I Language in perspective

‘Because people can speak, they think they can speak about language’

Goethe

Contents

1 By way of introduction
   1.1 Common ideas about language
   1.2 Some beginners’ questions
   1.3 Terms and notions
   1.4 The organisation of language

2 Characterising language
   2.1 Defining language
       2.2.1 Hockett’s design features
   2.3 Linguistic relativity

3 Analysing language
   3.1 Structural notions in linguistics
   3.2 Further structural notions
   3.3 Collecting linguistic data

4 On the origins of language

1 By way of introduction

Linguistics is the scientific study of language. For some two centuries it has been an established branch of science and very considerable strides have been made in researching how humans speak, how they acquire their native language and in documenting the known languages of the world. This knowledge has been arrived at by devising means for analysing human language and by adopting approaches based on notions which are likely to be quite different from the views which non-linguists will have about the subject. Thus one of the difficulties for linguists in their attempts to explain the nature of language to beginners is that the latter will always have ideas about language already. These derive from reflections on the more conscious parts of language such as vocabulary, on the widely held belief that language change is language decay, on ideas about what constitutes good and bad language, in particular in connection with style and social class, and on an inordinate reliance on the written word.
and finally on a general confusion of language itself with the people who use language. In all these areas linguists will find the ideas of nonlinguists need to be adjusted and set in a new direction. For example, linguistics is often more concerned with the less conscious parts of language such as the levels of sound and sentence structure, it views language change as an inherent and necessary aspect of the system seeing as how it consists of choices made by humans. The goal of linguistics is to describe language and certainly not to dictate usage to others and it is primarily oriented towards the spoken word, the spelling system of a language being a secondary phenomenon arising much later in history due to needs in society.

The purpose of language This is primarily functional, language is used to convey information or to express emotions, for example. But there is a strong social component to language as well. It is used to maintain social relationships and to identify with a certain section of society. This means that all human languages have two sides: an internal structure concerned with the organisation of linguistic information necessary for communication and an external aspect where the manner in which language is expressed carries social significance. When one considers the first aspect, the internal organisation of language, one can see that in the course of human evolution our ability to speak would appear to have become autonomous and self-contained. Not only that but the levels within language, those of sounds, words and sentences, would also seem to have become independent but with connections linking them. This modularisation is a distinct advantage to the organisation and maintenance of language and is the reason for treating the levels separately in books like the present one.

The goal of linguistics There is a common view that the purpose of linguistics is to provide tools such as those used in the teaching of languages or to offer means for providing remedies to language impairments. Useful as such applications certainly are, they only represent some of the concerns of linguistics. Furthermore, as applications they result from a previous concern with the nature and structure of human language. Without a general study of human language, applied linguistics cannot be developed. Examining the structure and principles of human language provides us with insights into its organisation which must necessarily precede any practical uses to which such insights might be put. Along with a willingness to concern oneself with notions of system and structure the beginner in linguistics should show a sense of history. Language is a phenomenon which exists in time and change is an inherent feature of all languages. This change is largely regular and shows certain limits in the directions it can take. The discovery of recurrent patterns and motivations in language change is one of the major advances in linguistics during the past two centuries.
1.1 Common ideas about language

In this section a number of ideas are presented which I have found common among nonlinguists during many years of teaching and working as a linguist. They represent some of the notions which the linguist is confronted with in introductory courses to linguistics and which need to be relativised in order to be acceptable in such a context.

**Primitive languages** A frequent belief among nonlinguists is that some languages are more primitive than others. Typical examples of such ‘primitive’ languages would be ones spoken in nonindustrialised regions of the developing world. This notion is definitely wrong. No language is primitive in the sense of being underdeveloped and demonstrably simpler in structure than others. Every language has a grammatical system which is adequate for those who speak it and a vocabulary which is appropriate for their needs. Of course a nontechnical society will not have words for the many phenomena of the modern industrialised world but that does not make such a language primitive.

Equally dialects are not more primitive versions of standard languages. In any country the standard stems form a dialect which has been promoted in society due to external historical events such as being the form spoken in the capital city (England, France, Spain), that used in great literature (Italy) or that employed in early key translations of the Bible (Germany, Finland). Of course with time standards come to have larger vocabularies and a greater range of styles because they are used by more people and in more contexts. However, anything which can be expressed in a standard can also be said in a dialect, if not as a one-to-one equivalent then always by means of paraphrase. Furthermore, a dialect may be richer in grammatical distinctions than the standard it is related to. With some languages, such as with present-day Irish, there is no accepted official form of the language but a group of dialects which differ considerably.

**Onomatopoeia** Although the notion may seem attractive, the number of words in a language which more or less directly represent the object or beings they refer to – this is known as onomatopoeia – is very small indeed. These are in the main restricted to the sounds which animals make or which are used with them such as moo (for cows), gee-up (for horses), miaow (for cats). Indeed the whole notion of onomatopoeia runs counter to a principle of language structure which demands that there be no necessary connection between the sounds of a word and what it represents. Having said this, one should note that languages sometimes have an indirect representation of phenomena of the outside world, for instance in English an initial *fl*- often indicates the movement of liquids or of objects through air as in flush, flow, flux; flip, fling etc. Other examples of similar meanings and phonetic form are slip, slime, slush, slob; bump, thump,
clump, stump. This phenomenon is called sound symbolism.

Iconicity Although onomatopoeia is not a dominant principle in today’s languages the question as to what extent linguistic structure reflects the organisation of the world is a valid one. This more indirect relationship between language and what is external to it is termed iconicity from icon meaning symbol, that is something which stands for something else.

A possible instance of this would be the Irish word bog /bɔːg/ ‘soft’ which only contains voiced or ‘soft’ consonants; the Irish word for ‘hard’ begins with a voiceless or ‘hard’ consonant: crua /kruː/. In a similar vein one could quote words for ‘large’ such as English huge, enormous, German groß, Irish móir which contain long back vowels (English big is a notable exception). Contrariwise words for ‘small’ often have a high front vowel: English little, Scottish English wee, German winzig ‘very small’, Irish bideach ‘very small’. In English there are two words which seem to confirm this hypothesis. The verb peep means to look through a small hole and was pronounced with the same vowel in the late Middle Ages, that is it did not undergo the major vowel shift to yield the diphthong as in pipe in Modern English which it would have if it had developed regularly. The adjective tiny has a common alternative pronounced teeny /tiːni/ which retains the former high front vowel indicative of smallness.

In sentence structure there are examples of iconicity. A correlation between the linear order of sentence elements and the temporal sequence of events is often found, e.g. When Fiona got up, she made herself a cup of coffee, i.e. first she got up then she made the coffee; Paddy hit Brian, i.e. Paddy initiated an action which resulted in Brian being hit. Statistically, there are more languages with the word order Subject + Object than with Object + Subject and the former usually corresponds roughly to the temporal sequence of events in a sentence. Such correspondence between the occurrence of elements in a sentence and the temporal sequence of events in the narrative is not by any means compulsory. Frequently, speakers wish to emphasise a certain element of a narrative and put it at the front to highlight it, e.g. Fiona made herself a cup of coffee when she got up.

Language reflects reality By and large it is true to say that languages have words for the objects of the world, the thoughts and feelings which its speakers experience. And to a certain extent it is the case that separate words for objects tend to reflect their relative importance for speakers. For instance, English has a special word for thumb, the finger on the inside of the hand which is at a slightly different angle from the others. But the equivalent on our feet, the big toe, does not have a special word for it. One could say that one uses one’s thumb more and one sees it more often and so there is a separate word for it. But not all languages work like that. Indeed some do not even have a separate word for ‘toes’ even, e.g. Irish uses a form méara coise ‘fingers of the foot’ (as does Turkish). In this case the pitfall is to imagine that the Irish (or the Turks) pay less
attention to their toes because they do not have a separate word for them.

Here is a parallel example from the realm of thoughts and feelings. English has a word for ‘remember’ but there is no special word to express what happens when one suddenly has a feeling again which one had in one’s past. This is a type of emotional remembering which is just as significant for humans as the non-emotional type. What speakers of course do is to use a phrase when a single word is not at their disposal, e.g. *I remember the feeling I had on the morning of my wedding day*. The danger here is imagining that the existence – or lack – of a special word for a particular matter somehow reflects its importance for the speakers of the language in question.

*Correct language* There is no such thing as correct language in any absolute sense as language is not in itself either right or wrong. However, in a given society there may be some external norms imposed on language which are used to decide what usage is socially acceptable and what is not. External attitudes to language and the nature of language itself should not be confused. Language is neither good or bad; such value-laden judgements are made by people on the basis of opinions which derive from social attitudes and prejudices.

The type of attitude which praises one type of usage and condemns another is generally termed ‘prescriptive’ in linguistics, that is it attempt to say what language should be like. The scientific study of language is, however, purely descriptive, that is it strives to say what language is actually like. For instance the idea that one should say *whom* for the object form of the personal relative pronoun *who* – as in *That’s the person whom we saw yesterday* – because this is more ‘accurate’, ‘elegant’ or whatever is entirely spurious. The *whom* form has been kept artificially alive over the past few centuries by the schoolmasters who have insisted on it. Spoken language shows that it is not necessary.

