to the development of a velar nasal [ŋ] before a velar plosive through assimilation. In the course of Middle English, the voiced velar plosive disappeared in word-final position after the nasal consonant, as in the words *young* or *sing*. These stages can be summarised as /sɪng/ → /sɪŋg/ → /sɪŋ/. The velar nasal allophone of /n/, then, became a separate phoneme, as evidenced by such minimal pairs [*see* PHONEMICS] as:

sin	/sɪn/
sing	/sɪŋ/

A phoneme may also split into multiple forms. Compare these developments in French:

Latin	*French*
	k/__w
/k/	s/__ [i, e]
	ʃ/__a

in such words as:

Latin	*French*	
quando	*quand* /kã/	'when'
centum	*cent* /sã/	'hundred'
campus	*champ* /ʃã/	'field'

Phonological split may also result in merger in which no new phonemes are created in the language. In most dialects of American English, for example, /t/ split into the voiceless stop [t] and the voiced flap [ɾ] in certain environments and [ɾ] merged with the similarly arising allophonic flap associated with the phoneme /d/. This gave rise to the homophony of *latter* with *ladder* and *bitter* with *bidder*.

Mergers may be **partial** or **complete**. If merger is complete, there is a net reduction in the number of phonemes in the language. Such is the case in some varieties of the non-standard London dialect Cockney (among many other dialects of English), where the two dental fricatives /θ/ and /ð/ have merged completely with /f/ and /v/, respectively. Hence, *thin* /θɪn/ is pronounced /fɪn/ and *bathe* /beɪð/ is pronounced /beɪv/. Four phonemes were reduced to two:

/f/	/θ/	→	/f/
/v/	/ð/	→	/v/

In African-American Vernacular English pronunciation in the USA, /θ/ merges partially with /f/, i.e. /θ/ → /f/ in all positions except word-initial. The form *with* is articulated as /wɪf/ but the word *thing* retains /θ/ as in /θɪŋ/ or /θæŋ/.

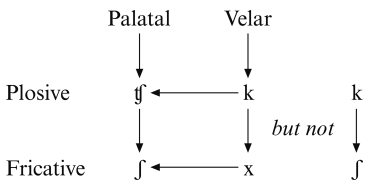
When a series of phonemes is systematically modified, such as /p/, /t/, /k/ → /b/, /d/, /g/, we may consider a wholesale shift to have occurred. A shift may be **partial**, when all the allophones of the phoneme do not participate in it, or it may be **complete**, when they do. The modification of long vowels in Late Middle English known as the **Great English Vowel**

Shift (see below) left no residue and appears to have been complete. The **First Germanic Consonant Shift**, in which /p/, /t/, /k/ → /f/, /θ/, /x/, however, left some of the voiceless plosives unaffected in specific environments, such as after /s/. Compare, for example, Latin *est* and German *ist* and see above.

Phonological processes that lead to allophonic variation and subsequent new phonemes generally occur one step at a time. The change of Latin /k/ to French /ʃ/, for example, in words such as *cane* /kane/ to *chien* /ʃjɛ̃/, did not happen directly, but instead involved two changes:

/k/	voiceless →	/ʧ/	voiceless →	/ʃ/	voiceless
	plosive		plosive		fricative
	velar		palatal		palatal

Phonological change usually takes place within the range of allophonic variation that varies by one feature. A phoneme /k/ might have allophones [t] or [x], which differ by one phonological feature, but not generally an allophone /ʃ/, which differs by two features. A change to /ʃ/ could be the result of either of the two allophones serving as intermediaries:



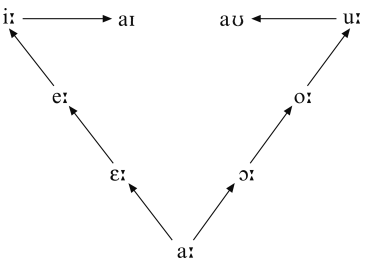
Non-phonologically motivated changes in pronunciation

Many phonological changes are not conditioned by the surrounding phonetic environments but are motivated by other factors relating to external forces, such as substratum influences, and internal forces inherent in the structural paradigmatic make-up of the language; it is often the case, however, that, obscured by time, these factors are no longer readily recoverable (though reasonable inferences can often be drawn on the basis of our knowledge of general patterns of language change). The **First Germanic Consonant Shift**, for example, occurred at a time in which there were no written records for the Germanic languages and under unknown circumstances.

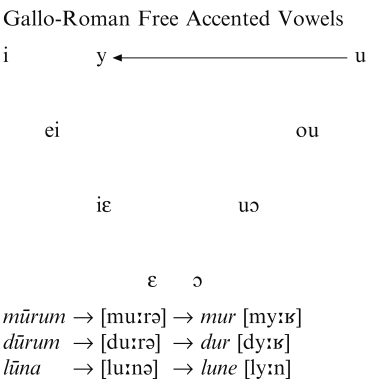
A major change in the history of English vowels took place at the end of the Middle English period (sixteenth century), in which the long tense vowels underwent a regular modification without the apparent assistance of an environmental stimulus. The modification is referred to as the **Great English Vowel Shift**.

Middle English	*Early Modern English*	
[miːs]	[maɪs]	'mice'
[muːs]	[maʊs]	'mouse'
[jeːs]	[giːs]	'geese'
[joːs]	[guːs]	'goose'
[brɛːken]	[breːk]	'break'
[brɔːken]	[broːk]	'broke'
[naːm]	[neːm]	'name'

The vocalic movement upward in which the high vowels diphthongised can be shown schematically as:



An upward pressure was also exerted on the back vowels of the Gallo-Roman language in about the ninth century during their evolution from Latin to French, and the high back vowel from Latin [uː], which had become [u], then shifted to [y].



Note [u] → [y] regardless of environmental position, so that explanations other than those involving conditioned change must be sought. One plausible interpretation of the event, based on paradigmatic considerations, suggests that, with the monophthongisation of Latin [au] → [ɔ] (*aurum* → *or* [ɔr]), which occurred prior to the change [u] → [y], the margin of tolerance, i.e. the physical space, between back vowels was not sufficient. The monophthongisation of [au] consequently forced upward pressure on the back vowels, and [u], the highest vowel, could go no higher and fronted.

The plosive and fricative consonantal structure of Early Old French of the eleventh and twelfth centuries consisted of the following phonetic inventory and relationships:

		Labial	*Dental*	*Pre-palatal*
Plosives	vl	p	t	ts
	vd	b	d	dz
		Palatal	*Velar*	
	vl	ʧ	k	
	vd	ʤ	g	
Fricatives	vl	f	s	
	vd	v	z	

(vl = voiceless; vd = voiced)

During the thirteenth century, the affricated palatal sounds became fricatives:

ć [ts]	→	s
ź [dz]	→	z
č [ʧ]	→	ʃ
ǧ [ʤ]	→	ʒ

The result of these changes was a later Old French system of consonantal sounds as follows:

p	t		k
b	d		g
f	s	ʃ	
v	z	ʒ	

The rationale for these changes has been sought in a tendency to reduce the overcrowded palatal zone and a leaning towards symmetry by reducing the five orders (labials, dentals, etc.) to four in accordance with the four series of plosives and fricatives.

In other attempts to explain phonological modifications that fall outside the realm of conditioned change, the notion of **substratum influence** has often been invoked. Certain words in Spanish, for example, developed an [h] (which has been lost in the modern language in pronunciation, but is still reflected in the orthography) where Latin had [f].

