Linguistics for Beginners

Basic Concepts
Linguistics for Beginners
Basic Concepts

TARIQ RAHMAN

OXFORD UNIVERSITY PRESS
## Contents

**Preface**  

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface</td>
<td>vii</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. A Brief History of Linguistics</td>
<td>8</td>
</tr>
<tr>
<td>3. The Study of Sounds</td>
<td>24</td>
</tr>
<tr>
<td>4. The Study of Sound Patterns</td>
<td>44</td>
</tr>
<tr>
<td>5. How Words are Made</td>
<td>53</td>
</tr>
<tr>
<td>6. The Study of Meaning</td>
<td>62</td>
</tr>
<tr>
<td>7. Introduction to Chomsky’s Theories</td>
<td>72</td>
</tr>
<tr>
<td>8. Transformational Generative Grammar</td>
<td>80</td>
</tr>
<tr>
<td>9. More about Grammar</td>
<td>101</td>
</tr>
<tr>
<td>10. Government and Binding Theory</td>
<td>111</td>
</tr>
<tr>
<td>11. Universal Grammar</td>
<td>117</td>
</tr>
<tr>
<td>12. Anthropological Linguistics</td>
<td>127</td>
</tr>
<tr>
<td>13. Sociolinguistics</td>
<td>140</td>
</tr>
<tr>
<td>14. Language Politics</td>
<td>162</td>
</tr>
<tr>
<td>15. Language Death</td>
<td>171</td>
</tr>
<tr>
<td>16. Educational Linguistics</td>
<td>176</td>
</tr>
<tr>
<td>17. Writing in South Asia</td>
<td>182</td>
</tr>
<tr>
<td>18. Notes for the Reader</td>
<td>198</td>
</tr>
<tr>
<td>19. Bibliography and Further Reading</td>
<td>199</td>
</tr>
</tbody>
</table>
Preface

This book is written to introduce South Asian students to linguistics. It is written in simple English, and the explanation is so clear and detailed that no previous knowledge of the subject is necessary to understand the basic concepts of linguistics. Indeed, the aim of this book is to teach linguistics to beginners who have never studied it and who do not have a teacher. Thus the book explains concepts in the kind of language which a teacher would use when explaining concepts to undergraduates or high school students. It is meant for the bachelors’ level but may be used in the ordinary two-year masters’ level in some South Asian universities which comes after a two-year BA and which does not have specialized courses to begin with. As this is not a specialized book on linguistics, there is a little of most of the concepts which South Asian universities require.

Of course, in order to make the beginners familiar with the most important aspects of a course in linguistics, there is much more of such core subjects as sounds, the production of sounds in the mouth, the description of sounds, the way words and sentences are constructed, and how there are rules governing these constructions. All these are areas of what may be called microlinguistics. But there are also important aspects of how language is used in society, how we create meaning, and how we express reality through language. This is called macrolinguistics and is often more interesting for people studying sociology, anthropology, computer sciences, politics, and history etc. The book does introduce some of these areas but, due to their importance, it gives more space to concepts without knowing which nobody will call you a linguist.

Linguistics—the scientific study of language—is taught in most of the major universities of the world. It helps us understand what human languages are and how they produce meaning. As languages are used in societies there is much that knowledge of languages and
how they are used in society can tell us about society itself. We can even understand some of the ways the brain itself functions by understanding how it produces the sentences of a language. In short, linguistics contributes towards the study of culture, behaviour, and thinking. People who work in artificial intelligence and robotics also use linguistics to provide them insights into how language is processed in the brain.

There are excellent departments of linguistics in India and a number of linguists of world stature. There are also departments of linguistics in Nepal as well as Sri Lanka. In Pakistan, courses of linguistics were introduced from 1987 onwards by the author and now there are degree courses in linguistics in several universities, though most of the courses they offer owe their origin to English Language Teaching (ELT). However, even so there are many students who study linguistics at the masters’ level in Pakistan. Unfortunately, there are very few books of an introductory nature for them in Pakistan.

The book which I used when I was teaching myself linguistics in 1987 (incidentally I had a PhD in English literature and was full professor then, but I had lost interest in literature for both research and teaching) was *An Introduction to Language* (Fromkin and Rodman 1974). In 1990, while doing a short course on language and the mind in the University of Cambridge, I met a really gifted lecturer, Dr Jean Aitchison, whose books were so clear that I would recommend them for anybody who is learning linguistics. A few months later in 1991, at the University of Sana’a, I met a colleague from India whom our family called Deedi but part of whose name Dr Chanda stays in my memory. She was trying to write a simple introductory book for students of English Language Teaching (ELT). I do not know if that book ever got written but the idea of a book which could teach beginners linguistics even without the mediation of teachers stayed in my mind. A few years later, while writing *Language and Politics in Pakistan* (OUP 1996; available in India in the Orient Blackswan edition of 2007), I wrote a small book called *An Introduction to Linguistics* (Lahore: Vanguard, 1987) which is the basis of this revised and updated edition. This, however, is a new book as almost everything has been changed in
it, though it does overlap with the earlier work in some places. I thank Najam Sethi, an old friend and publisher, who published *An Introduction to Linguistics*. I also thank Ms Ketaki Bose who encouraged me to rewrite it so as to produce a new edition for Indian students. Since I wanted the book to be available to Pakistani and South Asian students as well as others, I wanted it to be published in Pakistan in addition to India. I thank Oxford University Press for making this wish come true. I am especially grateful to Ms Ameena Sayid, Managing Director of Oxford University Press in Pakistan, for encouraging me to submit the manuscript. I also thank Mr Samuel Ray, Senior Editor, Oxford University Press for having taken pains to make the book see the light of the day. But for the work he and his colleagues put in the book would not have been before the readers today.

I realize there are many such introductory books, including some recent additions (Matthew 2003). However, I flatter myself that none of them cover as many aspects of linguistics and in as simple language as this one.

I have kept South Asian names and cultural references with a view to engaging the attention of students of all SAARC countries. However, being familiar more with the culture and publications of Pakistan and India, I apologize for there being more references to these two countries. I hope students in Pakistan and India, as well as other South Asian countries, will find this book useful. As the book is written in clear English and the jargon of linguistics is introduced to the reader with adequate explanation, it can also be read by the lay person. In fact, one of the reasons for writing it is to introduce the ordinary reader to linguistics.

This book has been divided into the following chapters. Chapter 1 defines what linguistics is, and Chapter 2 traces out the history of its development. Chapter 3 and 4 are about the sounds which combine together to make up a language (phonetics). Chapter 5 explains the way words are constructed (morphology). Chapter 6 is about the meaning of words or semantics. This is followed by five major chapters—7, 8, 9, 10, and 11—on the theories of Noam Chomsky, i.e. transformational generative grammar and the
theories born as it developed, such as the government binding theory and the minimalist approach. As these theories require a higher standard of knowledge than is assumed at this introductory level, the treatment is simplified and the more complex developments of recent years have not been touched upon.

Chapters 12 and 13 are on the social and cultural aspects of language which will be useful for students of anthropology and sociology, especially those interested in gender and language death. Chapter 14 is on language politics. It will be useful for those who are interested in the relationship between language, identity, and politics, and especially for those who write dictionaries and grammars of languages or make language policies. Chapter 15, on language death, is meant to create an understanding of one of the consequences of modernization and globalization, i.e. that these processes entail the disappearance of aspects of our culture and identity. Our sense of belonging is under an assault the like of which has probably never been experienced before. Chapter 16 is on educational linguistics, i.e. the teaching of languages and the political, educational, psychological, and economic perspective in which this teaching takes place. Chapter 17 is on writing in South Asia but there are some introductory sections on the history of writing in the world also. These chapters connect the study of language with the study of globalization, politics, language policy, public policy, and economics. In short, such awareness may bring about a deeper understanding of what human knowledge is, and how it can help the human race to live in peace and harmony. In chapter 18, while bidding good bye to the reader, a few hints are given as to what paths may be followed if one pursues linguistics as a field of study. The bibliography gives details of books which have been used in preparing this brief introductory guide. However, it also serves as a list of books for further reading. They will serve to give details of aspects of linguistics which have merely been touched upon in the text.

Tariq Rahman PhD
Islamabad
2010
Introduction

What is linguistics?

Do you know how many languages exist in South Asia? If not, go to the *Ethnologue* which is published by the Summer Institute of Linguistics. The latest online edition (2005) gives the following information.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number</th>
<th>Extinct</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>438</td>
<td>13</td>
</tr>
<tr>
<td>Nepal</td>
<td>126</td>
<td>3</td>
</tr>
<tr>
<td>Pakistan</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>Bhutan</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Maldives</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>


The world has about 6,912 languages. But it is not only numbers which tell us how important language—just the fact that we humans can convey meaning and write them down in symbols—really is. Without language we would have no conversation, no songs, no stories, no jokes, and no civilization as we understand the term. Indeed, perhaps we would be like intelligent monkeys but would we be human? One doubts that we would. Language distinguishes us from animals.

Just observe how we behave. People talk more than they do anything else. They generally talk when there is someone to hear them but sometimes even where there is nobody to do so. In literate societies words are also written and read. Almost everybody reads signs on the road, in shops, and on the TV, but there are also people who spend most of their waking hours writing or reading.
All this is language in action. Such an important thing—human language—is an object of interest in itself. Those who are interested in its study are called linguists and the scientific study of language is called linguistics.

**What do we study in linguistics?**

All human languages are made up of sounds arranged in such ways that they communicate meaning or express feelings or attitude. The sounds come first. In fact, when writing had not been invented, the sounds of languages existed. The study of sounds—i.e. how they are produced and described—is called phonetics. The study of how these sounds are arranged in a language, i.e. the sound-pattern of a language, is called phonology.

These sounds combine to make small units of meaning which are called morphemes. For example, *boy*, *girl*, *-ish*, and *-s* are morphemes. They express meaning on their own or in combination with other morphemes. Thus we can make the new words *boyish*, *girlish*, *girls*, and *boys* with these morphemes. The study of how words are made (i.e. the shapes of words) is called morphology.

When words combine together in sentences, they do so according to certain rules. These rules are the rules of syntax. If we want to know how sentences are formed in a language, we study its syntax.

Now that we have got our sentences, we should know what they mean. The study of meaning is called semantics.

Hence, these are the four main aspects of the study of a language: (1) phonetics and phonology (2) morphology (3) syntax (4) semantics.

Some people also add another study—that of language in real life situations. This is called pragmatics. Pragmatics deals with the way people actually use words in daily conversation. For instance, if someone asks me what time it is and I say the news have not started yet, that person will know what time it is only if he or she knows
that the news start at such and such time. The study of implications, the use of titles, words of politeness and rudeness etc. come under pragmatics.

Let us arrange these aspects—or levels as they are called (Crystal 1988: 82) in an order:

```
LANGUAGEx
   |STRUCTURE USE
   |   PRAGMATICS
   |SOUND GRAMMAR MEANING
   |PHONETICS PHONOLOGY
   |MORPHOLOGY SYNTAX
```

The linguist Leonard Bloomfield (1887–1949), one of the pioneers of linguistics, called these levels the structure of language (Bloomfield 1933). He recommended starting from phonetics and going on to phonology, morphology, syntax, and semantics. This kind of study would tell us a lot about the structure of a language. Such an approach to language—i.e. seeing it as a structure—is called structuralism.

**LANGUAGE AS SYSTEM**

Every structure is a system in itself. If you change one part of it the other parts change in order to adjust themselves. The first linguist to point this out was Ferdinand de Saussure (1857–1913) whom we shall discuss in more detail later. Saussure gave the name of *la langue* to the whole system of a language.
The knowledge of the whole system is in the mind of the native speaker. Without such knowledge sounds cannot be produced nor meanings understood.

Saussure’s other concept, *parole*, involves ‘the combinations by which the speaker uses the code of the linguistic system’ (Saussure 1916: 14). In other words it refers to what the speaker actually says, i.e. the sounds one produces. Now, of course, one may be in a hurry or under some pressure which may cause one to grope for words or stutter. Thus parole is often faulty, incoherent or inadequate whereas *la langue* is not. It cannot be because it is the knowledge of how the whole system of a language works. It is stored up in the brain but not all of it is used perfectly all the time (For Saussure’s ideas read Culler 1976).

If we want to know how the system works then we need not look at the actual performance of individuals. We can look at the laws by which sounds combine together (phonology); the way units of meaning combine to create words (morphology); and how words are placed one after the other to create sentences (syntax). Such studies of abstract laws or rules of the system of a language would be the study of *la langue*. This is what theoretical linguists do.

Saussure’s concept of *parole*, which refers to utterances of individuals, can also be called the performance of those individuals. Noam Chomsky (1928–), perhaps the greatest linguist in existence; used the term *performance* roughly for *parole*. For Saussure’s concept of *la langue*, Chomsky used the term *competence*. As competence is the knowledge of the system of a language, it is in the mind of a speaker of that language. Theoretical linguists are interested in this system in the mind but it is beyond the scope of this introductory guide.

So far we have talked about linguistic theory, the focus of the science of language. Some people call it *microlinguistics* because it deals with only one aspect, the description of language, and not with the other things language is connected with. These other things can be society, culture, biology, mathematics, literature,
politics, and so on. The subject dealing with the relationship of language with society is called **sociolinguistics**. If the emphasis is on culture, worldview, and preliterate societies, the subject is called **anthropological linguistics**. Those who are interested in language-learning and how the mind processes language study **psycholinguistics** and **neurolinguistics**. The latter is very interesting for biologists and medical scientists because it studies speech defects such as aphasia and the involvement of the brain in producing, understanding, and storing language. A pioneering book on the biological foundations of language by Lenneberg (1967) might repay reading. The brain has areas which store words and the rules which produce sentences. Most of us use the left side of the brain for linguistic work. We have about 12 billion neurons and they connect with each other as if electrical currents are passing through them. These days positron emission tomography (PET) can tell us which part of the brain is working. The principle is simple. If radioactive particles are given to a person in some form of sugar and the person does linguistic work, the part of the brain which consumes more energy will emit more particles. Thus we can find out which part of the brain does what kind of work. Unfortunately the scholarly work on this subject is a bit technical in nature. Yet one book which can be recommended for beginners, despite the fact that it has been superseded by new research (which is again a bit too difficult), is Peter Russell’s book called *The Brain Book* (1979). An advanced book on aspects of the mind-brain and language—which I recommend because it has been edited, among others, by a Pakistani linguist called Anjum Saleemi—is *In Search of a Language for the Mind-Brain* (Saleemi et. al. 2005). This book takes the reader into the highly interesting realm of philosophy as well as neurolinguistics and brain science.

Apart from biologists, computer experts, and mathematicians too are interested in linguistics. Thus there are branches of linguistics called computational and mathematical linguistics. Computational linguistics has given us the gift of seeing our South Asian scripts on the computer screen. We now have Urdu written in the *nastaleeq* script as well as Pashto and Sindhi in the *naskh* script on
everything including e-mail in Pakistan. In India we have many scripts including the *devanagari* script. To make the rules which put them there is a very complicated bit of computer science as well as linguistic knowledge. Recently insights from linguistics have been used by those who study the mind in general and artificial intelligence in particular. There are new intellectual ventures and **philosophical linguistics**, which is even older than the works of the Greek philosophers Plato and Aristotle on the subject, is being examined afresh. Highly complicated questions about how language helps people to construct reality are being debated. Style, including the use of style in literature, is being examined in **stylistics**. All these subjects may be called **macrolinguistics**, i.e. the study of language in relation to a number of other subjects. The following chart helps clarify what has been explained above.
In this book we will not cover all these subjects. However, an effort will be made to introduce you to microlinguistics and some aspects of language in relation to society and culture.
very society, and indeed almost every individual, has myths about language. In general these are myths about origin: that such and such language was created first by God or gods. Such myths are not testable, and because everybody does not believe in the same myth, they remain in the realm of faith-specific or culture-specific beliefs. In order to pursue scientific linguistics, we should aim at universality (i.e. our results should be valid for the whole world) and objectivity (i.e. everyone should reasonably believe our conclusions). An example of such an undertaking is Guy Deutscher’s book on the evolution of language (Deutscher 2005). The book can be understood by those who know basic linguistics. The arguments in the book are summed up below.

Basically, Deutscher argues that human languages have a complex structure but it was not crafted by some genius or a committee of geniuses in a short time. Instead, it arose through natural forces which keep changing all the time. Of course, it is the people who speak the language who change it. And why do they do that? Possibly because, among other things, they may want to borrow things from other cultures for which they have no words. A recent example is the borrowing of English words in our South Asian languages, especially in Urdu and Hindi. But there were also borrowings from other languages such as Portuguese (did you know that peon, almirah, and even achar is from that language?), and of course, the borrowings from Persian and Arabic are too many to be listed. However, it is not only the words which change. The pronunciation also changes since we may cut short a few sounds (economy) or decide to give a powerful impact to our words
(expressiveness). When we do the latter we may combine words together to create new words (like *fantabulous*!). Since our minds crave for order we keep creating patterns and so language keeps changing. The author sums his thesis admirably as follows:

> All that was required at the ‘me Tarzan’ stage were words for physical objects and actions (as well as two pointing words), and a few natural principles for ordering them. From these basic materials, the natural forces of change could have fashioned the structure of language in all its prolix splendour (Deutscher 2005: p.262).

Of course, nobody knows definitely whether there ever was a stage when people used only a few words like the Tarzan comics do (*I Tarzan; you Jane*). This is a hypothesis but it should be accepted as being true for the time being till we get a better explanation. This is how scientific research proceeds. Nothing is sacred in science, but whatever theory is accepted must explain things for the moment. Incidentally, this theory is very similar to the theory of the development of genes in the bodies of plants, animals, and humans. For a very clear account, read Richard Dawkins’ book called *The Selfish Gene* (1978). He too says that all these complicated forms of life grew as genes learned to produce photocopies of themselves in order to survive. But let us get back to language.

**The first language**

If language arose with the Homo Sapiens (*sapient* = wise) then it was born about 150,000 years ago probably in Africa, though it is possible that several humanoid groups arose in places within years of each other. This has been calculated on present-day variation of a form of DNA passed by the female line. The rate at which it mutates can be calculated and that gives us this approximate figure (Corballis 2002: 42). But we still do not know what this early humanoid spoke. There is, indeed, a theory that the earliest form of communication was the gesture. Michael Corballis has argued this in a highly interesting book (Corballis 2002).
Another question which people ask is whether there is a natural or ‘original’ language of humanity. The brief answer is that we do not know whether there is any such thing, and if there is, we do not know which one it was or is. People have carried out cruel experiments to find out what this language is. For instance, the Mughal king Akbar put children in a house where only deaf-mutes served them. They were not allowed to hear any human being talking. These children could not speak when they grew up. Children brought up by wolves and other animals (called feral children) do not know any human languages but do try to make sounds of animals. In short, human beings probably do not have any original or natural language. They merely learn the language which they hear around them.

ARE LANGUAGES EQUAL?

Yet another myth is that some languages are sweet while others are rough. Linguists do not classify languages in this manner because ‘sweet’ and ‘rough’ are subjective ideas. If we are talking about sounds we find the sounds called sweet in one language also present in a language which is considered rough. If, however, we are talking about intonation (the way we stress certain parts of words and our voice goes up and down) then we need to test this intonation pattern on a big sample of people in order to be able to claim that the terms ‘sweet’ or ‘rough’ are related to properties of speech itself.

Most people also agree on calling some languages superior and others inferior. They say that one language is more ‘logical’ or has more words than another which lacks these characteristics. Modern linguists hold the view that all languages are equal; none is superior to another. This does not mean that some of them are not better suited for a certain function. English, for instance, is better suited for space research or computers than a language which does not have the technical terms for such kind of work. This, however, is functional suitability not structural superiority. As far as the structure—phonemes, morphemes, syntax etc.—is concerned, we
find it difficult to set up a standard of excellence and then measure languages according to it. It may be true that certain sounds or patterns of stress may sound musical or pleasing to everyone but no scientific research has proved that any language is objectively ‘sweet’ or ‘beautiful’ for all human beings. Thus there are no scientific grounds for saying that any language is intrinsically (i.e. structurally) inferior or superior to any other language. As for functional suitability, any language can be changed to perform any function. For example, if we want Punjabi to be used for sophisticated scientific research, we can invent or borrow words to express new meanings. That is what the speakers of English did and that is what all language planners do everywhere when they want to use a language for science and technology.

Why then do people hold such strong myths about languages? In general these myths are expressions of their views about the world. If they look down upon someone or a certain culture, they also tend to look down upon that person or cultures’ language. In South Asia city dwellers look down upon villagers. Thus they also look down upon villagers’ languages. They regard them as inferior and even assert the view that they do not have a grammar. As you know, it is impossible to speak a language if its speakers do not have grammatical rules in their minds. Villagers do have the grammar of their languages in their minds but it may not be a written grammar. It may be unwritten not because it is inferior, but because villagers neither have power nor money, and without these things one cannot promote one’s language or publish books in it.

The persistence of these myths shows us that one must put one’s prejudices and beliefs aside when one makes a scientific study of language. If we look at the history of linguistics, we find that it is only recently that people became scientific in this field. This does not mean that there were no eminent linguists in earlier eras. There were many, and this chapter will introduce you to their thoughts.
**LINGUISTICS IN ANTIQUITY**

Most Western works on the history of linguistics begin from Plato (427–347 BC). However, it was in South Asia—to be precise, somewhere around Taxila in modern Pakistan—that the Sanskritic grammarian Panini lived between the seventh and fifth centuries BC, i.e. earlier than Plato. There might have been linguists even earlier than Panini but there is no known written work left behind.

Panini, however, left behind 4000 *sutras* (verses) on Sanskrit. His eight books, *Ashtadhyayi*, contain descriptions of Sanskrit grammar. Panini may be said to be the pioneer of phonetics because he tells us how sounds are produced, i.e. where the tongue moves in the mouth.

Plato, the great Greek philosopher, was more interested in semantics than phonetics. In his book *Cratylus*, some characters argue about the nature of language. One of them, named Hermogenes, argues that language is a product of social convention. This means that we use the word *dog* for a four legged animal of a certain kind only because of an arbitrary convention. We can agree to call it *kutta*, *sug*, *kulb*, *kutto*, *spai* or *chien*. No matter what we call it, the sounds we produce refer to this animal. There is no necessary relationship between the sounds and the object. This is known as the theory of the arbitrariness of language or the *conventionalist* view. It is also known as the theory of language being an agreed upon convention, an arbitrary code.

The opposite view, argued by Cratylus, is that words and the objects they refer to have an intrinsic relationship; that there is some essential core of meaning in the sounds we produce which refers only to certain objects in the world.

Socrates, the mouthpiece of Plato, seems to incline to this view though no final conclusion is reached. Later, the Greek philosopher Aristotle (384–322 BC) also agreed with this view in his easy called *De Interpretatione*. This is also known as the *essentialist* or *naturalistic* view.
The debate between the conventionalist and naturalistic views persisted throughout the medieval ages. In fact Ferdinand de Saussure, whom we shall discuss below, also contributed to it.

**LINGUISTICS IN THE MIDDLE AGES**

From the time of the Romans, linguists became more interested in grammar that anything else. Marcus Terentius Varro (127–116 BC), a Roman grammarian, wrote a book on Latin grammar called *De Lingua Latina*. The Romans were interested in the art of speaking forcefully and grammar was part of it.

This emphasis on speaking and writing (style) continued in Europe. In the medieval universities—Padua, Paris, Oxford, Cambridge etc.—grammar, dialectic and rhetoric were taught. As Latin was the language of scholars, it was considered superior. Thus when European scholars wrote grammars of modern European languages, such as that of English, they made these grammars like Latin grammar.

In South Asia too grammar became an artificial subject full of complicated rules and special terms. Modern Urdu grammars are even now full of Persian and Arabic terms which one has to memorize. Hindi grammars also have terms borrowed from Sanskrit which one has to memorize. These grammars do not describe what native speakers already know. They merely give artificial rules telling you to do this or not to do that. This is prescription and such grammars are called *prescriptive* grammars. This prescriptive tradition is part of schooling in the teaching of all languages, including English, in South Asia.

**COMPARATIVE LINGUISTICS**

In the nineteenth century linguistics progressed most in central Europe. European (especially German) universities were research oriented and their main object of research was comparing the words of one language with another. This method is also known as *philology* (*philos* means love and *logos* is word, i.e. love of words).
As the meanings of words change, these scholars had to study the origin of words (i.e. etymology) also. That is why they are known as historical and comparative linguists.

The main contribution of these philologists was the discovery that languages can be grouped in families. The languages of Europe and South Asia fall into the **Indo-European family** of languages. This was established by comparing words which were so similar that it seemed probable that they were derived from one word. This means that there was one parent language which gave birth to other languages as people moved away from their original homeland.

One of the main contributors to this kind of research was Sir William Jones (1746–1794). He was a judge in India and he made the Asiatic Society of Bengal. He established that Sanskrit is so similar to Latin and Greek that it could not be by chance. Instead, he argues, these three languages must have descended from one common ancestor.

Following this theory, the German linguist Franz Bopp wrote a book called *On the Conjugation System of the Sanskrit Language, in comparison with those of the Greek, Latin, Persian, and Germanic Languages* (1816). In 1818, Rasmus Rask, a Danish linguist, published his book *Investigation on the Origin of Old Norse or Icelandic Languages*. Later Jakob Grimm (of the fairy tales fame) also contributed to this field. Karl Brugmann (1897–1916) contributed his two volumes on the sounds and forms to the *Outline of Comparative Indo-European Grammar* in five volumes published between 1886 to 1893. This was an authoritative work and the idea of there being an Indo-European language became common.

The reasoning of comparative linguists can be understood by looking at Table 1 and 2. Both give words common between several European and Asian languages.
<table>
<thead>
<tr>
<th>Concept</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>father</td>
<td>plar pater pater pita pere father pidar peo</td>
</tr>
<tr>
<td>three</td>
<td>drae tres treis tri trios three sae tin</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Persian</th>
<th>Urdu/Hindi</th>
<th>Punjabi</th>
<th>Sindhi</th>
<th>Pashto</th>
<th>French</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>yak</td>
<td>ek</td>
<td>ik</td>
<td>hik</td>
<td>yau</td>
<td>un</td>
<td>vahid</td>
</tr>
<tr>
<td>2.</td>
<td>du</td>
<td>do</td>
<td>do</td>
<td>ba</td>
<td>dwa</td>
<td>deux</td>
<td>athnan</td>
</tr>
<tr>
<td>3.</td>
<td>sae</td>
<td>teen</td>
<td>tin</td>
<td>tre</td>
<td>drae</td>
<td>trois</td>
<td>salasa</td>
</tr>
<tr>
<td>4.</td>
<td>chahar</td>
<td>char</td>
<td>char</td>
<td>char</td>
<td>salor</td>
<td>quatre</td>
<td>arba</td>
</tr>
<tr>
<td>5.</td>
<td>punj</td>
<td>panch</td>
<td>punj</td>
<td>punj</td>
<td>pinzao</td>
<td>cinq</td>
<td>khamsa</td>
</tr>
<tr>
<td>6.</td>
<td>shish</td>
<td>che</td>
<td>che</td>
<td>cha</td>
<td>shpug</td>
<td>six</td>
<td>sitta</td>
</tr>
<tr>
<td>7.</td>
<td>huft</td>
<td>saat</td>
<td>sut</td>
<td>sut</td>
<td>owo</td>
<td>sept</td>
<td>saba</td>
</tr>
<tr>
<td>8.</td>
<td>husht</td>
<td>aath</td>
<td>uth</td>
<td>uth</td>
<td>ato</td>
<td>huit</td>
<td>samania</td>
</tr>
<tr>
<td>9.</td>
<td>noh</td>
<td>nao</td>
<td>no</td>
<td>nava</td>
<td>nava</td>
<td>neuf</td>
<td>tisa</td>
</tr>
<tr>
<td>10.</td>
<td>daeh</td>
<td>das</td>
<td>dah</td>
<td>dah</td>
<td>las</td>
<td>dix</td>
<td>ashara</td>
</tr>
</tbody>
</table>

In Table 2, notice that Arabic sounds for the same number are often completely different, whereas those in the other languages are similar. This is because Arabic belongs to a different language family, the Semitic family, whereas all the other languages mentioned above belong to the Indo-European language family.

August Schleicher (1821–1868), a German scholar, introduced the concept of a family tree theory. According to the family tree theory, languages of the same family descend from a common ancestor. This kind of tree was not constructed at once. Scholars spent years of painstaking research to provide evidence which made such constructions possible. The family tree in Figure 1 is for Proto Indo-European (PIE), i.e. our idea of what the original Indo-European language was.
Besides the Indo-European family of languages, there are many other families too. Table 3 gives the names of some major language families.

<table>
<thead>
<tr>
<th>Language Family</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semitic</td>
<td>Arabic, Hebrew, Amharic</td>
</tr>
<tr>
<td>Altaic</td>
<td>Turkish, Mongol</td>
</tr>
<tr>
<td>Sino-Tibetan</td>
<td>Mandarin Chinese, Wu Chinese of Shanghai, Languages of Tibet and Southeast Asia</td>
</tr>
<tr>
<td>Dravidian</td>
<td>Tamil, Telugu, Kannada, Malayalam</td>
</tr>
<tr>
<td>Bantu</td>
<td>African languages</td>
</tr>
</tbody>
</table>

Apart from these language families, there are other families too as mentioned above. Besides these there are some language isolates, i.e. languages which cannot be assigned to any family. One example of such an isolate is the Burushaski language spoken in Hunza, Yasin, and parts of Gilgit in Pakistan. For details of about 1000 languages listed according to their numbers of speakers, see Crystal (1988: pp.436-444).
Figure 1

The Family Tree for Proto Indo-European (PIE) Languages

PRIMITIVE INDO-EUROPEAN

- Germanic
  - Anglo-Frisian
    - English
    - Frisian
  - Dutch-German
    - Dutch
    - Afrikaans
    - German
    - Yiddish
  - Scandinavian
    - Breton
    - Welsh
    - Irish and Scottish
    - Gaelic
    - Oscan and Umbrian (dead languages)
    - Latin

- Celtic
  - Old forms: old Persian, Avestan and Pahlavi
    - Kurdish
    - Persian
    - Pashto
    - Baluchi
    - Tadzhiq
  - Old forms: Sanskrit and Pali
    - Hindi
    - Urdu
    - Bengali
    - Gujarati
    - Assamese
    - Sindhi
    - Oriya
    - Punjabi
    - Marathi
    - Dardic
    - Nepali
    - Sinhalese
    - Rajasthani
    - Bihari
    - Bhili
    - Pahari etc.