The conception of correct language is frequently linked with criticism of colloquial or regional varieties. Consider for a moment the two sentences: *Those days are gone. Them days are gone.* The sentences are identical in meaning so any disagreement about their acceptability is about their form. The second of the two is very common in many dialects of English and shows the use of a third person plural personal pronoun as a demonstrative. This additional use of the personal pronoun is not in any absolute sense ‘incorrect’, nor is it ambiguous or less accurate as some of the objections to it might suggest. There is also no ambiguity in a structural sense, i.e. which might make the sentence difficult to interpret, because in the position immediately before a noun the personal pronoun can only be interpreted as a demonstrative. Nor can one protest that *them* should not be used as a personal pronoun and a demonstrative. There are already many forms in English which realise different grammatical categories, e.g. *which* can be an interrogative or a relative pronoun, compare *Which linguist did you meet?* with *This is the book which we have to read for next week*. The conclusion is that any dissatisfaction with a phrase like *them*
days arises from prescriptive notions external to the structure of language.

Because the terms ‘correct’ and ‘incorrect’ imply value judgements, linguists avoid them and speak of a sentence or word as being acceptable or unacceptable to native speakers.

**Written language is superior to spoken language** This is another common misconception about language. For social reasons the written word is highly valued, for instance because of its status in contracts, legal documents, and official material in general. Furthermore, the written word is much more permanent than the spoken word. These aspects of the written word led to it being more highly valued in Western-style societies. However, from an internal, structural point of view, it is the spoken word which is more complex, intricate, sophisticated. Written language is codified and normally quite inflexible. Its range of uses is restricted to formal styles and whole areas of language use, such as intonation, are excluded from written language. Writing is always more formal and slightly more conservative than the spoken word because innovations come from colloquial language and take time to be accepted in the written form. There is always a time lag between change in spoken language and its acceptance in writing.

**Languages should be logical** One should not expect languages to be logical in any strict sense. Given that the function of language is primarily as an instrument of communication, then when this function is fulfilled that is all that matters. For example, a common expectation among those seeking logic in language might be that if there is a negative there should also be a positive. However, many forms in English show that this is not the case, compare *uncouth* which does not have an equivalent *couth* in the standard language. Negatives are not always the simple reverse of positive terms. There are two matters here: form and meaning. In English one has the adjective *real* but its formal negative *unreal* is not simply the opposite of the positive but has the meaning ‘hard to believe, most unexpected’. A similar pair is *possible* and *impossible* where the latter frequently means ‘highly unacceptable’ as in *impossible behaviour*.

Although languages are not logical in this simple manner the attitude that they should be can lead to changes. This has been the case with double negatives in English where formerly the assumption was made that two negatives should cancel each other out and form a positive (going on a Latin model). In the 18th century prescriptive grammarians, such as Bishop Robert Lowth (1710-1787), were responsible for disseminating fairly ridiculous ideas about language. One of these was that ‘two negatives make a positive’. This led to the prohibition on sentences like *I don’t like none of them* which then came to be replaced by *I don’t like any of them*. But this application of apparent logic misses the point about the actual function of double negatives: they intensify the statement they make rather than cancelling each other out. Despite the view in the standard, many regional forms of English have maintained this type of negation, notably
African American English but also colloquial forms of British and Irish English.

Random preferences in language It never ceases to amaze linguists just how many random and ill-founded ideas speakers have about language. There is a general pattern which one can recognise here. Wherever there is variation in language speakers decide – often unconsciously – when they become aware of the variation, just what form they prefer. Once an opinion has been reached, speakers offer all sorts of spurious reasons in support of it. To illustrate what is meant here consider the following true story.

The author has known a few people who happened to have had the surname Costello. In Ireland this word is stressed on the first syllable. Then he once met someone with this name (a fellow student) who insisted that the stress be put on the second syllable, i.e. Co'stello. This individual had very definite views about the use of initial stress on his name which he saw as a sign of Irish boorishness, the stress on the second syllable being a sign of European sophistication. The linguist would obviously say that the stress on the word is devoid of any value. The judgement any speaker may make about the position of stress results solely from nonlinguistic associations.

This case is not as trivial as it seems. Many of the likes and dislikes speakers have arisen for definitely nonlinguistic reasons. The most common dislikes stem from the association of forms with a group or possibly an individual with whom one disagrees.

A result of such rigid opinions is that people tend to wage war on their petty dislikes in language. Typical examples would be the prohibition on the use of final prepositions in English, as in This is the place they sent him to, or the putative correctness of I for me when the pronoun does not occur immediately before the verb, as in It's I who read the book or the use of hopefully in the sense of ‘we hope that...’ rather than ‘with much hope’: Hopefully, she’ll pass her exam.

Another notion common among nonlinguists is that a language should be protected from disturbing or corrupting influences from outside. This is a common attitude in French where official quarters feel that the ongoing influx of English loanwords is spoiling the language. However, by and large the general population do not take too much notice of such purist attitudes and words are borrowed if speakers feel the desire to use them. There does not have to be a need for them in the sense of ‘gaps’ in the vocabulary, for instance in French the phrase fin de semaine ‘end of week’ occurs but the English loanword weekend is generally preferred. However, a speech community is not always consistent in its preference for loanwords over native words, one well-known case being French ordinateur which is used rather than the English word ‘computer’.

1.2 Some beginners’ questions
Is linguistics a science? To answer this question one must first offer an explanation. There is no way of directly observing how an individual produces language, that is the object of study can only be considered by the output and not examined during production. Evidence in linguistics is gained indirectly. For instance, much knowledge concerning the structure of syllables has been gained from considering slips of the tongue. Evidence for grammatical structures comes frequently from the breakdown of language where speakers lose the ability to use some part of their language, such as the endings added to words (so-called agrammatism).

The indirectness of proof in linguistics raises the question about the status of empiricism (experimental evidence) in language studies. As language is ultimately symbolic human behaviour, it is not possible to give definite accounts of all its manifestations and neither can one predict its future development because there are too many uncertain or unknown factors or at least one does not know what relative weight competing elements in language will have in the course of further development. Having said this, one can now give a qualified answer to the above question: linguistics is a science as linguists use rigorously determined and objective criteria in the theories they propose. However, it is different from many natural sciences as it is not possible to directly observe many of the phenomena which are the object of investigation.

What is a rule? A rule is an explicit statement of a process in a language. Many different kinds of rule exist, e.g. rules governing the production of sentences, the formation and pronunciation of words. Given a natural human desire for order in any field of study, speakers expect rules in language to be water-tight. Unfortunately language does not work like that. There are rules of course, but most of them allow for various exceptions. The fact is that speakers can deal with a great deal of exceptions in language. And because languages develop over long periods of time and they are never overhauled and pruned back, so to speak, they tend to collect exceptions to rules as they go along. Not only is this true but many of the rules which do exist are invisible to native speakers: they are not conscious of them although they may well use them in their speech. Good examples of what is meant here are the rules governing the formation of acceptable sentences. Few if any speakers are in a position to verbalise the rules of sentence structure and yet we all use them in our native language. For example, verbs which take two objects frequently allow the direct object to be preceded by the indirect one rather than it coming afterwards with the preposition to: Fiona told Fergal the good news (for Fiona told the good news to Fergal). One might expect that a verb like explain, which also involves telling, would allow the word order ‘indirect + direct object’, but this is not the case: *Fiona explained Fergal the computer programme (a preceding asterisk indicates that a sentence or word is unacceptable in a language), the sentence must be Fiona explained the computer programme to Fergal.

A central question for work on rules in language is whether the
constructs postulated by linguists possess a psychological reality. If the rules concerning linguistic structure can be shown to have reality for speakers then they constitute a genuine insight into the nature of the human language faculty. There would seem to be no possibility of finding physical correlates to aspects of linguistic behaviour. The language faculty is intimately connected to consciousness and for this there is no direct physical correlate in the brain (or at least none has been found to date). The only connections that can be established stem from the area of language pathology. Here one can see that certain types of accidents, brain lesions or tumours can be associated with certain types of linguistic impairment. But all of this is very crude and the information gathered in this field is of relevance to the problem of the rehabilitation of accident victims but not as a major source of insight for the general linguist.

Given the above situation what one is left with is a language production mechanism in human beings which cannot be directly observed but only guessed at. Any evidence on the nature of this production ability is indirect. This fact has led to much discussion on the question as to whether linguistics is an empirical science or not. In the sense of direct observation of the phenomena of the field, linguistics is not empirical. However, the entire question of empiricism would seem to be a red herring and to derive from a false view of what constitutes evidence in a science. If no direct evidence is forthcoming in a science then indirect evidence must suffice. Here is an example from phonology and a slip of the tongue to show what this procedure looks like.

The theory of syllable structure assumes that a syllable is divided into a series of sections, i.e. syllables are entities in phonology which contain a structure of their own. The syllable is a unit which everyone recognises intuitively (that is knows how to count syllables but not how to define them). The syllable consists of an onset and a rhyme. The onset is basically everything before the vowel and the rhyme is the vowel and anything which follows, e.g. with spin, sp- is the onset and -in is the rhyme. The assumption here is that speakers organise the words of their language as syllables which consist of onsets and rhymes.

An area which typically furnishes indirect but powerful evidence for syllable structure is that of slips of the tongue. These are interesting because they show the language faculty faltering for some reason, usually psychological, such as nervousness or inattentiveness. The interesting point here is that the faltering would seem to betray the syllable structure which lies behind the utterances which we make. Here is a real-life example. One student complaining to another about study work she had to do exclaimed I couldn’t give a shamn about it. What would appear to have happened is that she was thinking of shit but decided to be more polite and just say damn. However, she took the onset of shit and merged it with the rhyme of damn, yielding sh-amn, the curious output form.