Latin	*Spanish*	
filium	*hijo* [ixo]	'son'
fabam	*haba* [áβa]	'bean'
folia	*hoja* [óxa]	'leaf'
feminam	*hembra* [émbra]	'female'
fumum	*humo* [úmo]	'smoke'

As the replacement of Latin [f] by [h] began in the north of the peninsula, where the Basques were in contact with Hispano-Roman speakers, and because Basque had no [f] sound, the hypothesis has been put forward that Basque speakers, upon learning the Hispano-Roman language, substituted their closest sound. According to this view, this sound was [ph] which subsequently became [h]. Those words not affected (cf. Latin *florem*, which became Spanish *flor*) were excluded from the change due to other factors, such as learned influences.

Diffusion of language change

Besides the study of mechanisms and processes of language change, the historical linguist must also be concerned with how changes spread throughout a speech community, as that too is part of the change's history. The vocabulary of a language may be modified by **lexical diffusion** in which a change begins in one or several words and gradually spreads in an essentially analogical fashion from word to word, with one serving as the model for the next, throughout the relevant portions of the lexicon. This therefore would be another non-phonologically motivated change in the pronunciation of a word. One such ongoing change can be seen in words such as *present*, which can be used as either a verb or a noun. At one time all such words were accented on the second syllable regardless of their status as noun or verb. In the period that gave rise to Modern English (sixteenth century), words such as *rebel*, *outlaw* and *record* began to be pronounced with the accent on the first syllable when they were used as nouns. Over the next few centuries

more and more words followed the same pattern, cf. *récess* and *recéss*, *áffix* and *affíx*. The diffusion process is still in progress, however, as indicated by the fact that many English speakers say *addréss* for both noun and verb and others use *áddress* as the noun and *addréss* for the verb. There are still many words that have as yet not been affected by the change, compare *repórt*, *mistáke* and *suppórt*.

Not all changes diffuse gradually through the lexicon. Some changes, especially sound change proper, affect all words in a given class at the same time. In some Andalusian dialects of Spanish, the phoneme /s/ has developed an allophone [h] in syllable-final position:

Standard pronunciation	*Andalusian*
[dos]	[doh]
[es]	[eh]
[mas]	[mah]

The change is regular and systematic, affecting all instances of syllable-final /s/ in the speech patterns of the individuals who speak this dialect.

Along with linguistic diffusion of change throughout the lexicon of the language, the linguist may also take into account diffusion of change throughout the speech community. A given speech modification begins in the speech habits of one or several individuals and spreads (if it spreads at all) to an ever-increasing number of people (a process that can be thought of as a kind of borrowing between dialects, with each speaker representing a 'dialect', that is, idiolect). Whether or not diffusion occurs may depend on the relative prestige of the people who initiate the change and their influence on the speech population, and on speakers' choices (largely unconscious) to model their speech on that of others they emulate or want to identify with (in the manner demonstrated by Labov 1963). If the prestige factor is high, there is a good chance that the innovation will be imitated by others. The loss of postvocalic /r/ in some eastern dialects of the USA was due to a change that originated in England and was brought to the New World by new settlers. Similarly, the adoption of the sound /θ/ in southern Spain (where no such sound existed) by speakers of the Andalusian dialect is due to their imitation of Castilian Spanish, the prestige dialect of Madrid and its surroundings.

Morphological and syntactic change

Effects of sound change on morphology

The effect of phonological change on aspects of morphology is evident in the restructuring of the plural forms in some English words:

	Germanic	*Old English*	*Modern English*
Sing	*mūs	mūs	[maʊs] 'mouse'
Pl	*mūsi	mīs	[maɪs] 'mice'
Sing	*fōt	fōt	[fʊt] 'foot'
Pl	*fōti	fēt	[fit] 'feet'

In these and examples like them, the process of **umlaut** or **mutation** operated to change the stem vowel [uː] → [iː] and [oː] → [eː] through the fronting influence of a following close front [i] which then disappeared. Subsequently, [iː] became [aɪ] and [eː] became [i] (see above), so that the modern forms show a phonetically unmotivated vowel change in the plural.

The influence of sound change on morphological structures may also be seen in the Old English system of nominal forms whose suffixes marked case and gender. Compare the Old English masculine noun *hund* 'dog'.

Old English

	Singular	*Plural*
Nom	hund	hund-as
Acc	hund	hund-as
Gen	hund-es	hund-a
Dat	hund-e	hund-um

Other nouns belonged to either masculine, feminine or neuter types distinguished on the basis of case endings, e.g., feminine *gief* 'gift' declined along the lines of *gief-u* in the nominative singular, *gief-e* in the accusative singular, etc.

Through phonological change, the case and gender distinctions of Old English were lost. By the fifteenth century, the /m/ of the dative plural suffix had been effaced and unaccented vowels of the case endings had been reduced to /ə/.

Middle English

	Singular	*Plural*
Nom	hund	hund-əs
Acc	hund	hund-əs
Gen	hund-əs	hund-ə
Dat	hund-ə	hund-ə

Previous distinctions between dative singular and dative plural, genitive singular and nominative plural, and so on, disappeared.

The distinction between singular and plural forms in Middle English was preserved by the continuance of the phoneme /s/, which survived also to mark the genitive singular forms. A genitive plural /s/ was added by analogy with the singular. The loss of case endings also obliterated the gender distinctions that were found among Old English forms. Sound change further modified the internal structure of morphemes such as *hund*, subject to the result of the Great English Vowel Shift, which diphthongised /u/ to /aʊ/ and resulted in:

Present-day English			
Singular		*Plural*	
hound	/haʊnd/	hounds	/haʊndz/
hound's	/haʊndz/	hounds'	/haʊndz/

Another such instance is the development of Latin into the Romance languages. Classical Latin contained six cases, which were reduced in the vernacular Latin speech of the Empire, and finally disappeared altogether in the Romance languages, with the exception of Romanian. Increasing stress patterns in Popular Latin gradually neutralised the differences between long and short vowels by creating long vowels in accented syllables and short vowels in unaccented syllables regardless of the original arrangement. With the concomitant loss of final *-m* in the accusative (by a regular sound change affecting final [m] in polysyllables), the nominative, vocative, accusative and ablative forms merged. The genitive and dative conformed to the rest of the pattern by analogy.

As in English, the loss of the case system brought on a more extensive and frequent use of prepositions and a more rigid word order to designate the relationships formerly employed by case functions.

	Classical Latin	*Popular Latin*	*French*
Sing			
Nom	*porta*	*porta*	*la porte*
Voc	*porta*	*porta*	*la porte*
Acc	*portam*	*porta*	*la porte*
Gen	*portae*	*de porta*	*de la porte*
Dat	*portae*	*ad porta*	*à la porte*
Abl	*portā*	*cum porta*	*avec la porte*

Word order, prepositions and articles

The developments of the Latin case system as the Romance dialects emerged provide a clear example of syntactic change. As long as relationships within a sentence were signalled by case endings, the meaning of the sentence was unambiguous. Compare the following Latin sentences.

Poeta puellam amat.
Puellam poeta amat. 'The poet loves the girl'
Poeta amat puellam.
Puellam amat poeta.

With the loss of case endings such as the accusative singular marker *-m*, subject and object would have become indistinguishable.

**Poeta puella amat.*
**Puella poeta amat.*

Consequently, one of the word orders, that in which the subject preceded the verb and the object followed, became fixed: *Poeta ama puella.*

This word order has persisted into the Romance languages, accompanied by the use of articles developed from Latin demonstratives (a further – and a rather common – morpho-syntactic and semantic innovation), and in Spanish by a preposition, *a*, to indicate personalised objects:

French	*Le poète aime la jeune fille.*
Spanish	*El poeta ama a la muchacha.*
Italian	*Il poeta ama la ragazza.*

More extensive use of prepositions also became an important factor in signalling other case relations such as possession, location, etc.:

Latin	*Puella rosam poetae in porta videt.*
French	*La jeune fille voit la rose du poète à la porte.*
Spanish	*La muchacha ve la rosa del poeta en la puerta.*
English	*The girl sees the poet's rose on the door.*

The changing phonological conditions in the Latin of the Empire also had a profound effect

on verbal forms. For example, compare Latin and French:

	Latin	*Old French*	*French*
Sing			
1	*cantō*	*chant(e)* [ʃãnt(ə)]	*chante* [ʃãt]
2	*cantas*	*chantes* [ʃãntəs]	*chantes* [ʃãt]
3	*cantat*	*chante* [ʃãntə]	*chante* [ʃãt]

The first-person singular [o] was lost, as were final consonants, and final unaccented vowels were weakened to [ə]. In the first-person singular an analogical [ə] was added by the fourteenth century.