- Italic
  - Demotic
  - Katharevousa

- Indo-Iranian
  - Iranian

- Balto-Slavonic
  - Baltic
    - Lithuanian
    - Latvian
    - (West Slavonic)
    - (South Slavonic)
    - (East Slavonic)
  - Slavonic
    - Polish
    - Czech
    - Slovak
    - Serbo-Croatian
    - Slovene
    - Bulgarian and Macedonia
    - Russian
    - Ukrainian
    - Byelorussian

- Hellenic
  - Albanian
  - Armenian
  - Tocharish

- Balto-Slavonic
  - (West Slavonic)
  - (South Slavonic)
  - (East Slavonic)
Twentieth century linguistics

Modern linguistics was born mainly out of the work of Ferdinand de Saussure who has already been mentioned. He was born in Geneva in 1857 in a family of educated people some of whom were eminent in their lifetime. In 1876 while studying in Leipzig, Ferdinand de Saussure studied comparative linguistics which was fashionable in those days. Like other linguists, Saussure too started his career with historical-comparative linguistics in the philological tradition. His book was originally written in French and the English translation of the title is Memoir on the Primitive System of Vowels in the Indo-European Languages was published in 1879. However, Saussure was not satisfied with it. On 4 January 1894, he wrote in a letter to Antoine Meillet,

‘The utter inadequacy of current terminology, the need to reform it, and in order to do that, to demonstrate what sort of object language is, continually spoils my pleasure in philology.’ (Culler:1976: p.15).

Such a terminology he made himself, though he did not publish it in his lifetime. In 1891, Saussure was offered a chair (i.e. a professorship) in Indo-European linguistics in the University of Geneva. It was here that his students were sufficiently impressed by his lectures to take down notes. After his death in 1913, two of his students collected the notes and in 1916 published a book entitled Cours de Linguistique Generale (A Course in General Linguistics). This is the book which created modern linguistics by presenting the following theories.

(1) **Linguistic signs are arbitrary**

If you remember the conventionalist–naturalistic debate from Plato’s *Cratylus* mentioned earlier, you will have no difficulty in understanding Saussure’s theory. Briefly, Saussure agreed with the conventionalist view. He said that language was merely an arbitrary system in which any series of sounds could mean anything arbitrarily.
The term linguistic sign (or simply sign) means the sounds which refer to some object (material or non-material). For example, *dog* is a series of sounds. It is a linguistic sign. In English it refers to a four-footed animal of the canine family. Thus, it *signifies* an animal. The animal it signifies is called *signifie*, while the sounds, or words, can be called *signifiant*. (See Table 4).

<table>
<thead>
<tr>
<th>Language</th>
<th>Signifiant</th>
<th>Signifie</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>dog</td>
<td>An animal of the canine family</td>
</tr>
<tr>
<td>French</td>
<td><em>chien</em></td>
<td>”</td>
</tr>
<tr>
<td>Arabic</td>
<td><em>kulb</em></td>
<td>”</td>
</tr>
<tr>
<td>Urdu/Hindi</td>
<td><em>kutta</em></td>
<td>”</td>
</tr>
<tr>
<td>Persian</td>
<td><em>sug</em></td>
<td>”</td>
</tr>
<tr>
<td>Pashto</td>
<td><em>spai</em></td>
<td>”</td>
</tr>
</tbody>
</table>

We see that the sounds which signify (*signifiant*) are different but the thing which is signified (*signifie*) is the same. Thus, argued Saussure, the relationship between the linguistic sign (*signifiant*) and the objects of the world (*signifie*) is arbitrary. We could have chosen any series of sounds to mean anything. There is nothing in the nature of sounds themselves that makes them fit for particular things. You might object that there are words which sound like the thing they describe such as the noises of animals. These are called *onomatopoeic words*. But even here different languages have different sounds for noises. (See Table 5).

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barking of a Dog</td>
</tr>
<tr>
<td>Urdu/Hindi</td>
</tr>
<tr>
<td>English</td>
</tr>
<tr>
<td>bhau bhau</td>
</tr>
<tr>
<td>bow wow</td>
</tr>
</tbody>
</table>
The onomatopoeic words in Malay are as follows:

<table>
<thead>
<tr>
<th>Onomatopoeic Lexicon</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>debuk</td>
<td>The sound of slapping or boxing.</td>
</tr>
<tr>
<td>debum</td>
<td>The sound of heavy things falling.</td>
</tr>
<tr>
<td>debung</td>
<td>The sound of drumming.</td>
</tr>
<tr>
<td>debur</td>
<td>The sound of waves breaking on the shore.</td>
</tr>
<tr>
<td>debus</td>
<td>The sound of flapping wings.</td>
</tr>
</tbody>
</table>


Barring these partial exceptions, Saussure’s theory is correct. This theory is very important because it led to the study of linguistic signs in relation to other things and eventually to the relationships themselves. These relationships can be expressed in mathematical language, which makes modern linguistics very precise and scientific (for details see Culler 1976:p. 19–23).

(2) SYNCHRONY VERSUS DIACHRONY

Saussure distinguished between synchronic and diachronic studies of human language. *Chronos* means time in Latin; *syn* means at the same time while *dia* refers to the passing of time. Thus *synchronic linguistics* is the study of the language system at this time while *diachronic linguistics* is the study of this system in history.

In Saussure’s time diachronic linguistics was emphasized. Nowadays synchronic linguistics is in fashion. It is felt that if we want to know about how a language system actually operates or how it is stored in the human mind, we should study that system as it is. This has made people resort to descriptions of language systems as they are at present.

(3) SYNTAGMATIC VERSUS PARADIGMATIC

Consider the following sentence.

(1) He is like a horse
Here the words which follow each other in a linear direction (i.e. a straight horizontal line) are He + is + like + a + horse. This relationship between the words is called a **syntagmatic** relationship.

Now let us look at the same sentence and what we can put in place of the words in it:

1. (1a) He is like a horse
2. (2) She was unlike the cow
3. (3) I am like one monkey
4. (4) You are like a boy
5. (5) We are unlike most cats

You can see that *he* can be replaced by words of the same kind. Similarly other words can be replaced by words of the same kind. In the slot of the verb ‘to exist’ (*is*), for instance, we can only put in other forms of that verb (i.e. *am, was, are* etc). We cannot put in words of another kind. These slots represent places which can be filled in by words related to each other. This relationship is called the **paradigmatic** relationship.

Thus out of the **paradigm** of pronouns we can choose any: *he, she, it, I, you, we* etc. Similarly we can choose words for other slots also.

### (4) **Langue versus Parole**

As this concept has been explained earlier, there is no need to spend more time on it. Briefly, *la langue* is the system of a language as a whole while *parole* are individual utterances.

### Structuralism

Ferdinand de Saussure’s work influenced the development of structuralism. Briefly, language is a system. It is a structure with its various parts connected with each other. In Europe, a well known school of linguistics is called the Prague school. It was represented
by Nikolay Sergeyvich Trubetskoy (d. 1965), and it carried on the structuralist tradition in linguistics. He also lectured in Paris, and therefore influenced French linguists like Meillet.

In America, the anthropological linguists Franz Boas (d. 1942), Edward Sapir (d. 1939), and Benjamin Lee Whorf were structuralists. However, their work will be dealt with in the context of anthropological linguistics. The only American structuralist linguist who will be mentioned here is Leonard Bloomfield (d. 1949). In 1924, Bloomfield reviewed Saussure’s book which influenced the development of linguistics in America. Chomsky’s concepts of competence and performance are very close to Saussure’s distinction between *langue* and *parole*.

Nowadays the ideas of Saussure are being promoted by the *Cahiers Ferdinand de Saussure* and there is also an Institute Ferdinand de Saussure. Some fresh material by Ferdinand de Saussure has also been published and interest in his work has been revived (Sanders 2005: 931).

Structuralism not only dominated linguistics but also came to dominate sociology, anthropology, and philosophy. We will study its application by Claude Levi Strauss to anthropology later. Here it may be mentioned that Emile Durkheim and Marcel Mauss, both eminent sociologists, used the ideas of structuralism to study society. Just as language is a system in which there are binary oppositions, so society too was classified according to binary oppositions. Thus a group of people could be ‘bride-givers’ (those who give their daughters in marriage) while those who are given the brides may be called ‘bride-takers’. They, in turn, may give their own girls in marriage to another group. In short various aspects of a culture are connected with each other in such ways that they cannot be fully understood in isolation. Such were the insights which affected all the social sciences greatly as a consequence of the ideas of the linguists Wilhelm von Humboldt (1767–1835) and Ferdinand de Saussure.
Humboldt was a German linguist who worked on the Basque language which is spoken in Spain. However, he is better known for his theory that there is some correspondence or similarity between the character of the speakers of a language and the structure of the language. This theory leads to racist stereotyping and it has long been abandoned. However, his work in philology did influence other linguists and he is considered an important figure in nineteenth century linguistics.

Bloomfield was also strongly influenced by the work of the psychologist Wilhem Wandt who believed in behaviourism, i.e. the study of behaviour. This meant that reference to mental categories was eliminated or reduced to a minimum. Since the mind itself could not be observed, it was considered unscientific to talk about what went on inside it. Behaviour, on the other hand, could be observed. Thus human behaviour was the focus of the psychologist’s study. In linguistics, this meant studying speech as if it was the response to some stimulus. Stimulus means what we put in and response means what comes out or what happens. For instance, if we apply the stimulus of extreme heat to an animal’s body, the response of the animal will be to move its body away so as to escape the pain. Bloomfield felt that meaning too could be seen as response to some kind of stimulus. Bloomfield’s book *Language* (1933) dominated American linguistics almost till the time of Chomsky. In this book, however, we shall mostly be concerned with Chomskian linguistics. However, let us first understand some of the basic concepts of theoretical linguistics. For this we turn to phonetics in the next chapter.
The study of sounds in human speech is called **phonetics**. It has the following branches:

1. Articulatory phonetics
2. Acoustic phonetics
3. Auditory phonetics

**Articulation** means producing sounds from the mouth. Thus, this branch of phonetics deals with the production of sounds, their description, and their transcription (i.e. writing them with the help of symbols).

**Acoustic** phonetics is the study of the physical properties of the sounds themselves. There are machines which give bands of frequencies for sounds. Other features of sounds can also be studied.

**Auditory** phonetics is the study of the reception of the sound by the ear and its processing in the brain.

In other words we study (1) how sounds are produced, (2) how they travel from the mouth to the ear, and (3) how they are perceived by the ear and the brain. In this book we shall only study **articulatory phonetics**.

**THE PRODUCTION OF SPEECH SOUNDS**

Let us first look at the mouth where speech sounds are produced. Diagram 1 will help you locate the main **articulators**, i.e. organs which help us in producing sounds.
Sounds are made by the air striking against the articulators. This air comes in the form of a stream from the lungs. This is known as the pulmonary stream. It is also known as the egressive stream, i.e. that stream of air which goes out of the mouth.

Sounds can also be produced by taking in air into the mouth. This is called the ingressive air stream. There are sounds in Sindhi and Siraiki which are produced by the ingressive air stream.

(1) Larynx

The larynx is a box-like space which allows the air from the lungs to pass through it.

(2) Vocal Cords

In the beginning of the larynx are vocal folds, or vocal cords. They are two bands of muscles lying opposite to each other (see Diagram 2).
(3) PHARYNX

It is a tube which begins above the larynx (i.e. the passage which goes down to the lungs). It is approximately 7cm long in women and 8cm in men. At the top it divides into two. One part goes into the nose and the other in the back of the mouth.

Diagram 2

When the air passes through, the vocal cords can open and close very fast. This is called vibration and it can be from 70 to 1000 times per second. This vibration is called voicing. If you say /z/ and keep your tongue in the same position saying zzzzz you will hear a buzzing sound. This is the vibration of the vocal cords. You can place your fingers on your throat and feel this vibration. If you want to switch it off say sssss and it will cease. Now you know that /z/ is a voiced sound and /s/ is an unvoiced sound. Other sounds are also distinguished like this.

(4) THE NASAL CAVITY

Let us look at the space in the nose. It is called the nasal cavity. It has air in it which can produce a kind of humming noise called resonance.

The nasal cavity can be made larger if the soft palate is lowered. As the soft palate is a movable bit of flesh it can divide or connect
the nasal and the oral cavities. If it is lowered the volume of air which will vibrate will be more. Now if the mouth is blocked and all this air passes out of the nose we get what is called a nasal effect. This is known as nasalization.

If your nose is blocked at the back of the nasal cavity, you will not be able to nasalize your sounds. If, however, your soft palate is permanently lowered, nasalization will increase.

(5) The oral cavity

Most of the articulators are in the cavity of the mouth or the oral cavity. Let us discuss them one by one.

(a) The lips

The lips can close together stopping the air. This action creates sounds which are called stops or plosives. They are called plosives because the air is stopped for a very short time and then goes out or in, i.e. explodes or implodes. Two such sounds are /p/ and /b/. Say them and feel how your lips close together. Speakers of Urdu–Hindi do not use implosive sounds in their languages. They are produced by sucking the air inwards. Try to say /p/ and pull the air in. You will produce an implosive sound. If you cannot do it, request a Sindhi friend to say /b/ as used for a goat in that language. You will hear an implosive /b/. The lips can also be rounded or brought close together without meeting each other. These actions too produce sounds which we will describe later.

(b) The teeth

The tongue can be behind the teeth or between them for certain sounds. The lower lip meets the upper teeth if we produce the sounds /f/ and /v/. 
(c) **Alveolar ridge**

Just behind the upper teeth is a slightly raised longish ridge called the alveolar ridge. Some sounds are produced when the tongue touches or comes near this ridge.

(d) **Hard palate**

It is the hard sloping place behind the alveolar ridge. If you take your tongue back from the alveolar ridge you can feel it.

(e) **Soft palate**

It is a soft fleshy place behind the hard palate. It is also called the velum and sounds produced by the tongue touching it or coming near it are called velar sounds.

(f) **Uvula**

It is a triangular piece of flesh hanging down from the roof of the mouth. You can see it in a mirror if you open your mouth keeping your tongue down. Some sounds are produced when it comes in contact with the tongue.

(g) **The tongue**

The tongue can move up and down and turn in the mouth. By doing so, it comes in contact with other parts of the mouth and helps in producing sounds. The tongue is divided in the following parts (see Diagram 3):

- (i) tip
- (ii) blade
- (iii) front
- (iv) middle
- (v) back
- (vi) root

Remember that the front of the tongue is not its tip. Behind the tip (or apex) is the blade and what is called the front is behind that. In fact, the front is almost towards the middle of the tongue. When the tongue is at rest the tip touches the teeth, the blade lies below the alveolar ridge, the front is below the hard palate, and the back is below the soft palate. The root joins the tongue with the back of the mouth.
Some books use the following adjectives:

- **apical**: to do with the tip of the tongue.
- **dorsal**: to do with the dorsem, i.e. front and back of the tongue.
- **lingual**: to do with the tongue (lingua is Latin for tongue).
- **laminal**: to do with the blade of the tongue.

**Diagram 3**

There are differences in terminology in linguistics and you are advised to read dictionaries of linguistics if you come across unfamiliar terms.

**(h) The glottis**

This is the place between the vocal cords. All glottal sounds—sounds coming deep from the throat—come from it.

**The classification of sounds**

We classify sounds by describing where they are produced in the mouth and how they are produced. The first is known as the **place of articulation** and the second the **manner of articulation**. The place of articulation may be the lips, the lips and the teeth, and so on. The manner of articulation means how the air stream comes out of the mouth. It can be stopped and then released, it can be released with friction, the lips can be rounded or opened, the vocal cords can vibrate or not, and a number of other things can happen.
Let us start from some of the commonly used sounds of English and South Asia languages which are articulated by the lips and go on to those which come out of the glottis. These sounds will be written in the symbols of the **International Phonetic Association (IPA)**.

The IPA was founded in Paris in 1886 by school teachers who wanted a notation for the pronunciation of foreign languages in their schools. It was based on the Roman alphabet because the people who made it were Europeans and this was the alphabet they used. However, since they wanted exactly one symbol to represent one sound they put marks (diacritical marks) on the symbols and brought some from the Greek alphabet to represent the sounds of many languages. Academics started working on the IPA with a view to understanding pronunciation in general as well as the pronunciation of individual languages.

One prominent scholar who has influenced British phonetics is Henry Sweet (1845–1912). Sweet was a don (faculty members at the universities of Oxford and Cambridge were called dons) at Oxford from 1901 to 1912. He was interested in spelling reform, old English, as well as phonetics. His *History of English Sounds* (1874) as well as *Handbook of Phonetics* (1877) influenced a whole generation.

Before we start a discussion on phonetics, let us review a few definitions. A sound is known as *phonema* in Greek. From this comes the technical term *phoneme* which means the smallest unit of sound of a language. Actually it is an abstract concept because we cannot pronounce most pure phonemes. If we try to pronounce a pure consonant we cannot do so. We have to add a vowel either before or after it. However, we have to talk of phonemes, units of sounds, in order to describe them.

Now let us define types of sounds. A **consonant** is a phoneme which is produced with some sort of obstruction in the air stream. A **vowel**, on the other hand, is produced without obstruction in the vocal tract.
CONSONANTS

Now let us classify the consonants of English and some South Asian languages:

Bilabials

Bilabial stops Labia means lips. /p/ and /b/ are sounds made by the lips. The place of articulation is the lips. The manner of articulation is that the lips are brought together, the air stream is stopped and then released with an explosion. Thus, /p/ and /b/ are called stops or plosives.

The vocal cords do not vibrate in /p/ but they do vibrate in /b/. Thus /p/ is a voiceless bilabial stop and /b/ is a voiced bilabial stop. If the air stream is taken in the stops are known as ingressive stops and written as follows: /łożyć/ and /ɓ/.

Bilabial nasal /m/ is also produced by the meeting of the two lips, but in this case the velum is lowered and the air comes out of the nose. The manner of articulation is nasal and /m/ is a bilabial nasal.

Semi vowel /w/ is produced by the lips rounding together. The manner in which it is articulated is very similar to vowels. The air comes from the lungs and flows out with very little obstruction except that which is offered by the rounded lips. That is why this sound is known as a semi-vowel or approximant.
Labio-dentals

Labio-dental fricatives  /f/ and /v/ are articulated when the upper teeth come in contact with the lower lip. The manner in which they are articulated is that the air keeps flowing out with friction. That is why they are known as fricatives: /f/ is unvoiced and /v/ is voiced.

Dentals

Inter-dental  Written as /θ/ and /ð/, they are fricatives articulated by the tip of the tongue coming between the teeth. They are fricatives so the air flows out with friction. /θ/ is unvoiced and /ð/ is voiced. These phonemes exist in English and Arabic, but not in Urdu, Hindi, Punjabi, Sindhi, Pashto, Bengali, Gujarati etc. They do, however, occur in Eastern Balochi. Thus most South Asian speakers of English use /θ/ for /θ/ as in think and /ð/ for /ð/ as in then.

Dental stops  When the blade of the tongue touches the back of the upper teeth, we produce dental phonemes. The air stops and is released. Hence these are dental stops or plosives. To show that they are dental, a dental diacritical mark is put below them. /t̪/ is unvoiced and dental /d̪/ is voiced. These are the ت and the د in the Perso-Arabic (Urdu) script.

They do not exist in English but do exist in French, Spanish, and Italian. They are very common in South Asian languages;
e.g. *teen* in Urdu means three, and *deen* means religion.

### Alveolars

**Alveolar stops**

/\(t\)/ and /\(d\)/ are articulated by the blade or the tongue coming in contact with the alveolar ridge. As they are stops, the air stream stops before being released. /\(t\)/ is unvoiced and /\(d\)/ is voiced.

**Retroflex**

The place of articulation remains the same but the manner is different. For retroflex sounds, the tip of the tongue goes up and curls back. These sounds are common in South Asian languages and distinguish speakers from Pakistan, India, Bangladesh, Sri Lanka, and Nepal from people of other countries. They are written in the IPA as /\(ʈ\)/ and /\(ɖ\)/ but for the sake of convenience we have represented them with a dot beneath them as /\(ṭ\)/ and /\(ḍ\)/.

It is interesting to observe that British English does not have retroflex phonemes so that their alveolar /\(t\)/ and /\(d\)/ are not the same as our retroflex equivalents. We do, however, pronounce *ten* and *den* with retroflex stops. They, in turn, pronounce our words beginning with retroflex stops in their own way, i.e. with alveolar ones.

The difference is most obvious when the English, the Arabs or the Iranians try to pronounce words with retroflex /\(ɽ\)/. They produce /\(r\)/ and not /\(ɽ\)/.

**Alveolar fricatives**

Written as /\(s\)/ and /\(z\)/, these are articulated by bringing the blade of the tongue close to the alveolar ridge. Being fricatives, the
air goes out with friction. /s/ is unvoiced and /z/ is voiced.

Alveolar nasal
Written as /n/, it is articulated by bringing the blade of the tongue close to the alveolar ridge. As the mouth is closed and the velum is lowered, the air is forced to go out from the nose. This makes it a nasal sound. It is voiced.

Lateral
Written as /l/, it is articulated by putting the blade of the tongue on the alveolar ridge. It is called a lateral because the tongue is pushed in a lateral direction, i.e. the left side of the mouth. You can verify this by saying ‘llll’ and sucking in air. Only the right side of the tongue will be cooled by the incoming air proving that the left side is close to the mouth.

Palato-alveolar
Fricatives
The middle of the tongue makes a hump which comes near the hard palate only slightly behind the alveolar ridge. The air goes out with great friction and a strident sound of sh is produced. This is the unvoiced version of the phoneme which is written as /ʃ/ while the voiced version is written as /ʒ/. The first is used in she while the second is in vision.

Palatal
Behind the alveolar ridge is the hard palate. Many sounds are produced with the tongue coming in contact with it. However, the one which is significant because it is used in English as well as South Asian language is /j/.
Palatal semi vowel or approximant

Written as /j/ in the IPA, this phoneme occurs in the beginning of *yes*. Remember it is not the first phoneme of *judge* which is written as /ʤ/. 

**Velar**

Velar stops or plosives

Behind the hard palate is the soft palate called velum. The sounds produced with the tongue coming in contact with (or very near) it are called velar sounds. The ones we need to know are /k/ and /g/. The tongue meets the velum and the air is stopped before it is released. /k/ is unvoiced and /g/ is voiced.

Velar nasal

The only velar nasal sound is *ng* written as /ŋ/. It is the sound produced by *ng* in *finger* and *singer* etc. In English and South Asian languages, it occurs only in the middle or end of words. In some other languages it occurs in the beginning of words.

Velar fricatives

The tongue comes near the velum and the air keeps flowing out with friction. These sounds do not occur in modern English but they are used in Urdu, Pashto, Persian, Arabic, and many other languages.

The unvoiced velar fricative is written as /x/ and is the first phoneme of the word *khan*. In Hindi it takes the form of /kh/ though some people do use /x/.

The voiced one, written as /ɣ/ is the phoneme depicted by *gh* in Afghan. If you want to hear the exact sounds of these symbols ask a speaker of one of the above mentioned languages to pronounce these words.
Uvulars

These sounds are produced with the back of the tongue coming in contact with the uvula. There are no uvulars in English but they do exist in Persian and Arabic.

Uvular stops

The sound produced by stopping the air with the back of the tongue coming in contact with the uvula is a uvular stop. The unvoiced one is written as /q/ and the voiced one as /ɢ/.

/q/ is used in Persianised Urdu which was spoken in parts of UP, such as Lucknow and Delhi. In Pakistani Urdu /k/ is used in its place. In Hindi too it is not an indigenous phoneme but it is used by some Hindi-speakers.

Pharyngeals

Fricatives

Sounds produced in the pharynx. The only one which you may come across is the fricative written as /ɦ/. It is the Arabic pronunciation of <ﺡ>. You may hear Arabic-speakers (or those who know the Arabic pronunciation) use this sound when they say Rahman. It sounds almost like Rakhman but it is actually /ħ/. The voiced pharyngeal fricative is written as /ʕ/. The sound of <ع> in Arabic is represented by it. The air is released with friction deep in the throat and the vocal cords vibrate. South Asians generally use a vowel instead of this sound which is not part of their sound inventory.

Glottals

The most commonly used glottal fricative is /h/ which is described as a voiceless
The study of sounds

glottal central fricative. It is actually a voiceless vowel with some friction in the glottis. There is a voiced version also written as /ɦ/. There is also a glottal stop written as /ʔ/. The air stops in the glottis when this is produced, and is released with an explosion.

The information given above can be summarized in the chart of consonantal sounds given in Figure 2. For more details about consonants read other books on phonetics such as O’Connor (1973), Ladefoged (1982), and Roach (1983).

**Figure 2**

| THE INTERNATIONAL PHONETIC ALPHABET |
| CONSONANTAL PHONEMES |
| PLACE OF ARTICULATION |
|---|---|---|---|---|---|---|---|---|---|---|---|
| MANNER | Bilabial | Labiodental | Dental | Alveolar | Post-alveolar | Retroflex | Palatal | Velar | Uvular | Pharyngeal | Glottal |
| Plosive | p b | t d | t d | c j | k g | q g | ʔ |
| Nasal | m ɱ | n | ɲ | ɳ | ŋ | N |
| Trill | B | r | | R |
| Tap or Flap | r | ɾ |
| Fricative | φ β | f v | θ ð | s z | s z | ç j | x y | χ ψ | h f | h h |
| Lateral fricative | | | | t ɾ | ɾ e | ɾ ɾ | ɾ ɾ | ɾ ɾ |
| Approximant | ʋ ɹ | ɻ j | ɰ |
| Lateral approximant | | | | ɾ | ɾ | ɾ | ɾ | ɾ |
| Ejective stop | p’ | t’ | t’ | c’ | k’ | q’ |
| Implosive | β ɭ | t ɭ | ɭ e | ɭ e | ɭ e | ɭ e |

Where symbols appear in pairs, the one to the right represents a voiced consonant.
VOWELS

As in the case of the consonants, vowels too will have to be classified with reference to the position of articulators in the mouth. Vowels are sounds in which the air is not stopped in the mouth. Thus it does not stop and explode (as in the case of fricatives). The tongue does move in the mouth while producing vowels, and the lips expand and contract. Vowels are determined by the tongue and the lips. The following three dimensions are used in the classification of vowels:

1. Height
The tongue can be made to go up or down in the mouth. If it is up, i.e. near the roof of the mouth, the vowel will be a high vowel. If it is low, i.e. near the lower jaw, the vowel will be low. If it is in the middle, the vowel will be a mid vowel. This is also called closeness. When the tongue is close to the roof of the mouth, i.e. high, the vowel is a closed vowel.

2. Frontness
The tongue forms an arch or hump when a vowel is pronounced. If this arch is made in the front of the mouth, the vowel is called a front vowel; if in the middle, it is a central vowel; and if in the back, it is a back vowel.

3. Roundness
If the lips are rounded, the vowel is a round vowel. If they are in the neutral position, it is neutral, and if they are spread, the vowel is a spread vowel.

Thus, the point of articulation of a vowel is the highest point on the arch of the tongue. The position of this point on the height and frontness dimension gives us the description of the vowel. To this we add information about lip rounding to get a complete classification of the vowel. The following chart summarizes this information:
The tongue can move in the space in the mouth forming an arch in front, in the centre, and at the back. It can either be low or in the middle or high in the mouth. Let us see some vowels and classify them according to where the arch is formed.

/ɪ/ While saying ee as in beet or seat (written as /ɪ/) the tongue is up near the roof the mouth. The arch is made in the front of the mouth. Thus it is a high front vowel.

/ʊ/ When you say oo as in boot or suit (written as /ʊ/) the tongue is high but the arch is made in the back of the mouth. Thus, it is a high back vowel.

/ɑ/ When you say /ɑ/ as in father the arch is still made at the back of the mouth but now the tongue is down in the mouth. This is a low back vowel.

You may not be able to feel the movement of the tongue in the beginning but if you say /ɪ/ and then /æ/ you will feel your lower jaw moving down. This means your tongue is going down in the mouth.

While saying /ʊ/ or /ɑ/ you will notice the tongue contracting. This will be because it will be arched at the back of the mouth.

Produce these sounds while noticing your tongue and its movement, and you will understand the meaning of height and frontness.

**Cardinal vowels**

The space of the mouth has an irregular shape but, for the sake of simplicity, let us imagine it to be a rectangle as given below. Let
us also imagine the position the tongue can occupy and where the arch can be made. There are, of course, many such positions but the extreme ones are as follows. We call them the cardinal positions and the vowels produced there are called the cardinal vowels.

Diagram 4

The first eight cardinal vowels which you will generally come across are given below:

1. /i/ high front (ee as in beet, seat, keep)
2. /e/ half high front (as in Urdu-Hindi /pet/, i.e. stomach)
3. /ɛ/ half low front (The English vowel in egg is close to it)
4. /a/ low front unrounded vowel (/æ/as in man is used in its place)
5. /ɑ/ low back unrounded vowel as in arms.
6. /ɔ/ half low back (as in British English saw /sɔː/)
7. /o/ half high back (as in Hindi–Urdu əʊ i.e. people /log/ or Scottish /bot/ for boat).
8. /u/ high back (oo as in boot, loot, root etc. are close to it. The German long /u/ is even closer).
These are the first eight cardinal vowels. The other eight are produced by changing the shape of the lips while everything else remains the same. Thus, if we round the lips for the first four cardinal vowels we get /Y/, /ø/, /œ/, /œ/. If we open the lips for cardinal 5 /a/ we get cardinal 13 which is /ɒ/ as in dog. If we open the lips (unround them) for the other three back vowels we get /ʌ/, /ɤ/1, and /ɯ/ which are cardinals 14, 15, and 16. Diagram 4 illustrates this clearly.

The cardinal vowels are extreme positions used to measure other vowels. They do not exist—or at least all of them do not—in their pure forms in most languages but the vowels which are like them are given the symbols designed for them. The vowel sounds in the brackets are like cardinal vowels. However, they are actually somewhat different than the sounds we use.

The vowels used in English and most other languages used in Pakistan are given below. You will notice that they are not all cardinal vowels or even near them. Thus new symbols are required for them. However, linguists often use the same symbols as they use for cardinal vowels. This can be a source of confusion so you should note the similarities and the differences.

/ı/ Front high vowel but lower than cardinal vowel 1. As in hit, did, bid etc. Note that the /ı/ does not have a dot on it.

/e/ As in bed. Note that we use this symbol for cardinal vowel 2 though this vowel is between cardinal vowels 2 and 3. Some people use the symbol /ɛ/ for it.

/æ/ Front low. It is slightly higher than cardinal vowel 4. As in sad, bad.

/ʌ/ a central back unrounded vowel near cardinal vowel 14 as in English cup, but.

/ɑ/ A short low back rounded vowel as in dog, hot etc. Cardinal vowel 13.