Exceptions to rules The point made in the last paragraph is that rules in natural languages are rarely without exceptions. But this fact does not hinder a language
or reduce its efficiency. It is more the student looking for clear-cut, exceptionless rules who is disturbed by the presence of exceptions. These do not invalidate a rule as can be seen from the following example. In the words divine and divinity, the vowels in the second syllable are different in each case, the vowel in divine is longer (technically called a diphthong) than that in divinity. This suggests that there is a rule in English whereby when a noun – like divinity – is formed from an adjective – like divine – the vowel in the second syllable is shortened. This ‘rule’ can be applied to other word pairs such as serene : serenity or audacious : audacity. However, it does not apply to all similar sets, for instance in obese : obesity the vowel in the second syllable is the same length in both words. The existence of this exception does not however invalidate the generality of the rule. Speakers simply remember that the rule does not apply to the obese : obesity pair. You can imagine that when young speakers of English come across this word pair for the first time they make a mental note of it not conforming to the otherwise regular rule of vowel shortening. Storing specific information for a word or words is what linguists called ‘lexicalisation’.

The range of processes Related to the idea of exceptions to rules is the expectation that a process can apply to any type of input. This is clearly not the case as, for example, with intensifiers in English. There are many instances where intensifiers are only found modifying certain words. Thus It’s so nice to read about linguistics is acceptable but *It’s so possible that she will come is not. Such distributions have established themselves in the language over time and it would seem that native speakers store such information as part of the entry for the word in the vocabulary in their heads, in what linguists calle the mental lexicon.

Another example of this phenomenon is the process of negative prefixation. For most adjectives in English it is possible to prefix un- which changes a word into the equivalent negative, necessary : unnecessary. But this does not apply to all adjectives; some are quite idiosyncratic in this respect, cheap : *uncheap; beautiful : *unbeautiful. In these cases the opposite meaning is expressed by a different adjective, i.e. cheap : dear, beautiful : ugly but this is not a water-tight account of the matter as there are words which have a regular negative and a different word with the same basic meaning, e.g. well : unwell : ill. It is interesting to see what happens in such cases. Two words with the same basic meaning tend to develop different shades of meaning (technically called ‘connotations’). In the case just mentioned we can see this has happened: unwell implies that one is not in the best of form (the word unwell is also more similar in its form to well) whereas ill means that one is at least temporarily in bad health.

What elements can occur together? Many of the above phenomena have to do with what words can combine with what others, often when one is a modifier
and the other is a word which is modified (technically called a ‘head’). A clear example of this can be seen with nouns where some co-occur with the comparative element *more*, while others cannot. Those that do are termed mass nouns, a rather flexible category. Hence *Do you want more milk, bread, pizza?* is acceptable but *Do you want more chair, bottle, plate?* is not. There is, so to speak, a syntactic side to this distinction. Mass nouns do not take an indefinite article but countable nouns do as can be seen particularly clearly from those nouns which can belong to either category: *Glass is a fragile substance* (mass noun), *I need a glass for the beer* (countable noun).

As always when trying to specify the regularities of language use there must be a caveat: the system may well be changing. Thus in English it used be the case that the word nonsense was solely a mass noun, e.g. *This is complete nonsense.* But increasingly this word is found with an indefinite article, i.e. it is changing its status and becoming countable as can be seen in sentences like *The whole matter is a nonsense.*

**Positive and negative expressions** Another area of apparent irregularity concerns the use of expressions in either a positive or negative form. While it is true that positive statements and structures are more frequent in language than negative ones there are some phrases which are only found in the negative, e.g. *He doesn’t give a damn about politics* but *He gives a damn about politics* is not acceptable. Another example would be *I haven’t a clue about linguistics* but not *I have a clue about linguistics.* Linguists say that such expressions have ‘negative bias’.

### 1.3 Terms and notions

**Linguistic terminology** The study of linguistics involves learning a whole series of new terms. Indeed the terminology is most often the main stumbling block for the student. This applies not just to new terms. In linguistics one comes across terms which have the same form as everyday words but the meaning is somewhat different. For instance, the term *accommodation* refers in linguistics to a process where speakers make their speech more like that of the people they speak to. This meaning cannot be directly derived from the general meaning of the word.

There are also many terms which are specific to linguistics. These are not devised arbitrarily but must always be justifiable on the basis of linguistic structure. Take as an example the word ‘determiner’. It is a label that allows one to link together definite and indefinite articles as well as demonstrative and possessive pronouns and treat them as a group. The justification of the term is that this group shows similarities in its behaviour in phrases, i.e. all its members can come before a noun and modify it as seen in such combinations as *the car, a car, my car, his car, that car, those cars,* etc. Using a single label for a class of words is a gain in simplicity: one needs fewer terms to describe the structure
and processes in a language.

**Abstract notions in linguistics** Analysis of language systems quickly reveals that there exist not only real, actual forms but also more abstract units which underlie these. This can be illustrated by a simple example. In English, there are two ways of pronouncing the word *butter*, with [t] (as in standard English), or with the sound we make when we are doing what is commonly known as ‘dropping our *t*’s’. In phonetics, this sound is known as the ‘glottal stop’ (because it is pronounced in the throat) and is transcribed as [ʔ]. Both pronunciations have the same meaning, ‘butter’, so making the distinction between the sounds [t] and [ʔ] in English does not change the meaning of a word. However, now consider the words *feel* and *veal*. Substituting the sound [f] for [v] (while keeping the rest of the words identical) actually produces a different word with a different meaning. Linguists express this difference by positing an abstract unit called the ‘phoneme’. Phonemes have the property of being ‘contrastive’, that is, they produce a contrast in meaning. The case of *feel* and *veal* is evidence for two phonemes in English, /f/ and /v/. Sounds that do not create a contrast in meaning – which are only alternative realisations of the same underlying form – are known as ‘allophones’. In the case of the two pronunciations of *butter*, one with [t] and one with [ʔ], we can see that the sounds in questions are allophones of the phoneme /t/ (in phonetics, allophones are transcribed with square brackets [], whereas phonemes are transcribed with slashes // – this is the subject matter of section ?? ?? below).

This phenomenon of abstract notion and actual realisation applies on all levels of language. Hence one has phonemes and allophones on the sound level, as just discussed. On the level of words we can see that some words represent different realisations of an underlying form. For instance, for the past tense of so-called ‘weak’ verbs there is an ending {d} (in morphology, the study of words and endings, curly brackets {} are used for underlying forms). However, this can appear as /t/ or /id/ depending on the end of the stem to which it is attached, e.g. *grieved* (with a final /d/), *liked* (pronounced with a final /t/) and *sorted* (pronounced with two syllables the second of which is /id/). So one can say that there is an abstract ending {d} which can be realised in three different ways in actual speech.

The principle of abstract notion and actual forms in speech or writing can be illustrated from the vocabulary of English as well. One can say that there is a verb *WALK* in English (the capital letters indicate that this is an abstract notion) and that it has the following realisations: *walk, walks, walking, walked*. Notation is important here because *WALK* is the abstract notion but *walk* is one of the actual forms as seen in a sentence like *We walk for an hour every day* (sample sentences and words are always written in italics).

The main point for the student to grasp at this stage is that while language manifests itself in actual speech or writing it has a more abstract structure
underlying this. The description and explanation of this structure is one of the primary goals of linguistics.

1.4 The organisation of language

**Modular organisation of language** Language is a phenomenon which evolved over tens of thousands of years and can be compared with parts of the human organism more than with man-made structures. In order to maximise efficiency and minimise the effect of damage to some part, language evolved into a system which is modular in its organisation. In this respect it can be compared to a part of human physical makeup like the immune system. Each module of language is self-contained with its own rules and representations – for instance, phonology (sound structure) and syntax (sentence structure) have quite different internal structures. There are of course interfaces between each module so that they can interact together and appear as whole in actual speech. What unifies all modules is the purpose of the system. Just as the immune system has the superordinate goal of protecting the body from infection, language has the primary function of enabling communication and ultimately all elements work towards this end. However, because language is a cognitive ability of humans it also has secondary functions. For instance, it has a frequent role as a carrier of social attitudes, something which has nothing to do with the simple communication of messages but which has been superimposed on the system. This is because speakers have always associated forms of language with those who use it and found that the system can be used to differentiate speakers socially without a loss of the primary function.

**Conscious and unconscious knowledge** When nonlinguists think of knowledge they think of conscious knowledge, i.e. of something which they can express and reflect on. For instance, if someone asks you whether you know how to play chess, then you would answer ‘yes’ if you felt able to list the rules of the game. Knowledge of chess would be seen as equivalent to the ability to use the pieces and explain the rules governing their movement to others. This is a typical instance of conscious knowledge as players are aware of the rules and reflect consciously on them when during the game.

Language is organised quite differently. Think for a moment of an iceberg. Nine tenths of it lies beneath the surface of the water. The one tenth which can be observed from above is comparable to the conscious knowledge of language which speakers have. This knowledge is generally confined to what are called open classes, to parts of language which can take in new elements. Open classes consists of nouns, verbs, adjectives and the like, that is of the elements of vocabulary which are not limited in number. Speakers’ awareness of language is greatest in this part of vocabulary: just think of the number of discussions you have heard on language which involve the use of such words, especially recent
formations like de-arrest or sex up. So-called closed classes are the opposite, they consist of sets of words, like prepositions, conjunctions or determiners, which are fixed in number and thus do not show a high level of awareness among native speakers.