The merger of verb forms in the French paradigm through sound change necessitated some manner of differentiating them according to person and led to the obligatory use of subject pronouns.

je chante
tu chantes
il chante

As the verb forms were clearly distinguishable in Latin by the endings, there was no need to employ subject pronouns except in special cases, a situation still to be found in languages such as Spanish and Italian:

	Spanish	*Italian*
1	*canto*	*canto*
2	*cantas*	*canti*
3	*canta*	*canta*

Not unlike sound change proper, morphological changes may proceed on a regular and systematic basis. The Latin **synthetic future**, for example, *cantabo*, 'I will sing', disappeared in all forms and was replaced by a new **periphrastic future** consisting of a verbal infinitive with *habeo* 'have' as an auxiliary; various reductions have led essentially to a new synthetic future in Romance languages, with new grammatical marking for future tense, for example, *cantare habeo* → *chanterai* [ʃãtre].

Analogical change

The effects of sound change may be offset by analogical formations that regularise forms on the basis of others in the paradigm. As discussed earlier, accented [á] in Latin became [ɛ] in French, as we see again in the following paradigm.

	Latin	*Old French*	*French*
Singular			
1	*ámo*	*aim(e)*	*aime* [ɛm]
2	*ámas*	*aimes*	*aimes* [ɛm]
3	*ámat*	*aime*	*aime* [ɛm]
Plural			
1	*amámus*	*amons*	*aimons* [ɛmõ]
2	*amátis*	*amez*	*aimez* [ɛme]
3	*ámant*	*aiment*	*aiment* [ɛm]

These forms undergo regular sound change into Old French, in which initial accented [a] became [ɛ] but remained as [a] in the first- and second-person plural, where it was in unaccented position. This led to an irregular (i.e. non-uniform) paradigm. During the transition from Old French to Modern French, however, the paradigm was regularised through analogy with the singular and third-person plural forms, obscuring the effects of the regular sound change and resulting in a uniform paradigm. Similarly, an orthographic *e* (cf. also *chante* in the previous section) was added to the first-person singular to conform with the rest of the paradigm.

In addition to paradigm-internal analogy, analogical pressures can be exerted from outside the paradigm. An example in Old English is the word for *son*.

	Singular	*Plural*
Nom	*sunu* 'son'	*suna* 'sons'
Acc	*sunu*	*suna*
Dat	*suna*	*sunum*
Gen	*suna*	*suna*

The plural forms had no [s] but the word has become *sons* in Modern English by analogy with other words that did make the plural with *s*, such as *bāt* (nom. sing.) and *bātas* (nom. plur.) which became *boat* and *boats*, respectively.

When sound change threatens to eliminate a well-entrenched grammatical category such as, for instance, singular and plural in Indo-European languages, adjustments may occur that preserve the category (albeit in a new phonological form). The previously mentioned loss of syllable- and word-final [s] in some dialects of Andalusian

Spanish, for example, also swept away the earlier plural marker in [s]. For example, compare:

Castilian		*Andalusian (Eastern)*	
Singular	*Plural*	*Singular*	*Plural*
libro	*libros*	*libro*	*libr*ɔ
gato	*gatos*	*gato*	*gat*ɔ
madre	*madres*	*madre*	*madr*ɛ
bote	*botes*	*bote*	*bot*ɛ

In compensation for the loss of the plural indicator [s], the final vowel of the word opened (lowered a degree), and the vowel lowering now indicates plurality.

Morphological differentiation was also a factor in the modifications of the second-person singular of the verb *to be* in the Romance languages. The distinction of second and third person in vernacular Latin was threatened by the loss of word-final /-t/; compare:

Latin	*sum*		
	es	→	*es*
	est	→	*es(t)*

The various Romance languages resorted to different strategies to maintain the distinction between the second- and third-persons singular. French distinguished them on the basis of pronouns that were obligatory in the language; Spanish borrowed a form from another part of the grammar no longer needed, namely the disappearing synthetic future; and Italian resorted to analogy of the second person with that of the first person by adding /s-/. For example, compare:

French	*Spanish*	*Italian*
je suis	*soy*	*sono*
tu es [ɛ]	*eres*	*sei*
il est [ɛ]	*es*	*è*

Some syntactic changes appear to be unmotivated by modifications in the phonological or morphological component of the grammar. In Old and Middle English, an inversion rule relating to the formation of Yes/No questions could apply to all verbs – for example, *They speak the truth* and *Speak they the truth?* During the sixteenth and seventeenth centuries, the rule changed to apply to a more limited set of verbs, those that function as auxiliaries. Disregarding the fact that the verbs *be* and *have* undergo an inversion even when they do not perform as auxiliaries, and ignoring here the details of the emergence of the auxiliary verb *do*, the change can be shown as follows:

Old construction	They speak.	→	Speak they?
	They can speak.	→	Can they speak?
New construction	They speak.	→	*Speak they? (replaced by Do they speak?)
	They can speak.	→	Can they speak?

Historical linguistics has only in recent years begun to investigate syntactic change in a systematic manner in conjunction with developments in the field of synchronic syntactic studies.

Lexical and semantic change

Besides changes in the grammar of language, modifications also occur in the vocabulary, both in the stock of words (**lexical change**) and in their meanings (**semantic change**). Words may be added or lost in conjunction with cultural changes. The many hundreds of words that once dealt with astrology, when the art of divination based on the stars and their supposed influence on human affairs was more in vogue, have largely disappeared from the world's languages, while large numbers of new words related to technological developments are constantly revitalising their vocabularies.

Some of the word-formation processes and other sources of lexical changes in English are:

compounding:	sailboat, bigmouth;
derivation:	uglification, finalise;
borrowing:	yacht (Dutch), pogrom (Russian);
acronyms:	UNESCO, RADAR;
blending:	smoke + fog → smog; motor + hotel → motel;
abbreviation:	*op. cit.*, *ibid.*, Ms.;
doublets:	person, parson;
back formation:	typewrite ← typewriter, burgle ← burglar;
echoic forms and inventions:	miaow, moo, splash, ping;

clipping:	prof *for* professor, phone *for* telephone;
proper names:	sandwich ← Earl of Sandwich (1718–92); boycott ← Charles Boycott (1832–97).

Changes in the meanings of words constantly occur in all natural languages and revolve around three general principles: **semantic broadening**, that is, from the particular to the general, e.g., *holy day* → *holiday*, Old English *dogge*, a specific breed → *dog*; **semantic narrowing**, from the general to the particular, e.g., Old English *mete* 'food' → *meat*, a specific food, i.e. flesh, Old English *steorfan* 'to die' → *starve*; and **semantic shift**, e.g., *lust* used to mean 'pleasure', *immoral* 'not customary', *silly* 'happy, blessed', *lewd* 'ignorant'.

The etymological meaning of a word may help to determine its current meaning. English words such as *television* or *telephone* can be deduced from their earlier Greek and Latin meanings with respect to the components (*tele* 'at a distance', *vision* 'see', *phone* 'sound'). Such is not always the case, however. Borrowed words as well as native forms may undergo semantic change so that etymological knowledge of a word may not be sufficient to assess its meaning. Compare the following:

English	*Latin*	
dilapidated	*lapis*	'stone'
eradicate	*radix*	'root'
sinister	*sinister*	'left'
virtue	*vir*	'man'

From the origin of *dilapidated*, it might be thought that it referred only to stone structures; *eradicate*, only to roots; *sinister*, to left-handed people; and *virtue*, only to men.