1 Remember that the symbol of the uvular voiced fricative is a bigger /ɣ/ than that for cardinal vowel 15 which is a small /ɣ/
/ǔ/ A short semi-high back rounded vowel between cardinal vowel 7 and 8. As in *put, look* etc.

/ɔ/ Mid central unrounded vowel as in English again. It is sometimes undistinguishable in speech from /ʌ/. It is called *schwa*. For details see Pullum and Ladusaw (1986: 44).

**Diphthongs**

Diphthongs may be defined as combinations of two vowels. The tongue shifts from pronouncing one to the other and in representing them we combine the symbols of the two vowels. The following diphthongs are used in English:

/ıǝ/ as in *beard*

/eǝ/ as in *scarce*

/ʊǝ/ as in *tour*

/eı/ as in *pain, paid*

/aı/ as in *tide, time*

/ɔı/ as in *voice, void*

/ʊo/ as in *home, road*

/aʊ/ as in *gown, house*

It should be noted that for /eı/ South Asian speakers generally use cardinal vowel 2 /e/. For example, in Urdu-Hindi we say /peːn/ and not /pein/ for *pain* (Remember /e/ is lengthened here. It is the IPA cardinal vowel 2, not the English vowel /e/).

Similarly we use /o/ for /ʊo/ Thus we say /hɒm/ and /lɔd/ for /hʊm/ and /lɔd/ when we pronounce the English words *home* and *load*.

**Transcription**

Now that we have learned the symbols for both vowels and consonants, we can write down the sounds of any language. This is called transcription. Transcription is of two kinds:
(1) Broad or phonemic   (2) Narrow or phonetic

In **broad transcription** we just write down the main phoneme generally used by speakers. We do not bother about its varieties or the minor changes the speaker may have made when pronouncing it. This kind of transcription is in slashes (/ /).

In **narrow transcription** the exact sounds are produced. This transcription is put in square brackets, i.e. [ ].

Let us see both kinds of transcriptions of the words *pen*, *ten*, and *car*.

**Broad**

/pen/
/ten/
/kar/

Here the actual way in which British or Pakistani/Indian speakers pronounce these words is not given, just the phonemes in them are represented.

**Narrow**

*South Asians*   *British*

[pen]     [pʰen]
[ten]     [tʰen]
[kar]     [kʰa]

Notice that the South Asian speakers use the retroflex /ʈ/ while the British use alveolar /t/. Similarly, we use /p/, /k/, and /t/ instead /pʰ/, /kʰ/, and /tʰ/.

Narrow transcription is more difficult than broad transcription because it is not easy to catch and represent every little difference a speaker may give to his or her production of sounds. Anthropological linguists and sociolinguists should learn to transcribe sounds as they are spoken.
The Study of Sound Patterns

The study of the arrangement or the combination of sounds in a language is called phonology. As such an arrangement may be called a pattern, phonology may be defined as the study of the sound-pattern of a language.

Distinctive features

To study phonology we shall have to study sounds in a little more detail. We defined phonemes as the units of sounds earlier but even phonemes can be seen in terms of certain oppositions. For instance, they may be voiced (vocal cords will vibrate) or unvoiced (vocal cords will not vibrate); they may be consonantal (air will be obstructed) or vocalic (air will not be obstructed); and so on. These oppositions are called binary oppositions. Either a characteristic is there or it is not there. When it is there we may use the sign of +, and when it is not there we may use the sign of −. Thus we say that a phoneme may be either + voice (i.e. voiced) or − voice (unvoiced) and so on.

These features or characteristics of phonemes distinguish them from each other. Thus they are called distinctive features.

According to Roman Jokobson and Morris Halle, two great linguists, there are ‘twelve basic oppositions, out of which each language makes its own selection’ (Jakobson and Halle 1956: 40). Let us have a look at some of the major binary oppositions in languages.
1. **Consonantal/non-consonantal [+ cons]**
Consonants, or consonantal phonemes, are produced with the air stream being obstructed in the mouth.

2. **Vocalic/non-vocalic [+ voc]**
Vocalics are produced without the air stream being obstructed in the mouth. However, there are some sounds which are not pure vowels but have some of the characteristics of vowels.

3. **Voiced/unvoiced [+ voice]**
In voiced sounds the vocal cords vibrate, in unvoiced ones they do not.

4. **Anterior/posterior [+ ant]**
Anterior sounds are produced in the front part of the mouth, i.e. at or in front of the alveolar ridge. Posterior sounds are produced at the back of the mouth.

5. **Coronal/non-coronal [+ cor]**
Coronal sounds are produced by raising the blade of the tongue towards the teeth or the hard palate. Non-coronals are produced without this action.

6. **Nasal/oral [+ nas]**
Nasal sounds are produced by the air passing through the nose while the lips are closed. In oral sounds the air passes through the mouth.

7. **Continuant/stop [+ cont]**
Sounds in which the air stream continues to flow are called continuants. In stops (or plosives) the air stops and then explodes. In continuants like /f/ or /v/ the air goes out with friction, because it is impeded but it is never stopped as in stops like /p/ or /b/.
8. **Strident/non-strident** [+ strid]

In strident sounds the air strikes two surfaces and a lot of noise is produced. In non-strident sounds the noise is less. Some of the stridents are /s/, /z/, /f/, and /v/ which produce high-intensity noise or friction.

9. **Low/non-low** [+ low]

In low sounds the tongue is low—right at the bottom—of the mouth. The tongue is in this position in all low vowels and in pharyngeals. All other sounds are non-low sounds.

10. **Tense/lax** [+ tense]

Only vowels are classified as tense or lax. It is said that in tense vowels the muscles of the tongue are more tense. The body of the tongue or its root gives the air less space to flow than in lax vowels. The root of the tongue may be drawn forward and its body may be raised. This is called **advanced tongue root** [+ ATR]. Some linguists consider +ATR same as + tense. It should be noted that tense vowels are longer than lax ones. For instance, /i/ is long and /ɪ/ is short so /i/ is + tense and /ɪ/ is – tense.

11. **High/non-high** [+ high]

The body of the tongue is high in the mouth (i.e. towards the palate) for high sounds. For non-high ones it is not.

12. **Back/non-back** [+ back]

When the hump made by the tongue is in the back part of the mouth, the sounds produced are called + back sounds. When it is not, we produce other sounds, i.e. – back sounds.

Now we have the means of classifying all the sounds of all languages through these distinctive features. The +/– or binary oppositions give us features which will allow us to understand the rules according to which sounds combine in different languages. Let us first, however, classify some of the sounds of English here. This
classification (Figure 3) is different from the one used by Jacobson and Halle (1956).

**FIGURE 3**

**CHART OF DISTINCTIVE FEATURES**

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>b</th>
<th>f</th>
<th>v</th>
<th>m</th>
<th>t</th>
<th>d</th>
<th>θ</th>
<th>s</th>
<th>z</th>
<th>n</th>
<th>tj</th>
<th>ɗ</th>
<th>j</th>
<th>ʒ</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Voc</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>– –</td>
<td>– –</td>
<td>– –</td>
<td>– –</td>
<td></td>
</tr>
<tr>
<td>Cont</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Cor</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Voice</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Nasal</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Strid</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
<tr>
<td>Ant</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td>+ +</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>g</th>
<th>l</th>
<th>r</th>
<th>w</th>
<th>j</th>
<th>h</th>
<th>η</th>
<th>x</th>
<th>y</th>
<th>q</th>
<th>ʕ</th>
<th>ħ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Voc</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cont</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Cor</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Voice</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Nasal</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Strid</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ant</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Lat</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Some linguists use other features for classification. For instance, they contrast sounds like /s/ with /z/ on the basis of the pressure exerted by the lungs. For /s/ the lungs push harder and the sound produced is sharper and stronger. It is called a **fortis** sound. In contrast /z/ is a **lenis** sound. Yet another way is to classify sounds is according to the part of the tongue which is used to produce them, i.e. apical (tip of the tongue); dorsal (tongue-back articulation). We may also classify them as being labial (produced by the lips or non-labial [±labial]).
Let us now classify some sounds with reference to these classificatory criteria:

\[
\begin{array}{llllllllll}
\text{p} & \text{b} & \text{m} & \text{t} & \text{d} & \text{n} & \text{k} & \text{g} & \text{g} \\
\text{stop} & + & + & + & + & + & + & + & + \\
\text{nasal} & - & - & + & - & - & + & - & - & + \\
\text{fortis} & + & - & 0 & + & - & 0 & + & - & 0 \\
\text{lenis} & - & + & 0 & - & + & 0 & - & + & 0 \\
\text{labial} & + & + & + & - & - & - & - & - & - \\
\text{apical} & - & - & - & + & + & + & - & - & - \\
\text{dorsal} & - & - & - & - & - & - & + & + & + \\
\end{array}
\]

*Source:* O’Connor 1973: 204

With the help of these distinctive features we can describe sounds without even using IPA symbols. Even more significant is the fact that we can write down laws or rules about the combination of sounds in a language or its sound pattern. These laws or rules are called **phonological rules.**

**PHONOLOGICAL RULES**

Let us now write down some phonological rules. They are written in the form of certain conventional symbols which are similar to those used in mathematics. You must learn the following symbols:

\[\rightarrow\] becomes

\[\emptyset\] nothing (or zero)

\[#\] word boundary

\[\#-\] word-initial

\[-#\] word-final

\[/\] in the environment of

\[\text{[]}\] Square brackets enclose the distinctive features of a phoneme.

\[\_\] in this place.
Most of the symbols are self evident but the symbols for word boundary and ‘in the environment of’ need explanation.

Words end and begin with phonemes. Both these phonemes are boundaries. When we say word-initial we mean that which comes before the first phoneme of a word begins. When we say word-final we mean after the last phoneme of the word ends. Consider the word below:

\[
d i g
\]

The phoneme /d/ is in the word-initial position and /g/ is in the word-final position. All that comes before /d/ is word-initial and is written as #-. All that comes after /g/, i.e. word-finally is written as –# (i.e. at the boundary after the last phoneme ends).

\[
# dıg#
\]

Now look at the symbol of ‘in the environment of’ which is /\/. It means in such and such conditions; when the following given conditions exist; when these sounds exist in this given pattern or occur in this manner. Then the sounds themselves are written in the sequence in which they occur.

Let us now write down a rule. First let us have an observation (an event, a phenomenon) which the rule will explain.

Observation: Speakers of Urdu-Hindi put an /ı/ before words like school, stool, spool, speak, street etc.

<table>
<thead>
<tr>
<th>British English</th>
<th>Urdu-Hindi Speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>/skul/</td>
<td>/ıskul/</td>
</tr>
<tr>
<td>/stul/</td>
<td>/ıstul/</td>
</tr>
<tr>
<td>/spul/</td>
<td>/ıspul/</td>
</tr>
<tr>
<td>/spik/</td>
<td>/ıspik/</td>
</tr>
<tr>
<td>/strit/</td>
<td>/ıstrit/</td>
</tr>
</tbody>
</table>

Objective: To make a phonological rule to express the above, i.e. to find out what rule the speakers of Urdu-Hindi have in their
minds which makes them put an /ı/ before the words given above.

*Method:* Notice that the first phoneme is the same in all the words, i.e. /sl/. The second is different, i.e. /kl/, /lt/, /lp/ etc. let us write down the distinctive features of all of them:

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>k</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Voc</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Voice</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Ant</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cor</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Cont</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Nasal</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Strid</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

We should now look for other examples of phonemes of the same kind and write them down in our list.

You will observe that the only relevant difference is that the first phoneme is *always* cont+ while the second one is cont− (i.e. a stop).

This means that Urdu-Hindi speakers add /ı/ in the word-initial position to all words where the consonantal cluster begins with a + cont while the second phoneme is a −cont.

Let us put it down through conventional symbols:

Ø → /ı/

This means that ‘nothing becomes /ı/’. In other words the speaker puts /ı/ or pronounces /ı/ out of nothing. There was no /ı/ but the speaker pronounces it or inserts it out of nowhere.

The environment is that after the words begin (# −) there is a cont followed by − cont consonantal sound. In such conditions the
speaker will pronounce an /ı/ before the word begins. The exact place where this /ı/ will appear is given by the dash (–) sign.

The rule is given as follows:

\[ \emptyset \rightarrow /ı/ \# -[+\text{cont}][-\text{cont}] \]

But now take the pronunciation of the word sphere /s ı ə r/. Hindi-Urdu speakers pronounce it as /ı s ı ə r/. In other words our rule does not cover all the possible cases. We need to refine it further. This refinement is as under.

\[ \emptyset \rightarrow /ı/ \# – [+\text{cons}] [+\text{cons}] \]

This means that a vowel is inserted by the speakers of Urdu and Hindi before consonantal clusters if such clusters occur in the beginning of words.

In the same words Punjabi speakers put the vowel /ə/ between the first and the second phoneme of such a consonantal cluster. You can work out the rule for these speakers, and for details of such rules in Pakistani and Indian English, see Rahman (1990: 31–32).

**Allophones**

Besides the sounds given in the chart, most South Asian languages also have aspirates. These are /bh/, /ph/, /dh/, /th/, /dh/, /th/, /kh/, and /gh/ etc. they are used in Urdu-Hindi words like bhalu (bear), phool (flower) etc.

In English some phonemes such as /p/, /k/, and /t/ are aspirated in the beginning of words but the aspiration in this case is less than in the case of Hindi-Urdu. Another important point is that in English the meaning remains the same even if you do not aspirate the phonemes e.g. king has the same meaning whether you pronounce it as /kɪŋ/ or /kʰɪŋ/. In Urdu-Hindi, Punjabi, and Sindhi the meaning changes according to aspiration, e.g. /pul/ means a bridge in Punjabi but /phul/ means a flower.
In English /k/ and /kʰ/, and /p/ and /pʰ/ are allophones. In Hindi-Urdu /k/ and /kh/ or /p/ and /ph/ are different phonemes. If two sounds create two meanings, they are different phonemes in that language. The words they make are minimal pairs. If they do not create different meanings, they are allophones.

Phonology gives the rules according to which sounds are organized in a language and you will find it a highly interesting study. It is far more technical than this brief chapter can tell you so be prepared to devote serious attention to it if you want to master it.
To study how words are made is called **morphology**. *Morphe* means shape in Greek and *logos* refers to knowledge, the study of something etc. Thus morphology is the study of the shapes of words, i.e. how words are formed in a language. Words are the building blocks of the house of language. For a modest dwelling in the English language one needs about 13,000 words, for a big house about 60,000, and if you want a palace then you would need over 120,000 of them (Akamajian et. al. 2001: 11). But people who learn many words, whether to read the works of Shakespeare or play scrabble, do not learn how these words are formed. Your aim as a student of linguistics is to find out precisely this.

**Morpheme**

A **morpheme** is a unit of meaning which makes words. Just as a phoneme is a unit of sound in a language, we use the term morpheme for those parts of words which mean something on their own or when they are used with other morphemes.

Look at the following examples:

<table>
<thead>
<tr>
<th>Words</th>
<th>Morphemes making the words</th>
</tr>
</thead>
<tbody>
<tr>
<td>boy</td>
<td>-boy</td>
</tr>
<tr>
<td>man</td>
<td>-man</td>
</tr>
<tr>
<td>unman</td>
<td>-un -man</td>
</tr>
<tr>
<td>unmanly</td>
<td>-un -man -ly</td>
</tr>
<tr>
<td>unmanliness</td>
<td>-un -man -li -ness</td>
</tr>
<tr>
<td>boyish</td>
<td>-boy -ish</td>
</tr>
</tbody>
</table>
Morphemes like \(-boy\) and \(-man\), which have meanings on their own, are called FREE morphemes. They can stand on their own without being combined with other morphemes.

Morphemes like \(-ish\), \(-un\), \(-ly\) (or \(-li\)) have no meaning on their own. They do indicate some concept but are not used alone. They are called bound morphemes as they are used with other morphemes. They are therefore bound with other morphemes to give meaning.

As you can see, long words are made up of free and bound morphemes. There is a well known game in which inordinately long words are made up in this way. Here is an example:

**HYPERPOLYSYLLABICSESQUIPEDALIANISM**

Here are the morphemes used for making the word: *poly* means many and *hyper* means great or big. *Sesquipedal* means being one and a half feet long. It is itself made up of other morphemes, i.e. *ped* (foot). You can find out the meaning of the other morphemes and come to the conclusion that the word means: THE HABIT OF USING ONE AND A HALF FEET LONG WORDS!

**CLASSIFICATION OF MORPHEMES**

Let us see how linguists classify morphemes. Besides the major categories of free and bound morphemes given above, linguists use the following classification:

![Morphemes Diagram](attachment:morphemes-diagram.png)
**Free morphemes**

Free morphemes are lexical items or commonly used words like *boy, girl, horse* etc. They mean something on their own but cannot be broken down into lesser units of meaning.

The functional free morphemes can be used on their own but they serve some function. They may give a relationship or refer to something or be used as fillers.

Let us look at the sentences below.

1. The door is *in* the house.
2. Open *the* door.
3. */dervaze ko kholo/.*
4. */bua kholo/.*
5. */dervaza kholo/.*
6. *Do* open the door.

In sentence (1) *in* serves the function of indicating a location. In sentence (2) *the* indicates a particular door. In the Urdu-Hindi sentence (3) *ko* is a filler. It has no equivalent in English. The Punjabi sentence (4) as well as the Urdu-Hindi one (5) say the same thing without using *ko*. Similarly *do* in sentence (6) is an empty filler. The sentence would give the sense if *do* is omitted but the urgency or the emphasis would be missing. So fillers are not always empty of all meaning. They may also convey emotion and attitude. All these morphemes are called **functional free morphemes**.

Other examples of functional morphemes are:

- **Articles** *the, a*
- **Demonstratives** *this, that*
- **Conjunctions** *and, or*
- **Quantifiers** *all, most, some, few*
- **Prepositions** *to, from, at, with*
BOUND MORPHEMES

Derivational morphemes derive, or make, a new word. The pure derivational morphemes derive a word which is always in a different grammatical class from the original one. Let us take the common morphemes -able and -ish and add them to words.

eat + able = eatable \[eat is a verb but eatable is an adjective or a noun\]

boy + ish = boyish \[boy is a noun but boyish is an adjective\]

Here the newly derived words are in a different grammatical category from the original ones.

Affixes too derive new words but generally do not change the grammatical categories of the original ones. To affix means to attach. If the affix is attached before the word it is called a prefix. If it is attached after a word it is called a suffix. If it is attached inside a word, it is called an infix. English and most Pakistani languages do not have infixes. However, Brahvi—a language of Balochistan—has infixes.

Here are some suffixes, prefixes, and infixes:

a-, auto-, re-

a + moral = amoral \[not concerned with morality. Both are adjectives\]

auto + biography = autobiography \[both are nouns\]

re + print = reprint \[both are verbs\]

/pǝ/ in Brahvi is an infix meaning not. Its use is as follows:

/xǝlniŋ/ to beat

/xǝlpǝniŋ/ not to beat.

In these cases the grammatical class of the new words remains unchanged.
**Inflectional morphemes** never change the grammatical category of the words to which they are attached. They are added according to the grammatical rules of the language and are used mostly in inflectional languages (i.e. languages in which meaning is expressed by changing some morphemes after the main roots of words).

The past tense morpheme –*ed* in English is used in many verbs: *jump + ed = jumped; dance + ed = danced; and save + ed = saved.* However, there are other ways of expressing the past in English also. Latin, on the other hand, is much more inflectional than English.

The morpheme –’s (as in *the girl’s*) behaves both like an inflectional and a derivational morpheme. Thus there are problems with all classifications which must be kept in mind.

**Neologism**

The process of creating new words is called **neologism**. This is sometimes a political activity in the sense that the morphemes (or words) chosen are selected to emphasize identity (ethnic, religious or nationalistic) rather than comprehensibility. For instance, we find borrowings from Sanskrit in modern Hindi, and from Persian and Arabic in modern Urdu. These do not make those who use these styles of language intelligible to ordinary people but that is not the aim of those who have put in these words. Their aim is to emphasize identity. Sanskritized Hindi is symbolic of the Hindu identity while Persianized Urdu stands for the Muslim identity. Indeed, one major reason why these two styles of one language (old Hindi or Hindvi) drifted apart from the 18th century onwards is because of this politicized neologism. If you want details of this process see Amrit Rai’s excellent book *A House Divided* (1994). If you want to read about neologism in Pakistan by those who express their ethnic identities through coining words rooted in the ancient forms of their languages see the present author’s books Rahman (1996) and Rahman (1999).
But not all neologism is politically inspired. Adolescent slang in America uses the word *geek* for an 'intellectual'. This word will be accepted in the language one day as new words often are. In other cases, new things are created and we need words for them. Sometimes old words are given new meanings and sometimes new ones are created. For example:

- **Radar**  
  *radio detecting and ranging*
- **Laser**  
  *light amplification by stimulated emission of radiation*
- **URL**  
  *uniform resource locator*
- **DOS**  
  *disc operating system*
- **HTML**  
  *hypertext markup language*

Words are also made by blending them such as *brunch* (breakfast and lunch), *motel* (motor and hotel), and *edutainment* (education and entertainment), or they may be borrowed from another language. *Kindergarten* from German means the garden of children (i.e. children’s class); *croissant* is from French; *avocado* and *guava* are both from the Aztec language, and so on. Of course Hindi-Urdu have a lot of borrowing from Persian, Arabic, Turkish, Portuguese, and English on an Indic base which has a large Sanskritic base mixed with some other languages of South Asia.

**Allomorphs (or morphs)**

The morpheme is written in a fixed way. However, its actual pronunciation is not the same in all situations. These different, or variant, forms of morphemes are known as allomorphs or morphs.

Let us take the English plural morpheme (which is inflectional) -s. It is written as -s in all contexts but can be pronounced /s/, /z/ or /ɪz/ according to the sounds which come before it. See the words given below:
As you can see, the morpheme is conventionally written as -s in all cases. However, it has three allomorphs /s/, /z/, and /ız/. They are used according to the rules of pronunciation, or phonological rules, in our minds. When the word ends on an unvoiced phoneme such as /t/, the allomorph /s/, which is unvoiced, is added. When the word ends on a voiced phoneme such as /g/, the voiced allomorph /z/ is added. When the ending is a sibilant, the kind of fricative in which the hissing noise is very sharp, the allomorph used is /ız/. When the singular and the plural is the same as in sheep, we may say that a zero allomorph has been used, i.e. it has no pronunciation.

These rules are part of the language system and speakers are not consciously aware of them. Because we want to study language in a systematic manner, we create names for these different ways of pronouncing an ending which is written in the same way.

It is because we want to create uniform rules that the concept of the zero morph is used. It is merely another way of saying that there is an allomorph which has no pronunciation. This allows us to add the morpheme -s before all the words of English no matter how they are actually pronounced.

This desire for using one conventional symbol for pluralization (i.e. -s in English) can lead to complications. For instance, look at child – children; mouse – mice; ox – oxen; man – men and foot – feet. We may say:
child + s = children
mouse + s = mice
ox + s = oxen
man + s = men
foot + s = feet

This creates regularity as far as our use of the conventional symbol for the morpheme -s is concerned. But the pronunciation is so different that we have to give special rules of how the allomorphs are derived. The study of the sounds of morphemes involves, as we have seen, both phonological and morphological rules. It is therefore called morphophonemics or morphophonology. In such cases, complicated morphophonemic rules are used to represent the derivation of words.

**Morphophoneme**

Some linguists believe that it is useful to think of the variant pronunciations of a morpheme as a morphophoneme. Taking the example of the pluralization morpheme in English again:

<table>
<thead>
<tr>
<th>Morpheme</th>
<th>Allomorphs</th>
<th>Morphophonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>/s/</td>
<td>/z/</td>
<td></td>
</tr>
<tr>
<td>-s</td>
<td>/ iz/</td>
<td>/ z/</td>
</tr>
<tr>
<td>/ø/</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The morphophoneme is an abstract concept like the morpheme itself. The actual phonemes which give the plural sounds are the allomorphs. However, the morphophoneme can be used as a symbol. It can be written to indicate the class of all the allomorphs. Then we can write the rules which will give its actual phonetic value in different environments (i.e. the way it is pronounced after a voiced ending or an unvoiced ending etc.)
To conclude, morphology is an important aspect of a language. It can be seen as a process or an arrangement. If it is seen as a process, a form is taken as a basic or underlying form, and other forms are derived from it according to certain processes. If it is taken as an arrangement of items (morphemes), the rules for this arrangement are studied. Nowadays, most linguists use both these concepts to study the shapes of words in different languages. For more details, read books on morphology such as Matthews (1974).
The study of meaning is called **semantics**. *Semaino* means ‘to mean’ or ‘to signify’ in Greek, and –*ics*, like –*logos*, is used to indicate that this is a branch of learning. Since meaning is so crucial, not only linguists but philosophers are also interested in it. In fact philosophers have been much more concerned with it in the past than linguists.

Our main concern in this introductory book is with the way linguists study meaning. As it is a complex subject we shall attempt nothing more than to explain some of the terms used in the subject. After that we shall touch upon those aspects of it which will be of use for further study.

**Some basic terms**

Most people talk of the meaning of words. However, the term word is rather inexact. Let us examine the concept die. It can be expressed as follows:

1. **die**  
   He died.
2. **to lose life**  
   He lost his life.
3. **to pass away**  
   He passed away.
4. **to kick the bucket**  
   He kicked the bucket.

As you can observe, not one word but two are used in examples (2) and (3), i.e. *lose* and *life*; *pass* and *away*. In example (4) three words are used *kick*, *the*, *bucket*. Moreover, *die* and *died* are the same concept but expressed in two variant forms.
In short, the underlying concept is one but more than one word, as well as variants of a word, are used for expressing it. If we want to talk about the underlying concept alone, we should use some other term and not the term *word* for it. Linguists use the term **lexeme** to mean a unit of meaning. Some people use the term **lexical items** for the same concept. In this book, the term lexeme is used for a unit of meaning. Remember the term morpheme does not mean a single concept. It is that which makes what we call words. Thus the idiom *pass away* has words made up of several morphemes but it expresses only one concept—‘to die’ (to cease to remain alive).

The other terms we shall come across are: **synonymy**, **antonymy**, **hyponymy**, **polysemy**, **homonymy** etc.

**Synonymy** refers to the sameness of meaning. A dictionary generally gives you synonyms. For instance, the synonyms of *beautiful* are *pretty*, *handsome*, *good looking*, *personable*, *beauteous*, *comely*, *pulchritudinous* etc. However, if we use them without reference to what or whom we are talking about, we shall create misunderstandings. See the sentences in columns A and B.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>My horse is personable.</td>
<td>My horse is beautiful.</td>
</tr>
<tr>
<td>This man is pretty.</td>
<td>This man is handsome.</td>
</tr>
<tr>
<td>That boy is beauteous.</td>
<td>That boy is good looking.</td>
</tr>
<tr>
<td>The heroine is pulchritudinous.</td>
<td>The heroine is pretty.</td>
</tr>
<tr>
<td>The scene is comely.</td>
<td>The scene is beautiful.</td>
</tr>
</tbody>
</table>

The sentences in column A do not use the synonyms in their precise shade of meaning. For instance, if a man or boy is called *pretty*, it would indicate that they are effeminate. *Pulchritude* is used by philosophers, and if we use it for a woman, we would appear unnatural.

However, the underlying concept is the same and the precise use is determined by other factors.
Antonymy refers to opposites. The antonyms of good is bad, fat is thin, beautiful is ugly, and so on.

Hyponymy refers to the notion of inclusion. For example, when we say animal we include tigers, cats, and horses under this main category. The word at the top (i.e. animal) is the superordinate term, whereas the words under it are co-hyponyms.

Remember that your world view will be reflected in the way your language classifies the world. People who believe that animals and human beings belong to different categories will not put human under the superordinate term animal. For them living creatures will be the superordinate term:

Polysemy refers to a lexeme having more than one meaning (poly = many). Thus chip in English means a piece of fried potato, an electronic device, and a piece of wood. We use the same sounds to refer to different concepts.

Homonymy refers to two or more lexemes having the same shape. Bank, for example, is a verb meaning ‘to rely upon’. It is also a place where money is kept as well as the side of a river on which
one can walk. The context makes it clear in the example given below:

*I bank upon the Habib Bank situated on the left bank of the river Ravi to give me a loan.*

**Homophone** refers to words having the same sound but different spelling (*threw* and *through*).

**Homograph** refers to words having the same spelling but different sounds. *Wind* means air movement and is a noun. It also means to turn into one direction; to finish, i.e. *wind the clock; wind up your work* etc. The pronunciation is different.

\[
\text{wind (n) — } /w\ ı\ n\ d/ \\
\text{wind (v) — } /w\ aı\ n\ d/
\]

Apart from these words you will also encounter words like **deixis**. These are lexemes which can be understood only with reference to the speaker’s position in space or time. Actually *deixis* comes from a Greek word for pointing. Thus deictic forms point out. **Personal deictics** are *you* and *I*. They point out people. **Spatial deictics** are *this*/*that; here*/*there; come*/*go; bring*/*take*. They point out where the speaker is in space in relation to others. What is near the speaker is *here*, what is far is *there*, and so on. **Temporal deixis** refer to time. They are *now, yesterday, and tomorrow*, as well as *then* and *now*. They also mark tenses. They tell you how time is perceived by the speaker.

**Semantic components**

You have seen how the concept of binary opposition helped us to divide a sound into its features, i.e. if it was a consonant we called it + cons – voc. If it was voiced we called it + voice, and so on. In the same way, some lexemes too can be seen as being made up of binary distinctions.
In short, a boy would be:

```
boy
+ human
+ animate
– adult
+ male
```

but what about a dead boy? We would have to make the term animate minus i.e. – animate. As David Crystal points out, meaning is so complex that such an analysis is not always possible (1988: 107). However, this concept is used in componential analysis which you will study later.

**THE PROBLEMS OF SEMANTICS**

Although we have explained some of all the terms given above, we do not have a semantic theory on which everyone can agree. Part of the problem is that the term meaning is far from clear. In fact C. K. Ogden and I. A. Richards published a book in 1923 which had the title *The Meaning of Meaning*. It had about twenty two definitions of meaning (1923: 186–87) but even this list is not exhaustive. Much later Chomsky said that ‘meaning’ tends to be used as a catch-all term to include every aspect of language that we know very little about (1957: 103).

Later Katz and Fodor (1964) tried to present a semantic theory which could fit in with Chomsky’s concept of generative grammar. According to them a native speaker has a dictionary of lexical items of his or her language in the mind and a system of rules as to where they will be used. These rules tell the speaker where to put in a word. They are called *projection rules*. For instance, unless we are talking about the world of make-believe, we cannot say:

* The stone talked to me.