Apart from certain types of words there are other areas of language of which we have unconscious knowledge, for instance syntax (sentence structure) or phonology (sound structure) and the underlying elements and structures of these areas. People do not think of the rules of syntax when they speak. But there must be such rules otherwise the form of our syntax would be completely random. So the assumption is correct that there are many rules governing well-formedness in sentence structure but these rules cannot normally be listed by native speakers, i.e. they cannot verbalise them unless the speakers in question have received specialist training in linguistics. An important point to note in this connection is that unconscious knowledge is as valid and substantial as conscious knowledge. And it must exist, otherwise our speech would be incomprehensible. Here is a comparison to illustrate what is meant: if you think about how you walk, then you probably imagine yourself putting one foot forward, shifting your body weight onto this and then moving the other foot forward and then shifting to this other foot. So much would be the equivalent of conscious knowledge of language. But walking requires a very intricate interplay of nerves, muscles and tendons as well as feedback to the organ in the head responsible for balance to ensure that one remains upright during the action. The neurological and muscular activity involved is not visible to the person walking and is similar to the unconscious knowledge of language which is active ‘in the background’ when we speak. This knowledge must be there, but it is right for the human organism to keep it hidden from speakers as too much consciousness would render speech too difficult – the main thing is that the unconscious knowledge works properly which it does nearly all the time. The only exceptions being unusual situations, temporary ones such as nervousness, tiredness or inebriation when one’s speech is somewhat uncoordinated. There are also more lasting disturbances, known from language pathology, which can arise after an accident or brain disease such as a tumor or a stroke.

Finally readers should bear in mind that unconscious knowledge about language is in part innate – the universals of language which we inherit and which are encoded genetically – and in part it is acquired in the early years of childhood. Unconscious knowledge works very well which is why a language which is acquired early in life is learned well. After puberty the ability to acquire knowledge so efficiently declines rapidly which is why adults have such difficulty in learning a foreign language.

What do speakers know about language? The distinction between unconscious and conscious language not only refers to what forms and structures are well-formed but also to what is not possible in a language. Imagine you had to invent a name (in English) for a new brand of dog food and someone suggested
the word *fnoppy* to you. It is unlikely that you would accept it because you know (unconsciously) that in English /fn-/ is not a permissible syllable onset (although it existed in Old English (e.g. *fneosan* ‘to sneeze’) and is allowed in other languages like Russian). In fact you know (again unconsciously) that only a fricative (a sound produced without interrupting the air flow fully) produced at the same point in the mouth can occupy the slot before a nasal sound here: *snoppy* would be a permissible sequence but *thnoppy* or *shnoppy* would not as neither *th-* nor *sh-* is produced at the same point as *n*. Now the fact that English speakers cannot formulate the restriction on syllable onsets simply means that this knowledge is unconscious. It cannot be verbalised by nonlinguists. But that does not diminish its validity.

Take another example, from syntax this time. The following sentence would not be accepted by speakers of English: *Saw I on the strand her*. The reason is simply that they know that in sentences representing statements the subject precedes the verb, i.e. *I saw ...,* and that the direct object always precedes a prepositional phrase, i.e. *...her on the strand*. You would never think of producing such a sentence because when speaking you automatically avoid structures which are ill-formed in English. However, the order of sentence elements varies greatly across languages. For instance, in Irish the order of elements in the ill-formed sentence of English just quoted is quite normal as can be seen from the Irish equivalent of this sentence: *Chonaic mé ar an trá í*, literally ‘saw I on the strand her’. Such examples show that there is variation in word order across languages, that speakers recognise this and that they unconsciously know (from language acquisition in early childhood) what the order is for their native language, that is what order will result in well-formed sentences acceptable to other native speakers.

**Summary**

- The *goal* of linguistics is to provide valid analyses of language structure. Linguistic theory is concerned with establishing a coherent set of independent principles to explain phenomena in language.

- There are no *primitive* languages. Each language is adequate for the society which speaks it, be this industrialised or not.

- *Onomatopoeia* (using sounds to indicate meaning directly) is not a major principle in language although symbols (*icons*) may be present on a more abstract level.

- There is no such thing as *correct* language in any absolute sense. Language is *neutral* and should not be the object of value judgements. Nonlinguists often confuse language and attitudes to those who use language.

- *Written* language is *secondary* and derived from spoken language. Despite its status in western societies, written language is only of marginal interest to the linguist.
Linguistics is a *science* although the evidence for assumptions about the structure of language is rarely direct. Linguists are more concerned with designing general models of linguistic structure rather than with searching for proof in any strictly empirical sense.

Language consists largely of *rules* which determine its use. There are, however, many exceptions. Native speakers can deal with a large amount of *irregularity* which is stored in the *mental lexicon*.

Knowledge of language refers to many *abstract structures* such as those of sentence types or systemic units such as phonemes or morphemes.

Language would appear to be ordered *modularly*, i.e. to consist of a *set of subsystems*, which are labelled ‘linguistic levels’, such as phonology, morphology or syntax.

Most *knowledge* about language is *unconscious* and cannot be accessed directly. The task of the linguist is often to demonstrate the existence of this unconscious knowledge, describe it and to suggest models for analysing it.

### Further reading


### 2 Characterising language

Having hopefully cleared the ground in the previous sections one can now turn to the strictly linguistic task of characterising language and defining linguistics.

*What is language?* This is the first question which is naturally posed in an introductory course on linguistics. When defining language a main aim is to...
delimit it from other systems of communication, be they animal ‘languages’ or artificial languages. Here the main emphasis is on the specific characteristics of human language. Looked at the other way around, one can say that the use of language is one of the defining characteristics of humans. The origin of human language will be discussed at the end of the current section and the nature of language change will be dealt with later on in this book (see section ???).

What is linguistics? The answer to this question is basically quite simple: linguistics is the scientific study of language as noted at the beginning of this chapter. However, some time will have to be spent on considering what aspects of language can be gainfully studied and how one is to go about this. These matter will be discussed in section 3 Analysing language below and throughout the book as a whole.

2.1 Defining language

Normally there is tacit agreement among linguists as to what constitutes language. However when pressed on the matter, they find it difficult to come up with a single definition which satisfies everyone. By their very nature, definitions of language try to compress into a single sentence the essential elements which are characteristic of human language. Below a number of such definitions by well-known linguists of the twentieth century are presented. There is common ground among all of them but the emphasis is different in each case.

1) Edward Sapir “Language is a purely human and non-instinctive method of communicating ideas, emotions and desires by means of voluntarily produced symbols”.

2) Robert Hall [Language] “is the institution whereby humans communicate and interact with each other by means of habitually used oral-auditory arbitrary symbols”.

3) Robert Robins [Languages] “are symbol systems ... almost wholly based on pure or arbitrary adaptability”.

4) Noam Chomsky “From now on I will consider a language to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements”.

The last definition above is a technical definition of the formal aspects of syntax and is less comprehensive than the other ones. It adds a dimension to the definition which is particularly important in modern linguistics, see (4) below. The common core among these statements can be summarised as follows.

1) Language is a system of communication
2) It involves sounds with arbitrary symbolic value
3) It is used by humans
4) It is a rule-governed system which is open-ended

One can of course go into more detail on the question of what is language and, having look at some definitions in broad outlines, a closer look at some features of language can now be taken.

*Language is a system of communication* This fact is fairly obvious. Despite the secondary functions to which language can be put to – for instance as a carrier of social attitudes, see above – it remains primarily a sign system for conveying information, i.e. a semiotic system.

*Language is stimulus-free* As opposed to most animal communication systems, human language does not need a trigger such as danger or the search for food or the desire for procreation. In essence, we can speak without any external motivation.

*Language is structure-dependent* Language does not consist of a string of random elements. The elements of language – sounds on the level of phonology, words on that of syntax – are arranged in a certain meaningful order determined by the rules of the language. If the elements are not, then the structures, words or sentences, would be incomprehensible.

*The relationship of words to concepts/objects language is arbitrary* Readers should note that the word ‘arbitrary’ is used in linguistics to denote a relationship between linguistic signs (words) and what they stand for (concepts which typically refer to objects in the outside world) and that this relationship is not fixed or determined by the nature of the objects. Of course individuals do not change the signs (words) used in their language – here the relationship is set by convention but one should remember that for instance there is no reason why a cow should be referred to as [kau] in English, after all *vache* [vaʃ] in French, bó [boː] in Irish and *korova* [koˈroʊə] in Russian sound very different and nonetheless seem to speakers of these languages to be entirely appropriate as the word for this animal. The question of arbitrariness will be returned to below (see 3.1 *Structural notions in linguistics*).

*Language shows duality of structure* One of the major principles in the organisation of language is that it involves two levels of structure, one of units and one of elements use to build these units. Take the structure of words as an example. These consist of sounds which in themselves have no meaning. For instance, one cannot say that /p/, /v/ or /n/ have a meaning but the combination /pun/ pun does.

*Language consists of discrete elements* The sounds of a language must be kept apart clearly, that is they are discrete in the technical sense. In English one cannot use a sound which is intermediary between /p/ and /b/ as this would not be sufficiently separate from both of these. This applies equally to vowels. Again in English one must distinguish clearly between the vowels is *bid, bed, bad, bud* and *bush*. The difference between the vowels in the second and third words is especially important for English and many Europeans have difficulty
here, often using the same vowel for both. What happens here is that these speakers do not treat these vowels as discrete, i.e. separate, elements in the sound system of English.