Words, then, do not have immutable meanings that exist apart from context. They tend to wander away from earlier meanings and their semantic values are not necessarily clear from historical knowledge of the word.

Changes in the material culture, sometimes called **referent change**, have an effect on the meaning of a word, as is the case of the English word *pen*, which once meant 'feather' (from a root **pet* 'to fly'). This name was appropriate when quills were used for writing but remained when pens were no longer feathers. Similarly, the word *paper* is no longer associated with the papyrus plant of its origin.

Social and cognitive aspects of language change

As the earlier discussion of the diffusion of change suggests, social factors such as prestige and group identity can play an important role in language change. Social factors come into play in other ways too. For instance, language change often comes about through the socially motivated phenomena of **taboos**, **metaphor** and **folk etymologies**. The avoidance of particular words for social reasons seems to occur in all languages and **euphemisms** arise in their place. For instance, instead of *dies* one may use the expression *passes away*, which seems less severe and more sympathetic. Or one *goes to the bathroom* instead of the *toilet*, but does not expect to take a bath – even dogs and cats may go to the bathroom in North America. Elderly people are *senior citizens* and the poor are *underprivileged*. Like all social phenomena, taboos change with time and viewpoint. In Victorian England the use of the word *leg* was considered indiscreet, even when referring to a piano.

Taboos may even cause the loss of a word, as in the classical Indo-European case of the word for 'bear'. A comparison of this word in various Indo-European languages yields:

Latin	*ursus*	Old Church Slavonic	*medvedĭ*
Greek	*arktos*	English	*bear*
Sanskrit	*ṛkṣaḥ*	German	*Bär*

The presumed Indo-European ancestor of the Latin, Greek and Sanskrit forms was *Hṛk'þos. Avoidance of the term is thought to have occurred in the northern Indo-European regions, where the bear was prevalent, and another name (employed, perhaps, not to offend it or as part of a hunting taboo against speaking the name of the prey) was substituted in the form of Proto-Germanic **ber-* 'brown', that is, 'the brown one'. In Slavic the name invoked was *medved-*, from Indo-European **medhu* 'honey' and **ed* 'to eat', that is, 'honey-eater'.

Taboo may also account for seeming irregularities in phonological change. The name of the

Spanish town of Mérida, for example, did not undergo the usual syncope of the post-tonic vowel as did other Spanish words of the *veride* → *verde* 'green' type, presumably because the result would have been *Merda* 'dung', a word that would have inspired little civic pride.

Unaccustomed morphological shapes in a given language are often replaced by more familiar ones through a cognitively based process of **reinterpretation**. Loan words are readily subject to this process, as they are often unfamiliar in the adopting language. Reinterpretation of forms typically involves making a connection with phonetically and semantically similar forms already in the language, a process generally known as **folk etymology**, in that speakers impose an analysis on (i.e. give a synchronic etymology – or parsing – for) these otherwise unanalysable forms. One example involves the Middle English word *schamfast*, which meant in Old English 'modest', that is, 'firm in modesty'. To make the word readily parsable, the infrequent form *fast* (in the meaning found in *hold fast*) was changed to *face* and the word came to be *shamefaced*. Middle English *berfrey* 'tower', with nothing to do with *bell*, has become *belfry* and is associated with a *bell tower*. Words may also change their shapes due to resegmentation, such as Middle English *a napron*, which was misconstrued as *an apron* so that the noun became *apron*. Similarly, Middle English *nadder* became *adder*.

Among other characteristics of variation or style in language that may lead to semantic change (metonymy, synecdoche, hyperbole, emphasis, etc.), **metaphor**, a kind of semantic analogy, appears to be one of the most important aspects of linguistic behaviour. It involves a cognitive transfer through a similarity in sense perceptions. Expressions already existent in the language are often usurped, giving rise to new meanings for old words – for example *a galaxy of beauties*, *skyscraper*. Transfer of meanings from one sensory faculty to another occurs in such phrases as *loud colours*, *sweet music*, *cold reception*, and so on.

Linguistic borrowing

The possible effects of contact between speakers of different languages must be considered in any aspect of language change. When a community of speakers incorporates some linguistic element into its language from another language, **linguistic borrowing** occurs. Such transferences are most common in the realm of vocabulary, where words may come in and disappear with little consequence for the rest of the grammar. The borrowing language may incorporate some cultural item or idea and the name along with it from some external source; for example, Hungarian *goulash* and Mexican Spanish *enchilada* were taken into English through borrowings, and the words *llama* and *wigwam* were adapted from American Indian languages.

When words are borrowed, they are generally made to conform to the sound patterns of the borrowing language. The German word *Bach* [bax], which contained a voiceless velar fricative [x], a sound lacking in most English dialects, was incorporated into English as [bɑk]. English speakers adopted the pronunciation with [k] as the nearest equivalent to German [x]. In Turkish, a word may not begin with a sound [s] plus a plosive consonant. If such a word is borrowed, Turkish speakers added a prothetic [i] to break up the troublesome cluster. English *scotch* became Turkish [iskotʃ] and French *station* appears in Turkish as [istasjon]. Latin loan words in Basque encountered a similar kind of reconditioning: Latin *rege* became Basque *errege*, inasmuch as Basque words did not contain a word-initial [r-].

Only in relatively rare instances are sounds or sequences of sounds alien to the adopting language borrowed. The word-initial consonant cluster [kn-] does not occur in native English words, having been reduced to [n] in the past and persisting only in the orthography, but the word *knesset* 'Israeli parliament' from Hebrew has been taken over intact.

Borrowing is one of the primary forces behind changes in the lexicon of many languages. In English, its effects have been substantial, as is particularly evident in the extent to which the common language was influenced by Norman French, which brought hundreds of words into the language relating to every aspect of social and economic spheres, e.g.:

> *Government and social order:* religion, sermon, prayer, faith, divine;

Law: justice, crime, judge, verdict, sentence;
Arts: art, music, painting, poet, grammar;
Cuisine: venison, salad, boil, supper, dinner.

For the historical linguist, borrowings often supply evidence of cultural contacts where vocabulary items cannot be accounted for by other means. The ancient Greeks, for example, acquired a few non-Indo-European words, such as *basileus* 'king' and *plinthos* 'brick', presumably from a pre-Indo-European substrate language of the Hellenic Peninsula, along with certain non-Indo-European suffixes such as *-ḗnai* in *Athḗnai*.

Onomastic forms, especially those relating to **toponyms** such as names of rivers, towns and regions, are especially resistant to change and are often taken over by a new culture from an older one. Compare, for example, *Thames*, *Dover* and *Cornwall*, incorporated into Old English from Celtic, and American and Canadian geographical names such as *Utah*, *Skookumchuck* and Lake *Minnewanka*.

A sampling of the broad range of sources that have contributed to the English lexicon is: *bandana* (Hindustani), *gimmick* (German), *igloo* (Inuktitut [Eskimo]), *kamikaze* (Japanese), *ukulele* (Hawaiian), *zebra* (Bantu), *canyon* (Spanish), *henna* (Arabic), *dengue* (Swahili), *lilac* (Persian), *xylophone* (Greek), *rocket* (Italian), *nougat* (Provençal), *yen* (Chinese), and many others.