The sentence is syntactically perfect but in the world we live in stones do not talk. Thus the noun *stone* has different properties from the noun *girl* (who can talk). This is how the semantic
properties of nouns and verbs are projected onto the syntax and, in a sense, pre-determine the structure of a sentence.

This is how Katz and Fodor define the entry for the word *play*:

1. To move swiftly, erratically; to flutter; to dart to and fro; to vibrate.
2. To perform an activity for pleasure.
   To perform in the character or part of; to act.
   The stage representation of a drama; a performance.

Every item is marked for its syntactic category i.e. is it a noun, verb, preposition etc. These categories are called **grammatical markers**. The sense it conveys, or the semantic label we put it under, is called the **semantic marker**. The way its meaning, or what it refers to, is distinguished from its other senses is called its **distinguisher**. Here is the entry for *bachelor*.

Adapted from Katz and Fodor 1963: 496.
Here, the grammatical marker is noun, the semantic markers are human, animal, male etc., and the distinguishers are the meanings given in square brackets.

Initially, as you have seen, Chomsky concentrated on syntax. But later it became clear that semantics has a close relationship with syntax. The theory of generative semantics denies that syntax is separate from semantics. Thus, concepts determine syntax at the deep level and are not inserted in a given structure later. These theories are by no means final. Research on language is going on and new theories are being proposed. If you are interested in the subject you should keep abreast with them.

**Meaning in Philosophy**

Although meaning has been discussed since time immemorial by philosophers—see Plato’s *Cratylus*—it became a focus of philosophy in the twentieth century. The German mathematician Gottlob Frege and the English philosopher Bertrand Russell were both interested in it. However, their major interest was in a logical language (such as the symbols used in mathematics) and not in natural human languages.

The major problem of philosophical linguists was whether the words we use refer to something. The word *Islamabad*, for instance, refers to a city in Pakistan. But all words do not have such referents in the world. For instance, what do words like *love*, *ask*, *consistent*, *justice* etc. refer to in the world? We have some idea or thought in our mind, and our word (called *symbol* below) refers to the ‘thing’ or ‘concept’ (referent) through our mind. Ogden and Richards gave the following triangle for this concept:

```
   Thought

       Symbol — Referent
```

(Source: Ogden and Richards 1923: 99)
Philosophers also tried to understand meaning with reference to truth value i.e. is a sentence true or not? The Polish-born mathematician Alfred Tarski defined Truth according to the following formula:

\[ X \text{ is true iff } P \] (iff means if and only if).

\[ X = \text{sentence (i.e. description or statement).} \]

\[ \text{iff = if and only if.} \]

\[ P = \text{a variable telling us about something in the world.} \]

\[ \begin{align*} 
  & \quad X \\
  \text{Coal is black} & \quad \text{P} \\
  \text{(iff coal is black)} 
\end{align*} \]

But in natural languages curses, blessings, orders, requests may neither be true nor false.

Some philosophers, therefore, talk about the use of words. The philosopher Ludwig Wittgenstein called language a game and compared a word to a piece used in a game of chess. This will be explained in a little more detail below.

First, however, let us briefly describe the life of Ludwig Wittgenstein whose theories had such a profound influence on some aspects of linguistics. Born in Vienna on 26 April 1889, he studied mathematics and natural sciences in Vienna from 1903–1906. From 1906 to 1911 he studied Mechanical Engineering in Berlin and Manchester.

The most important event of his life from the point of view of his philosophical career is his arrival in 1911 to Cambridge to attend Bertrand Russell’s lectures in mathematical logic. But Wittgenstein soon got bored of Cambridge and also developed differences with Russell, so he moved to Skjolden, Norway in order to pursue his reflections on the philosophy of language. However, when the First World War started in 1914 he abandoned philosophy to join the Austrian army thus becoming Britain’s enemy. In 1918 as a prisoner of war he kept working on his book later published as *Tractatus* (Wittgenstein 1922) which he sent to Russell.
Being the eccentric he was, he now resolved to become a school teacher and that too in a primary school in rural Austria. This position he filled from 1920 till 1925 when he was thrown out for having boxed a pupil’s ears till he bled. For some time he was an equally unsuccessful architect too but this too came to nothing. Oddly enough it was not wealth he was after because he refused his share in his father’s considerable business fortune giving everything to his sisters. Even more unexpectedly this failed school teacher now wrote to his Cambridge friends that he was coming over to the University and that they should provide him with means of subsistence. They did, and in 1929 Wittgenstein settled down in Trinity College Cambridge and submitted *Tractatus* for a doctorate. The account of his telling Russell and G.E. Moore, the occupant of the chair of philosophy at Cambridge, is so hilarious that you should not miss reading it in his biography. In 1939 he succeeded Moore as the Professor of Philosophy at the University of Cambridge and it was only because he himself resigned in 1947 that he ceased to hold that exalted academic position. Wittgenstein never married and had no house of his own, but he was lucky in his friends because till his death on 29 April 1951 he never lacked a roof and somebody to take care of him (Malcolm 1958).

It was this man who influenced linguistics, especially pragmatics, through his books. In *Tractatus* his main focus is on the truth or falsity of sentences. He says that everyday language is an impure vehicle of thought. The basic assumption is that reality does exist and thought is the interpretation of this reality which is expressed through language. Thus, ‘the limits of my language are the limits of my world’ (*Tractatus* 5.6). The ideal language would be logical symbolism. However, in his later book *Philosophical Investigations* published after his death in 1953, he relinquishes the search for an ideal logical language.

In *Investigations* he looks at words in everyday use. *Language-games* refers to rules for playing with words as we have rules for games. When we sing we play a language game. The theory of language-games mentioned above can be understood by looking at
the way we use a word. According to Wittgenstein ‘the meaning of a word lies in its use’. Speakers and readers use words according to the situation which has its conventions, rules, and norms of appropriate behaviour. For instance, if we meet acquaintances we ask them how they are. They reply that they are fine. The game is to adhere to this ritual question and answer. The words in this game mean only an exchange of noises to convey to the other person that we acknowledge their presence. If someone actually replies truly to the query then it would be deviation from the rules of the game.

The idea of language-games forms the basis of pragmatics. G. Ryle, P. F. Strawson, and J. L. Austin, who developed the idea of speech acts, were all influenced by Wittgenstein (Battaner-Moro 2005: Vol 2, 1179–1181).

According to John L. Austin’s theory of speech acts, saying something was doing something. All requests, commands, promises, warnings, declarations have illocutionary force—that which pertains to some form or aspect of action.

This brief account of the philosopher’s concern with meaning is not meant to be exhaustive. On the contrary we have not even touched upon most of the theories in this field. Another fascinating aspect of meaning is that it tells you a lot about your world view—the way you see reality. This is an aspect of it which will be taken up again. Meanwhile, if you want to learn more about semantics you should refer to the books given in the bibliography. To begin with read Leech’s book entitled Semantics (1974).
Introduction to Chomsky’s Theories

Noam Chomsky was born on 07 December 1928. His father was a professor of Hebrew and Jewish Education at Grate College, Philadelphia. Chomsky was initiated into linguistics by Zellig Harris, (who was teaching at the University of Pennsylvania) and completed his PhD from the same university in 1955. Since 1961 he is a full professor at the Massachusetts Institute of Technology, one of America’s leading research universities. Since 1966, he has been Ferrari Ward professor of Linguistics, and in 1967 he was made Institute Professor of Linguistics and Philosophy at MIT. He is now considered the most distinguished and original linguist in the world. He is also famous as a radical intellectual.

Chomsky became a radical in politics with a left-wing orientation during his youth. He is now famous as an articulate and vitriolic critic of America’s neo-imperialist policies. Although brought up in a religious Jewish family, he is also a critic of Israel. His books on international relations, media, and philosophy are just as erudite and scholarly as those on linguistics.

He has received many honours, including doctorates from all around the world. In 1988 he received the Kyoto Prize in basic science which is given for subjects not covered by the Nobel prize committee. He has also delivered the John Locke Lectures at Oxford University in 1969 and the Bertrand Russell Memorial lectures at the University of Cambridge. He has also lectured in India and Pakistan on political subjects (Lasnik and Otero 2005).

This chapter introduces some basic ideas of Chomsky in very simple words.
INTRODUCTION TO CHOMSKY’S THEORIES

GRAMMAR

The rules governing the way words are placed after each other to create sentences are dealt with in syntax. In the following three chapters, we will mainly refer to Noam Chomsky’s theories about syntax in particular and linguistics in general. Let us begin with a few definitions.

First, let us define grammar. You have been studying grammar at school. It consisted of certain rules which were meant to make you write and speak a language correctly. Your teachers told you what to do. They prescribed rules such as (1) Do not say ‘It is me’. Say ‘It is I’ and (2) Do not end a sentence with a proposition. This is called prescriptive grammar.

Some rules taught by teachers are not necessary at all. For instance, ‘It is I’ is considered pedantic and people normally use ‘It is me’. Moreover, excellent writers of English end sentences with prepositions. These rules come from Latin grammar. As the language of the scholars of Europe was Latin, they were so impressed by it that they thought other languages too should follow these rules (see Palmer 1971: 15-27). You will observe that in Urdu many people follow the rules of Persian and Arabic. For example, the plural of ustād (teacher) in Urdu is ustād. But some people use the pedantic usataza which follows an Arabic pluralization rule. Likewise modern Hindi grammar follows Sanskritic rules.

We are not talking about prescriptive grammar in this book. Most linguists, including Chomsky, whose theories will be discussed in this book, are interested in descriptive grammar. Like all scholars their real interest is in how people form sentences: What rules are there in the minds of native-speakers which enable them to create sentences?

Like all scientists, our job is merely to describe these rules. We are not concerned with anybody’s notions of good and bad language. Thus all varieties of language, including the varieties of language spoken by uneducated people, are part of our study. All languages have grammar. In fact without having some rules for producing
sentences, they could not have been produced at all. Thus it is meaningless to say that such and such language does not have a grammar.

Now look at the following two sentences of English.

(1) I went home
This arrangement or order is known as syntax, and some rule in our brain/mind tells us that the correct order is as in (1) and not as in the following sentence.

(2) *Went home I
Obviously (1) is a sentence in English, and your mind, can only tell you that it is correct (or well-formed) if you are a native speaker of English or know English quite well. Given any utterance in your own mother tongue, you will be able to judge whether it is well-formed or not. This judgement will be largely based on intuition. Thus we have to depend on the judgements of the native speaker of a language to determine the ‘well-formed-ness’ or otherwise of sentences/utterances. Since we take sentences/utterances produced by native speakers as data when we attempt to describe a language, it is clear that our grammar is not meant to teach anyone how to write correctly. Our aim here is to understand how the mind works when it creates language. This can be understood by an analogy with computers which are fed with instructions which are processed in order to give answers. In the same way, the human brain also appears to be acting under certain instructions which we shall call rules or principles. These rules process words and parts of words to produce speech (utterances) or writing (sentences). Chomsky is mainly interested in these rules and principles and feels that we can understand how the mind actually works if we can find out how it produces language. This means that Chomsky’s approach to language is mentalist, i.e. it is based upon understanding mental phenomena. Indeed, one of the aims of his theories is to understand the mind itself. Chomsky himself says:

Personally I am primarily intrigued by the possibility of learning something, from the study of language, that will bring to light inherent properties of the human mind (Chomsky, 1972: 103).
Chomsky’s other aims

To understand the mind, however, is the ultimate goal. To attain this goal, Chomsky wants to understand what language is. Now language obviously has a structure, that is, some kind of systematic connection between the ways words combine together to create meaning. Clearly you cannot put words together in any random order, nor can you produce any sounds you like if you want to make sense. The whole language is like a system or structure, and one of Chomsky’s aims is to understand this structure. In other words, he wants to develop a theory of language structure. Such a theory seeks to answer questions such as:

- What are natural, or human, languages?
- Are they different from artificial languages, such as those used in computers?
- Are there any common or universal common features, and is it possible to construct a Universal Grammar?

Along with these questions, Chomsky is also interested in the theory of language acquisition, i.e. how small children learn the language they hear most. Both these theories will help us understand how the brain/mind works when it learns or produces human language.

Particular grammar

The goals given above cannot be reached unless we have a great deal of data at our disposal. Our data in this case is obtained from human languages, such as English, French, Urdu, Persian, and so on. If we describe these languages fully we shall understand the structure of particular languages. This might help us to understand the structure of language in general. These particular descriptions are also called grammars. These grammars should cover phonology, morphology, syntax, and semantics. However, most people only talk about syntax in grammar. Notwithstanding this fact, the ideal grammar of particular languages will describe the phonology, morphology, syntax, and semantics of that language.
**COMPETENCE AND PERFORMANCE**

Native speakers of a language, however, carry the grammar of their particular language in their minds. They know the structure of their language, although of course they are rarely consciously aware of the extent of their knowledge, nor can they describe it adequately. However, the knowledge which they have, which is assumed to cover the whole system of the language, is called *competence*. Thus particular grammars are merely reflections of the rules and principles, as well as the lexemes of the people who speak it.

In practice however, people do not use the complete system of the language they are speaking. They use only some words in a certain combination, whereas, of course, they could use many other words in longer or shorter sentences. Often, they also pause and break sentences off in the middle. This actual use of language is quite different from the speaker’s potential competence; it is called *performance*.

Let us now summarise the two definitions in Chomsky’s own words given in his famous book *Aspects of the Theory of Syntax* (1965): *Competence* is ‘the speaker-hearer’s knowledge of his language’ and *performance* is ‘the actual use of language in concrete situations’ (Chomsky 1965: 4).

**LANGUAGE ACQUISITION DEVICE (LAD)**

We have been talking about rules and principles over the last few pages. But what exactly are these? In this section we shall try to answer this question. Remember that we mentioned that producing a sentence in a language may be compared to the process of getting results out of a computer. The computer is programmed and arranges items according to the instructions given to it. It is equipped to process information in a certain way. Suppose the human brain too has some device which can help it to process language; what would it do? This device, which Chomsky called the *language acquisition device*, (or LAD for short), would have the capacity to arrange the lexical units of any human language
according to some general universal principles. Then, there would be variation according to the language which we are using. These variations would not be haphazard, but would be controlled or governed by certain rules. This would make the LAD’s work easier since it would already have universal principles ‘programmed’ into it, and all it would have to do would be to learn the rules of different particular languages. Since this principle of economy is seen in all natural processes, Chomsky maintains that this may also be how the human brain works (for an easy discussion of these theories, see Lyons 1970: 122).

**Rules and Principles**

The principles common to all languages (or universal principles) are general patterns of language use. We can say that all languages must be governed by certain rules and must have a structure. They cannot be haphazard. The rules of a language determine how it is produced. Without rules no sentence, however simple, can be produced, so that we can say that language is always rule-governed. Consider the following simple sentences:

(1) Neha ran to the school.
(2) Neha school *bhagi*.

Sentence (1) is in English and (2) is in Hindi-Urdu. Now we simply cannot put these sentences in any order we like. We also cannot give a general rule that one particular word, for example, *school*, will always come at the beginning of the sentence. We shall have to consider the structure of the sentence. This means that we must consider whether a word is a noun, verb, a preposition, and so on. These are syntactic categories, and structure depends upon them. In other words, no matter which language we use, any rules which produce well-formed sentences in that language must be structure-dependent. Rules will apply to the structure of the language which is determined by nouns, verbs, adjectives, and so on—that is, by syntactic categories. This structure-dependence is therefore a universal principle. It applies to all human languages.
Let us now come to the rules of English and Hindi-Urdu, as demonstrated in sentences (1) and (2). It is clear that in (1) the verb *ran* comes before the place where the subject (*Neha*) went. In Hindi-Urdu however, the verb *ran* (*bhagi*) comes after *school*. Moreover, in English the verb *ran* is not marked for gender; whether Neha is a boy or a girl, we shall use only *ran*. But in Hindi-Urdu it is clear that we are talking about a girl because we put the morpheme -i at the end. If we had used -a, then the subject would have been masculine. This adding of -a or -i is an example of *inflection*, and if we are describing the morphological rules of Hindi-Urdu we will have to give the information that the gender governs the morpheme used for inflection in verbs. This means that *universal principles* remain the same, but the rules vary from one language to another. Our task is to discover the rules of particular languages as well as the universal principles of all languages. In this way we can try to find the answer to questions such as ‘What is language?’ and ‘How does the brain work when we produce and understand it?’

**Different kinds of rules**

We know that languages have the following dimensions: (a) phonological, (b) morphological, (c) syntactic, and (d) semantic. Thus we would need to take account of at least these four types of rules in order to describe a language.

(a) **Phonological rules** specify how words and combinations of words are pronounced.

(b) **Morphological rules** specify how words are built up of small units of meaning, i.e. morphemes.

(c) **Syntactic rules** specify how words combine to form sentences.

(d) **Semantic rules** tell us the meaning of words and combinations of words.

This chapter has defined certain key concepts in modern grammatical theory as developed by Chomsky and his associates.
Through these definitions and their explanations, we hope that you will have grasped the preliminary concept that grammar is the reflection of all that the native speaker knows about his or her language. Through describing the particular grammars of different languages, we will be able to postulate a universal grammar, which will give us the principles common to all languages, all of which are rule-governed, and these rules are phonological, morphological, syntactic, and semantic. The aim of the Chomskyan linguist is to discover these rules with a view to understanding the nature of language itself, and in the final analysis, the properties of the human mind, which seems to be especially equipped to understand and produce language.
The term ‘transformational generative grammar’ is used to refer to Noam Chomsky’s theories about syntax which have been touched upon in the last chapter. These theories were first put forward in a book entitled *Syntactic Structures* which was published in 1957. If you want to know more about Chomsky’s ideas on this as well as other aspects of his thought read the French linguist Mitsou Ronat’s interviews of Chomsky in a book entitled *On Language* (Chomsky 2002). However, here is a summary of these views. Basically, Chomsky tried to find certain rules which would create well-formed sentences of a language. A sentence was defined as being ‘well-formed’ if the native speakers of that particular language felt that it was correct or grammatical. As Chomsky wanted to create rules of a general kind about language he had to use what is seen by many as mathematical terms. For instance, he defined a language as follows:

I will consider a language to be a set (finite or infinite) of sentences (1957: 13).

But he also said that it was not mathematics he was doing. For instance, he told Mitsou Ronat:

It turns out that the way to ‘speak precisely’ is by formalization, but it would not be correct to consider that as mathematics. For example, some variety of recursive function theory provides the means, in principle, to express linguistic rules. But up to that point, this is formalization, not mathematics.

(Chomsky 2002: 125)
What he meant by his definition of language was that any language, whether English or Urdu or Hindi or any other you can think of, was made up of a number of well-formed sentences. These sentences could be infinite, since it is possible to keep making new sentences. However, at any given time (for example, at the moment you are reading this), the number of sentences is actually finite. That is, at this moment they can actually be counted, though of course more sentences will be produced after the fixed time up to which they are counted. According to Chomsky, his grammar is generative, since it can generate or create an infinite number of sentences. It is called transformational since a basic or simple sentence like:

(1) I read the book

can be changed or transformed into a number of sentences with either the same meaning:

(2) The book is being read by me.

or with different meanings:

(3) Do I read the book?

(4) I read the book, don’t I?

(5) I do not read the book.

What we have done is that we have transformed the basic sentence by adding words, deleting words, and, above all, by the movement of words. These changes also take place through specific rules, and we call these rules transformational rules. Thus grammar generates and transforms sentences. It can therefore be called transformational generative grammar, or TG for short.

**Properties of transformational grammar (TG)**

TG has the following properties:

(a) It will generate only the well-formed or grammatically correct sentences of a language. It will not generate a
sentence (that is to say, a syntactic structure) which is ill-formed or incorrect, since it is meant to create those rules and principles which are in the mind or brain of a native speaker. These rules and principles are by definition about that form of a language which native speakers consider correct or well formed.

(b) This grammar will have **recursive rules**. This property of recursiveness is the capacity of a rule to be applied again and again in order to generate an infinite set of values. In this case, by ‘values’ we mean new combinations of words which are grammatically correct. In other words, recursive rules allow us to generate an infinite number of well-formed sentences of a language. Let us see how recursive rules work. Consider the sentence:

(6) **This is the boy.**

Now it can be made longer by adding

(7) **who rode the horse.**

But the rule which added (7) can also be applied to (8), and we can add

(8) **that chased the fox.**

and again

(9) **that chased the jackal.**

But if we keep adding new sentences beginning with ‘who’ or ‘that’, we will never finish the sentence, even if we run out of words and have to begin with boy or horse again. This kind of change is called **embedding**. In this, we put in (or embed) a smaller sentence or several small sentences in a large sentence. But recursive rules can work in other ways too. One could even keep adding **and** to create an endless list. The point to remember is that recursiveness accounts for the endless creativity of language. Because we can apply a rule again and again, we can create an infinite number of
sentences out of finite number of words (that is to say, out of a list of finite lexical items).

DEEP AND SURFACE STRUCTURE

In sentences (1) and (2), you notice that the first sentence is simpler or more basic than the second. Let us reproduce them:

(1) I read the book.

(2) The book was read by me.

According to Chomsky, sentences may be present in the mind/brain at two levels. Simple, basic or kernel sentences consisting of ideas or rough meanings, must be present at a deep level. The way these sentences are formed at that level is known as deep structure. However, we cannot actually ‘see’ sentences in the brain at that level, we only hear spoken sentences or see written ones at the surface level. The structure of these sentences is known as surface structure.

At this point you must understand clearly that all we can actually see or hear is the surface structure. The deep structure (D-structure) is an abstract concept, i.e. a theoretical concept that cannot actually be seen or touched or felt. But why do we talk of D-structure or D-level, then? Is it only to make syntax more difficult? The answer is that we formulate certain hypotheses to explain certain facts that we can observe. A hypothesis is just a guess, but it is the guess of an expert in that particular branch of knowledge. Thus it has a possibility of being correct, and even if it proves to be wrong, it helps researchers find out more about the facts being investigated. The deep level is assumed since we think that the brain/mind would first assemble certain concepts, that is to say, a basic sentence. This would be operated upon by rules and changed into more complex sentences. We think this would be more economic for the brain or mind. However, we could be wrong. In science any theory can be proved to be wrong. At the moment, however, we assume that sentences are formed at some deep level in the mind or brain. They are very basic, and can be called kernel sentences at
that level. These kernel sentences are then operated upon by certain rules so as to produce surface sentences. The following simple diagram illustrates this:

```
DEEP STRUCTURE
  ├── Kernel sentences
  │    └── Transformational rules
  │        ├── Surface sentences
```

This diagram is very simple and therefore ignores certain complications which we shall come across during the course of this study.

**Symbols used in writing rules**

A symbol represents something else. For example, the national flag stands for the honour and unity of a nation. In the language of science there is a need for symbols. We can choose any letter, for example, a, b, > (is greater than) or + (plus) as a symbol of anything. In linguistics too, we use symbols when we write down rules. The writing down of rules is known as **formalization**, that is to say, explaining something in the form of a precise formula. Since Chomsky was greatly interested in mathematics, he expresses rules in a quasi-mathematical manner. Here are some of the symbols and the basic rules which can produce the sentences of English:
S = sentence (e.g. I walk home)  
N = noun (e.g. Tania, he, they, boys…)  
V = verb (e.g. walk, talk, smile…)  
Adv = adverb (e.g. quickly, slowly, clearly…)  
Adj = adjective (e.g. good, green, pretty…)  
P = preposition (e.g. on, in, to…)  
Art = article (a, the)  

As you know, the terms ‘noun’, ‘verb’, and so on, are all syntactic categories. You must have read about these basic categories when you studied the parts of speech at school. Now we shall introduce you to some terms which you probably did not meet at school. These are:

NP = noun phrase  
VP = verb phrase  
Adv P = adverbial phrase  
Adj P = adjectival phrase  
PP = prepositional phrase  

THE NOUN PHRASE

A noun phrase is also a syntactic category, like a noun. It can have more than one kind of noun (that is, a proper name, a common name, or a pronoun); it can have several nouns joined together; it can have nouns modified by articles; or some other words which can determine the nature or the quality of the noun, or specify it more accurately. A noun phrase can also have an adjective qualifying the noun. The following are all noun phrases:
With only one proper noun (PN)  

**Sameer**  

NP has only PN

With a common noun (CN)  

**boy**  

NP has only CN

With a pronoun (ProN)  

**he**  

NP has only ProN

With two proper names  

**Iqbal and Sonia**  

NP has PN and PN

With noun and determiner  

**a boy, the boy, this boy, that girl, those boys**  

NP has determiner + N

With adjective qualifying the noun  

**good boys, a bad boy, the bad boy**  

NP has determiner + adj + N

We can see that the noun takes the same place as the noun phrase does. They belong to the same syntactic category and can fill the same slot in a sentence. For example:

(1) SAMEER ran home  

NP = PN

(2) THE BOY ran home  

NP = CN

(3) HE ran home  

NP = ProN

(4) THAT BOY ran home  

NP = Determiner + N

(5) THAT BAD BOY ran home  

NP = Determiner + Adj + N

If we do not use the term noun phrase (NP), we cannot express the fact that proper nouns, common nouns, pronouns, and nouns specified exactly or modified by adjectives (and so on) belong to the same syntactic category. That is why we use the term noun phrase even when we have only a single noun, as well as when we have many nouns or a modified noun. The definition of an NP is that its key constituent is a noun. We call this key constituent a **head**. If there is anything else in that phrase, it will tell us
something more about that noun, and is said to be **dominated** by it.

(a) NP  (b) NP  (c) NP

\[
\begin{array}{c}
N \\
| \\
| N \text{ det} N \text{ det} \text{ adj} N \\
| \\
| \text{Sameer} \text{ a} \text{ boy} \text{ a} \text{ good} \text{ boy} \\
\end{array}
\]

(a), (b), and (c) are all NP’s. In all of them the most important constituent is a noun, and, if there is anything else, it tells us more about that noun.

**Other syntactic categories**

Just as an NP has a noun at its head, a verb phrase (VP) has a verb as its head. In the sentence *Sameer ran home*, ‘ran home’ is a VP. The verb which is at its head is *ran*. This VP can contain a number of other things, for example, it could have a noun (*home*), an adverb, (*quickly*), or a prepositional phrase (*to his mother*) after it. But all other phrases will come under the head or will be dominated by the head.

For example:

(6) Sameer ran.
(7) Sameer ran home.
(8) Sameer ran quickly.
(9) Sameer ran home to his mother.

In all the above examples *ran* (that is, the verb) is the most important thing. Everything else is just additional information telling us more about the verb. Thus everything else will be dominated by the verb. Thus a VP will certainly contain a verb at its head, although it can contain much more under that head.

Similarly, in a prepositional phrase (PP) the preposition dominates; in an adverbial phrase (Adv P) the adverb dominates; and so on.
**Phrase structure rules**

As we have seen, phrases have definite structures. For example, an NP consists of at least a noun. It can have a determiner, an adjectival phrase, and so on. Similarly, a sentence also has a structure. If we can find a formula, or a formal way of writing how these structures are made or generated, we can write down the phrase-structure rules (or PS rules) of a language. Let us therefore write down the PS rules of English.

\[
\begin{align*}
S(\text{entence}) & \rightarrow NP + VP \\
NP & \rightarrow \{\text{art, (adj), N, PN}\} \\
VP & \rightarrow V + NP + (PP) + (Adv) \\
PP & \rightarrow \text{Prep} + \text{NP} \\
N & \rightarrow \{\text{boy, Sameer, dog…}\} \\
N & \rightarrow \{\text{Ashwariya, Shabana, Rome…}\} \\
\text{Det} & \rightarrow \{\text{a, the, that, this, these…}\} \\
\text{Adj} & \rightarrow \{\text{pretty, bad, ugly…}\} \\
V & \rightarrow \{\text{run, wrote, will think…}\} \\
\text{Prep} & \rightarrow \{\text{with, to, near…}\} \\
\text{Adv} & \rightarrow \{\text{quickly, recently, yesterday…}\}
\end{align*}
\]

In the PS rules the following symbols have been used:

- \(\rightarrow\) consists of
- + and or followed by (This is omitted by some people).
- \{\} curly brackets contain elements out of which only one should be selected.
- ( ) brackets contain constituents which may be omitted.
Now let us generate the sentence: *Sameer ran.*

As it is a sentence, it contains an NP followed by a VP. Let us make a labelled tree diagram of it:

**Tree diagram**

A tree diagram is only a convenient way of showing how the different constituents are arranged in a sentence. It tells us which categorical constituent dominates which, and gives us a pictorial model of the sentence.

\[
\begin{align*}
S & \longrightarrow \text{NP + VP} \\
\text{NP} & = \text{N (i.e. Sameer)} \\
\text{VP} & = \text{V (i.e. ran)}
\end{align*}
\]

![Tree Diagram](image)

It is clear here that the S node dominates all the other nodes. It dominates the NP and the VP. It is just another way of saying that the sentence is a larger syntactic category which contains a noun phrase and a verb phrase. Similarly the NP node and the VP node **dominate** all the other things which come under them. Consider the tree diagram for the following sentences:

(10) Geeti ran very fast.

(11) Geeti ran to the lady.
Here, *fast* is an adverb. It tells us more about the verb *ran*, or modifies it. Thus it comes under an adverb phrase (adv P). But we have another modifier of fast which is *very*; this is an **intensifier**. The Adv P node will dominate the intensifier also. However, all that the Adv P does is to tell us something about the verb *ran*, thus the VP node dominates the Adv P node.

Here, as you can see, the VP node dominates the PP node, which in turn dominates the second NP node. Since certain nodes dominate other nodes in the tree diagram, we say that is hierarchically arranged.
**Other Ways of Showing Syntactic Structures**

Making a tree diagram is a convenient way of showing the syntactic structure of a sentence. There are, however, some other ways of giving the same information. For instance, we can use brackets. Below the brackets we can write down the name of the category we have enclosed in the brackets. This is called ‘labelling’, and labelled brackets can give the same information as the labelled tree diagram, see below:

(12) Sheela ran.

First step: Put the noun and verb in brackets and label them.

\[
\begin{array}{cc}
(\text{Sheela}) & (\text{ran}) \\
\text{N} & \text{V}
\end{array}
\]

Second step: Put N inside bigger brackets labelled NP and V inside a bracket labelled VP.

\[
\begin{array}{cc}
( (\text{Sheela} ) ) & ( (\text{ran} ) ) \\
\text{NP} & \text{VP}
\end{array}
\]

Third step: put both the NP and the VP inside a larger bracket labelled S.

\[
\begin{array}{cc}
( ( (\text{Sheela} ) ) ) & ( ( (\text{ran} ) ) ) \\
\text{S} & \text{NP} & \text{VP}
\end{array}
\]

As you can see, this method uses many brackets and can be confusing. It is, however, useful if you want to save space. The tree diagram is generally simpler to use, but it takes more space.