Discreteness requires that one has an exact realisation of each sound in the language/variety which one speaks. This is the essential difference between noise and the sounds of human language. Noise can vary at random but sounds of language must hit their target closely otherwise they are in danger of being confused.

Language is productive The number of utterances one can make in a language is not limited. For instance, new sentences are produced by taking one of a limited set of sentences structures and filling it with words from one’s vocabulary. By these means one can produce a theoretically unlimited set of sentences.

In word-formation one can also see this principle at work. Take the example of the ending -wise which is used to make adjectives from nouns. This can be applied productively to virtually any noun, irrespective of whether the new word already exists or not, e.g. Computerwise the department is well equipped.

What enables productivity is the application of rules to any input element to gain a new structure, be it syntactic or lexical. Care should be taken here not to confuse productivity with frequency. For example there is a method of forming plurals in English by altering the vowel of the stem as in man : men; foot : feet, tooth : teeth, goose : geese; mouse : mice. This is called umlaut (an originally historical process common to nearly all Germanic languages). Umlaut is not productive in English although it occurs with nouns which are in the core of the language’s vocabulary such as names for animals, humans and parts of the body. But it is never used with new words or borrowings from other languages.

2.2.1 Hockett’s design features

One of the most comprehensive treatments of the characteristics of language as a whole was made by the American Charles Hockett (see Hockett 1960). This set is referred to as ‘design features’ and represents an effort to characterise human language comprehensively. The more obvious of these features are part of general discussions of language and have been dealt with in the previous section. However, in the following additional features enumerated by Hockett are listed and commented on briefly.

Vocal-auditory channel This refers to the parts of the human physiology which are used for the production and perception of speech. In this context one should note that the organs of speech are not primarily designed for language production (they all have some other more basic function, like chewing for one’s teeth). Our system of hearing was not originally developed to perceive speech (but just the sounds of our environment).

Rapid fading Spoken language dies away quickly thus freeing the
channel for the next message. This increases the quantity of signals per unit of time. Contrast this with animal communication devices such as pheromones (chemical secreted to evoke a response in other members of a species).

*Broadcast transmission* This refers to the fact that the transmission of sound is omnidirectional, i.e. one does not have to face one’s partner in conversation and furthermore information can be conveyed in darkness. This does not apply to communication systems based on gestures.

*Directional reception* This permits the location of the speaker by the hearer.

*Creativity* Users of language are free – within the limits of social acceptability – to devise new forms in a language (see discussion of productivity at the end of the previous section).

*Semanticity* This refers to the existence of associative ties between elements in the linguistic system and the things or situations in the real world.

*Reflexiveness* The capacity to use language to talk about language itself. This is part of the general ability of humans to reflect on themselves.

*Prevarication* The possibility of deliberating telling untruths and talking nonsensically, i.e. language can be used in ways which are at variance with its primary function of communication.

*Displacement* The ability to refer to things or events remote in time or space.

*Learnability* The ability to learn another language (or more than one simultaneously). This would seem to be a consequence of the language system which does not per se put an upper restriction on either the nature or the amount of language which can be learned.

*Interchangeability* Refers to the ability of any human to act as transmitter and receiver alternatively. This does not necessarily apply to animals, e.g. bees convey information during their dance, but they do not receive feedback from others.

*Complete feedback* What is meant here is that speakers can perceive their own signal totally (and thus monitor it more carefully). This is obviously not the case where facial gestures are involved, e.g. with chimpanzees.

*Continuity theory* This refers to the view that language evolved over a long period of time and not as the result of a relatively sudden event.

*Iconicity* The use of symbols which bear a recognizable resemblance to the object their denote, see the discussion in 1.1 *Common ideas about language* above.

*Gaps and prohibitions* A gap is a permissible form in a given language which happens not to be attested. Prohibition refers to the exclusion on principle of some forms from a language. Consider the case of *blick* which is a possible (but unattested) form in English with *bnick* which is illegal due to the non-existence of initial clusters with a stop and a nasal.

In general human language can be characterised as displaying *efficiency* and *distance*. Efficiency is achieved by rapid fading, discreteness, broadcast
transmission, etc. Distance on the other hand tends to remove language from the physical environment in which it is produced. Here the essential aspects are displacement, reflexiveness, prevarication, interchangeability and arbitrariness.

Finally a distinction between types of rules in language is necessary. It is correct to assume that language is rule-governed behaviour, otherwise the output of linguistic activity would be random and language acquisition would be impossible as would sensible communication of any kind. Nonetheless not all rules of language are of equal importance. Here one can distinguish two main types of rule.

*Constitutive rules* These actually constitute the activity in question. For instance, the rules of chess form part of the game of chess.

*Regulative rules* These refer to the regulation of an activity which exists independently of the rules themselves, e.g. the rules of etiquette which regulate social behaviour.

### 2.3 Linguistic relativity

During the twentieth century linguists in America were concerned with describing the remaining native American languages before they died out. There are several hundred of these languages belonging to a couple of large families, the exact composition of which is still not decided on. Because they are unrelated to languages in Europe they are structurally very different these. This fact led many linguists to reflect on the possible influence which the structure of language has on thought. Two linguists in particular are associated with this idea and have given their names to a well-known hypothesis: Edward Sapir, the first major American structuralist and his student, the anthropologist Benjamin Lee Whorf. The view that language substantially influences thought is known as the *Sapir-Whorf Hypothesis* or the *Linguistic Relativity Hypothesis*.

This hypothesis has a strong version, namely that language determines thought and it is this which Whorf apparently adhered to. He assumed that the structure of western languages is quite similar and termed this Standard Average European. The structure of the native American languages which he investigated, above all Hopi (from Arizona), is radically different from European languages. For instance, the verb system of Hopi does not have a past tense form and the number system only includes lower numbers.

The proponents of the *Linguistic Relativity Hypothesis* maintain that such structures in language influences the thought of native speakers. The critics of such an extreme view state that distinctions which are not formally encoded in a language – or which are encoded differently – can always be expressed by paraphrase. A superficially similar example from European languages to show the absurdity of the extreme view. In French, German, Russian, Italian, Spanish, etc. there is the phenomenon of grammatical gender. To conclude from this that the speakers of these languages think of inanimate objects as male or female is absurd. Grammatical gender generally overlaps with natural gender within the
set of animate beings but not otherwise. It would show a basic misunderstanding to equate grammatical with natural gender for objects in these languages. A closer look at the native American languages, which Whorf originally considered, shows that they also have the option of paraphrase or have availed of other methods to overcome their seeming deficiencies. Hopi, for instance, has borrowed numerals from Spanish and uses these for high numbers.

One should be careful not to draw any hasty conclusions on the basis of what words exist in one language and do not in another. In German there is a separate word for animals when they eat, *fressen*, which is found alongside the term for eating when done by humans, *essen*. In English on the other hand there is only one term for both animals and human, *eat*. It would, however, be foolish to draw the conclusion that eating by animals has some special significance in German culture beyond that in England just because there is a special word for it.

Although the strong form of the *Linguistic Relativity Hypothesis* is obviously not tenable there is some truth to it which is why it surfaces again and again in linguistic literature, for example in introductions to linguistics like the present one. There have been many recent assessments of the hypothesis, see Lucy (1993).

**Summary**

- Linguists vary in their definitions of language. However, all agree that language is a *system of vocal signs* with an *internal structure* and used for the purposes of *human communication*.

- The *relationship* between signs and what is symbolised is *arbitrary* but fixed by *social convention*. The *system* is *stimulus-free* and *non-random*. It shows a *duality of structure* in having *building blocks* (phonemes) and *units* consisting of these (words). A small number of building blocks permits a large number of meaningful units.

- Languages vary greatly in their form and this has led some linguists to imagine that one’s native language determines the way one thinks. This extreme opinion is rejected nowadays.

**Further reading**


3 Analysing language

3.1 Structural notions in linguistics

The study of linguistics has two main aspects. Firstly, it is an attempt to understand the internal structure of language. Secondly, it endeavours to account for the way in which social significance is superimposed on this structure, i.e. how speakers manipulate language – if only unconsciously – to make a social statement of some kind, frequently either identifying themselves with those they are in contact with or dissociating themselves from them. The use of language in society is the topic of a separate chapter later in this book (see ??? below). First of all, one must consider the structure of language.

In the course of the history of linguistics various scholars have made a number of distinctions which help one to understand better the manner in which language is organised. The most eminent scholar in this respect is the Swiss-French linguist Ferdinand de Saussure (1857-1913) who taught in Geneva at the end of the nineteenth and the beginning of the twentieth century. He did not write any books on general linguistics but after his death his pupils put together notes taken during his lectures into a book and published it posthumously as Cours de linguistique générale ‘A course in general linguistics’ (1916) from which the following notions stem.

**Synchrony and diachrony** Saussure’s distinctions are based on dichotomies. The first and foremost of these is that of synchrony [ˈsinkrəni] and diachrony [doiˈækroni]. Synchrony is the investigation of language at one particular point in time (which may, but does not necessarily have to be, the present). Diachrony is the investigation of language over time. It is essential to grasp here that diachronic linguistics is secondary to synchronic linguistics. Indeed diachronic investigation is just the investigation of several synchronic ‘slices’ ordered one after the other. The reason why the structuralist looks at the matter like this is to stress that, even in diachronic studies, the changes which are noted arose due to factors concerning the system in each successive synchronic stage of a language.

```
+-----------------------------------------------+  > time axis
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
|                                             |
| synchronic ‘slices’ (points in time)        |
```

**Langue and parole** Saussure’s second major dichotomy is that between langue and parole. By langue is meant the system of the language; by parole is meant actually speaking a language (the term langage is used to refer to language as a
defining component of society). These two notions are close to those used by Noam Chomsky in the early sixties, namely competence and performance. There is a difference between langue and competence, however. The former stresses the system of a language as the common core of linguistic knowledge in a community of speakers (here Saussure was under the influence of leading nineteenth century sociologists like Émile Durkheim) while the latter refers to the abstract ability of a speaker to produce and recognise correct sentences in his/her own native language on the basis of structures which he/she has abstracted during the period of language acquisition in his/her childhood.