The social contexts in which linguistic borrowing occurs have often been referred to as the **substratum**, **adstratum** and **superstratum**. When a community of speakers learns a new language that has been superimposed upon them, as would have been the case when Latin spread to the provinces of Spain or Gaul, and carry traces of their native language into the new language, we have what is commonly called **substratum influence**. The French numerical system's partially reflecting multiples of twenty, for example, may have been retained from the Celtic languages spoken in Gaul prior to the Roman occupation, that is, from the Celtic substratum. **Adstratum influence** refers to linguistic borrowing across cultural and linguistic boundaries as would be found, for example, between French and Spanish, or French and Italian or German. Many words for items not found in the cultures of English colonists in America were borrowed from the local Indians under adstratum conditions, such as *chipmunk* and *opossum*. Influences emanating from the **superstratum** are those in which linguistic traits are carried over to the native or local language of a region as the speakers of a superimposed language give up their speech and adopt the vernacular already spoken in the area. Such would have been the case when the French invaders of England gradually acquired English, bringing into the English language a number of French terms.

The degree of borrowing from language to language or dialect to dialect can be related to the perceived prestige of the lending speech. Romans, great admirers of the Greeks, borrowed many words from this source, while the Germanic tribes in contact with the Romans took up many Latin words. The English also borrowed greatly from the French after the Norman Conquest, when the French aristocracy were the overlords of England.

Sometimes only the meaning of a foreign word or expression is borrowed and the word or words are translated in the borrowing. Such conditions are referred to as **loan translations**. The English expression *flea market* is a translation of the French *marché aux puces*. The word *telephone* was taken into German as a loan translation in the form of *Fernsprecher*, combining the elements *fern* 'distant' and *Sprecher* 'speaker'.

While borrowing across linguistic boundaries is primarily a matter of vocabulary, other features of language may also be taken over by a borrowing language. It has been suggested that the employment of the preposition *of* plus a noun phrase to express possession in English, e.g., *the tail of the cat* versus *the cat's tail*, resulted from French influence: *la queue du chat*. In parts of France adjoining Germany, the adjective has come to precede the noun, unlike normal French word order. This is due to German influence, e.g., *la voiture rouge* 'the red car' has become *la rouge voiture* (cf. German *das rote Auto*). Such structural borrowing is especially evident in cases of sustained intimate contact involving bi- or multilingualism, where structures from one language a speaker uses 'bleed' over into the other language. The spread of finite (person-marked) subordinate clauses in languages of the Balkans (Greek, Albanian, Bulgarian, Macedonian, etc.) is a case in point.

Language reconstruction

The systematic comparison of two or more languages may lead to an understanding of the relationship between them and indicate whether or not they descended from a common parent language. The most reliable criterion for this kind of genetic relationship is the existence of systematic phonetic congruences in specific morphemes coupled with semantic similarities. Since the relationship between form and meaning of words in any language is arbitrary, and since sound change is reflected regularly throughout the vocabulary of a given language, the existence of concordances between related languages, or lack thereof, becomes discernible through comparisons. Languages that are genetically related show a number of **cognates** – that is, related words in different languages that descend from a common source.

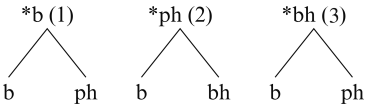
When the existence of a relationship has been determined, the investigator may then work with cognate forms to reconstruct the earlier form of the relevant languages, or the common parent, referred to as the proto-language, in order to extend the knowledge of the language in question back in time, often even before written documentation. Reconstruction makes use of two broad strategies: the phoneme that occurs in the largest number of cognate forms is the most likely candidate for reconstruction in the proto-language (this is a special case of Occam's Razor, a principle of scientific investigation that says to choose the simplest solution, all things being equal); and the changes from the proto-language into the observable data of the languages in question are plausible only to the extent that such changes can be observed in languages currently spoken or derived from well-known phonetic principles.

A phoneme that occurs in the majority of the languages under consideration but nevertheless cannot be accounted for in the daughter language by a transition from the proto-language based on sound linguistic principles should not be posited in the proto-form. For example, if a majority of languages had the sound [ʧ] and a minority contained [k] in both cases before the vowel [i], one would reconstruct the phoneme /k/ and not /ʧ/, by virtue of the fact that /k/ before /i/ has often been seen to become /ʧ/, while the reverse seems never to occur or at least is phonetically unlikely.

Thus, there are cases where it may not be reliable to use a statistical method. Given the following languages and cognate forms:

Sanskrit	*bharāmi*	bh-
Greek	*pherō*	ph-
Gothic	*baira*	b-
English	*bear*	b-
Armenian	*berem*	b-

the predominance of [b-] suggests that it is the most likely candidate for the **proto-sound**. On the other hand, assuming that the simplest description is the best one and that phonological change occurs one step at a time, we might note that, given the various possibilities,



changes (1) and (2) require at least two steps to derive one of the reflexes ([b] → [p] → [ph], [ph] → [p] → [b]), while change (3) requires only one step for each reflex, i.e. loss of aspiration and devoicing, respectively. The sound [bh-] appears to be the logical candidate for the proto-sound based on Occam's Razor. Further enquiry would also show that Gothic and English reflect a common stage with [b-]; that is, one has to take sub-grouping of related languages into consideration. The predominance of [b-] in three of the five languages is then somewhat deceptive in terms of comparative reconstruction.

Latin	*pēs*
Greek	*pous*
Sanskrit	*pad-*
Old High German	*fuoz*
Old English	*fōt*
Church Slavonic	*noga*

If we compare the words for *foot* in the Indo-European languages, we could disregard the form *noga*, given its considerable distance phonetically from the other putative cognates, as being from another source (actually, it once meant 'claw') and consider either *[p] or *[f] as the initial proto-sound. As the Germanic branch

of Indo-European has [f] where other languages have [p], we posit the proto-sound as *[p] and deduce a shift from *[p] to [f] in Germanic.

Through examination of the vocabulary of other related languages of the Indo-European family, such as Umbrian *peři* 'foot', Latvian *peda* 'sole of foot', Church Slavonic *pesi* 'on foot', we could posit the proto-vowel as *[e].

Considerations in establishing the earlier form of the final consonant might come from the Latin genitive form *pedis*, from the Greek genitive *podos*, Gothic and Old English *fôt*, among others. The proto-consonant in root-final position seems certain to have been a dental plosive ([t̪] or [d̪]). Noting that Germanic languages generally have [t] where other Indo-European languages (Latin, Greek, Sanskrit) have [d], compare Latin *decem*, Greek *deka*, Sanskrit *daśa* and English *ten*, we might conclude that the proto-language had *[d], which became [t] in Germanic. The proto-word for *foot* can now be constituted as *[ped-], a non-attested hypothetical construct posited for the proto-language.

In reconstructing the phonological forms of an earlier language, the linguist will also be concerned with the possible motivating factors underlying the change as these will often give some insight into the direction of the modification and ultimately help to establish the proto-form. Among the following Romance words one can readily see the influence exerted by environmental conditions that led to modifications in some of the languages.

Spanish	*Portuguese*	*Italian*	
agudo	*agudo*	*acuto*	'acute'
amigo	*amigo*	*amico*	'friend'

The appearance of voiced plosives [b, d, g] in earlier Spanish and Portuguese, contrasted with their voiceless counterparts in Italian, suggests that the voiced surrounding (between vowels) gave rise to the voiced consonants and that Italian has preserved here a more conservative or older stage of the language. There is no obvious motivation for the process to have occurred the other way around, with the voiced sounds becoming voiceless in voiced surroundings.

Some features of a proto-language are beyond recovery through reconstruction. The identification of proto-sounds or grammatical and syntactic characteristics of an unwritten parent language after complete loss through merger or other means in the descendent languages may simply not be possible. Without written records of the period, we could not identify or reconstitute vowel quantity in proto-Romance (Latin) speech. The phonological distinctiveness of vowel quantity in Latin is obvious from such words as *dĭcō* 'I dedicate' and *dīcō* 'I say', but the modern descendent languages display no such oppositions in vowel quantity.