We have learned so far that there are phrase structure rules which determine the structures of phrases, i.e. the head of an NP is N, the head of VP is V, and the head of AP is A (adverb). Suppose we want to reduce the number of phrase structure rules—how can we do this? Obviously, we will have to integrate all the rules under one rule. But why should we want to do this? Because it is scientific to have as few rules as possible (the ‘minimalist’ approach).

Let us start with VPs. We know that VPs have V and other items (NP, PP, PP…) under it. The maximal projection is VP but under
it there are projections of V labelled as V-bar written as \( \bar{V} \) or \( V' \). Consider the sentence:

(13) I will run down the garden

(14) I am sure you will do so

Now *do so* substitutes *run down the garden*. This constituent phrase is also a small verb phrase. But, being under a big VP, it is a VP-bar or \( V' \).

Before \( V' \) there may be a **specifier**.

(15) We have all run down the garden.

*All* is a quantifier. It occupies the **specifier** position \([\text{spec, VP}]\). It combines with \( V' \) to form a \( V'' \) projection. The \( V'' \) is the same as VP.

In short the structure of \( V'' \) is as follows:

\[
V'' \rightarrow \text{spec} - V'
\]

\[
V' \rightarrow V...
\]

\( V'' \) is the maximal projection of \( V \) and this is what we called VP earlier.

Now let us extend this kind of structure to other phrases. First, to NPs.

(16) The meeting of the artists in the studio went well.

This is how we used to parse such a sentence.

```
NP
  /   \
/     \
|      |      |
N      PP    PP
  /   \
/     \\
|      |
the meeting of the artists in the studio
```

Suppose we say now:
(17) The one at night did not go well.

We have substituted *one* for *the meeting of the artists in the studio*.

Now let us use the idea of the specifier. As you can see *the* specifies a particular meeting. Thus we can parse it as follows:

\[
\text{N}'' \rightarrow \text{Spec; N'} \\
\text{N'} \rightarrow \text{N; XP}
\]

We can propose a similar model for adjective phrases, propositional phrases etc. This is called the X-bar theory and it can be summed up as

\[
\text{X}'' \rightarrow \text{Spec; X'} \\
\text{X'} \rightarrow \text{N; X; YP}
\]

X stands for N, V, A or P. so now we have a general format instead of PS rules for each of these nodes.
We can sum up the X-bar theory saying that the lexical head of a projection is a zero projection ($X^0$). Complements combine with $X$ to form $X'$ projections as do adjuncts. The topmost or maximal projection is formed with a specifier. When the specifier position is not filled in by a word it is considered as filled in by zero.

Now let us go ahead to consider sentences ($S$) as projections. In our earlier analysis the $S$ was on its own. In this new analysis we will propose that there is a complementizer position before $S$. Comp positions are clearly obvious in sentences beginning with wh-words (whether, what, when etc.).

(18) Whether you will do it is doubtful.
(19) What you will do is doubtful.
(20) When you will do it is doubtful.

In all the sentences above the position before you is occupied by wh-words. The general structure is:

```
S'  
\___ Comp\wh-words\___ S\___ NP\___ Aux\___ VP
\      \              \you\will          
```

Now $S'$ and $S$ are called exocentric categories not endocentric ones. This means that they do not get projected from heads like N, V, etc. They are composed of several units next to each other. This means we need a head for them. If we get such a head we will have a schema common to the other categories we have been advocating above. Let us call this inflection (INFL or I for short). Now let us see what the I slot can be filled in with. First let us see whether we can argue that all sentence nodes are under INFL.
Auxiliary Nodes

These are nodes dominated by AUX. This means it will have tense (past, present) and modals to mark the future (shall, will etc.).

But remember that verbs inflect for person and number also. This is less for verbs in English, but French and Farsi have many forms. Let us compare Urdu-Hindi forms with English ones.

Speaking = bolna

<table>
<thead>
<tr>
<th>S.no</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mai bola</td>
<td>Mai boli</td>
<td>I speak</td>
<td>I speak</td>
</tr>
<tr>
<td>2</td>
<td>Tum bole</td>
<td>Tum boleen</td>
<td>You speak</td>
<td>You speak</td>
</tr>
<tr>
<td>3</td>
<td>Tu bola</td>
<td>Tu boli</td>
<td>You speak</td>
<td>You speak</td>
</tr>
<tr>
<td>4</td>
<td>Vo bola</td>
<td>Vo boli</td>
<td>He speaks</td>
<td>She speaks</td>
</tr>
<tr>
<td>5</td>
<td>Hum bole</td>
<td>Hum bole</td>
<td>We speak (pl)</td>
<td>We speak (pl)</td>
</tr>
<tr>
<td>6</td>
<td>Aap bole</td>
<td>Aap boleen</td>
<td>You speak (pl)</td>
<td>You speak (pl)</td>
</tr>
<tr>
<td>7</td>
<td>Vo bole</td>
<td>Vo boleen</td>
<td>They speak (pl)</td>
<td>They speak (pl)</td>
</tr>
</tbody>
</table>

As you can see in Urdu-Hindi the verb is inflected –i, een, -e but the pronouns are the same. Moreover, there are three ways to refer to the second person aap (politest), tum (intimate; also less polite than aap), and tu (impolite or very intimate). English has only one word you instead of these three. In English, moreover the pronouns for male and female are different for the third person (he and she) but not in the other instances. As for the verb, it does not inflect for gender at all. As for the tense it has only two forms.

In Farsi the verb inflects for person, singular and plural but not for gender. Moreover the pronoun also does not indicate gender.

Man amadam I came (singular both male and female)
To amedi You came (singular: m and f)
Oo amad He/she came
Ma amadaem We came
Shuma amadeed You came (plural as well as polite: m and f)
Anha amadand They came (m and f)

In short, there are six forms of the inflection of the verb and yet we need not clarify whether we are talking about a boy or a girl.

To be grammatically correct agreement rules should be followed. If one says *I* then we should say *speak*; if we say *he* then we should use the form *speaks* and so on. No matter how many forms that particular language may have, the agreement (called AGR) feature is necessary. The inflection, we sum up, has tense, agreement, and modals. So tensed clauses have $^{+\text{Tense}}_+\text{AGR}$ and we can see that sentences are either tensed or infinitives. The following are tensed sentences.

(21) I will run in the garden.
(22) I am running in the garden
(23) I run in the garden

The INFL contains:

Sentence (21) contains *will* (modal indicating future intention).
Sentence (22) contain *am* and –*ing* i.e. the verb *to be* and an indicator of an action continuing at present.
Sentence (23) contains past tense of the verb *run*.

Now let us take a sentence without tense. Such a sentence is as follows:

(24) I intend to run in the garden.

As you can see, *to* is the **infinitival marker**. The infinitive form of the verb introduced or marked by *to* does not have tense. This too is in INFL. *To* has neither tense nor agreement. We can call it $^{–\text{Tense}}_–\text{AGR}$.
So finite clauses have [+tense] and [+AGR], and infinitival clauses have [-tense] and [-AGR]. All clauses then have [±tense] and [±AGR] in the INFL.

So if we assume that INFL is the head of S, then S will be a projection of INFL phrase (or IP). Now, using our previous schema of using specifier, we can re-write S as:

\[ I'' \rightarrow \text{Spec}; I' \]

\[ I' \rightarrow I; VP \]

Let us parse a sentence of the \( S \rightarrow NP + VP \) under this schema.

(25) I will drive the car.

Now let us go back to complementizers. As you can see the complementizers *that* and *if* go with finite clauses and *for* goes with infinitival ones. As for *whether*, it can choose both types of clauses.

If there is no complementizer, we do not say that the slot does not exist at all. We say that the slot is there but it is empty. You will understand why we do this. It is because our schema remains the same but in some cases we have zero and in others we have some words.

Complementizer phrases (CP) may be seen to head complementizers as in the other schemes.

(26)
In this case the CP is empty but we can fill it in with *whether*:

(27) Whether I will run in the garden is doubtful.

Thus it is better to assume that there is a CP position but it is not always filled in. Of course, in the light of our previous schema, there will be a specifier too. So the structure will be:

If you have the question word *will* as in:

(28) Will I run in the garden?

Then *will* moves to the specifier position. In such a case *whether* will not be this node. We cannot have:

(29) * will whether I run in the garden.

This is because the node can have only one word not more than one. But we can have this sentence:

(30) When will I run in the garden?

The tree of this is as follows:
The auxiliary moves from one head to another. This is called **head-to-head movement**. We will not, however, go into the details of these movements here.

This means that question words move to [spec CP]. In main clause questions, they precede the inverted auxiliary. In subordinate clauses, however, wh-words do not occur with the complementizers *that* or *whether*. This is true for English as far as the words quoted above are concerned. However, the principle is a general one. We have it in our heads and we modify it according to the language we speak.

The question is, what kind of structure do we have in the mind? Answers to such questions come from theory, i.e. we speculate that the structure will be like this. Then we compare it to another model. Then, whichever model gives the minimum complexity must be adopted. Let us look at some diagrams.

(31) I am walking

Let us have all the models of branching which are possible here. Here are the possibilities:

(31a)
Now let us have a theory which allows only binary branching.

If we have a principle in mind it will be: binary branching is allowed. All others are ruled out. This principle is easy—it is a minimalist principle—rather than some principle which allows all kinds of branching. The kind of structural principles we have been talking about are part of what is called **minimalism** in grammar. An easy introduction to minimalist grammar is Andrew Randford’s *Syntax* (1997). You are advised to read it along with more recent works.
In the last chapter, you have been introduced to transformational generative grammar. In this chapter, let us try to understand more grammar. Let us take up the simplest definition of grammar: it is a device which can create a well-formed sentence in a human language. Let us now take a simple sentence:

(1) The girl followed the boy.

Now let us forget the phrase structure (PS) rules we read earlier and try to guess how this sentence was created. The simplest guess (or hypothesis) is that words have been placed one after the other:

The + girl + followed + the + boy.

This is a linear model because words come one after another in a straight line. It can also be called a chain model, since words are connected to each other like links in a chain. It is a static model since nothing can move from one place to another. It is also called a finite state grammar since it is static, and therefore cannot change state. This simple model of grammar can be created by imagining the flow of an electric current from one point to another, creating a sentence as it goes. For example:

A The B girl C followed D the E boy F

Now, if you start the ‘current’ at B, and stop it at C, the sentence will be
Similarly, if you want to create more than one sentence you will have to create loops so that the current can be passed through one circuit or the other as you like.

Let us try to generate the sentences given below by this model of grammar:

(3) The girl followed the boy.
(4) The boy followed the girl.
(5) The boys followed the girls.

The models will be as follows:

For (3)

```plaintext
A The B girl C followed D the E boy F
```

For (4)

```plaintext
A The B boy C followed D the E girl F
```

For (5)

```plaintext
A The B boys C followed D the E girls F
```

But it would be extremely uneconomical to use a new model for every sentence. Let us see what is common in these models and try to make only one model which will produce all these sentences. This model will be:

```plaintext
A The B girl C followed D the E boy F
```

More complex sentences would require even more loops and the grammar would not be efficient because it would create sentences we do not want: for example, (6) The boys followed the boy, (7)
The boy followed the boy, (8) The girl followed the girls, and so on.

**Non-linear models**

Because of these defects, Chomsky abandoned this linear model of grammar in *Syntactic Structures* (1957). As the title indicates, he concentrated on **syntax**, i.e. the arrangement of words in a sentence, and stated that this arrangement was subject to rules. We have already looked at phrase structure (PS) rules, which, he said, generated basic or deep structures. These deep structures were then subjected to change by transformational rules (T-rules) which created **surface structures**. In *Aspects of the Theory of Syntax* (1965), Chomsky changed his grammatical model to look something like this:

![Diagram](attachment:grammar_diagram.png)

To put it very simply, the idea is that the **base** contains PS (phrase structure) rules which create a kind of skeleton for the syntactic categories (NP, VP, and so on) to fit in one after, or under, the other. Then comes the **lexicon**, which is a dictionary or collection of words. Every word is specified in such a way that only that word can be fitted into the PS rules correctly. This creates a kind of **deep structure**. Now T-rules are applied, which change the sentence so that it comes in a changed form to the **surface level**. Now it takes on meaning, and it is the **semantic component** which gives it that
meaning. It can be spoken, or written in symbols which refer to spoken sounds, which is the **phonological component**. This is a simple model of TG which is presented in Chomsky’s books mentioned above. The theory behind this model is slightly more complex, but it has been simplified here to make it understandable. This theory is called the **standard theory**.

As you can see, the PS rules can generate sentences in a fixed order. If this were the only way of generating sentences, we would need an infinite number of PS rules to generate an infinite number of sentences. That would not be feasible. Thus, according to Chomsky, only the simplest sentences are generated through PS rules. Even these kernel sentences are generated at a deep level. These are then changed through movements of constituents. This movement transforms sentences into different kinds of sentences. That is why they are called transformational rules. To illustrate transformation through movement, consider the following sentences:

(9) The boy followed the dog.

The following sentence means the same thing:

(10) The dog was followed by the boy.

These second sentence (10) must be a transformed form of the basic sentence (9) at the deep level. Some transformational rules (T-rules) have produced the second, passive structure. Let us number the NPs and VPs:

<table>
<thead>
<tr>
<th>The boy</th>
<th>followed</th>
<th>the dog</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP₁ = 1</td>
<td>VP = 2</td>
<td>NP₂ = 3</td>
</tr>
</tbody>
</table>

The structure is 1 + 2 + 3. This is called the structural description (or SD).

Now we need a rule which will bring 3 in place of 1, and 1 in place of 3. We also need to put in ‘was’ and ‘by’. The transformation is shown by a double arrow (→→). This change is called the structural change (or SC).
So, \( SD = 1 + 2 + 3 \)

\[
SC = 3 + \text{was} + 2 + \text{by} + 1
\]

\[
1 + 2 + 3 \rightarrow 3 + \text{was} + 2 + \text{by} + 1
\]

(10a) The dog was followed by the boy

As you can see, this T-rule has moved constituents and inserted ‘was’ and ‘by’ in this case. The passive transformation is a little more complex, but it has been kept very simple here so that you can understand the concept of the T-rule better.

**Grammar as a Machine which Generates Sentences**

To sum up, you can imagine grammar as a machine or a device for creating sentences. It is in the mind of the speaker and represents his or her knowledge of the language. It contains rules for creating simple sentences at the deep level which are then transformed by transformational rules into more complex sentences.

Now let us see what kind of arguments can be advanced in favour of their existence. One of these arguments is that in some sentences movement seems to have occurred, i.e. it appears that some constituents have been moved from their original places. Let us take an example:

(11) I put the book on the table.

Now, ‘put’ is a verb which is followed by an NP, that is, something which can be put somewhere. This NP is followed by a PP (prepositional phrase), beginning with *in, on, inside, outside*, and so on—that is, something which tells us where the thing has been put. In other words, for a verb like ‘put’, the lexicon must specify what constituents follow it.

\[
\text{put} = V \text{ (followed by) NP, PP}
\]

Thus if we omit either the NP or the PP after ‘put’, we will produce a grammatically ill-formed sentence. For example:
(12) * I put on the table. (Here the NP following the V has been left out).

(13) * I put the book. (Here the PP following the NP after the V has been left out).

Now consider the following sentence:

(14) What did I put [----] on the table?

In this sentence, ‘put’ is not followed by an NP. However, we have just stated that ‘put’ is subcategorized in the lexicon as being a verb followed by an NP and a PP. These subcategorisation facts are constant. Thus, what has happened is that the NP after put has moved away from its original position at some stage. But the very fact that it has moved away must mean that there was an original position for it to move away from. Thus, we assume that there is a deep level at which it was there, and that certain T-rules have moved it away. Some T-rule also brings ‘what’ to the beginning of the sentence. In short, rules move parts of phrases from one place to another in sentences. We will try to understand this in more detail later.

**Transformations**

We said earlier that T-rules move, add or delete constituents in a sentence. First, we must find out if the sentence has got the kind of structure to which rules can apply. This is its structural description (SD). If this structure is such that these rules can apply, then we can use the rules. Let us now study only one transformation—the passive transformation.

If you understand this T-rule, you will not only learn more about T-rules, but also about argumentation in linguistics. Let us then take the passive transformation. We will see how a sentence, which states a proposition, changes into the passive form. But before we do that, we will go back to our PS rules. We have covered some of them, but we have not covered the expansion of the auxiliary.

AUX ————> Tense {present}
This rule simply means that a verb, for example *walk*, must have a marking for tense. If it is in the past tense, (i.e. + past), we will read it as *walked*; this addition of the morpheme –*ed* inflects the verb for the past. If it is in the present tense (i.e. + present), we will read it as *walks*.

Other helping verbs (or auxiliaries) can be verbs, such as *will*. This will give us the future tense too. Such words are called modals and their list includes will, shall, can, would, must, may, and so on. So another expansion of AUX will be:

\[ \text{Aux} \rightarrow \text{Modal} \]

But, whereas all verbs must be marked for tenses, modals are not necessary in all phrases. As modals are optional, they are put into brackets. Our rule should be read as follows:

\[ \text{Aux} \rightarrow \text{Tense (Modal)} \]

But verbs can also be in the progressive aspect or the perfect phase. For instance:

(15) He is reading.

is in the progressive aspect. It shows that an action is in progress (or in the progressive phase). Now take the sentence given below:

(16) They have been reading.

Here the perfect is shown by *have* and the *-en* shows the past participle ending. In other words, to show that a verb can be in either the progressive or perfect form, we need two optional elements in the AUX. Speakers of English have chosen the following formulas to show these two elements:

Progressive \[ \rightarrow \text{(be + ing)} \]

Perfect \[ \rightarrow \text{(have + en)} \]

These formulas are the only ways of showing that in the progressive form, there is the verb *to be* and the ending –*ing*. In the perfect form, there is *have* or *has* and the ending –*en*. The formula is somewhat misleading, since the past participle of ‘read’ is not
formed by adding –en. Here the past participle is merely, read (/red/). The –en in the formula therefore simply means ‘make this into the past participle, whatever realization that might have’. This realization is determined by other rules which give the exact pronunciation of every morpheme. These are called phonemic rules. Let us now add another optional expansion of AUX to our list:

\[ \text{AUX} \rightarrow (\text{be + ing}) (\text{have + en}) \]

The full expansion of AUX will be:

\[ \text{AUX} \rightarrow \text{Tense (Modal) (have + en) (be + ing)} \]

**THE PASSIVE TRANSFORMATION**

Let us now come to the passive transformation and look at one specific example:

(17) The man may have opened the door.

The passive form will be:

(18) The door may have been opened by the man.

The structural description (SD) of (17) is as follows:

\[ \text{NP1 + AUX + V + NP2} \]

The T-rule given below, which makes the passive transformation, can be applied to this SD:

\[ \text{NP2 + AUX + be + en + V + by + NP1} \]

Notice that be + en and by have been added to this rule. Let us now expand this as follows:

NP 1 and (have + en) + by are added in certain places.

This creates a deep level sentence which will be brought to the surface level.

D-Level: the door present may + have + en be + en open by the man
Applying Auxiliary Inversion Rule (Auxiliary is inverted. It is also called **affix hopping**): The door may + present have + be + en open + en by the man

Applying the morpho-phonemic rule (i.e. the way we pronounce the past of *open*, etc.), we do not say *openen*, we say *opened*. We call the present form of *have* simply *have* and the past form of the verb *to be*, *been*. So we have:

(18a) The door may + have + been + opened + by the man

i.e.

(18b) Surface Level: The door may have been opened by the man.

This might seem to be a very artificial and roundabout way of deriving a passive sentence. However, the principle is quite significant. It is that complex sentences are not created like simple sentences. Simple sentences, or kernel sentences, are created at the deep level and they are changed to complex sentences by transformation rules.

To read Radford’s book you do not need any more knowledge of syntax than you already possess after having read this book. The conclusion of his detailed treatment of V movement, I movement, NP movement, WH movement etc. is the intriguing rule, **alpha movement**: Move alpha (where alpha is a category variable, i.e. designates any random category you care to choose) (Radford 1988: 537).

This rule is intriguing because it appears to be a recipe for syntactic anarchy. Can we really move any category anywhere in a sentence? The answer is that this is a rule of Universal Grammar, whereas in every language there are particular rules which prevent certain kinds of movements. These restrictions are the ones which linguists have to find out and write down. For instance, the structure must be preserved. The rule for this is written as follows:
Structure-preserving principle: Major grammatical transformations are either root or structure-preserving operations (Radford 1988: 539).

A root transformation is one that applies to a root clause, i.e. the main or independent clause. A non-root transformation is one in which the transformation is such that it can be generated by the grammar in that position.

In short, if we extend this principle further, we come to the conclusion that all transformations are structure-preserving. This would make sense even if you do not go into the rather complicated arguments of linguists, because sentences would not make sense unless there were some rules to tell us how we can change them. In order to understand how such arguments are given, you should read the detailed chapter on alpha movement in Radford (1988: 527-584).

Next, you may read Yule (1985), Lyons (1970), and Radford (1988). If you want to read more about transformations and Chomsky’s theories, the next step would be to read Radford (1997), before going on to Hageman (1991) and more recent works.
In order to understand how the mind puts words in a sentence in order, we will consider some more theories. One of them is the government and binding theory, which is based on Chomsky’s work of the 1980s. However, the simplified account given in this chapter is based upon Liliane Haegeman’s excellent and accessible explanation of the theory (Haegeman 1991). Let us begin by looking at the idea of dominance.

**DOMINANCE**

If we say that a node dominates another node, we mean that it occurs higher up the tree than the latter, and is connected to it by an unbroken set of lines. A more technical definition is as follows:

Node A dominates node B if and only if (iff) A is higher up in the tree than B and if you can trace a line from A to B going only downwards (Adapted from Haegeman 1991: 122).

In the sentence below:

(1) I will leave my horse.

One can represent this in the tree-form as follows:
Here Node A (S) dominates node B (NP), as it is higher up the tree-than NP and because one can trace a line from S to NP going only downwards. S also dominates VP like it dominates NP with no node in between. This is immediate dominance. Here both NP and VP are sister nodes.

This gives us a definition of government.

Government: A governs B if

(a) A is a governor

(b) A and B are sisters (Adapted from Haegeman 1991: 123). Governors are heads and if x is a head and it governs Y, then x head-governs Y.

Another aspect of governance is the notion of C – command. The rule of C-command is:

Node A C-commands node B iff

(a) A does not dominate B and B does not dominate A.

(b) The first branching node dominating A also dominates B.

Consider the following tree:

```
  S
 /|
/ |\nB  A
 /\  /
NP VP
```

If we want to know which node VP C-commands, we should move up till we reach the first branching node and then we come down the NP node. Then NP is C-commanded by VP. Similarly NP C-commands VP. Government, by this definition, turns out to be a matter of ‘mutual c-command’—something which politicians would never agree with! Let us put this new definition in the following form:

**Government** (revised definition): A governs B iff
(a) A is a governor.

(b) A C-commands B and B C-commands A.

Refining this further, there are barriers to government. Chomsky has written about them in Chomsky (1986 b: 8). To put it simply, if there is a node in between the two nodes, then it is a barrier to government. Let us explain this further.

**Government**

If the node X is fully projected, this is called maximal projection. The term **M-command** is used when X is a maximal projection. Let us define government again:

A C-commands B iff A does not dominate B, and every X that dominates A also dominates B.

Here X is not the first branching node (this would be strict C-command). But now X is the maximal projection. Then we say that A M-commands B. Now we turn to Chomsky’s definition of government.

A governs B iff A M-commands B and no barrier intervenes between A and B. Maximal projections are barriers to government.

Governors are heads (Chomsky 1986:8).

Let us now turn to binding. It is also based on Chomsky’s work of the 1980s. It is based on a simple notion: how are references to an NP interpreted? For example, how do we interpret a sentence like

(2) *My uncle* hurt *himself*.

Of course, the reflexive pronoun ‘himself’ is bound or indexed to ‘my uncle’, but how do we understand this? Or rather, how do we think the mind must have processed this idea?

First, quite clearly, the mind must be indexing the two versions of the nouns ‘my uncle’ (NP₁) and ‘himself’ (NP₂). In this case NP₁
is the antecedent of ‘himself’. We can say they are co-indexed and give them the same number so that we know they refer to the same person.

(2a) My uncle, hurt himself.

The two noun phrases are co-indexed or bound with each other. The antecedent is the **binder** of the reflexive.

There are barriers to binding also. For instance, look at the following sentence:

(3) My uncle thinks that I have hurt himself.

Here the reflexive (‘himself’) is not locally bound. The antecedent (‘uncle’) is too far away from it and it is not in the same clause. Thus, it cannot bind the reflexive. To bind the reflexive, the antecedent must be in the same clause. This is known as the **clause-mate condition**. Let us state it in more detail.

**Principle of reflexive interpretation**

A reflexive must be bound by a clause-mate antecedent. **Binding** can then be defined as follows:

A binds B iff

(a) A c-commands B.

(b) A and B are co-indexed.

Binding by antecedents is known as **A-binding**.

Besides reflexives, we have reciprocals which are bound to their antecedents. The following sentence illustrates this.

(4) My two uncles like each other.

Here, ‘my two uncles’ is the antecedent of ‘each other’. The term **anaphor** is used to refer to both reflexives and reciprocals. NPs, then, are of three types:
(1) R-expressions. These have independent reference. They do not need an antecedent
(2) Anaphors
(3) Pronouns
If we say:
(5) Uncle likes her.
Here ‘uncle’ is an R-expression. It is free. And so is the pronoun ‘her’.
We have already given examples of anaphors and reflexives above.

BINDING THEORY
Let us sum up the Binding theory.

First Principle: An anaphor must be bound in its governing category.
Second Principle: A pronoun must be free in its governing category.
Third Principle: An R-expression must be free everywhere.

But the governing category is not so easy to decide. Let us discuss it in a little more detail.

We said earlier that a reflexive must be bound by a clause-mate antecedent. But suppose there is a sentence such as:

(6) My uncle believes himself to be the best.
Here ‘my uncle believes’ is one clause and ‘himself to be the best’ is another clause. Thus, the reflexive is not a clause-mate of the antecedent. Yet, the sentence is grammatically correct. This means that our earlier rule was wrong.
The new rule is as follows:

A reflexive must be bound inside a clause that contains its governor.

According to this rule, the reflexive need not have an antecedent in the same clause. If the governor of the antecedent and that of the reflexive are the same, then the reflexive will be bound to its antecedent.

(7) \[\text{My uncle} \quad \text{believes} \quad [\text{himself} \quad \text{to be the best}]\]

As we can see the verb ‘believes’ case-marks the reflexive ‘himself’. This means that the reflexive is governed by the verb ‘believe’ and so we can extend its domain. By this extension, we bind ‘my uncle’ with ‘himself’.

The theory we have outlined above helps us understand how sentences are processed in the brain. These are theories and they keep changing because this is how science moves forward.

However, the account given above is very simple. If you want to study the theory in detail, read Liliane Haegeman’s (1991) account of it.
Universal grammar (UG) is defined by Chomsky as follows:

[It is]...the system of principles, conditions, and rules that are elements or properties of all human languages—the essence of human language.

(Chomsky 1976: 29)

This definition is fairly simple and does not need any further explanation. In fact, we have examined some of the rules and principles of one particular language, English, in the previous chapters. Just as English has rules and principles, so do all other languages. The linguist’s task is to find out if all languages have some basic principles which are universal. These principles would be shared in common and may be called universal grammar.

**The Principle of Structure-Dependency**

We have touched upon the idea of structure-dependency earlier. We said that the fact that languages have structure and movement proves that we must be able to understand this structure. For example, consider the sentences below:

(1) I gave sweets to him.
(2) Who gave the sweets to him?
(3) *Main ne us ko mithai di.* [Urdu-Hindi]
(4) *Kis ne us ko mithai di?* [Urdu-Hindi]

In both questions (2) and (4), the word which introduces the question (‘who’ and *kis*) comes at the beginning of the sentence.
This is not a random process, but comes from wh-movement, which depends upon the structure of these languages. Wh-movement takes place in other languages too. In fact, we can say that it is a universal feature of human languages and all languages are structure-dependent. Since human languages are structure-dependent, we give rules like.

\[ S \rightarrow NP + VP \]

These rules have been refined and changed as you have read in the discussion of the x-bar theory given earlier. However, for the sake of simplicity, we will use the older rules.

In this rule, we mention categories (i.e. NP and VP) and we give their relationship in the structure of a sentence. If there was no such principle, we would have given some other kind of rule, for example, ‘Place the words in alphabetical order’. But that will not allow us to generate grammatical sentences. It is essential, therefore, to understand that structure-dependency is the only theory that helps us to understand how languages function.

**THE HEAD PARAMETER**

**Parameter** may be defined as a variable which may be given a series of values. In other words, it is something which can change in value, but which does refer to a ‘heading’ or measure under which similar things can be placed. For example, a tailor’s parameters would consist of headings such as length of leg, length of arm, width of shoulder, and so on. All of these parameters would vary from client to client, but all of his clients would possess these measurements.

Head, as we saw earlier, means the ‘head of a phrase’. For instance, in the NP *a girl*, the head is the N (= girl). In talking about the head parameter, we are concerned with the place of the head in various languages. The shared or universal fact is that all languages have phrases with heads, i.e. NP, VP, PP, and so on, but the head may be either at the beginning of the phrase or at its end. Thus, the head parameter varies from language to language, from head-
left, to head-right in structure. Let us look at the placing of the head in Urdu-Hindi and English below.

(5) The picture is good. [English]

(6) Tasveer acchi hai. [Urdu-Hindi]

In (5), the VP = ‘is good’ and the V = ‘is’. This V comes at the head which is on the left. In the Urdu-Hindi sentence (6), the VP = ‘acchi hai’ and V = ‘hai’; that is to say, the head comes on the right.

Similarly, according to Cook (1988: 7), Japanese is also a language in which the head comes at the end of the VP. This feature is common to Urdu-Hindi, Punjabi, and so on, but it is not shared with English. Thus, in the choice of placing the head in a phrase (i.e. filling the head parameter), UG gives us two possibilities: the head can be first or last in the phrase. Individual languages give specific rules as to whether it should be put first or last in that particular language. People who learn that language learn that specific rule.

**THE PROJECTION PRINCIPLE**

We have touched upon this principle in the chapter on semantics, but let us study this in more detail. We know that every word, or
lexical item, has properties which decide whether it is suitable or unsuitable for playing a certain role within the syntax of any given sentence. Thus, we can give the general principle that:

The properties of lexical entries are projected onto the syntax of the sentence.

To explain this further, let us take the example of the verb *put*, which we used earlier. As we saw earlier, *put* is always followed by an NP and a PP of location.