The distinction between langue and parole is important for locating the roots of language change. Variation on the level of parole is usually the source of language change (there are other cases which result from reanalysis of structures during childhood). When a change which starts as a matter of variation becomes obligatory and when it is no longer reversible (with a later generation) then one can speak of language change. It then becomes part of the langue of a community of speakers. For instance, the adverb alive comes from a much earlier phrase on life, as does asleep from on sleep. The preposition at the beginning was blurred in speech and at some stage was no longer recoverable for a later generation and so became part of the langue.

Signifiant and signifié For Saussure a language is a sign system. He uses the term sign, however, in a manner which is somewhat unusual for the linguist. A sign in a language is the relationship between a concept and an acoustic image, i.e. between a signifié (that which is signified, i.e. the concept) and a significant (that which signifies, i.e. the acoustic form, the word in a language). From this account it is clear that Saussure sees the concept as something mental which is independent of language to start with. Take as an example the concept of ‘tree’. For Saussure we have a notion of tree which is independent of language. The sign then is the link between a mental concept and a word in a particular language. This view has been the source of much discussion in later work. For many linguists the sign is the mental concept itself. However, this understanding of a sign would make our mental concepts part of a language. A basic tenet of Saussure’s thinking is that thought is independent of language and related to the latter by a system of arbitrary signs. These signs allow languages to link the mental concepts which speakers have (and which are not language-specific) to actual word forms. The relationships which are then established for a particular language are agreed upon by social convention and acquired by children in their childhood. The fact that language acquisition is so thorough a process can lead to a feeling that the word forms used for certain concepts are unalterable. However, this is not the case. A simple comparison with another language will show that this feeling is shared by speakers of other languages although the word forms are quite different. One can see within a single language that the words used for a concept can alter in the course of a single generation, just consider all the new terms for existing concepts which arose in order to remove linguistic
discrimination, e.g. *chairman / chairwoman > chairperson > chair* ‘head of a group’. Some changes were initiated to have a more neutral connotation, e.g. *prostitutes* are now referred to as *sex workers*. Other instances strive to add respectability to a matter, e.g. when *body guards* are called *protection officers* or *old-age pensioners* are termed *senior citizens*. Often equality is gladly suggested, even if this does not exist, e.g. using *economic co-operation* for *developmental aid*. On a slightly ridiculous level one has of course the excesses of political correctness, e.g. *small in height > vertically challenged, overweight > horizontally challenged*, although some of these formations, such as *visually challenged* for individuals with poor eyesight, have become accepted usage.

In connection with the present discussion it is important to observe that the Saussurean sign is the relationship between a word form and a concept and not between a word form and an object. For instance, although the word *tree* denotes an object in the real, non-linguistic world, the sign is the link between the acoustic form *tree, /triː/*, and our mental concept of a tree.

Saussure sees the relationship of signifiants (signs) to each other as linear, that is there is a principle of concatenation between words which holds in all languages and realised as a temporary sequence in speech. This view has been subject to criticism by many later linguists, such as Roman Jakobson, who stress the internal structure of segments in a language (words or below these, phonemes) and also that relationships existing between elements need not be linear but hierarchical.

<table>
<thead>
<tr>
<th>‘signifiant’</th>
<th>‘signifié’</th>
</tr>
</thead>
<tbody>
<tr>
<td>word in language</td>
<td>extra-linguistic concept referring to object in world</td>
</tr>
</tbody>
</table>

The view of the relationship of the signifiés to the words of a language as arbitrary is one of the major insights in linguistics. Saussure’s ideas on this have precedents which reach back to the Middle Ages and beyond. However, none of his predecessors formulated the matter with such rigour and precision. This type of relationship is what gives human language its specific quality as a communication system. Once language no longer had to reflect a word’s contents in its form (basically just onomatopoeia, see above) it was free to develop new forms which needed only to be consistent with the internal rules for word structure in the language and could be set in relationship to objects, ideas, feelings outside of language by arbitrary convention.

*Paradigm and syntagm* The fourth major dichotomy is that between *paradigm*
[ˈpærədaim] and syntagm [ˈsɪntæm]. A syntagm is a series of linguistic units arranged horizontally (e.g. a phrase or sentence). Each syntagm consists of a number of ‘slots’ into which various elements can be placed. A paradigm is then the set of elements which can occupy a single slot in a syntagm.

| Paddy | built himself | a | new | house |
| Mary  | bought | the | large | car |
| Colm  | taught himself | some | simple | language |

There is a further related use of the word ‘paradigm’ meaning the set of inflectional forms which a lexeeme (e.g. a noun or verb) can have. Thus ‘the verbal paradigms of Irish’ would mean the set of Irish verbs and all the various forms which each individual verb shows.

3.2 Further structural notions

Apart from the four Saussurean dichotomies discussed in the previous section there are others which are part of the bedrock of linguistic analyses. A number of these are discussed now and will be referred later, particularly in the sections on morphology and syntax.

Closed class This refers to those elements or forms in a language which are limited in number. For instance, the distinctive sounds of a language are limited, a figure of not much more than 40 such sounds is a typical average. Other examples are the group of prepositions, the number of verb forms all constitute closed classes. These are acquired in early childhood, are retained fully throughout the rest of one’s life and are virtually unalterable (though instances of language change in this area can lead to slight shifts).

Closed classes tend to contain polyfunctional elements. The motto here would seem to be ‘make maximum use of fewest elements’. For example, the set of morphological endings is limited in a language, but its elements often have several functions. In English the inflectional /-s/ signifies (1) genitive singular, e.g. Fiona’s hairslide (note that the apostrophe has no phonetic value) (2) plural of nouns, e.g. hundreds and thousands and (3) third person singular in the present tense of verbs, Fiona likes baking. In Irish the ending -(e)ach /-əːx/ covers a wide range of functions, e.g. (1) as a nominal ending, e.g. marcach ‘rider’, (2) an adjective iontach ‘wonderful’ or (3) a genitive ending for many feminine nouns, e.g. méid na cánach ‘the amount of tax’. The simple element a /ə/ has a whole range of functions from possessive pronoun, e.g. a ceann ‘her head’ to a relative pronoun, e.g. An cailín a bhuil suim aige inti ‘The girl that he is interested in’.
Characteristics of closed classes
1) small number of units
2) polyfunctional
3) acquired in early childhood
4) low or non-existent awareness for lay speakers

Open class This is a group of elements which can change in size, by adding new elements and of course by losing others. The typical example of an open class is the set of nouns, verbs, adjectives and adverbs. If one reflects for a moment one recognises that nearly all the new words which one learns are members of one of these word classes. The vocabulary of all speakers fluctuates throughout their lives.

An essential feature of an open class is that it enjoys a high degree of consciousness among speakers. This stands to reason: if one can add or remove items from a class, then one is aware of this. Furthermore, this means that when nonlinguists reflect on change in their language they invariably mention words which have been added or have been lost. Speakers rarely note ongoing structural changes to their languages.

Markedness This is a principle of language structure whereby pairs of features, seen as oppositions, are given different values (by linguists) as marked or unmarked. In its most general sense, this distinction refers to presence versus absence of a particular linguistic feature. For instance, there is a formal feature marking plural in most English nouns. The plural is therefore ‘marked’, and the singular is ‘unmarked’. The reason for postulating such a relationship becomes clear when one considers the alternative which would be to say that the features simply operate in parallel, lacking any directionality. Intuitively, however, most linguists would seem to prefer an analysis whereby dogs is derived from dog rather than the other way round – in other words, to say that ‘dogs is the plural of dog’, rather than ‘dog is the singular of dogs’. Most of the discussion of markedness centres on the extent to which there is intuitive justification for applying this notion to all such oppositions. But there are cases where the plural is more common than the singular, e.g. twins. Here one can say that the plural is unmarked and the singular marked, i.e. a sentence like Fiona is a twin is less likely to occur than, say, Fiona and Nora are twins.

It should also be mentioned that the assumption that one item is unmarked might be a result of traditional conceptions in a society. For example, who is to say that prince is the unmarked item in the pair prince/princess?

Several other interpretations of the notion of markedness are found in the literature, where the concept of ‘presence versus absence’ does not readily apply. One interpretation relates markedness to frequency of occurrence, as when one might say a falling intonation pattern was unmarked (as in most statements), compared with a rising one (that used in questions), because the latter is somewhat less common. Another interpretation is found in the semantic
analysis of lexical items, where pairs of items are seen as unmarked and marked respectively, on the grounds that one member is more specific than the other (e.g. *dog*/bitch, where the latter is marked for sex - one can say *male/female dog*, but these adjectives are inapplicable to bitch). A third, related sense occurs when the distribution of one member of an opposition is restricted, compared with the other: the restricted item is then said to be marked – several comparative sentences illustrate this, e.g. *How tall is Brian?* (whereas *How short is Brian?* is abnormal). Markedness is also used in grammatical theory to denote structures which are less usual than others, viewed cross-linguistically. Thus the word order VSO, found in Irish and Welsh for instance, would be regarded as marked because it is less common than SOV and much less frequent than SVO (the basic word order of English).