Similarly, the proto-language, Latin, had a system of **synthetic passive** forms, e.g., *amor*, *amaris*, *amatur*, etc., 'be loved', which left no trace in the Romance languages, where **analytic passives** developed as in Spanish *soy amado* and French *je suis aimé* 'I am loved', in conjunction with the Latin verb *esse* 'to be' and the past participle of the main verb. Without written records, the synthetic constructions in Latin, the Romance proto-language, would remain virtually undetected.

While the **comparative method** is the most powerful tool for reconstruction, another – **internal reconstruction** – may be utilised when comparative information is not available, or when the goal is to reconstruct earlier forms of a single language. The primary assumption underlying internal reconstruction is that many events in the history of a language leave discernible traces in later stages of the language. An examination of these traces can lead to a reconstruction of linguistic processes of change and thus to a reconstructed form of the language prior to events that changed it. By way of example, we can look at a few related forms in Spanish from the point of view of internal methods.

[nótʃe]	*noche*	'night'	[nokturnál]	'nocturnal'
[ótʃo]	*ocho*	'eight'	[oktagonál]	'octagonal'
[dítʃo]	*dicho*	'said'	[diktaθjón]	'dictation'

There is an alternation among these related words between [tʃ] ~ [kt] but no apparent motivation for a change such as [tʃ] → [kt], while, on the other hand, [kt] → [tʃ] would not be unexpected. The velar [k] was pulled forward into the palatal zone by anticipation of dental [t] (assimilation) to become [j] and then the [t] was palatalised by the preceding [j], i.e. [kt] → [jt] → [tʃ]. We can now reconstruct the forms in [tʃ] as [kt]:

*nókte
*ókto
*díkto

The undeciphered ancient Iberian language of Spain's Mediterranean coasts, known only from inscriptions and not yet found to be related to any other language, contains the following lexical forms:

baite	*baikar*
baiti	*bainybar*
baitolo	*baiturane*

Since the sequences *kar* and *-nybar* appear in other words, they are assumed to be separate morphemes; compare *balkar*, *antalskar*.

This suggests an alternation between *bait* and *bai*, in which the forms (allomorphs) occur as follows:

bait + vowel
bai + consonant

or

bait → *bai*/ _consonant

We are now in a position to reconstruct **baitkar* as an earlier form of *baikar*,**baitnybar* as an earlier form of *bainybar*.

The reduction of the sequences *[tk] to [k], *[tn] to [n], [tt] to [t], is in accordance with the phonotactics of Iberian, which does not display sequences of plosive plus consonant as part of the language.

The results of this method of internal reconstruction are not verifiable, however, unless corroborating evidence can be found. In this case, we note that Basque has a form *bait* which, when combined with *-gare*, becomes *baikare*, similarly, *bait-nago* → *bainago*, *bait-du* → *baitu*, avoiding sequences alien to Basque and suggesting an affiliation between the two languages.

Linguistic palaeontology

The lack of cognate forms of a particular word in related languages may suggest that the earlier and common stage of the languages in question had no such word and linguistic differentiation occurred before such a word was needed to represent the relevant idea or cultural entity. For example, few words for metals are common to the Indo-European family of languages. This kind of information means to the practitioner of **linguistic palaeontology** that words for these items were unknown in the proto-language, which, therefore, must have broken up during the period of pre-metal usage or Neolithic times. Conversely, the various cognates for names of trees such as 'beech' suggest that the word existed in the proto-speech and that the homeland of the speakers was located in a region where these trees grew.

The lack of specific words in the parent language for grains and vegetables but many words for animals, both domestic and wild, suggest a heavy reliance on meat. Words relating to the level of the family are abundant, but those indicating a higher social order or political structure are not evident. Information of this kind may be used to reconstruct the cultural ambience and the geographical location of the proto-speakers.

Pitfalls abound, however, in the study of linguistic palaeontology; besides the fact that words may change their reference (a *robin* in England is not the same species as a *robin* in the USA), they are also readily borrowed from language to language. The word *tobacco*, common to the Romance languages, could easily lead to the false conclusion that the Romans smoked. The word itself appears to have spread from Spanish and Portuguese to the other Romance languages at a much later time.

Genetic classification of language

A major result of historical and comparative linguistic investigation has been the mapping of the world's languages into groupings of related languages, called 'families', and sub-groupings within these families. When a given language has been shown to belong within the folds of a particular grouping as defined by linguistic relationships indicating a common descent from an earlier source language (a proto-language), it is said to have been classified genetically. (This use of 'genetic' has nothing to do with DNA or biological genetics but rather reflects the meaning of the Ancient Greek source for the word, i.e. 'having to do with origins'.) A useful method for

expressing genetic relationships is the family-tree diagram consisting of the parent (proto-)language as the starting point and branches indicating the descended 'offspring' languages (to extend the metaphor of a biological family tree).

Genetic classification has shown that the vast majority of the languages currently spoken in Europe belong to one of four families: Indo-European, Uralic, Caucasian and Basque. In addition, some 300 or more other language families have been recognised around the world. It may well be that some reduction of this number is possible, in that some families may form higher-order 'phyla' with other families, but such moves are often controversial and not warranted by the methods mentioned here (e.g., rigorous application of the comparative method, which depends on an assumption of relatedness if it is to work).

Indo-European

The Indo-European family extended from Europe to India and in recent times has spread over much of the globe, including North America, South Africa, Australia and New Zealand as well as a number of pockets around the world. It is the most thoroughly investigated and best-known family of languages today and is derived from a hypothetical parent called **Proto-Indo-European**, thought to have been spoken in the fifth millennium BC (see Figure 1). Judging from the distribution of the various Indo-European languages, their migratory chronologies, and from archaeological evidence (Kurgan culture), the parent language is thought to have originated in the region of the Black Sea, though much is controversial about this issue.

The major groupings of the Indo-European family of languages are shown below. The **Germanic** branch of Indo-European has been divided into three subgroups: East Germanic languages are now extinct but the best known is Gothic, for which written texts exist from the fourth century AD. The North Germanic or Scandinavian branch includes Icelandic, Norwegian, Swedish, Danish and Faroese. West Germanic contains German, Yiddish, Dutch, Flemish, Frisian, Afrikaans and English. Afrikaans is a descendant of Dutch spoken by the early white settlers of South Africa, the Boers. Frisian is spoken along the northern coast of the Netherlands, the north-western coast of Germany and on the Frisian Islands. English is derived from the languages of the Angles, Saxons and Jutes, Germanic tribes of northern Germany and southern Denmark who began settling in England in the fifth century AD.

The once-widespread **Celtic** languages, extending from the British Isles to the Anatolian peninsula, are now generally extinct except for those surviving in the British Isles and Brittany. The Continental Celtic languages are best known from Gaulish, spoken in France, and Hispano-Celtic (also known as Celtiberian), of Spain and Portugal, which have bequeathed some documentation. The insular branch has been segmented into two groups – Brythonic and Goidelic – of which the former includes Welsh and Breton, and the latter Irish Gaelic and Scots Gaelic. Breton is an offshoot of now-extinct Cornish, spoken in Cornwall up to the eighteenth century.

Prior to about the third century BC, linguistic relationships on the **Italic** peninsula are obscure, but clearly attested after this time as belonging to the Indo-European family are the two groups Sabellic (best represented by Oscan and Umbrian) and Latino-Faliscan. Latin, in time, displaced the other languages on the peninsula and gave rise to the Romance group of languages.