(7) I put my car in the garage.

Here *put* is followed by the NP (= my car) and the PP (in the garage). Just saying

(8) *I put.
(9) *I put the car.

or

(10) *I put in the garage.

would be ungrammatical. Thus, we can sub-categorise *put* as follows:

\[ \text{PUT} (= V) \rightarrow \text{NP PP \{loc\}} \]

This is to say, *put* is a verb which must be followed by an NP and a PP of location.

In the same way, *like* is a verb which must be followed by an NP because you must like someone or something.

(11) I like mangoes.

whereas just saying

(12) *I like.

would be wrong. Thus, we can give the following information for *like*:

\[ \text{LIKE} (= V) \rightarrow \text{NP} \]
But if you say
(13) I smile.
(14) I go.

you will be correct. *Smile* and *go* do not need an NP to follow them. They are called **intransitive verbs**. Now let us turn to verbs, both transitive and intransitive, and talk about the way they are projected onto the syntax.

Verbs are said to have arguments. The term **argument** comes from logic but it is used in linguistics in the following way. Let us consider the verb *defeat*.

1. The Zombo defeats the Zombi.

In this ‘defeats’ has two arguments: the noun ‘the Zombo’ and the noun ‘the Zombi’. In traditional grammar, you could parse the sentence into subject and predicate. The predicate defines some relation between the two nouns. This is a two-place predicate or a two-argument predicate.

Now if we say
(16) I shouted.

then ‘shouted’ is a one-place or one-argument predicate.

In traditional grammar, as you will recall, ‘defeats’ is a transitive verb and ‘shouted’ is an intransitive verb. Transitive verbs are followed by a noun or a clause. Intransitive verbs can stand on their own. They do not need to be followed by nouns or clauses etc. For example:

(17) * I am fond
(18) I am fond of horses.

The verb ‘fond’ needs to be followed by something, i.e.

\[ \text{PP (of } \text{NP ((horses)))} \]
In addition to these two, traditional grammar recognizes **ditransitive verbs**. These require a direct object and an indirect object. The verb *give*, for instance, is ditransitive.

(19) I will give my pen to you.

In this example ‘my pen’ is the direct object and ‘to you’ is the indirect object. We can represent this as: give (NP ((PP))).

We can also have another NP instead of the PP. For instance:

(20) I will give you my pen.

which can be written as: give (NP ((det + NP)))

This consideration of argument structure leads us to what is called **theta theory** (or θ-theory). Let us consider the arguments of the verb *kill*.

(21) The snake killed the mouse.

*Kill* has two arguments: ‘the snake’ (NP₁) and ‘the mouse’ (NP₂). The snake is the **agent** because it has agency or power to take an action (that of killing). The mouse (object) undergoes the action of being killed and is the **patient** of this activity.

In short, one has to look for the meaning of verbs in order to understand how they are related to their arguments. These relations are called **thematic roles** or **theta roles** (θ-roles). It is the meaning of the verb which gives sense to its arguments. This is called marking the arguments. In simple words, this marking, or theta-marking (θ-marking), determines whether the argument will be agent, patient, or something else.

Some thematic roles (θ-roles) you will come across are:

a. **Agent (or actor)**: The one who initiates the action (for example, the action of killing in (21) above).

b. **Patient**: The one who undergoes or suffers the action (for example, the mouse who is killed in (21) above).
c. **Theme:** The person, thing, phenomenon, state of being etc. moved by the action (for example, same as above in this case).

d. **Experiencer:** That which experiences some state, psychological or otherwise, expressed by the predicate (for example, same as above in this case).

e. **Beneficiary (or benefactive):** The entity which benefits from the action expressed by the predicate. The above example has no beneficiary but the following one does:

(22) The snake gave the mouse to the cat.

Here ‘the cat’ is the beneficiary. Of course, ‘the snake’ is the ‘agent’ and the ‘mouse’ is the ‘theme’.

f. **Goal:** The entity towards which the activity expressed by the predicate is directed. For example:

(23) The mouse fell near the cat.

In this case the ‘mouse’ is the theme and the goal is ‘near the cat’.

g. **Source:** This refers to the entity from which something is moved.

(24) The cat brought the mouse from the snake

Here ‘the snake’ is the source from which ‘the mouse’ is moved.

h. **Location:** This refers to the place in which the action or state expressed by the predicate is located. For instance:

(25) The mouse is in the mouth of the cat.

Thematic roles (θ-roles) are assigned by the meaning of a verb. In short, the information given in a word (lexical information) is represented in the way the sentence is constructed (syntactical
representation). This is called the projection principle which is expressed as:

Lexical information is syntactically represented. Further, each argument is assigned one and only one theta role.

Thus, words themselves determine where they can be placed in the syntax. Thus, when we put the words from the lexicon into the syntax to create the structure of sentences, those words fit in according to their lexical properties. That is why we say that lexical entities project their properties onto the syntax.

**Types of Universals**

We have given only three universal principles so far. The question that arises now is whether a principle which is not found in all human languages can be called a universal principle of grammar or not. For instance, movement, examples of which have been given in English and Hindi-Urdu, does not exist in the same form in all languages. In Japanese, question words (what, how, when, etc.), are not moved to the beginning of the sentence as in English. However, some other kind of change (or movement) does take place. As far as this syntactic movement is concerned, we can call it a parameter which varies from language to language. It is still universal, though all languages may not have it. Since it is universal, there cannot be a human language which will allow movement to take place without structure-dependency. If movement takes place, it will be structure-dependent, otherwise it may not occur at all.

This means that we can classify languages according to which principles of UG they share. These are called **implicational**, **statistical**, or **Greenbergian universals**: Cook gives the example of the presence of relative clauses in sentences (1988: 18-19). **Relative clauses** are clauses which are introduced by *that* or *who*. Data collected for a number of languages suggests that there is an order according to which languages take relative clauses. They can have only subject relative clauses, i.e. the subject of the clause is related to the noun.
(15) Nelson Mandela is the man who remained in jail.

The relative clause ‘remained in jail’ and its subject ‘the man’ are related to the noun ‘Nelson Mandela’. All languages permit this kind of relative clause. Some languages also permit relative clauses in which the object relates to the noun, and so on. This means that languages can be arranged in a descending scale according to the kind of relative clause they take:

<table>
<thead>
<tr>
<th>Relative Clause Type</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject relative clause</td>
<td>All languages</td>
</tr>
<tr>
<td>Object relative clause</td>
<td>Most languages</td>
</tr>
<tr>
<td>Indirect object relative clause</td>
<td>Many languages</td>
</tr>
<tr>
<td>Object of preposition relative clause</td>
<td>Some languages</td>
</tr>
<tr>
<td>Genitive relative clause</td>
<td>Few languages</td>
</tr>
<tr>
<td>Object of comparison relative clause</td>
<td>Very few languages</td>
</tr>
</tbody>
</table>

This represents a hierarchical model in which we can arrange languages on the basis of one variable—in this case according to what kind of relative clause they can take. It is also called the accessibility hierarchy. Note that this can also be represented as follows:

Subject > Object > Indirect Object > Object of preposition > Genitive > Object of comparison. (Here ‘>’ means ‘greater than’ or ‘more than’). This means that all languages allow subject relative clauses, many allow object ones, and so on.

**The language faculty**

As we have seen before, language is structure-dependent. But there are areas of the mind which are not structure-dependent. Thus, we are able to think about things such as poetry and mathematics. Children can learn how to walk. These activities seem to be different from learning a language. The kind of logic used in mathematics is different from the kind of principles used in language. Thus it may be said that the human mind possesses the faculty of acquiring language, and that this is a unique faculty, and is not just part of normal intelligence as such. As we have seen before, Chomsky calls this faculty the Language Acquisition Device
Now let us see what light Universal Grammar sheds on LAD.

When we learn a language, we use the LAD, which, according to Chomsky and his associates, uses the principles of UG. But these are general principles, and every language has its own particular rules. The rules merely determine the choices with reference to the values set out in the parameters. For instance, we have the UG principle that there is a verb (V) at the head of a VP. Then we have the choice that the verb (V) will either be at the left or at the right of the VP. The rules of Urdu-Hindi say that the head of the VP will be on the right, and the rules of English say that it will be on the left. Thus, when children acquire a language, they make use of the principles of UG resident in the LAD in their brains, and then set the values of the universal parameters according to their language. Thus, they learn the rules of their own language.

The basic thing to remember is the general principle of UG: We all possess certain basic rules for processing human languages in our brain at birth. When we hear a particular language, we set the ‘switches’ of our brain according to the particular rules of that language. Since the basic rules are already in the brain, language learning becomes easy for human beings.
The study of culture through language; the relationship between language and culture; and how world view is related to language are some of the concerns of anthropological linguistics (or linguistic anthropology).

Franz Boas, the famous American anthropologist who studied Native Americans (or Red Indians as they were called earlier), wrote an article suggesting that anthropologists should know the languages of the people they study (Boas in Hymes 1964: 15–26). The obvious reason is that if s/he depends on an interpreter, s/he will not be able to correct the bias of the interpreter or any particular view which the interpreter might have acquired from the people whose language s/he is interpreting. Another reason, which is crucial, is that language expresses thought:

It has been claimed that the conciseness and clearness of thought of a people depend to a great extent upon their language (ibid 17).

Boas thought that European languages have more terms for abstract concepts so that they can express abstract, philosophical ideas more easily than primitive languages. Nowadays, linguists do not agree with such views because thought is too complex a phenomenon to be explained by any one factor. Anyway linguistics became a part of ethnology—the study of peoples’ culture, world view, and behaviour (i.e. cultural and social anthropology).

Language was part of the study of culture even before Boas laid emphasis upon it. Sir Edward Tyler, a British anthropologist, gave

---

2 S/he means she or he. It has been used for the generic masculine pronoun he because feminists have raised the point that the use of ‘he’ alone ignores them.
importance to it in his books *Primitive Culture* (1871) and *Anthropology* (1881). R. G. Latham, another Englishman, had already written on the importance of language. Later, the famous British anthropologist Malinowski wrote ‘Meaning in Primitive Languages’ (1923) and *Coral Gardens and Their Magic* Part 2 (1935). In both works, he was concerned with the differences between civilized and primitive languages. (See his supplement to Ogden and Richards’ *The Meaning of Meaning*, 1923). As we have seen, Ferdinand de Saussure influenced thinking in both sociology and anthropology. As discussed earlier, Saussure differentiated between *la langue* (the system of the languages, which is an abstract thing) and *parole* (concrete utterances or instances of use of the system). We shall study later how the anthropologist Claude Levi-Strauss used Saussure’s ideas.

In the United States, much work was done on the American Indian languages, not only in order to understand the languages themselves but also to find out about the culture of the native tribes (ethnology, ethno-semantics) through language. Franz Boas was a pioneer in the field. He edited the *Handbook of American Indian Languages* (1911) and pioneered courses at Columbia University. Edward Sapir, another American linguist, was Boas’s student. Other eminent linguists like Alfred Kroeber and Bloomfield were influenced by him.

Edward Sapir compared languages and his contribution to theories about the influence of language on thought (the Sapir-Whorf hypothesis) is well known. Benjamin Lee Whorf is also associated with this theory. Anthropological linguistics is still developing in the USA, though the Chomskyan tradition is better known.

We have seen how structuralism came to dominate linguistics. The structuralists emphasized upon the system—an abstract entity—but this emphasis was not accepted by some socio- and anthropological linguists. For Hymes, it was the speech act (*parole*) which was important (1964: 63).
Malinowski saw language as a mode of action, not an instrument of mere reflection (1923: 312). His approach is called functional. The function of an institution is the part it plays in fulfilling the needs of individuals. Language maintains social solidarity. Even idle chatter or meaningless formulaic utterances keep the channels of communication open. This is called phatic communion and its purpose is to bind people together.

Society can also be seen from the point of view of interacting persons or groups. This is interactionism, and the ethnography of speaking—how people actually use language in society—is in this tradition.

THE AIMS OF ANTHROPOLOGICAL LINGUISTICS

The aims of anthropological linguistics, as described by Dell Hymes are as follows:

1. Evaluation of differences and similarities among languages (especially exotic ones).
2. The significance of linguistic patterns for the basic outlook of a people (i.e. their world view).
3. The relation between a people’s vocabulary and their other interests.
4. How speaking enters into norms of interaction among persons.
5. How the motives of play and art are manifested linguistically.
6. The relation between levels or varieties of speech on the one hand and types of community and their boundaries on the other.
7. How social factors enter into linguistic change.
8. Modes of classifying and interpreting resemblances among languages, especially their historical import (1964: xxiv).
Thus, linguists came to focus upon worldview, ‘acts of speech’—how one uses language in the social context—or the ‘ethnography of speaking’.

**Levi Strauss’s use of linguistic concepts**

Let us now see how Levi Strauss used some of the concepts of linguistics to study culture.

One of Levi Strauss’s papers called ‘Structural Analysis in Linguistics and in Anthropology’ gives us an insight into his methods. He says that the concepts of phonology can be used to study social systems such as kinship phenomena. As we have seen, the relations between terms are very important in phonology. The features voiced and voiceless (+ voice and –voice) are in binary opposition. They distinguish many phonemes such as /p/ and /b/ etc. Similarly, two elements may be in complementary distribution, i.e. they are the same thing but they take different forms. Thus, allophones are the same thing. /p/ has two allophones, /p/ and /pʰ/. We use /p/ in the end or middle of a word, and /pʰ/ in the beginning, but they do not change the meaning.

These concepts of binary opposition and complementary distribution etc. may be used in describing social systems such as kinship relations. One of the ways in which this is done is componential analysis, which will be touched upon later. Here I shall concern myself with Levi Strauss’s study of the avunculate.

By the term avunculate, Levi Strauss meant the concept of the relationship between uncles and nephews/nieces. An uncle can be one’s (i.e. the ego’s) father’s brother or mother’s brother. There may be a social system according to which this uncle’s relations with his nephews/nieces may be strict, or informal and affectionate. The uncle’s attitude as part of a system, and not only because of his temperament, is what Levi Strauss is interested in.

To understand this, we can take the example of the institution of the relationship between a newly married bridegroom and his wife’s sisters in Pakistan and many parts of India. The man (doollah or
jeeja ji) can joke with his wife’s sisters and girl cousins (salian) and vice versa. However, he is supposed to be strictly formal with his parents-in-law. This joking relationship is allowed by society and in some cases a day is fixed for it. On this day, the girls and the groom indulge in playful ragging which sometimes leads to frivolous horseplay (this is called chauthi, i.e. the event of the fourth day in some communities).

This kind of system was studied by a number of anthropologists including Radcliffe-Brown and Malinowski. It is this which Levi Strauss takes up for analysis using concepts borrowed from phonology.

If the maternal uncle is feared and exercises authority on his nephew, the relations between them are strict. When this happens, the relations between fathers and sons are familiar and affectionate. If the relations of fathers and sons are strict, those between nephews and maternal uncles are familiar.

In short if one pair of relations is known, the other can be deduced. According to Levi-Strauss:

The two groups of attitudes form, then, as the phonologist would say, two pairs of oppositions (Levi Strauss 1945: 45).

Thus +/– can be used to refer to this relationship.

But this is not all. These relations can further be seen in relation to the relations between men and their wives, and brothers with their sisters. Here, too, there are binary oppositions.

Societies which allow tender relations between husband and wife (i.e. the public exhibition of intimacy), insist upon strict relations between brother and sister. The reverse is also true: if brother and sister are intimate and tender, husband and wife are supposed to be formal and strict in public.

Now, if the relations between uncle/nephew and brother/sister are tender (+ positive), then the relations between father/son and husband/wife will be strict (- negative). Roughly speaking, this is the position in some areas of Pakistan and parts of India.
This can be depicted in a chart as follows:

```
+-----------------+-----------------+
| Wife            | Husband         |
+-----------------+-----------------+
| Sister          | Father          |
|                 |                 |
+-----------------+-----------------+
| Uncle           | Nephew          |
+-----------------+-----------------+
| Brother         | Son             |
```

The law is:

The relation between maternal uncle (mamun) and nephew is to the relation between brother and sister, as the relation between father and son is to the relation between husband and wife.

Thus if $mu + n \rightarrow +$ then $b + sis \rightarrow +$

$f + s \rightarrow -$ then $h + w \rightarrow -$  

(mu = maternal uncle; n=nephew; b=brother; sis=sister; f=father; h=husband; s=son; w=wife)

**COMPONENTIAL ANALYSIS**

Kinship systems can be understood in general through the application of linguistic concepts. A kinship system contains, in essence, a certain relationship. The relationship can be by marriage, by descent or by consanguinity (being brother and sister or cousins etc.).

The idea of using binary oppositions to contrast meaning is the basis of componential analysis. We have already come across this
idea in semantics. Let us now see how it is used in understanding aspects of culture.

Goodenough and Lounsbury, for example, have constructed studies of kinship systems using componential analysis (Goodenough 1956; Lounsbury 1964).

Before we deal with componential analysis, let us understand the vocabulary, or metalanguage, used in it. Here are some basic terms:

1. **Designata**: It is a linguistic expression which designates a class of images or concepts. The singular is designation and the plural is designata.

2. **Significata**: The significata are the necessary or definitive attributes of this designated class. They are the properties of the things which form that class. Other things which do not have these properties are excluded.

3. **Denotata**: The denotata are the things we refer to.

There are other attached or associated meanings of the linguistic expression called **connotata**, but componential analysis is not concerned with them.

Let us take the series of sounds ‘dog’. This linguistic expression designates a class of images. The attributes of the members of this class are that they are animals with four feet, a long snout with teeth, and a tail. They are domesticated and they bark. They are members of the canine family, which wolves also belong to. However, they are different from wolves, jackals etc. They are also different from cats and horses etc. These are the significata of dogs. Thus, the significata are the necessary attributes of a dog. So when we use the word ‘dog’, it denotes animals having the attributes mentioned above as significata. All animals which belong to the canine family, which bark, and are domesticated are called dogs.

The term ‘dog’ also connotes faithfulness or a low despicable nature and may be used for human beings. That is its connotation.
Let us now list the denotata of ‘uncle’. If we take all kinds of people called uncle. They are:

mother’s brother
father’s brother
mother’s sister’s husband
father’s sister’s husband etc.

Let us now abstract the essential distinctive features:

Uncle  + male
       + senior  [i.e. of a senior generation]
       + 2  [i.e. 2 degrees of genealogical distance from ego]
– lineal  [i.e. one has not descended from him]
– marriage tie  [i.e. ego is not connected by marriage ties, but ego’s parents may be].

As you can see, if any of these terms are changed, we get someone else who cannot be called ‘uncle’. The word or words used for ‘uncle’ may be absent or may mean different things but we can find out what a term actually refers to.

Before going ahead, however, let us explain the abbreviation we will use in this analysis. The capital letters are used as follows:

F = father, M = mother, B = brother, S = sister, H = husband, W = Wife

The small letters stand for:

s = son, d = daughter

Now let us see how these symbols and used in componential analysis:

The anthropologist Goodenough wrote about the American Indian (Iroquois) society in which one’s sex determined whether one’s brother’s daughter was to be called ‘niece’ or ‘daughter’. This is very different from the Indo-European languages, but it is a way
of classifying reality which becomes clear in terms of componential analysis. The formulas are:

1. For senior generation kin:
   Is there a sex equivalence between ego’s linking parent and alter?

2. For junior kin:
   Is there a sex equivalence between ego and alter’s linking parent?

3. Same kin of the same generation:
   Is there a sex equivalence between ego’s linking parent and alter’s linking parent?

In these formulas, the key terms are defined as follows:

- **sex equivalence** = having the same sex. For example, nephews and uncles have the same male sex.
- **ego** = the person from whom a relationship is traced.
- **alter** = the person who is referred to in the kinship term.
- **linking parent** = the parent who provides a link. For instance, in father’s brother as well as father’s brother’s son etc. the linking parent is *father*.

Let us take the senior generation first. The question whether the term *father* can be used for more than one male of this generation can be determined by answering the question: Is there a sex equivalence between ego’s linking parent and alter?

Let us take a boy as ego. The relation will be traced from him. Let the alter be his father (i.e. the biological mate of his mother). The linking parent too is the father. As we can see the linking parent and the alter are the same person. We can say that there is a sex equivalence between the linking parent and alter (both being male). Thus, the boy will call his real father as father.
A girl too will call her father as father since the linking parent and alter is the same person in this case too.

Now let us see how many other relatives are linked by the linking parent, i.e. father. These are:

- F B: Father’s brother
- F Mss: Father’s mother’s son’s son
- F F Bs: Father’s father’s brother’s son
- F M Bs: Father’s mother’s brother’s son
- F Fss: Father’s father’s son’s son
- F F F Bss: Father’s father’s father’s brother’s son’s son etc.

All these males are linked by the male linking parent to the ego (the boy). Since there is a sex equivalence between the linking parent (father) and the alter (FB etc.), all these males will be called father by him.

Now let us take the terms for junior kin. The question is who shall be called son, daughter, nephew or niece. The question to be asked is whether there is sex-equivalence or non-equivalence between ego and alter’s linking parent.

Now suppose the ego is a man. The alter is a boy of the younger generation (< generation + male – adult). This man is linked to the boy by his brother—the boy being his brother’s son (Bs). The ego (man) and alter’s linking parent (his brother) are both males. Thus there is sex-equivalence between both. Hence, the boy will be called his son.

Now consider the same man again. Let the alter again be a boy of the younger generation. However, this boy is the son of his sister (Ss). The ego (man) and alter’s linking parent (his sister) are of different sexes. Thus, there is no sex-equivalence between both. Hence, the boy will be called his nephew and not his son.

Now let us take the case of a woman, i.e. ego is female. The alter is her own biological son. The ego (woman) and the alter’s linking
parent (her own self) are of the same sex. Thus, there is sex equivalence between the two and her son will be called her son. However, her brother’s son becomes a nephew, while her sister’s son is again her son.

The terms are given as follows:

For male ego:

son : s, Bs, MSss etc.

For female ego:

son : s, Ss, MSds etc.

Let us take the last case now—that of the same generation (i.e. = generation). Here, the question to be considered is whether there is a sex equivalence between ego’s linking parent and alter’s linking parent.

If the ego is a boy and the alter is a boy who is his father’s brother’s son (FBs), the linking parent is the father. As the ego’s linking parent (his father) and the alter’s linking parent (the alter’s father) are of the same sex, he will be called brother. If the alter was his father’s sister’s son, then he would have been called cousin and not brother, because the linking parents would be of different sexes.

**Lounsbury’s Use of the Term ‘Parallel’ in Componential Analysis**

According to Lounsbury, we can simplify our formulas even further by using the term + parallel instead of sex equivalence. The idea on which + parallel is based upon is called complementary distribution. It is often used in phonology. The theory is that if two or more linguistic elements are (1) similar in form, and (2) non-contrastive in function, they are variant forms of the same element.

Thus, as we saw in the chapter on morphology, the morpheme –s is pronounced as /s/, /z/, and /ɪz/ in cats, dogs, and buses respectively. These are variant forms of the same morpheme in complementary distribution. This means only one can occur in a given situation.
They are not distributed at random, but are distributed according to fixed laws. Thus, after the unvoiced ending /t/, only the unvoiced variant /s/ occurs. After the voiced ending /g/, the voiced ending /z/ occurs. After the hissing sound (sibilant) /s/, only /ız/ occurs. They complement each other being the same thing, but their distribution follows certain laws.

In the same way, the sex equivalences too are really the same thing but it has three kinds of distributions: (1) for senior generation (>), (2) for junior generation (<), and (3) for the same generation (=). If we call it + parallel, we are using one term for something which is in fact a single phenomenon. By + parallel we mean equivalence of sex and by—parallel we mean non-equivalence of sex. But the formula is a little more specific:

+ parallel means equivalence of sex between the two kin of the generation above ego or alter (whichever is junior).

– parallel means non-equivalence of sex between the two kin of the generation above ego or alter (whichever is junior).

As we have made a single formula to cover all the relationships considered earlier, it appears somewhat complex. However, it is quite clear if we work it out.

Let us take only one example: father

It is > generation + parallel + male

Let us concentrate only on + parallel. According to the definition, there should be equivalence of sex between the two kin of the generation above ego or alter (whichever is junior).

Let us try out: F, F B, F M Ss, F F Bs only

ego = a boy

alter = his biological father
The junior is ego in this case (the boy). The two kin of the generation above the ego are the father and the linking parent who is also the father. Thus, the father will be called father.

Now take FB. The junior is again the ego. The two kin of the generation about the ego are the linking parent (his father) and the father’s brother. Since they are of the same sex (+ parallel), FB will be called father.

Now take FMSs. The junior is the ego, who is the boy. The two kin of the generation above the ego are the linking parent (his father) and the father’s mother’s sister’s son. They are of the same sex. Hence, the alter will be called his father, and so on and so forth. If you work out all the formulas carefully, you will understand them. Some of them are given below:

- mother : kin > generation + parallel – male
- uncle : kin > generation – parallel + male
- son : kin < generation + parallel + male
- cousin : kin = generation – parallel
- sister : kin = generation + parallel – male

(Adapted from Leech 1974: 246).

Kinship analysis has a mind teasing quality but that is because of the number of relationships involved. In any case, most societies do not have such complex formulas of kinship. The point to remember is that linguistic theories can be used to understand kinship systems.
Languages are spoken by human beings who live in societies. The study of language in relation to society is called sociolinguistics. This is a vast subject and the present chapter cannot even introduce all aspects of it, let alone do justice to them. This is because there are so many subjects one can investigate that only a very brief list of some important ones is possible here. Among these subjects are: the varieties of language; language and world view; language and gender; linguistic politeness; and identity and language planning. Some of these subjects, like the relationship of language with identity and language planning, are so vast that they are covered under a different sub-heading of their own. Ethnic identity and aspects of language planning are covered under language politics in this book. Subjects not covered here are the language of schooling and issues of the medium of instruction which are focused upon in educational linguistics. The other issue not dealt with here is the analysis of conversation and the ethnography of communication. Ethnography of communication is also part of anthropological linguistics. Indeed, there are overlaps between anthropological and sociolinguistics because both focus upon human beings living in interaction with each other.

Let us then understand some basic concerns of sociolinguistics. First, do human languages vary? That is, when we say ‘the English language’ do we mean just one set of words in a single pronunciation and a grammar which does not vary at all? If not, then what is English? Let us understand the theory of varieties of language before we answer this question.
**Varieties of Language**

What we call languages are merely abstractions. Going back to our example of English, have you ever seen it? Does it exist? Well, dictionaries exist and books of grammar exist but the spoken language came first and these written products came much later. When it was a spoken language only, speakers existed and what they spoke could be called English if other speakers felt they were using the same sounds for the same objects. But, while doing so, speakers ignored the minor differences in the way other speakers pronounced words or used grammatical rules. If these differences are minor—such as the pet words a person might use—we ignore them or call them personal idiosyncrasies. Thus every individual uses a different (however minor the differences are) form of a language which we call his or her *idiolect*.

But sometimes we can understand the other speakers while being aware of quite clear differences in pronunciation, meanings, and even grammar. These are called the *varieties* of the same language. The varieties can be classified as below:

![Language Diagram]

The variety of a language according to its use is dependent upon its function and is called *register*. If you use it to teach medicine, you will put in it the special language of medicine (its technical jargon), derived ultimately from Latin and Greek. Thus, a fat person will be called ‘obese’ and inability to sleep will be diagnosed as ‘insomnia’.

Similarly there is a theoretical language—called a metalanguage—for all academic subjects as well as different professions. Thus, in
the language of the British Indian army ‘RV’ was used for ‘rendezvous’. But ‘rendezvous’ means meeting place, while in the register of the British Indian army it meant the place and the time of the meeting.

Remember that the register is meant for a certain use. It is not used outside that particular subject or profession. Thus, only a pompous or incompetent person will write and speak to everyone in the jargon of his profession or specialization. The best public lecturers are those who can hold people’s attention by not using the register of their specialization.

A special situation arises when a language has a specific high variety (H) which is used in all formal domains, while ordinary conversation is in the low variety (L). This concept was called diglossia by the linguist Charles Ferguson (in Hymes 1964: 429–439). Ferguson referred to the use of Arabic in the Arab world, Modern Greek, Swiss German, and Haitian Creole. He made the point that the H variety is nobody’s mother tongue nor is it used in ordinary conversation by anybody. It is learnt in schools and used in the class rooms, offices, courts, places of worship etc. The H variety has got great prestige because it is the language of classical literature and religious texts. Moreover, it unites people at least at the formal, written level. However, the L variety is the real living language of the common people. It should, however, be remembered that L is not made up of one language but has many varieties. We now take up the case of varieties according to the user.

As mentioned earlier, the idiolect is the special variety used by an individual. The dialect is the variety of a language used by a geographical area. Thus, the Pashto of South Waziristan, which is different from that of Peshawar, may be called a dialect of Pashto.

The important thing about dialects is that they are mutually intelligible. People who speak one dialect can understand the speakers of another dialect. It is, in fact, a variety of the same language and not a different language.
However, there are two views about dialects: (i) the traditional or popular view, and (ii) the modern linguistic view.

(i) The traditional view

According to this view, the standard language is the one which is used by educated people while writing books, lecturing, and in formal conversation. It is also used in the domains of power such as higher education, administration, judiciary, armed forces, and politics.

The dialects are spoken by uneducated people. They are stigmatized—i.e. seen in a negative way—by educated people. They are considered rough (uncouth), rustic, sub-standard, and inferior. They are considered intrinsically inappropriate for education and literature etc. Some people consider them distorted forms of the standard language, which is called the ‘real’, ‘actual’ or ‘pure’ language. They are also area-bound because they are used only by the people of a certain area. For instance, Maldivian is used in the Maldives islands.

The traditional view is hierarchical, i.e. people tend to regard some languages, or rather varieties of them, as superior to the others. They develop contempt for the speakers of these stigmatized varieties. Indeed, the fact is that they already have contempt for the speakers and this is transferred to the language they speak. Here is an example of such an attitude.

Slovenly and careless of speech, these Gullahs seized upon the peasant English used by some of the early settlers and the white servants of the wealthier colonists, wrapped their clumsy tongues about it as well as they could, and, enriched with certain expressive African words, it issued through their flat noses and thick lips as so workable a form of speech that it was gradually adopted by the other slaves and became in time the accepted Negro speech of the lower districts of South Carolina and Georgia.

Gonzales 1922:10. Quoted from McDavid 1980:78
The Gullahs were the African-Americans who were brought as slaves to the American cotton-growing states of the south. The writer’s contempt for the Africans is obvious and it is not for the language alone. It is also for their ‘thick lips’ and ‘flat noses’ and their dark complexion. Moreover, there is also contempt for the peasants from England and white servants, i.e. all the subordinated, working class groups from all societies. Arguably, it is this contempt for these groups which extends to their speech.