Naturalness This is a notion which was first used in linguistic theory to explain why there is a preference for some element over another cross-linguistically, i.e. among several not necessarily related languages. For instance, among the world’s languages back vowels tend to be rounded and front vowels tend to be unrounded. This is a statistical tendency but not an absolute universal. The linguist would then label a front unrounded vowel as ‘natural’ and a back unrounded vowel as ‘unnatural’ in this statistical sense. ‘Unnatural’ elements tend to be lost if they occur or never to occur in the first place, e.g. English has lost the front rounded vowels it once had, the sounds as in French *rue* ‘street’ and *peu* ‘little’; Irish never had these. French has nasal vowels, as in *bon* ‘good’, *chanter* ‘sing’ and *vin* ‘wine’. Polish has the latter two vowels as well but other European languages do not have any nasal vowels.

The notion of naturalness is also linked up with first language acquisition. Take for example the case of final devoicing, the pronunciation of all stops and fricatives at the ends of words as voiceless. Some linguists maintain that it occurs with all young children and is only reversed when they notice that English has an ‘unnatural’ setting for this feature because it distinguishes between voiced and voiceless obstruents at the end of words as in *cease* [sɪːs] vs. *seize* [sɪːz].

The idea of naturalness has been applied to other levels of linguistics as well, for instance in morphology it is used to characterise ideal situations, e.g. where nouns have an unalterable stem and where the singular is not marked and the plural has a single identifiable suffix as in English plurals like *boat*: *boats*. Words like *mouse*, *tooth*, *foot* are classified as unnatural because they form the plural by altering the stem rather than adding an ending. This notion has proved useful in explaining language change. For instance, it can account for the tendency of strong verbs to become weak, e.g. *help* which used to have a change in stem vowel in the past (*helpan*: *halp*: *geholpen* in Old English).

Oppositions Many of the distinctions in linguistics are based on the notion of oppositions. These are usually binary, that is they can have one of two values. A
simple example would be non-graded antonyms such as *dead* : *alive* where the first means ‘not alive’ and the second ‘not dead’. Such binary oppositions are found frequently in phonology, /p/ and /b/ are sounds which are opposed to each other in the presence or absence of voice. This recognition led in the 1930s to the theory of distinctive features outlined most clearly in *Principles of Phonology* (1939) by Nikolai Trubetzkoy (see the section on phonology below). Oppositions can be neutralised under certain conditions, for example the opposition of voiceless and voiced sounds is neutralised for /g/ and /k/ in Irish which is why the name *Pádraig* ‘Patrick’ is actually pronounced like ‘pawrick’ or the word *Nollaig* ‘Christmas’ is pronounced like ‘nullick’.

**Redundancy** This is an important feature in language and refers to the fact that more information is contained in a message than is in fact necessary. In the sentence *The books are on the table* the noun is marked for plural and so is the verb. For an adequate interpretation of the sentence only one instance of marking is required. Some languages are content with this, Swedish has one form for the verb in the present tense (*är*), the marking on the noun or the pronoun indicates number and person adequately, e.g. *Studentarna är i rumet* ‘The students are in the room’, *Hon är i staden* ‘She is in town’. In inflectional languages – Russian, German, Irish, Latin, for instance – there is considerable redundancy due to the many endings. Consider Irish *bun an ghleanna* ‘bottom of the valley’ which marks the second noun *gleann* ‘valley’ for the genitive (by changing *g* to *gh-* and adding -*a*) although its position immediately after another noun which qualifies it automatically assigns it to the genitive case.

Although inflectional languages have a lot of redundancy it is a feature of all languages to varying degrees. The reason for this is probably that it helps understanding in less than optimal situations, e.g. when there is a lot of extraneous noise, or poor transmission of the linguistic message. To take the first example above: if you did not hear the /s/ on the end of *books* then you could still grasp that a plural is meant by the special plural form of the verb *are*. The amount of redundancy in a language is thus a result of the tension between economy of grammatical forms and the communicative need to have easily understandable utterances in all situations.

### 3.3 Collecting linguistic data

There are a number of methods of collecting the data which one might want to use in a linguistic investigation. Before discussing these a distinction must be made which is important when examining data. **Metalanguage** is a term for language used to talk about language. The text of this book is an instance of metalanguage (from Greek *meta* ‘above’, i.e. language ‘above’ what one is talking about). **Object language** is what is examined by the linguist; it is the actual data of an investigation.
Scholarly investigations up to the beginning of the twentieth century very often tended to be subjective in their choice of data. This led scientists in various disciplines to consider means of rendering such investigations objective. In psychology the school of behaviourism arose which had a great impact on linguistics, particularly in America between the two world wars as evidenced by Leonard Bloomfield’s famous book *Language* (1933). According to behaviourism the only legitimate data is that gained from direct observation of human subjects. Applied to linguistics this meant that the only valid linguistic data was that gained from observing linguistic behaviour. A consequence of this was the rejection of any kind of speculation concerning linguistic knowledge or ability which was not directly observable. In its extreme form, linguistic behaviourism maintained that the conclusions to be drawn about language are those which are directly derivable from a body of linguistic data.

*The source of data* Any linguistic investigation will involve data as these are what the linguist makes statements about. An important consideration here is ensuring that data are as valid and general as possible. Valid data are those which are accepted by a majority of speakers of a language or, in the case of earlier stages of a language, those which are attested most widely. It is essential in this connection that linguists does not influence data by their procedures of selection. When collecting data there are basically three sources.

1) *The intuitions of linguists* This is only permissible where the linguist is a native speaker of the language in question. But even then it often occurs that many of the structures which are central to the analysis being made are not universally accepted by all speakers. This problem has in particular beset analyses of English syntax over the years.

2) *Elicitation of data from native speakers* If the linguist is not sure what structures are valid in a language then he/she may choose to interview a representative selection of native speakers on this issue. For instance, some verbs in English take a verbal complement with an infinitive and others take a participle. To determine what options are valid for a particular verb one could elicit responses from test persons by presenting them with templates like the following and asking them to fill in the empty slot.

*Fiona considered ____ home  Fiona wanted ____ home*

The answers would probably be as indicated below (a tick shows the acceptable and an asterisk the unacceptable sentences to speakers of present-day English).
Fiona considered going home  Fiona wanted to go home
*Fiona considered to go home  *Fiona wanted going home

Such instances are fairly clear cut. However, there are other similar structures where variation is to be found as in the following example.

Fiona was wanting to take her exams in the autumn.

acceptable o  unacceptable o

Here the interesting question is what percentage of interviewees accept the structure – want in the progressive form – and what do not (indicated by ticking the box). Any survey where there is less than full agreement on acceptability can indicate if language change is taking place. In such a survey the linguist would need to interview people from various sections of society, of different ages and both genders.

3) Consultation of a corpus Especially for earlier stages of languages using a corpus is the best method of attaining reliable data for one’s investigation. A corpus is in essence any ordered set of data, usually in written form (some audio corpora are available but only as far back as tape recordings go). The criteria for the compilation of a corpus are always specified in advance: a corpus could be a collection of works of a single genre, such as drama, or works from a given epoch or works which share some structural or stylistic property such as translations or personal letters or indeed a corpus could consist of the works of a single author. A corpus should also be as comprehensive and objective as possible. Nowadays many corpora for different languages and varieties are available on CD-ROMs for use with personal computers, e.g. The Helsinki Corpus of English Texts (Kytö 1993) or A Corpus of Irish English (Hickey 2003). There is also a special clearance centre for corpus research at the University of Bergen in Norway.

Objections to corpora Not all linguists agree with the use of corpora for linguistic analysis (type (3) above). The basic objection is that a corpus contains at once too much and too little data. Too much in that much of the data is sometimes unsuitable for investigation. It needs to be cleaned up before being analysed. Too little in that often the phenomenon which one is interested in is not represented in the corpus at one’s disposal (or only insufficiently so).

Advantages of intuitions The second source of data above has the advantage that one can test particular aspects of language which one is interested in by eliciting data from native speakers. Indeed what one frequently wishes to do is examine structures which are changing and see which speakers tend to accept and which tend to reject them. Elicitation is furthermore useful to test for the permissibility
of structures which a grammatical model may generate, thus providing vital information about the syntax of a language.

**Disadvantage of one’s own intuitions** The first source of data above has the disadvantage that there is no external control over the type of intuitions which a linguist may use for an investigation. This is particularly true when the linguist avails of his/her own intuitions. The danger of just accepting the data which fits one’s assumptions looms continuously.

## 4 On the origins of language

**The speculative tradition** There is a long tradition of speculating on the origin of language. This goes back for centuries and includes somewhat strange views, from simple religious ones that language is god-given to others that language originated in the imitation of animal sounds. The search for the first language has also led to many unreasonable views throughout history. The notion that there was once a single language, from which all others derive, has been particularly tenacious. Again there were many religiously guided views here such as the one that Hebrew – as the language of the Old Testament – was the original language. This particular view is connected with the legend of the Tower of Babel a name given to a story from the Old Testament (Genesis 11: 1-9) which recounts that the descendants of Noah began building a tower on a plain in Babylonia, intending to reach heaven. God was angered by this arrogance and foiled the plan by introducing different languages among the builders thus halting the project. Babel is a Hebrew word (*Babhel*) from Babylonian *bab-ilì* ‘gate of God’, possibly influenced by Hebrew *balal* ‘confuse’. The notion that Hebrew was the original language even led to the view that if children had no contact with anyone in their childhood they would end up speaking Hebrew.