Indo-European speakers of what was to become the **Hellenic** or **Greek** branch entered the Balkan peninsula of south-eastern Europe apparently sometime early in the second millennium BC, and at a later time we can speak of two main groups: East Greek, called Attic-Ionic, the

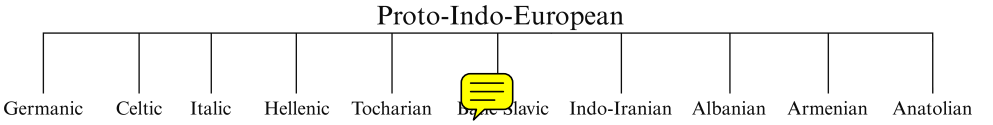


Figure 1

languages of Attica and much of Asia Minor, and West Greek. All modern Greek dialects except Tsakonian are descendants of the Hellenistic koiné, based largely on Attic, the speech of classical Athens.

Tocharian is a group of two Indo-European languages, forming their own subgroup, recovered from manuscripts of the seventh and eighth centuries AD. It was once spoken in what is now Chinese Turkestan.

The **Balto-Slavic** branch is composed of two main subgroups, Baltic and Slavic. Lithuanian, Latvian (or Lettish) and the now-extinct Old Prussian make up the **Baltic** languages, situated along the eastern coast of the Baltic Sea. Lithuanian contains an elaborate case system much like that established for the parent Indo-European language.

The **Slavic** branch is composed of three subbranches: East, South and West Slavic. East Slavic consists of Russian, Ukrainian and Byelorussian, the latter spoken in Belarus (capital Minsk) to the west of Russia, while South Slavic is composed of Bulgarian, Macedonian, Slovenian, and Bosnian, Croatian and Serbian (formerly called 'Serbo-Croatian' but now reflecting the various nation-states that emerged out of the former Yugoslavia). The West Slavic branch includes Czech, Slovak, Polish and Sorbian (Lusatian).

The **Indo-Iranian** branch was carried to India and Iran and consisted of two main branches: Indic and Iranian. The former appeared as Sanskrit, which subsequently evolved into the various Indo-European languages of India and Pakistan, such as Hindi, Urdu, Bengali and Gujarati, while the latter evolved early into the Avestan and Old Persian dialects. Various Iranian languages are in use today and include Pashto, Persian, Kurdish and Ossetic, among others.

Forming its own branch as well is **Albanian**, spoken since ancient times in the southern Balkans and now found in Albania and parts of Greece, Macedonia and southern Italy. Its putative relationship to the poorly known ancient Illyrian or Thracian languages is disputed and rests on slender evidence at best.

Located primarily in the Caucasus and northeastern Turkey, the **Armenian** language, attested from the fifth century AD, also continues its own line of descent as a separate branch of Indo-European.

Indo-European migrations into the **Anatolian** peninsula gave rise to Hittite and the related Luwian, Palaic, Lydian and Lycian languages. All are now extinct.

There are many other extinct languages such as Illyrian, Thracian, Ligurian, Sicil and Venetic, whose scanty documentation points to membership in the Indo-European family, but their affiliations are unclear.

Uralic

Consisting of about twenty languages, the Uralic family is spread out across the northern latitudes from Norway to Siberia. There are two major branches: Samoyedic and Finno-Ugric. The former is spoken in Russia and Siberia; the latter includes Hungarian, Finnish, Estonian and Lappish. They are primarily agglutinating languages [*see* LINGUISTIC TYPOLOGY] with an extensive system of cases. The proto-language may have been spoken in the northern Ural mountains about 6000 BC. The earliest texts are from the twelfth century AD, a Hungarian funeral oration.

Caucasian

The languages of the Caucasus area are often referred to as the 'Caucasian languages' but in fact this is a geographic designation; there are some thirty-five languages in the area, in three recognised language families: North-east Caucasian (including Abxaz and Kabardian), Northwest Caucasian (including Chechen-Ingush) and South Caucasian (better known as Kartvelian, including Georgian). The languages are characterised by glottalised consonants, complex consonant clusters and few vowels. The earliest texts are in Georgian, a Kartvelian language, and date back to the fifth century AD.

Languages of Asia

Language families indigenous to Asia are Altaic, Sino-Tibetan, Austro-Asiatic and Dravidian.

Though controversial, a wide-ranging language family has been posited for many of the languages of Turkey, Russia, China and Mongolia, and possibly also Korea and Japan. This

'**Altaic**' family comprises some thirty-five to forty-five languages, in three main branches: Turkic, Tungusic and Mongolian, though some specialists include Japanese and Korean in the family as well. The family is characterised by agglutinating structures and some languages by vowel harmony. The earliest Turkish texts, the Orkhon inscriptions, date from the eighth century AD.

Second only to Indo-European in number of speakers, the **Sino-Tibetan** family contains about 300 languages in two major branches: Tibeto-Burman and Sinitic (Chinese). The Sinitic branch encompasses northern and southern groups of languages. The principal language of the north is Mandarin, and those of the south are Cantonese and Wu. Tibeto-Burman languages are found in Tibet, India, Bangladesh and Burma. The region contains great linguistic diversity and, as yet, the overall linguistic picture is unclear. The languages are generally tonal [*see* TONE LANGUAGES].

The **Austro-Asiatic** family consists of about 150 languages, in two major groupings: Munda, which includes languages of central and north-east India; and the larger Mon-Khmer group with Cambodian (Khmer), Vietnamese and many others of Cambodia and Vietnam, Burma and southern China. These languages are characterised by complex vowel systems, and some (e.g., Vietnamese) by tones. The Mon-Khmer branch may have been a unified language in the second millennium AD. The earliest texts date to the sixth century AD.

Found mainly in southern India, there are about twenty-three **Dravidian** languages. The most important, in terms of number of speakers, are Telegu, Tamil, Kannada and Malayalam. Dravidian peoples appear to have been more widespread once, but were displaced southward during the Indo-European incursions into northern India. The languages are commonly agglutinating and non-tonal, with retroflex consonants and word-initial stress.

Languages of Africa

The number of distinct languages spoken throughout Africa is estimated at about 1,000, all of which belong to one of the four language families: Afro-Asiatic, Niger-Kordofanian, Nilo-Saharan and Khoisan.

Afro-Asiatic, often referred to by its older name of Hamitic-Semitic, is a group of languages spoken mainly across the northern half of the continent and throughout the Middle East, and consists of about 250 languages divided into six primary branches: Egyptian, now extinct except for the limited use of its descendant, Coptic, in religious rituals; Cushitic languages of Ethiopia, the Sudan, Somalia and Kenya; Berber, once widespread across the northern regions of the continent but now primarily restricted to pockets of speakers in Morocco and Algeria; Chadic, spoken in the region of Lake Chad and distinguished from the other groups through the use of tones; Omotic, considered by some to be a branch of Cushitic; and Semitic, the branch responsible in large part for the displacement of the Egyptian and Berber branches, spoken throughout the Middle East, across North Africa and in Malta. The three best-known members of this branch are Arabic, Hebrew and Amharic. Pharyngeal sounds and consonantal roots characterise many of the languages.

The **Niger-Kordofanian** language family covers much of the southern half of the African continent and embodies many more languages than Afro-Asiatic. Of the two main branches, Kordofanian and Niger-Congo, the latter consists of especially numerous sub-branches. The languages are typically tonal (except Swahili) and usually agglutinating in structure. Perhaps the best-known subgroup of Benue-Congo, itself a branch of Niger-Congo, is Bantu, which consists of over 100 languages, including Swahili, Zulu and Kikuyu. Found primarily in East and Central Africa, the **Nilo-Saharan** family contains several subgroups and about 120 languages. They are generally tonal and nouns are often inflected for case. This family is still relatively unexplored. Some of the languages are Masai (Kenya), Nubian (Sudan) and Kanuri (Nigeria).