This example is from America of the 1920s but we in modern South Asia consider several of our languages inferior and make fun of them. Think about which language(s) you consider superior and which you think of as inferior. Then sit down and try to understand whether the language, or variety, you consider inferior is spoken by people who do not have a high social status in the eyes of the community you belong to.

One major objection to the traditional view of languages is that it is related to bias and encourages prejudiced attitudes in society. Let us now come to the other view.

(ii) The modern linguistic view

According to this view, the standard variety of a language is also a dialect. Thus all dialects are basically varieties which are mutually comprehensible. A language is, then, merely the sum total of all its dialects. This can be represented as follows:

\[ L = d_1 + d_2 + d_3 + \ldots + d_n \]

(where \( L \) = language and \( d_1 \ldots d_n \) are its dialects. The number may vary, so we write ‘n’ which may represent any number of dialects.)

As you will observe, dialects are real in the sense that people actually speak them. The language (L) is an abstraction because nobody speaks it. Everybody speaks a dialect. If someone speaks what we call the ‘standard language’, even that is a dialect of the abstract entity called ‘the language’.
If you look at the historical development of languages, you will discover that the standard variety was developed by language planning. The decision to make one dialect the standard was a political one—it happened to be the language of those who had power and money or for some other such non-linguistic reason. Thus, the language of the upper class of English around London was standardized. Standardization means fixing spellings, pronunciation etc. and writing dictionaries, grammars etc. Thus, languages are standardized for non-linguistic reasons.

Another point you should keep in mind is that all dialects are equal. Their structure is similar and there are no scientific reasons for calling one of them superior or inferior. However, the standard variety is meant to serve many functions—being used in education, law, commerce etc.—and has a larger vocabulary than other dialects. This means that the standard variety is functionally more useful for the purpose of modern life. However, if any other dialect had been standardized and modernized, it too would have been functionally just as useful.

As linguists, we say that everybody speaks or writes a dialect or variety, and by language we mean the sum total of all mutually intelligible dialects or varieties of a language. If we want to use the term ‘dialect’ in its traditional sense, we may call it a non-standardized variety of a language.

You should also remember that for non-linguistic reasons—ethnic identity, nationalism etc.—mutually intelligible forms of speech may also be called languages. Thus Hindi and Urdu are mutually intelligible (except in their extreme Sanskritized or Persianized varieties) but are called different languages. The languages of Sweden and Norway etc. are also quite similar but are called languages for nationalistic reasons. This may not be strictly correct in linguistic theory, but it is an established practice.

Now are we in a position to identify the English language? Well, but we come to the strange conclusion that there is no such thing as English at all. There are several varieties which are mutually
intelligible which we add together to call their sum total—English. There is, for instance, British Standard English, American Standard English, Australian Standard English, South African Standard English, New Zealandian Standard English, and Canadian Standard English. These are the standardized varieties of native-speakers’ Englishes. But then within Great Britain, there are several non-standardized varieties (dialects) such as Cockney, Scottish English, Yorkshire’s English etc. In the USA too, there are different varieties of English as the book *Varieties of American English* (McDavid 1980) testifies. And varieties are not associated with region only; they are also associated with class. In England, the upper classes speak what is called RP (Received Pronunciation). In America, the English of African Americans is called ebonics or African Vernacular English (AVE), and it is different from the English of middle class, white Americans. But the story does not end with native-speakers only. English is the second language of the educated elites of South Asia also. They use it with varying degrees of competence, but rules of their own first languages interfere with the way they speak and write English. These are non-native varieties of the language. In short, the idea of the English language makes sense only if we call it the sum total of all its varieties. But if it is the sum total of all its varieties, then there is no one monolith or entity which we can call English. In short, languages do not exist; only their varieties exist. Languages are an abstract concept; varieties are concrete systems which can be studied.

**LANGUAGE AND WORLD VIEW**

One of most interesting theories about culture and language came from Benjamin Lee Whorf (1897–1941) and Edward Sapir (1884–1939). The theory which is associated with their name is called the *Whorfian hypothesis* or the *Sapir-Whorf hypothesis*. It is also called the *theory of linguistic relativity*. It has two forms: (a) strong linguistic determinism, and (b) weak linguistic relativity.

Form (a) asserts that language completely determines thoughts. Thus, your world view is so strongly determined by your language
that you live in a different universe from speakers of other languages.

Form (b) argues that language makes the expression of some thoughts easier than others. Thus, we can understand other cultures but may find it difficult to express the concepts of these cultures in our language.

According to some linguists—(Hill 1988: 15)—Sapir and Whorf did not give the strong form at all. Some of their statements were a bit exaggerated and people thought they had given the strong form. As for linguistic relativity, there have been many experiments to confirm to what extent human languages are variable and how we can relate this variability to culture. The basic idea is to find out whether language reflects the world view of the culture which created it and uses it. This is, in fact, the cultural determinism of language. Then there is the question of finding out whether a language makes us think in a certain way, i.e. determines our world view? The second question, at least in its extreme form, has been ruled out. Our experiments will then confirm or disconfirm only the weak form of linguistic determinism.

One of the major ways in which linguistic relativity was tested was by comparing colour terms in different languages. Berlin and Kay (1969) gave 329 coloured squares to speakers of 98 languages with instructions to arrange them into the basic colours of their languages. By a basic colour they meant something which was not a mixed colour like ‘reddish-brown’ or ‘olive’. They found that colours were found distributed in languages according to certain rules. If a language had only two colour terms, they would always be black and white. If a third existed, it had to be red. The next category of languages had a choice between yellow or either blue or green (Berlin and Kay called it ‘grue’). The next category would have blue and green. Then brown was added and the last category had 11 colour terms and these were purple, pink, orange, and grey. This phenomenon was called implicational universals by the two researchers. The idea was that colour terms is a universal phenomenon and is predictable, so that if there are three terms you
can be sure that this implies they would be black, white, and red, and not blue, purple, and orange. This implies that the human mind is programmed to think in this manner and that the strong version of the Sapir-Whorf hypothesis is wrong. Our perception is not determined by our language alone. It is also determined by some common human ways or abilities of understanding reality (cognitive universals). However, speakers do categorize what they see in colour terms which they possess. Human beings see the same colours but refer to them the words of classification they possess. This supports the weaker version of the Sapir-Whorf hypothesis that language does influence perception.

**Language and Colour Terms**

If a language does not have a name for a colour, its speakers consider it the shade of another colour for which there was a name. In Pashto, for instance, the word *sheen* is used for green and blue. This does not mean, however, that green and blue are seen as exactly the same. They are seen as different shades of a colour which comes under the broader category of *sheen*. Sometimes, one could refer to grass to give the idea of greenness or to the sky for blueness. However, the way the colour spectrum (i.e. the outside world) is cut up is somewhat different from a person whose language has different terms for blue and green. Likewise, we consider scarlet, pink, shocking pink, purple, and other shades of redness as red. However, another culture might have a language with words which regard them not as shades of red but different colours. Then the speakers of this language will define pink and scarlet clothes not as shades of red but different colours.

The order is as follows:

\[
\begin{align*}
\text{white} & < \text{red} < \text{green} \quad \text{or} \quad \text{yellow} < \text{blue} < \text{brown} < \\
\text{black} & \quad < \\ 
\text{purple} & < \text{pink} \quad \text{or} \quad \text{orange} \quad \text{or} \quad \text{grey}
\end{align*}
\]

The sign $<$ means that the colour (s) towards which the tip of the arrow points will be there in a language containing the colour on
the broad end of the arrow. Thus, a language which has brown will
certainly have blue and the other colours on its left.

Let us order them according to the chart of Berlin and Kay:

<table>
<thead>
<tr>
<th>Type of Language</th>
<th>Number of Terms</th>
<th>Colours</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Two</td>
<td>white, black</td>
<td>Jale (New Guinea highlands)</td>
</tr>
<tr>
<td>2</td>
<td>Three</td>
<td>white, black, red</td>
<td>Tiv (Nigeria)</td>
</tr>
<tr>
<td>3</td>
<td>Four</td>
<td>white, black, red, green</td>
<td>Hanunoo (Philippines)</td>
</tr>
<tr>
<td>4</td>
<td>Four</td>
<td>white, black, red, yellow</td>
<td>Ibo (Nigeria)</td>
</tr>
<tr>
<td>5</td>
<td>Five</td>
<td>white, black, red, green, yellow</td>
<td>Tzeltal (Mexico), Pashto (Pakistan and Afghanistan)</td>
</tr>
<tr>
<td>6</td>
<td>Six</td>
<td>white, black, red, green, yellow, blue</td>
<td>Plains Tamil (India)</td>
</tr>
<tr>
<td>7</td>
<td>Seven</td>
<td>white, black, red, green, yellow, blue, brown</td>
<td>Urdu and Hindi (South Asia)</td>
</tr>
<tr>
<td>8</td>
<td>Eight, Nine, Ten or Eleven</td>
<td>white, black, red, green, yellow, blue, brown, purple, and/or pink, and/or orange, and/or grey</td>
<td>English</td>
</tr>
</tbody>
</table>

Adapted from Leech 1974: 235

As you can see, types 3 and 4 are different only in one respect. In
type 3, green is the fourth colour, whereas in type 4 it is yellow.
Both are four-term languages. Similarly, there can be no ordering
in purple, pink, orange, and grey. Any set of colours can be there
in any language.

Berlin and Kay also gave an evolutionary hypothesis: that languages
evolve from having fewer terms for colours to having more. The
idea is that as society becomes more complex, it needs to
differentiate colours more precisely so that what were considered
shades of a basic colour are given the status of colours. Thus, new
terms for colours are added to the vocabulary.

Can you think of additions of colour terms in Punjabi, Tamil,
Singhalese, Nepalese, or any other language of South Asia? Arrange
the colour terms in these languages and see if you confirm or
disconfirm Berlin and Kays’ hypotheses.
KINSHIP TERMS AND WORLD VIEW

Let us see whether our world view is reflected in our kinship terms. In English, we have just the term ‘cousin’. If we want to be precise, we can call girls ‘female cousins’ and boys ‘male cousins’. See the English sentence below.

(1) I met my cousin.  (English)

Sentence (1) does not tell you anything whether this cousin was male or female and related from father’s or mother’s side. Now see sentence (2).

(2) Main khala zad bahen se mila.  (Urdu)

Literal translation: I met my mother’s sisters’ daughter (who is like a sister for me).

Meaning: I met my mother’s sister’s daughter.

Sentence (2) makes it clear that the speaker met a female related to him in such and such a way. This only implies that language, such as the possession of colour terms or kinship terms, facilitates the expression of certain thoughts. It does not make it impossible for us to express them but we may not have terms corresponding to what we have in our mind. This is what makes translation such a difficult thing. This supports the weak version of linguistic determinism but not the strong one.

What is especially significant is that we can find out something about a culture through its language. We can find out about its world view for instance. If there are many terms for ‘cousin’, as there are in Siraiki, it must be a culture where the family, or kinship, is given great importance. In other words, it would not be a modern individualistic (Western) kind of culture.

LANGUAGE AND GENDER

Sex is a biological reality, i.e. human beings are mostly born as male and female. Some very few individuals may be born as having the physical features of both sexes. Yet others may be psychologically
inclined to feel like females, even when they have male sexual organs and vice versa. However, these are exceptions. Gender, on the other hand, is how sexual differences are played out in societies. It refers to the way a female or male is expected to behave. In our traditional South Asian cultures, there was even a niche for eunuchs, hermaphrodites, and transvestites or people who, for any reason, physical or psychological, did not conform to the expected gender roles of males or females. In short, sex is biologically determined, but gender is constructed by us human beings in our societies. Why we say we construct it is because when we bring up our children, we tell them that boys do not keep long hair and girls get their ears pierced to put earrings in them. We say men do not wear pink clothes and girls do not go about whistling at boys. In other words, we socialize our children into the expected gender roles we want them to accept and play out in our societies.

As anyone with any historical knowledge will tell you, boys and men did have long hair in Europe up till the eighteenth century. In South Asia, Sikh men do not cut their hair so they also have long hair. Some Muslim Pashto-speaking tribesmen in Waziristan have long hair. They also put kohl in their eyes, which is thought to be something effeminate by people living in the city. In the past, in some parts of the world men did get their ears pierced and one can see a number of faqirs with ear rings in South Asian cities. Some sadhus and malangs also wear bangles and rings for good measure. We also know that purple was the colour of kings, and some bishops and cardinals sometimes wear shades of red even now. In other words, these stereotypes of masculinity which we take to be fixed are only recent phenomena. They are European constructs of the 19th century and not universal or natural in the sense that not all human beings believe them to be true now nor did they so in the past.

The gender roles which we enforce upon children are accepted as being natural by them, but the fact that they differ from society to society and time to time suggests that they are largely constructed. According to the anthropologist Margaret Mead, who studied
tribes of New Guinea between 1931 and 1933, the gender roles of men and women there were different from Western societies. The Arapesh men, for instance were cooperative, unaggressive, and caring while Mundugumor women were aggressive and ruthless. In the Tchambuli tribe, the sex roles were as follows:

We found a genuine reversal of the sex-attitudes of our own culture, with the woman the dominant, impersonal, managing partner; the man the less responsible and emotionally dependent person.

Mead 1935:190

In short, the men were like our stereotype for women. As she put it in the preface for the 1950 edition, the men ‘are catty, wear curls, and go shopping, while the women are energetic, managerial, unadorned partners’ (Mead 1935: Preface to the 1950 edition).

Those who have read tales (dastans) will remember that the men in these stories are romantic lovers who faint away often only by looking at the beauty of the beloved. They cry easily as powerful emotions grip them and they are apt to woo their mistresses through long speeches in verse quite as much as performing daring feats for them. The rakes (bankas) of colonial Lucknow wore very fine clothes and went about scented as, indeed, did the dandies of Paris in the 18th century and London of the decadent 1890s which is called the fin de siecle (end of the century). In fact, wearing some kinds of scents is considered auspicious by Muslims and men put on itar on Eid and when they visit the tombs of the great sufi masters like the Dargah of Sheikh Nizamuddin Auliah in Delhi or that of Sheikh Bahauddin Zakariya in Multan.

Gender, then, is constructed. Language plays a role in constructing gender. In our male-dominated societies, this role is negative as far as women are concerned. Following is a list of devices that are used in English for registering bias against women. They are taken from Lia Litosseliti (2006: 14–15). This list can be used for our South Asian languages as the examples given below indicate.
1. **Sex specification**: The sex is specified when one mentions a trade, profession or role. The English term ‘authoress’ is given as an example of this. We now use the more neutral ‘author’ without specifying the sex of the person in question.

2. **Gratuitous modifiers**: These are said to diminish prestige and draw attention to the sex. We often use the term ‘lady doctor’ in our hospitals so that when just the word ‘doctor’ is used it is assumed that we are talking about a man. This makes women feel that while it is normal for a man to be a doctor for them it is so unusual that their sex is specified. Similarly the term male nurse is used since just nurse refers to a female nurse.

3. **Lexical gaps or under-lexicalization**: Not having enough words to describe the same kind of behaviour in men and women. For instance, both men and women can be promiscuous. Yet, the terms for women of this kind are far more than those for men.

4. **Semantic derogation**: A word when used for a woman gets degraded while equivalent terms for men do not. Just look at the term ‘mistress’ which is the equivalent of ‘master’. But mistress has come to mean a woman kept for sexual purposes, whereas master still means one who owns or is in control of something.

   Incidentally, this happens to terms for low-prestige professions also. For instance, the term ‘mehtar’ means exalted in Persian. But when it came to be used for sweepers in parts of north India, it came to be degraded. The term ‘jamadar’, who was an official with considerable prestige, also came to be used for sweepers and was so degraded that the army officials called jamadars were embarrassed by it and the term ‘subedar’ has been used for them since many years in Pakistan.

5. **Asymmetrically gendered language items**: Single words used for women without similar words for men. For instance, the title ‘Mrs’ was used for married women, but there was nothing to show that a man was married. Similarly, ‘Miss’ was an unmarried woman but an unmarried man, in common with a married one, was referred to as ‘Mr’. This was objected to by feminists because it
meant that women were identified in relation to men (i.e. through marriage with them) but men were not.

Other such items are ‘chairman’, ‘policeman’, ‘fireman’ etc. These refer to the male gender and now they have been changed to refer to females also, such as ‘chairperson’.

6. Connotations of Words: A word may mean one thing literally but have negative implications. The first is its denotative meaning and the second its connotative meaning. For instance, the word ‘hot’ means having a high temperature. However, its connotation is that the person is sexually attractive. For a girl to be called ‘hot’ will be taken as an insult in most situations in South Asia. Even the terms ‘girl’ and ‘girlish’ may imply immaturity rather than just femaleness.

Basically, this bias is called sexist language, i.e. language which discriminates against women on account of the gender roles expected from them and the stereotypes these have given rise to. Recently, feminists have started opposing such kind of words because they depict women as sex objects or as mothers and nurturers. They also trivialize and judge women. This means that they do not take women seriously and pass judgments upon them related to their gender, rather than related to what they have or have not done. For instance, if men talk loudly, they are called friendly, but if women do so, they are stigmatized as non-serious chatterers.

An attempt to create an unbiased language, at least as far as English is concerned, is the opposition to androcentric generics (andro = man; centric = centred upon; and generics are general terms such as ‘mankind’, ‘manpower’ etc.) The feminist argument is that such generics refer to men on the assumption that women are automatically included. However, this makes women appear as special cases or makes them invisible. Thus, the words which they substitute instead are ‘humanity’, ‘personnel’ or ‘workforce’ etc. This campaign has succeeded in so far that scientific journals issue guidelines not to use sexist language. However, the ideological
forces which subordinate women, as scholars have pointed out are changing rather slowly (Cooper 1989: 14–21).

Most languages have ways to show gender. For instance, in English if we say:

(1) She went.
(2) He went.
(3) It went.

The pronouns she and he refer to the female and male gender respectively and it refers to neutral. In Urdu-Hindi, the sentences would be like this:

(4) Vo gaya.
(5) Vo gayi.

and we cannot have the equivalent of (3) because everything has a grammatical gender in Urdu-Hindi. So if we say:

(6) Garmi gayi. (The summer went)

then garmi has a feminine grammatical gender, while jara /dʒa ra/ is masculine. So if we want to say The winter went, we would say:

(7) Jara gaya.

However, another word for winter is sardi, and it is feminine. The sentence would then be:

(8) Sardi gayi.

Now consider Persian.

(9) Oo raft (She/he went)

In this language, neither the inflexion of the verb nor the pronoun indicates the gender. In short, different languages indicate gender differently, but the relationship between this and the way women are actually treated in society is too complex to be determined by language alone. Just because Persian refers to man and women in
the same manner does not mean they are treated exactly alike in society.

Another aspect of gender and language is the study of **Women’s Language (WL)**. The classic study of this is by Robin Lakoff who says it has certain special features such as tag-questions (isn’t it?), rising tone on declaratives, words like ‘charming’ etc. (Lakoff 1975). The idea was that women are weak as a collectivity, and therefore use the language of persuasion rather than that of authority and strength. Later some scholars suggested that women are more cooperative as conversationalists and care for the ego of the other person (Coates 1996; Holmes 1995; Tannen 1990). However, in Western sociolinguistic writings, these things are not seen as virtues. They are associated with women being weaker and subordinate. This makes many women want to speak more aggressively like they think men do.

It seems that the kind of style we associate with the traditional upper class culture of Lucknow of a century ago had some of the features of WL. It was considered polite and gentlemanly, not effeminate or womanly. This area needs more research.

However, some features of WL do stand out in South Asia at least among traditional sub-cultures. Evidence of this comes from a genre of Urdu poetry called rekhti (not rekhta which is another word for Urdu itself). This poetry was written by men in order to enjoy themselves at the expense of women. Thus they present stereotypes, and not real women. However, real women of this period in certain social circles did use the words reproduced below:

- hariyan: used to show disapproval for a man or thing
- zanakhi: female friend
- naoj: to show disapproval
- mardua: man (pejorative use)
- ooi: oh!
- bae: oh!
The WL shown in Urdu writings is not politer than the language of men. It does, however, have special words used only by women and fewer Persian and Arabic words than men of the same social classes used. That is why the reformers of that period such as Altaf Hussain Hali and Maulana Ashraf Ali Thanvi urged that women should be taught standard Urdu, the style used by educated gentlemen, rather than the *begmati zaban* (the language of ladies) (Minault 1998: 69 and Minault 1984). In short, WL was not seen by Indian Muslim reformers as a necessary attribute of gender. They saw it as a bad habit which education and training could change.

While these stigmatized and much ridiculed markers of WL in north India have disappeared, some special features remain even now. In Pakistan and India, women use ‘*Uf*, ‘*Hae*, ‘*Hae main mar gayi*’, ‘*Hae Allah*’, ‘*Hae Ram*’, and so on.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Meaning</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Uf</em></td>
<td>Oh</td>
<td></td>
</tr>
<tr>
<td><em>Hae</em></td>
<td>Oh</td>
<td></td>
</tr>
<tr>
<td><em>Hae main mar gayi!</em></td>
<td>Oh! I am dead!</td>
<td></td>
</tr>
<tr>
<td><em>Hae Allah!</em></td>
<td>Oh God!</td>
<td>(Muslim women)</td>
</tr>
<tr>
<td><em>Hae Ram!</em></td>
<td>Oh God</td>
<td>(Hindu women)</td>
</tr>
<tr>
<td><em>Hae Guru Ji!</em></td>
<td>Oh God</td>
<td>(Sikh women)</td>
</tr>
</tbody>
</table>

Such expressions tend to differentiate the sexes and, more significant in terms of gender roles, they are interpreted as indicators of lack of power and courage. They are used in distressful situations which are supposed to be faced with stoicism by men. Because of these associations, these expressions are not used as much by Westernized women as they are by those who live in traditional sub-cultures.

Another aspect of language about women in parts of South Asia is that women are concealed from unrelated men. This segregation, whether in the form of taking the veil or modest segregation from men, is part of traditional sub-cultures in this part of the world. This is expressed in language by the circumlocutions for ‘wife’
which exist in Urdu, Punjabi, Pashto, and English as used in Pakistan:

- **Urdu**: *ahle khana; ghar vale; andrun-e-khana*
- **Punjabi**: *kar vale; karon; bachche*
- **English**: *family; ladies (not lady)*
- **Pashto**: *kor valan; kor*

Verbal politeness is expressed by males by not mentioning the other’s young female relatives. For instance, one man asking another about his wife’s health will either not make the enquiry at all or, if he does so, it will be in some such form (as evidenced in Pakistan):

1. *Family theek hai?* (Is the ‘family’ all right?)
2. *Ghar vale theek hain?* (Are the people of the house fine?)
3. *Hamari bahen theek hain?* (Is my sister all right?)
4. *Aap ki Mrs theek hain?* (Is your Mrs all right?).

Language and gender can be studied in relation to the media, the workplace, the classroom, and in discourses of various kind. The scope of research in this field is tremendous because South Asian societies are in flux and there is so little research on such issues.

**Linguistic Politeness**

Politeness is expressed through language as well as non-verbal expressions such as smiling, bowing, bending the body, folding hands, inclining the head downwards, and looking pleased. These gestures are called body language, and they express attitudes and emotions of various kinds. In this section, however, we will give a brief introduction only to linguistic politeness, not to non-verbal expressions of it.

The classical study of linguistic politeness is Brown and Gilman’s famous work on the pronouns of power and solidarity (1960). Basically, they argue that there are two basic forms of pronouns: *tu* (T) forms and *vos* (V) forms. Both occur in Latin and the former stands for *tu/ tum* in Urdu-Hindi while the latter is the equivalent
of *aap*. They claim that the *tu* form was used for all till the Roman emperor became so exalted above other people that the plural form (*vos*) came to be used for him to show his special status. When people use the *T* form with each other, they show solidarity or intimacy; when they show the *V* form they are showing respect but also distance. When one uses *T* but receives *V* in response, then the user of *T* is less powerful in relation to the user of *V*. This is non-reciprocal use of *T* and *V*. In our hierarchical societies, children use *V* forms for parents and elders, and receive *T* forms instead. Servants also use the *V* form and receive the *T* form from employers and so on.

As democracy and egalitarianism increased in Western societies, the non-reciprocal use of *V* and *T* became less fashionable. Nowadays, English has lost the *T* form (*‘ye’*) and ‘you’ is used by everybody. In French, however, the *T* and *V* forms remain, but the *T* form is used among friends to signal intimacy and equality. It is mostly reciprocal. In the few instances where it is not reciprocal, it is confined to the family circle with parents using *T* and receiving *V* from their children. However, in public one uses *V* for the working classes as well as younger people.

Besides *T* and *V*, the use of honorifics and titles with names and first names themselves are part of linguistic politeness. The phenomenon of using first name (*FN*) versus title and last name (*TLN*) has been studied with respect to American English. This study relates the use of *FN* with informality (Brown and Ford 1961; Erwin-Tripp 1972). Another study analyses the address systems in several languages, suggesting the use of *T/V* and such other devices for expressing linguistic politeness as ways of expressing power. However, while the basic reasons for choosing one set of linguistic markers rather than another remain the same (i.e. the desire to show intimacy or differences of power etc.), the fashions for doing so keep changing (Braun 1988). But now the question arises as to what is politeness.

Scholars have answered this in terms of ‘adequacy’ i.e. a form of address which is in accordance with the rules of the community in
question is considered adequate or polite (Braun and Schubert 1988:49). Brown and Levinson define politeness in terms of ‘face needs’. The want to be unimpeded is ‘negative face’, while that of being approved is ‘positive face’ (Brown and Levinson 1987: 63). My own definition of politeness is:

Politeness is that which is considered gratifying by the addressee. It is a matter of subjective interpretation and is not necessarily bound to verbal or non-verbal signals without reference to context.

Rahman 1999:191

This definition emphasizes the subjectivity of the interaction. This means that if young people meet friends who greet each other with shrieks and abusive words, it will not be rudeness because it will mean that the friends are really intimate with each other. Also, if you meet your professor and address him by his first name when he belongs to a generation when young people used the honorific Ji or Sahib after the name and a title (Professor or Doctor etc.) before it, he will be right to interpret it as rudeness and insensitivity. On the other hand, a professor in a Western country who wants his students to address him by his first name will interpret this as an indicator of being at ease or relaxed with him.

In my article on this subject, I have presented the hypothesis that a small English-using elite in South Asian cities is using only the first name for older people without being sensitive to their norms of politeness. They are doing it in the name of equality, informality, and ease of interaction. However, they are doing it from a position of power because they are Westernized and highly educated, and serve in institutions (NGOs, Think Tanks, corporate sector etc.) which can afford to be indifferent to indigenous norms of behaviour (Rahman 1999: 183–223).

Yet another aspect of the relation between linguistic politeness and power is that terms of prestige change according to the power of the community which uses them. For instance, when the British started ruling South Asia, the older titles and forms of address
which carried respect earlier, became less prestigious, and the British forms came to be endowed with prestige. A list of these terms is given below:

**INDIAN AND BRITISH TERMS IN THE NORTH WESTERN PROVINCES IN THE 1850s**

<table>
<thead>
<tr>
<th>Hindus</th>
<th>Muslims</th>
<th>British</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lalaji (Baniya or Kaesth)</td>
<td>Mianji</td>
<td>Master Sahib</td>
</tr>
<tr>
<td>Nai Pandhe (barber teacher)</td>
<td>Mullahji</td>
<td>Teacher</td>
</tr>
<tr>
<td>Panditji</td>
<td>Munshiji</td>
<td>Sir</td>
</tr>
<tr>
<td>Guruji</td>
<td>Sheikh Sahib (Sunni)</td>
<td>Miss (for females)</td>
</tr>
<tr>
<td>Maharajji</td>
<td>Mir Sahib (Shia)</td>
<td>Madam (for females)</td>
</tr>
<tr>
<td>Bajpeiji</td>
<td>Mirza Sahib (Mughal)</td>
<td>–</td>
</tr>
<tr>
<td>Parohatji</td>
<td>Khan Sahib (Pathan)</td>
<td>–</td>
</tr>
<tr>
<td>Daoji</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Vaed Ji</td>
<td>Hakeem Ji</td>
<td>Doctor Sahib</td>
</tr>
<tr>
<td>–</td>
<td>Qazi Ji</td>
<td>Judge/Justice Sahib</td>
</tr>
<tr>
<td>–</td>
<td>Darogha Ji</td>
<td>Thanedar Sahib</td>
</tr>
<tr>
<td>–</td>
<td></td>
<td>Police Officer SHO</td>
</tr>
<tr>
<td>Diwan</td>
<td>Wazir</td>
<td>Minister</td>
</tr>
</tbody>
</table>

*Source: Reid 1852: 81-82 and other sources.*

The study of the relationship of society with language is fascinating. The field is so vast that it will require years of research to be able to understand it. This chapter was meant to introduce you to the basic concepts and to stimulate your appetite for further research of your own.
Politics is defined as the pursuit of power. Power is needed by human beings in order to do things. First, to secure food, security, shelter, and produce children. Secondly, to go beyond those necessities of survival itself to obtain pleasures such as beautiful things, prestige symbols (such as posh cars), and pleasurable experiences. If you have power, you can make your life pleasant. Moreover, you can make your village, town, and country more comfortable. For instance, if you get political power by getting elected as a municipality head, you can construct roads and schools for your area. If you become the prime minister, you can improve the lives of millions of people. However, people tend to blame the pursuit of power, because people do it at the expense of others and do not help anyone but themselves or their relatives and friends. Let us remember, however, that one cannot do any good to anyone if one does not have power. Moreover, weak people are bullied and made miserable. Power enables us to preserve our pleasures. It is for pleasure that one seeks power. But, of course, it should be sought within rules, otherwise we end up fighting wars which reduce everybody’s pleasure. Indeed, modern weapons are so destructive and poisonous that if we fight for power, we could kill the human race itself.

Language and power

But what has language got to do with power? Well, it connects with power in so many ways that one is surprised. First, we use a language or languages in our government, judiciary, administration, education, commerce, media, films, and the corporate sector. If we
know these languages, we can get jobs in these domains, but if we do not, we will be left out. Now these domains of work are called domains of power because modern states and the private sector exercise power through them. The languages of power are the languages used in them. As you can appreciate, the major language of power in South Asia is English. If you know English very well, you are much likely to enter the domains of power. If you know the other languages used in the lower domains of power, you will get jobs but not very good ones. If you do not know any of the languages used in any of the domains of power, you will probably not even be reading this book. If, therefore, you or your parents are keen that you learn English, then it is because they want to empower you. And, if whole groups want to empower themselves, they learn the language of power. As Muslim rulers, including the Moghuls, used Persian, this was the language which the people of India learned if they wanted to work in the domains of power. The Kaesth class among the Hindus became very good at Persian and have left behind works of literature in it. When the British started ruling India, the Indians learned English. Raja Ram Mohan Roy, for instance, requested the British to teach English to the Indians. Similarly, Nawab Abdul Latif and Sir Syed Ahmad Khan wanted Muslims to learn English because, in their view, this was the best way of empowering themselves.