The divine origin of language has long been adhered to. In the Judaeo-Christian tradition language is seen as a gift from God to mankind, indeed language and God seem to be intimately linked as seen in the famous quotation from the start of the Gospel according to St John: ‘In the beginning was the Word and the Word was with God and the Word was God’. Other religions have similar beliefs, for instance in Hinduism the creator of the world was Brahma but language came from his wife Sarasvati. The reason for divine origin theories is probably that man was, and often is, regarded as the zenith of evolution and language is the distinguishing feature of man from animals so this quality must have been given to him by God. This also accounts for the tradition of special language varieties for religious ceremonies and the development of ritual language for similar purposes. The power of language in this respect has always been recognised and is exploited in incantations and in name-giving customs. In the latter area one can see how language became highly significant for a community. One’s name is an important individual attribute and knowing or
not knowing names can represent power in a group so the ceremony of giving names is equally important.

However, all the speculative views of language origin are more of a hindrance than a help to the modern student of language. Indeed this was realised since at least the nineteenth century. When the scientific study of language began and institutions and journals were founded for academic discussion, the question of the origin of language was frequently banned because of the degree of quackery hitherto associated with it.

When did it all start? Language is a unique property of humans. However, many animals have communication systems which in principle are similar to language, i.e. they are used to convey information between members of a species. For instance bees use a special dance in which information about a source of nectar, its size and distance from the hive is transmitted by movements in space by the bees, their ‘dance’. Whales have an acoustic system which consists of noises sent out beneath the water to other whales. Other senses can and have been used for communication. For example many insects exude pheromones, scents with a certain signal value for a member of a species, normally to attract females. These systems – however fascinating they may seem to us because they are so different to ours – are all quite restricted in the amount of information which can be conveyed and are quite rigid. Except for the song of some imitative birds, like certain types of thrush, and a few other cases animal communication systems are generally inflexible.

In the line of evolution the humans are at the top along with other higher primates, notably gorillas and chimpanzees with whom they share various qualities such as (possible) upright movement, long arms, fingers, body hair rather than fur and a general similar appearance. We are first cousins of chimpanzees and gorillas but not directly descended from them. There was a split on the common path of evolution about six million years ago and the hominids continued to develop as a separate branch.

The species *homo sapiens sapiens*, the modern humans, started to evolve after about 200,000 years ago in eastern and southern Africa. From the point of view of anatomy these were modern humans, for instance they had the long necks which are associated with the production of speech. They also used tools as implements. *Homo sapiens sapiens* started spreading from Africa to other parts of the world. By 60,000 they appear to have reached Australia. In Europe, the first *homo sapiens* turned up around 80,000 ago as *homo sapiens neanderthalensis* but this line did not continue. The Cro-Magnons from the region of the Dordogne in south-west France are closer to modern humans. These were cave dwellers are engaged in wall painting and used complex tools.

One can also take it that language probably began to evolve about 200,000 years ago as well. In keeping with many evolutionary developments one can assume that there was a slow beginning, a period of rapid expansion and a later slowing down – this corresponds to an S-curve which is a common type of
development for natural phenomena. By 50,000 years ago language probably had reached a stage where it had all the functions which we know from languages today. This means that by the time recorded history starts – not more than 2,000 years BC – human language already had all the structural characteristics and functions of languages today. The societies of 4,000 years ago were certainly less complex than present-day ones but the languages used were not less primitive.

Where and why did it start? Unfortunately, we do not know what the impetus was for the rise of human language. One view, which has received favour among scholars recently, is that the first humans to use language came from east Africa. The reasoning is as follows. There was a major earthquake in east Africa which resulted in the Great Rift Valley which runs from the Red Sea down to the Madagascar Strait and which separated off the higher primates into one group which continued to live in the lush rain forests of western and central Africa and another group which was left stranded in the semi-arid regions of east Africa. The inhospitable conditions here led to the necessity for the higher primates to use their intelligence more than their western cousins. Part of their adaptation to the more severe conditions of their environment would have been the development of a system of communication amongst each other which evolved into human language. This view is strongly supported by archaeological evidence: chimpanzees are found only to the west of the Great Rift Valley while humanoid fossils are confined to the regions east of this.

There are many theories about the process of language evolution, for instance that gestures preceded language, seeing as how humans frequently accompany speech with hand movements. It is also uncertain just how language evolved, as a slow-quick-slow development (S-curve) or by leaps and bounds.

Humans have a considerably larger brain than other higher primates and there is a relationship between this fact and our ability to speak, though it is not possible to specify exactly what the correlation is. Furthermore, it is uncertain whether the large brain stimulated the development of language or whether language led to an increase in brain size. It may well be that the two aspects of human evolution went hand in hand, that is language development was approximately parallel to general cognitive development in humans.

Language as an evolutionary phenomenon There are a number of steps in the evolution of language which are significant for its special nature as a communication system. The first is independence of environment by which language came to be dissociated from the physical habitat of its users. The second, as mentioned above, is where language split up into a modular system in which each module or subsystem achieved quasi-autonomy and hence was shielded from interference from other modules. This represents a kind of optimisation, seen in other bodily functions, such as the digestive system, which is achieved in the fullness of time.
As language is not a human artefact it is not subject to aesthetic considerations which are often superimposed by humans on the objects they create. Instead it must fulfil its ultimate function of communication which may be done even if residue from earlier stages is still present or where there are fuzzy edges to categories.

Like other evolutionary phenomena, language became genetically encoded over a very long period of time. In this respect it can be compared to the way in which most animals have experience from previous generations of their species encoded in their genes, allowing this experience to be transmitted to following generations. This is why animals know how to behave – for instance how to fly or what plants to eat – without being instructed by their parents. For humans, language experience would seem to have led to the genetic encoding of universal features which are common to all languages but specific to none. The details of single languages are what are acquired by children in the first five or six years of their lives. This view of an inherited core and an additional amount gained from one’s immediate surrounding in childhood is the essence of recent linguistic theory which is concerned with determining what elements of language structure are universal, i.e. common to all languages, and hence most likely to be innate.

Primary and secondary functions In the course of evolution many primary bodily functions acquire a secondary function which is superimposed on the first. A good example is urination which is the process of ridding the body of liquid waste. Because of the individual smell of urine it came to be used by many animals for delimiting their territory vis-à-vis others of their species.

Many features of human physiology point to an adaptation for language. For instance, the larynx of humans is particularly low in the neck. This has a physiological disadvantage in that we could theoretically choke when we eat as we must close off the top of the trachea each time we swallow. However, the advantage for speech is that we can produce a large number of sounds with a considerable volume. The sounds of human languages – the great range of vowels and consonants – is only possible given the shape of our mouths, the flexibility of our tongues and the additional use of the nasal cavity and the switching on or off of vocal fold vibrations which is an essential feature for the consonantal systems of all languages.

Another feature deserving of comment is the highly flexible human tongue. To fulfil the primary function of moving food around the mouth and swallowing it our tongues would not have to be so dexterous – just think of the tongues of dogs or cats which are quite limp but nonetheless sufficient for their primary purpose. So it would appear that the agile tongue muscle of humans developed because of the secondary function of speech production.

In more general terms one could also point to the phenomenon of neoteny. This refers to the extended childhood of humans which allows them to develop their faculties, such as language, more fully than if they reached rigid
adulthood more quickly. There is also a sense in which humans are also born ‘too early’ (from a purely physical point of view). A simple reason for this could be to allow the child to pass through the birth canal with its fairly large head without too much difficulty. However, a consequence of this is that the child enters the outside world quite early and can receive and process language and other stimuli from this source from a very early point in its physical development.

Qualitative leaps in the development of language Whatever about the speed of development of human language there must have been a number of leaps in evolution which led to language attaining new qualities. A number of these qualitative leaps can be mentioned here, especially as they characterise human language vis a vis animal systems of communication. The first leap must have been the break in the relationship of signifier and signified, that is the removal of a necessary connection between the sounds of language and what they denoted. Another similar leap must have been the rise of duality of patterning (see above), the principle whereby all languages consist of building blocks from which units are constructed, i.e. individual sounds from which words are made. This principle greatly increased the power of human language. With a small number of building blocks (sounds) it was possible to make a virtually infinite number of units (words). This principle was extended to encompass syntax as well: with a limited set of structures it became possible to construct a potentially infinite number of sentences. Other developments in language are more closely connected with the structure of human cognition and may have been dependent on advances in this area. For instance the ability of language to show locational and temporal displacement, i.e. to refer to places and times other than the here and now. The ability to abstract away from the present is associated on a linguistic level with irrealis constructions, what we do in English with if-clauses or verbs like suppose, imagine, etc. This feature in language is closely connected with the use of abstract nouns which testify to the imaginative power of human cognition.

Summary

• There is long tradition of speculating about the origin of language. Most of this was and is unscientific as it does not apply stringent principles of historical continuity and interrelations.

• Modern humans have existed for about 200,000 years and after 50,000 BC language had probably developed all the structural properties which are characteristic of it today.

• Language is an evolutionary phenomenon which is continually adapted to the communicative needs of its speakers.

• The organs of speech are biologically secondary but their rise has led
to a *specialisation* such as the great flexibility of the tongue or the relatively deep larynx which distinguishes humans from higher primates.

**Further reading**