Squeezed by Bantu expansion from the north and European expansion from the south, **Khoisan** speakers of approximately fifteen languages are now pretty well restricted to areas around the Kalahari Desert. This family, unlike any other, is characterised by clicks of various kinds which function as part of the consonantal system. A few neighbouring languages of the Bantu sub-branch, such as Zulu and Xhosa, have borrowed these clicks from the Khoisan

languages. They are also characterised by tones and nasal vowels.

Languages of the Pacific

Some 2,000 languages are (or were) spoken in the Pacific region (including the Indian Ocean and Australia), representing several language families and geographical groupings, about a quarter of the world's languages.

Austronesian, with c. 1,200 languages (perhaps the world's largest family, vying with Niger-Congo for that honour), extends from Madagascar to Easter Island and from Taiwan to New Zealand. Proto-Austronesian was spoken in Taiwan, where some ten indigenous Formosan Austronesian languages are/were found. The large Malayo-Polynesian branch (which used to be the name of the whole family) contains the languages outside of Taiwan, among which are the Philippine languages and the large Oceanic branch, whose members include among others Polynesian and Fijian languages.

The c. 750 **Papuan** languages include most of the non-Austronesian, non-Australian languages of the Pacific region, most in New Guinea (Papua New Guinea and Indonesia's Irian Jaya province), but some also in Alor, Bougainville, Halmahera, New Britain, New Ireland and Timor. Papuan languages do not represent a genetic grouping (language family), but opinion varies on their classification. For conservative classifiers, they fall into some eighty families; a commonly cited less conservative figure is sixty families; and even the most optimistic do not see being able to reduce the figure to less than twenty-five distinct families.

There are or were c. 200 distinct **Australian** languages – some cite 200–300 (all remaining ones highly endangered except c. twenty). They represent some twenty-five distinct language families. The large **Pama-Nyungan** family (c. 175 languages, in twenty-seven branches) covers 90 per cent of the country, with the several other families limited to far northern Australia.

Several questions of classification remain to be resolved, and there exist several controversial hypotheses of more distant, broader-scale groupings. For example, many believe all Australian languages are related, which is plausible, but it has not been possible to demonstrate this with standard linguistic methods. **Tasmanian** languages are also often thought to be distantly related to Australian languages, but this cannot be demonstrated, perhaps due to the long separation and poor quality of most of the surviving information on Tasmanian languages. The controversial Indo-Pacific hypothesis from Joseph Greenberg, however, has largely been abandoned. He argued that most of the non-Austronesian languages of the Pacific from the Andaman Islands to Tasmania, but excluding Australia, were genetically related. Most of these are Papuan. Specialists in these languages have rejected this hypothesis. Weak hypotheses of various sorts have attempted to link Austronesian with the likes of Ainu, Eskimo-Aleut, Indo-European, Sino-Tibetan, Japanese, Austro-Asiatic (including Munda and Mon-Khmer) and Austro-Tai (Austronesian with Tai-Kadai). None of these is accepted today.

American Indian languages

While many relationships remain unclear with regard to Amerindian languages in the northern hemisphere, the following families have been identified, to which most of the languages belong: **Eskimo-Aleut**, **Algonquian** (north-east USA and Canada), **Athapaskan** (Alaska, western Canada and south-western USA), **Salish** (Pacific north-west), **Wakashan** (Vancouver Island), **Siouan** (Great Plains), **Uto-Aztecan** (Mexico), **Muskogean** (south-eastern USA), **Iroquoian** (eastern USA), **Yuman** (Baja California), **Mayan** (Mexico and Guatemala). It is estimated that nearly 400 distinct languages were spoken in North America in pre-Columbian times, 300 of these north of Mexico. Today, about 200 survive north of Mexico, but many of these are near extinction.

Along with the languages of the Pacific, South American linguistic relationships are the least documented in the world, and estimates run from 1,000 to 2,000 languages, although only about 600 are actually recorded and 120 of these are extinct. Three major South American families which account for most of the known languages have been posited: **Andean-Equatorial**, whose principal language is Quechua; **Ge-Pano-Carib**, extending from the Lesser Antilles to

southern Argentina; and **Macro-Chibchan**, covering some of Central America, much of northern South America and parts of Brazil.

Some language isolates

In some cases, a single language has no known relationships with other languages and cannot be assigned to a family. When this occurs, the language in question is called an **isolate**. Some languages that have not been related to any other are Basque (spoken in north-eastern Spain and south-western France), Ainu (of northern Japan), Kootenay (British Columbia), Gilyak (Siberia), Tarascan (California) and Burushaski (spoken in Pakistan). There are also the extinct Sumerian, Iberian, Tartessian and many other languages known only from inscriptional material.

J. M. A., H. C. D. and B. D. J.

Note

[1] This entry is based in part on the entry by James M. Anderson in Edition 1 and 2 of this Encyclopedia.

Suggestions for further reading

Anttila, Raimo (1989) *Historical and Comparative Linguistics*, 2nd edn, Amsterdam: John Benjamins.

Campbell, Lyle (1997) *American Indian Languages: The Historical Linguistics of Native America*, Oxford: Oxford University Press.

— (2004) *Historical Linguistics: An Introduction*, 2nd edn, Cambridge, Mass.: MIT Press.

Campbell, Lyle and Mixco, Mauricio (2007) *Glossary of Historical Linguistics*, Edinburgh: Edinburgh University Press.

Campbell, Lyle and Poser, William (2008) *Language Classification: History and Method*, Cambridge: Cambridge University Press.

Clackson, James (2007) *Indo-European Linguistics: An Introduction*, Cambridge: Cambridge University Press.

Fortson, Benjamin (2004) *Indo-European Language and Culture: An Introduction*, Oxford: Blackwell.

Hock, Hans Henrich (1991) *Principles of Historical Linguistics*, 2nd edn, Berlin: Mouton de Gruyter.

Hock, Hans Henrich and Joseph, Brian D. (2009) *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*, 2nd edn, Berlin: Mouton de Gruyter.

Joseph, Brian D. and Janda, Richard D. (eds) (2003) *Handbook of Historical Linguistics*, Oxford: Blackwell.

Labov, William (1994) *Principles of Linguistic Change, Vol. I: Internal Factors*, Oxford: Blackwell.

— (2001) *Principles of Linguistic Change, Vol. II: Social Factors*, Oxford: Blackwell.

McMahon, April M.S. (1994) *Understanding Language Change*, Cambridge: Cambridge University Press.

Mallory, James P. and Adams, Douglas Q. (1997) *Encyclopedia of Indo-European Culture*, London: Routledge.

Mallory, James P. and Adams, Douglas Q. (2006) *The Oxford Introduction to Proto-Indo-European and The Proto-Indo-European World*, Oxford: Oxford University Press.

Ramat, Anna-Giacolone and Ramat, Paolo (1997) *The Indo-European Languages*, London: Routledge.

Sihler, Andrew (2000) *Language History: An Introduction*, Amsterdam: John Benjamins.

Steever, Sanford B. (ed.) (1998) *The Dravidian Languages*, London: Routledge.

Thomason, Sarah G. (2001) *Language Contact: An Introduction*, Washington, D.C.: Georgetown University Press.

Trask, Robert Lawrence (1996) *Historical Linguistics*, London: Arnold.

History of grammar

The grammars that concern linguists today have developed on the basis of a long tradition of describing the structure of language which began, in the West at least, with the grammars written by classical Greek scholars, the Roman grammars largely derived from the Greek, the speculative work of the medievals, and the prescriptive approach of eighteenth-century grammarians (Dinneen 1967: 166; Allen and Widdowson 1975: 47). These early grammars also form the basis for many grammars in use in schools in both native- and foreign-language teaching. In particular, the adaptation of Greek grammar to Latin by Priscian (sixth century) has been influential.

Priscianus major and Priscianus minor

Priscian's work is divided into eighteen books. The first sixteen, which the medievals called