As we have mentioned earlier, power is also expressed through pronouns such as the different forms of ‘you’ (aap, tum, and tu), and it is part of gender as expressed through language. Let us now take language planning (LP) as one way of using language or manipulating language for political aims, i.e. power-oriented aims. The term ‘language planning’ was popularized by the famous linguist Einar Haugen. He defined it as follows:

The activity of preparing a normative orthography, grammar, and dictionary for the guidance of writers in a non-homogeneous speech community (1959: 8).
In short, language planners make alphabets, write dictionaries, and tell people what rules to follow in writing or speaking a language.

Some people divide the activities of language planners under different headings. These are: (1) status planning, (2) acquisition planning, and (3) corpus planning.

**Status Planning**

The term means determining the status of a language. Status means whether it will be used as a national language, a provincial language, the language of primary education, the language of higher education, the language of administration, and so on. Such decisions are generally taken by governments and enforced by governmental officials. These officials, whether civilian or military, make up the apparatus of the state.

Hence, we can say that status planning is the same as language policy. This policy is enforced by the state. If someone opposes it, the state may use force to counter this opposition. Sometimes, however, the state compromises and the policy is changed.

**Acquisition Planning**

‘Acquisition’ means acquiring or learning a language. If a language is taught or used in the TV and radio etc., then people acquire it. If someone makes a policy to increase the number of people who acquire a language, then it is known as acquisition planning. Nowadays, this phenomenon is also known as language shift (LS). Sometimes, people learn the language which is required for getting good jobs, social prestige, and political power so well that they forget their own language. In some cases, a groups’ own language is completely forgotten. This phenomenon is known as language death. In Ireland, Irish Gaelic, which was the language of Ireland before the British conquest of the island, died. People shifted to English. In Wales, too, a similar thing happened. However, in Wales, some people tried to revive the use of Welsh once again.
This phenomenon is called reversing language shift (RLS). We will study this phenomenon in detail in the next chapter.

Some people consider acquisition planning a part of status planning. Their argument is that the decisions about using a language for education and in the media is, indeed, status planning. Since this increases the number of people who use these languages, status planning is the same thing as acquisition planning.

**CORPUS PLANNING**

Corpus planning means making planned changes in the language itself. For instance, we may want to change the script of a language, introduce new words, fix the spellings of words, or eliminate some words. The two major activities of corpus planners are (1) standardization and (2) modernization.

**Standardization** means fixing a standard. In pre-modern communities there are no fixed norms or standards. Some people speak one variety of a language while others speak another one. In fact, the language changes after a few miles. Even if the language is written, the spellings are not fixed. Other rules of writing—where to put commas, where to put full stops etc.—are also changeable. When a society gets modernized, it needs a fixed standard. The printing press prints millions of written pages which are read by many people and it would be expensive to have different spellings in different places. Moreover, it would create confusion in an industrialized society where the printed word is used all over the country. Thus, one variety of the language is standardized while others are not. Spellings are fixed and the rules of punctuation etc. are decided. These standard linguistic norms are written in dictionaries, books of grammar, and manuals. They are taught in schools and other training institutions. In short, the whole education system of the modern state tries to eliminate diversity while creating uniformity.

The other major activity of corpus planners is modernization. As we can see, standardization too is part of modernization. Thus,
modernization of vocabulary can also be called neologism or lexical modernization. It refers to creating new terms for concepts or things which did not exist in the language. As you read earlier, every language is meant to express the reality of the culture in which it was created. The Siraiki and Arabic languages have many words for ‘camel’ because camels were important for the people who spoke these languages. The societies which make aircrafts, cars, rockets, and computers also made words for them. They change the meanings of existing words or borrow words from other languages to express new ideas. This is lexical modernization or neologism.

There are other activities of corpus planning too—such as purging a language of some words, changing the script, making a new script, and so on. If you want to learn about it in detail, read Cooper (1989).

**Language Planning and Politics**

Apparently, all the activities described above are about language. However, a little reflection will tell you that they are also connected with politics. The Westernized elite in both India and Pakistan use English. They know English so well that it is to their advantage to keep using it in the domains of power.

In France, the French Academy made Parisian French the standard language for the whole of France. The political advantage of this move was that all the other varieties of French as well as the other languages spoken in France became sub-standard rustic dialects. Nobody could attain a high social position if they knew only these languages and not the standard French. Secondly, the French ruling elite could claim that the whole of France was united by the French language. So, on the one hand the ruling elite consolidated its rule all over France by claiming that everyone spoke only French, while on the other hand, since the village dialects were considered inferior, the elite kept the masses away from power. Of course some clever students could learn standard French but this was not easy,
especially until the twentieth century, because earlier school education was not so easily available.

**Ethnic politics and language**

The policies of the ruling elite are sometimes challenged by those who are not in power. The challengers are sometimes called the proto-elite. They are generally led by intellectuals and educated people. These people want a symbol for mobilizing support. Sometimes this symbol is religion, while in most instances it is language. Language gives a sense of unity; of being one. It is one of the things which give identity. Those who oppose the state use language to form a pressure group. Some people feel that if their own language is used in the domains of power, they will be at an advantage. This is how ethnicity or ethno-nationalism is created.

Some people feel that ethnicity is a throwback to the past. They argue that pre-modern people used to define themselves in ethnic and religious terms, while modern people are nationalists. This theory is called the *primordialist* theory of ethnicity, and it claims that language is a primordial bond which is asserted because it has never been lost. Other people argue that it is modernity which makes people assert their ethnicity. There is a lot of competition for jobs and power in the modern state. If a group feels that it is not getting its proper share of goods, services, and power, it feels wronged. Language may be a symbol used by the leaders of the group to mobilize people and create a sense of common identity. These mobilized people become a pressure group. In short language-based ethnicity is a product of modern conditions. This is the *instrumentalist* theory of ethnicity.

It is true that modern conditions create the desire to assert ethnicity, but people in language movements may have both primordialist (or sentimental) and instrumentalist (or rational) motivations. People may feel that their culture—of which language is a symbol—is being taken away from them. This is a sentimental reason. They may also feel that they will be left behind another
group in the competition for power and money. These are instrumentalist reasons.

If you look closely at the language-based ethnic movements in Pakistan and India, you may be able to explain them through these theories (for India see Phadnis 1989; for Pakistan see Rahman 1996).

**CORPUS PLANNING AND POLITICS**

It has been mentioned earlier that people may be emotionally attached to their script. They may regard it as part of their identity and may be unwilling to change it. It should also be mentioned that when new words are formed, the morphemes which make them are also borrowed according to one’s political views. Thus, when modern Hindi was developed, Hindu language planners put in words of Sanskrit origin in it in order to make it a symbol of Hindu identity. The official policy in Pakistan is to emphasize the Islamic and Pakistani identity. Hence, the new terms made in Pakistani Urdu are borrowed from Arabic and Persian. For instance, ‘white paper’ could have been translated as *sufaid kaghaz* in Urdu. However, it has been translated as *qartas abiaz* (*qartas* = document, paper; *abiaz* = white—both from Arabic). On the contrary, the ethno-nationalist language planners use morphemes of their indigenous languages to express their native roots. Here are examples of new words created by ethno-nationalists in Pakistan:

**PUNJABI**

<table>
<thead>
<tr>
<th>Word commonly used</th>
<th>Punjabi Coinage</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>lafz</em></td>
<td><em>akhar</em></td>
<td>word</td>
</tr>
<tr>
<td><em>salana</em></td>
<td><em>varhe var</em></td>
<td>yearly</td>
</tr>
<tr>
<td><em>bhejna</em></td>
<td><em>ghalna</em></td>
<td>to send (a letter)</td>
</tr>
<tr>
<td><em>sailab</em></td>
<td><em>harb</em></td>
<td>flood</td>
</tr>
<tr>
<td><em>taqreeb</em></td>
<td><em>ikath</em></td>
<td>gathering</td>
</tr>
</tbody>
</table>
## PASHTO

<table>
<thead>
<tr>
<th>Word commonly used</th>
<th>Pashto coinage</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>sumreen dub mako</td>
<td>submarine</td>
<td></td>
</tr>
<tr>
<td>vai jaaz alvatak</td>
<td>aeroplane</td>
<td></td>
</tr>
<tr>
<td>profesar pohand</td>
<td>professor</td>
<td></td>
</tr>
<tr>
<td>dastkhat las lik</td>
<td>signature</td>
<td></td>
</tr>
<tr>
<td>rasmulkhhat lik dud</td>
<td></td>
<td>script</td>
</tr>
</tbody>
</table>

## BALUCHI

<table>
<thead>
<tr>
<th>Word commonly used</th>
<th>Baluchi coinage</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>mauzu</td>
<td>bangapp</td>
<td>subject</td>
</tr>
<tr>
<td>lavz, labz</td>
<td>gal</td>
<td>word</td>
</tr>
<tr>
<td>lahja, boli</td>
<td>galwar</td>
<td>dialect</td>
</tr>
<tr>
<td>kitab</td>
<td>wanagi</td>
<td>book</td>
</tr>
<tr>
<td>film</td>
<td>tamur</td>
<td>film</td>
</tr>
</tbody>
</table>

## SINDHI

<table>
<thead>
<tr>
<th>Word commonly used</th>
<th>Sindhi coinage</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>brake</td>
<td>hawai brake</td>
<td>air brake</td>
</tr>
<tr>
<td>born</td>
<td>narghat</td>
<td>air born</td>
</tr>
<tr>
<td>automatic</td>
<td>khudkar</td>
<td>automatic</td>
</tr>
<tr>
<td>seedha hath</td>
<td>sajir</td>
<td>right hand drive</td>
</tr>
<tr>
<td>ulta hath</td>
<td>kabir</td>
<td>left hand drive</td>
</tr>
</tbody>
</table>
To conclude, Language Planning (LP) is related to politics. That is why controversies about languages, scripts, and words are not only about these things but about things much deeper than these. The real dispute is about way of life, way of thinking, and who is going to get what, and not only about language.
The average life of human beings on this planet is said to be between 5 to 10 million years. While it is not known exactly when human languages were invented, all known societies have always used them. It is estimated that there were 10,000 to 15,000 languages in prehistoric times. Languages have always died when their speakers have died or started using other languages. Indeed, even very powerful languages, which spread over large areas of the world, have faded out and now we owe it to linguistic historians like Nicholas Ostler to tell us something about their rise and decline (Ostler 2005 and 2007). But nowadays, the pace of the disappearance of human languages has increased. We still have about 6,000 plus languages but we are losing them day by day. You must have noticed in the first chapter where the languages of South Asia are listed, some extinct languages are also mentioned. Indeed, the Summer Institute of Linguistics, whose figures I have quoted, does not mention any extinct language for Pakistan. However, some researchers have confirmed that at least one Pakistani language (Badeshi) is dead, another (Domaaki) is about to die, and three (Chilliso, Gowro, and Ushojo) are getting so weakened that they will die unless something is done to strengthen them (Rahman 2006: 81–82). In short, we may actually have more languages on the way to extinction than the reports on languages tell us.

**Do languages die or are they killed?**

What is ‘language death’? As previously mentioned, it occurs when the last speaker of a language dies. This happens when the speakers have shifted to another language. That is why the phenomenon is
also called **language shift** (Fishman 1991). Other people argue that people do not shift from their mother-tongue to another language, unless there is pressure upon them to do so. It may not be that they are forbidden to speak their own language (though this has happened in history), but it does happen that there is no education and no jobs in their language. Or, as is usually the case, their language is looked down upon and treated with contempt. When this happens people learn the language which is used in jobs, the education system, media, commerce, and respectable company. This means that the language does not die, it is ‘killed’, i.e. conditions are created in which people have no benefit from learning it and start considering it inferior themselves. That is why Tove Skutnabb-Kangas, a linguist who is a great supporter of language rights, calls this phenomenon ‘linguistic genocide’ (Skutnabb-Kangas 2000: 312). As the very group whose language is being killed wants to learn the killer language because it is so useful to empower them as individuals and as a group, it is also possible to see this as ‘linguistic suicide’. However, the usual term for this phenomenon is language death and that is what we will use in this chapter.

**Modernity and language death**

Why languages are dying so fast is because of the impact of modernity and globalization. Modernization makes societies mobile and educated. They leave their villages and crowd into the cities. Here, they generally acquire jobs where they are exposed to one language which is either the national language or some other dominant language of power. The schools teach it; the media uses it; and they themselves have to learn it in order to continue living the new life they are trying to adjust into. Globalization has increased all these trends. The means of communication have improved and people are much more mobile than they were ever before. People also travel abroad and are exposed to the Internet and TV far more than ever before. Thus, more and more people are shifting to the dominant languages and leaving their own language behind in the village—and eventually, even the village
gets swamped by so many modern things that it no longer is the village they remember from their childhood.

In South Asia, English, Hindi, and Urdu are the major languages which put pressure on the smaller languages. In other parts of the world, too, there are many dominant languages which threaten minor and weak languages. This is now a major concern of linguists who have started writing on the phenomenon of language death. In the year 2000 alone, there were several major books on the subject (Two fairly easy ones to understand are Nettle & Romaine 2000; Crystal 2000). The UNESCO published a position paper supporting education in minority languages in 2003 and a conference on multilingual education was held in Bangkok on this subject in November 2003. The second major conference on the same subject was held in July 2008 in Bangkok, where a large number of participants described how minority languages were being used to teach little children. This is one way of saving a language from dying out.

**Can we do something to save languages from dying?**

Should we save languages from becoming extinct at all? After all, there is a view that having many languages is costly, divisive, and inconvenient. One has to learn them and translation costs a lot of money. Besides, one needs reading material, teachers, and jobs in these languages; otherwise they could not be learned. As such, according to this point of view, it is better to become monolingual. Those who advocate this in a multilingual country often support the official national language and deny ethnic identity as well as the languages on which it is based. Those who support a world language favour English monolingualism or English-only policies. They are often found expressing antagonism to the tradition and literature of all other than the English-speaking peoples or their colonial clones in the former colonies. One reason for not favouring the domination of any one language is precisely this: it leads to contempt for others and helps in the domination of weak and vulnerable groups.
David Crystal has given the following five major reasons for preserving the linguistic diversity of the world:

1. We need diversity.
2. Languages express identity.
3. Languages are depositories of knowledge.
4. Languages contribute to the sum of human knowledge.
5. Languages are interesting in themselves. (Crystal 2000:27–67)

Arguments in favour of linguistic diversity take other forms also. These have been presented with reference to ethics, philosophy, and psychology by scholars. You can read Skutnabb-Kangas (2000: 249–284) for a summary of their views.

You will understand, of course, that states do not value diversity or the assertion of identity because they are afraid that, carried to an excess, both might break them up. The fact is that this excess occurs only when the rulers suppress the minorities and the suppression of their languages is part of it. For instance, the Kurdish language is suppressed and punished in Turkey, but this has not made the Kurds become reconciled to Turkish rule. Indeed, this has only hardened their resistance to Turkish domination.

Some linguists are now trying to write the grammars and dictionaries of languages which are threatened with extinction. There is also a book by a famous sociolinguist called Can Threatened Languages be Saved? (Fishman 2001). (This is a sequel to an earlier book called Reversing Language Shift, Fishman 1991). Fishman discusses a number of strategies for saving languages. Ten years later this is what the author had to say about this issue:

So can threatened languages be saved? This question now has an informed, though uncertain, answer: Yes, more of them can be saved than has been the case in the past, but only by following careful strategies that focus on priorities and on strong linkages to them, and only if the true complexity of local human identity, linguistic competence and global interdependence are fully recognized.

Fishman 2001:481
One of the ways of saving languages is to educate children in them. That way, they understand concepts better and, what is more important, come to respect their language and their identity. But, of course, the teachers and the other grown ups they come across—the role models—must show respect for their languages and culture. If they do not, the pupils will regard their language as nothing more than an additional and unwelcome burden. That is where language policy is connected with educational policy, and both are connected with social psychology and the art of teaching itself.

Some people suggest that those who specialize in language death and reversing language shift should study educational linguistics as well as writing grammars in disappearing languages. In fact, this whole branch of learning is sometimes called preventive linguistics, i.e. that branch of linguistics which will help us to prevent the death of languages. Such a person is a practitioner and a language activist as well as a scholarly linguist. If this is what you want to be, South Asia is an excellent place for you.
Educational linguistics refers to the use of language in education. This is a very vast subject indeed. While discussing language planning (LP), we referred to acquisition planning (Cooper 1989: 157–163). Acquisition planning is not simply language teaching, but it is the most important part of it. In addition to teaching, one can plan to increase the number of people who know one’s language through other means: radio, television, films, songs, sports, games, and so on. But, of course, teaching—especially in schools—is the most important way of spreading a language. But teaching a language is by no means a simple matter. It is so complicated that there are branches of applied linguistics dealing with foreign language teaching (FLT), second language teaching (SLT), the teaching of classical or dead languages, and so on. In the South Asian context, most people going to college must have heard of subjects like ELT (English Language Teaching), ESP (English for Special Purposes), and TEFL (Teaching of English as a Foreign Language) etc. In Pakistan, the Society of Pakistan English Language Teachers (SPELT) was created with the pioneering leadership of Zakia Sarwar in 1984 (Rahman 2002: 316).

Debates in educational linguistics

There are several debates in the field of educational linguistics. First, there is no clear consensus on what a ‘second’ and a ‘foreign’ language is. German is obviously a foreign language in Pakistan, as it is not used in any domain. If you want to study it, you have to go to those specialized institutions which offer courses in it. But
English is used in the media, educational institutions, the officer corps of the armed forces, superior judiciary, high-level bureaucracy, and so on. Many people actually use it spontaneously in private conversation because they are very fluent in it. Is it then a ‘second’ rather than a ‘foreign’ language? The consensus of most ELT theoreticians is that English is, indeed, a second language in most of South Asia and it should be taught through the techniques called TESL (Teaching of English as a Second Language). This is what has been happening in Pakistan and India since the last quarter of a century. However, the fact remains that our societies are deeply divided and the class exposed to English is only found in the urban areas and even then only at the upper socio-economic levels. After all, it is only the rich and the powerful who are exposed to English at home and in the school, and not the ordinary people for whom maybe Urdu or Hindi is a second language and English is a foreign language. However valid such kind of questions may be, the consensus of educational linguists is that if a country uses a foreign language in many domains, then it is a second, and not a foreign, language. So, in the final analysis, we settle these things by convention and consensus, even if there are nagging doubts. This does not mean that anything is finally settled, because the scientific spirit requires us to keep investigating, keep asking questions, and keep up our inquisitiveness.

**Do we learn other languages better if we start with our own?**

Another question is whether we learn other languages better if we start learning them as early as possible or graduate to them after having learned our first languages (mother tongues)? English-medium schools in most cities of South Asia tell us proudly that they do not allow the children to speak the ‘vernaculars’. Even small children are forced to speak in English and this is considered a great distinction. The logic is that they must be immersed in English—the target language—from early childhood, otherwise they will never master it. As for the mother-tongue, why they will
acquire it anyway, since it is all around them. This apparently commonsensical approach is refuted by many linguists, including the researchers at the United Nations Educational Scientific and Cultural Organization (UNESCO), which says in its latest report:

1. UNESCO supports *mother tongue instruction* as a means of improving educational quality by building upon the knowledge and experience of the learners and teachers.

2. UNESCO supports *bilingual* and/or *multilingual education* at all levels of education as a means of promoting both social and gender equality and as a key element of linguistically diverse societies.

3. UNESCO supports language as an essential component of *inter-cultural education* in order to encourage understanding between different population groups and ensure respect for fundamental rights (UNESCO 2003).

There is plenty of evidence at least about one thing: children do not learn their languages very well if they begin with a foreign language. South Asian children exposed to English from childhood not only do not know their own languages except at the most elementary spoken level, but also express contempt for them. This is regrettable since the languages of this subcontinent have some of the classics of world literature in them. But you find many English-medium students actively ignoring, shunning, and denigrating the significance of the rich literatures of our part of the world.

The other part of the claim, i.e. that children learn the target language better if they are taught no other language but that from early childhood, is not true either. Well, of course there are people in India and Pakistan who speak English very well. Their writing skills in that language is as good as the best writers of English anywhere in the world and they produce literature in that language (see Rahman 1991). Yet this does not prove that they would not have excelled in English if they had first learned their first languages at school just as well as they learned English. Indeed, in a study of
language learning, Tove Skutnabb-Kangas argued that it is a part of linguistic imperialism to teach another language at the expense of one’s first language. This is subtractive bilingualism (i.e. it takes away or subtracts a language), whereas it is only additive bilingualism (i.e. that which adds a language without taking away one’s first language) which is welcome. She equates this kind of imperialism with racism and argues that it infringes upon the rights of children to know and respect their languages. She states:

Promoting diversity is an accepted principle in declarations. Still Western language policies have to a large extent been based on false either-or thinking (you need to choose between languages, you cannot have both this language and that language and maybe others too). It has promoted subtractive rather than additive language learning and subtractive spread of English and other dominant languages: the learning of a dominant language has been presented as necessarily happening at the cost of a dominated language, instead of in addition to it. Diversity is killed.

The central question is whether it is easier to learn only English in order to be good in it, or is it better to learn one’s first language along with English or even before one learns English. The authors argue for multilingual education (MLE) for social justice. They say that when only a few dominant languages are taught and all others excluded from the school system, there is injustice. People are marginalized and feel inferior and the rate of literacy increases very slowly. Even more to the point the psychological cost of this exercise in discrimination is appalling. In the words of the authors:

Why then are mother tongues neglected despite persuasive evidence to the contrary? As the voiceless minorities suffer the sinister exclusion of mother tongues, the silent elites enjoy the pre-eminence of dominant languages such as English. In the post-colonial world, ‘the killer languages’, including English, thrive at the cost of other languages, and in many countries the myth of English medium superiority is propagated to the detriment of the poor and the marginalized. English and other ‘killer languages’ set in motion a hierarchical pecking order
of languages which severely disadvantages the other languages, those of the indigenous peoples and minorities, in particular. And yet, it does not have to be so. In a true multilingual system, all languages can have their legitimate place: mother tongues, languages of regional, national, and wider communication.

Mohanty 2009:7

Look at the first sentence. Yes, there is evidence that if the mother tongue is taught first, other languages, such as English, are learned efficiently. The idea that we must start with English is not supported by research. However, research in the social sciences, including language-learning, is very complex. Human beings are complex themselves and circumstances keep changing in unknown ways, so you will find people making contrary claims in the research literature. My own view, however, is that power is a very important component. If a community is powerful and its language is prestigious, it is learned more easily by everyone concerned. But if the community is less powerful and the language is despised or marginalized, then it is more difficult to learn in school situations. Human beings are impressed with power and glamour and are simply more self-motivated to learn a prestigious language. This does not mean, however, that a less prestigious language should not be taught. Indeed, if anything, it should be given more time, and more reading material as well as radio and TV programmes should be produced to sustain it. Ours is a multilingual world and if it is to be a just one, then we must care for language rights, diversity, and pluralism. For this, we must teach all languages in an additive manner, beginning with the first language as far as possible.

**The solution?**

This is obviously a political and economic issue. After all, someone has to decide which first languages are to be taught and who will pay for the reading material and the teachers etc. These are ultimately political questions, as my books on the politics of language learning in Pakistan (Rahman 1999: 88–102; Rahman 2002) suggest.
In a nutshell, if English is taught well to a few elitist children, they will be qualified for the best jobs in most South Asian countries. But when this happens, most other children will find English a sieve or obstacle which forces them to remain at the bottom, unless they are very bright or work much harder than their English-medium counterparts. Thus, the class boundaries are maintained by English, while it appears that we are talking of merit and not class advantage. But what goes unstated is that English is a kind of cultural capital (a concept given by the French sociologist Bourdieu 1991: 55), which the upper classes possess and which they use to dominate the society culturally and intellectually. Thus, educational linguistics connects at some level with political linguistics and language policy. This is one of the great beauties of macrolinguistics—it provides new insights into politics, economy, society, culture, and education.
Do you know that South Asia is home to very ancient writing systems? The Taxila Museum houses a number of specimens in very ancient scripts. One of these is the Aramaic script which grew out of the Phoenician script and was very influential for over a thousand years. This will be discussed later in more detail.

The inscription is described as follows (unfortunately, the words ‘Aramaic’ and ‘emperor’ have been misspelt here):

**Aramic Inscription.**

This inscription found from the masonry of a wall at Sirkap (Taxila) is the only example of the use of Aramic (from which Kharoshti Originated) in the north west region of Pakistan. It mentions the ruler Priyadarsi and his queen and sons from which it seems that it was used in the first half of the 3rd century B.C. by the Mauryan Emperor Ashoka who ruled over Taxila at that time.
Let us, however, introduce you to the basics of writing before we discuss the scripts now used in the world. As you will agree, writing is very important in the modern world. It creates a permanent record (i.e. memory) in a certain language by conveying ideas through visible marks. These marks combine together to create meaning according to various principles. Let us look at these principles below.

**Marks carrying meaning**

If the visible marks convey meaning directly, we call the system in which they are used a semasiographic system. This system is used even now in pictures. For instance, a road sign has the picture of a running child. This conveys the meaning that children may be running across the road and the drivers should slow down. Of course the picture leaves a lot to the driver’s knowledge about the world. Someone from Mars may not know exactly what the picture means and whether any action is required or not. Some instances of semasiography are given below:

<table>
<thead>
<tr>
<th>The basics of Blissymbolics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blissymbolics is a graphic, meaning-based communication system. Some of the symbols are pictographs. They look like the things they represent.</td>
</tr>
<tr>
<td><img src="image1" alt="house" /></td>
</tr>
</tbody>
</table>

You will notice that semasiographic systems are rather limited. You can hardly convey difficult messages through them. Perhaps that is why glottographic systems of writing developed (*Glotta* is tongue and *graphein* is writing in Greek). These systems provide visible representations of speech, i.e. sounds which make sense. Glottographic systems are used for all kinds of languages, all the
way from Urdu and Hindi to Chinese. However, the writing system of different languages are based on different principles.

Some glottographic systems are called **logographic**. As *logos* means sense, logic, idea or concept, it appears as if sense or idea is directly represented in writing. This is the general view about the Chinese language. However, it is not true. Let us go into some details about this language. Words consist of units of sounds (phonemes) as well as units of meaning (morphemes). If we represent morphemes, then we will have a **phonographic** and segmental system of writing. A segment, as you will recall, is a vowel or consonant, i.e. any phoneme.

The writing system of the Chinese language is predominantly logographic. In this language, morphemes are represented by separate symbols or marks. These marks are also called **graphemes**. In Chinese, as it happens, a word is generally a morpheme. Thus, we can say that almost every word will be represented by separate graphemes. This means that we would need to learn several thousand different shapes of graphemes to be able to communicate in written Chinese. This is obviously a disadvantage for anyone who is learning Chinese.

The advantage is that the written script unites the speakers of the various dialects of Chinese. In fact, these dialects are so different that they are not always mutually intelligible. This means that they are like different languages (if you remember, dialects are defined as being mutually intelligible). But, because the symbols in writing do not represent pronunciation, the same symbols are used. Thus, a word is written in the same way, although it is pronounced differently in different dialects of Chinese.

**Marks carrying sounds**

The phonographic principle is the one to which we are accustomed in our South Asian languages and English. The idea is that every sound should be represented by a grapheme—a visible mark. These graphemes make up the alphabet (the word comes from *alpha* and
beta, which in the Greek script stood for ‘a’ and ‘b’). Ideally, every sound (i.e. segment) should be represented by one and only one grapheme. In fact, this generally does not happen. English is quite notorious for having only five graphemes (the letters a, e, i, o, and u) to represent over sixteen vowels and many diphthongs. It also has graphemes which give different sounds at different places. Look at <c> and <k>. They stand for the phoneme /k/ in the words king and cut. But in cent the letter <c> stands for the sound /s/ so that cent is pronounced as /s e n t/. Similarly, <g> has two sounds and <x> represents a combinations of three sounds (/e k s/) at places as in the word rex which is pronounced as /r e k s/.

Urdu and Hindi use the Arabic and the Devanagari scripts respectively. Both are phonographic, but there are letters which are not pronounced. Urdu also has some letters borrowed from Arabic which are pronounced in the same way. For instance, the phoneme /z/ has the following graphemes in Urdu <ﺽ>, <ﻅ>, <ﺫ> and <ﺯ>. Hindi too has instances of several graphemes for the same phoneme.

On the other hand, the Urdu script, as well as all scripts based on the Arabic system of writing, lack graphemes for vowels. The Devanagari script too lacks some vowels, though it does have more symbols for them than English. The reasons for this will be discussed later. Another problem in both the Devanagari and the Arabic scripts is that the graphemes take on different shapes when they are joined together. In other words, the grapheme <ﻡ> in Urdu which represents the phoneme /m/ takes the shape <ﻡ> in the initial position, while it becomes <ﻡ> in the end of words.

**The Script of Urdu**

The script of Urdu is based on the script of Arabic, but it is written somewhat differently. The Arabic script in which most Arabic books are written in South Asia is called naskh while the script in which most Urdu books are written is called the nastaleeq. Nastaleeq is the script of Persian which is a modified form of the Arabic
script. It can, therefore, be called the Perso-Arabic script. In Pakistan, Punjabi, Siraiki, Hindko, Balochi, and Brahvi use the nastaleeq script. The naskh is used for writing Sindhi and Pashto. However, all these languages have modified the graphemes of the Arabic script to represent the phonemes of their languages. This process is also going on in the case of Burushaski, Shina, Wakhi, Balti, Khowar, and other minor languages in Pakistan.

Besides representing the segmental features of languages, the non-segmental features of languages such as stress, tone, and rhythm can also be represented by diacritical marks.

**Other Scripts and the Principles They are Based Upon**

Apart from representing segments, a script can also have graphemes for syllables. For instance, the word *only* has two syllables *on* and *ly*. We can have only two graphemes (instead of the four we have in English) to represent this word. Such scripts are called syllabic scripts.

A third kind of phonographic script is called featural. Featural scripts are somewhat unique. Features are the + – signs we use for the qualities of sounds. A sound could be a consonant and we call it + consonantal, – vocalic. It could be voiced and we call it + voice, and so on and so forth. Now, if there is a writing system which has graphemes showing these or other features of sounds, it would be a featural script.

There is, in fact, such a script. It was made in Korea in the fifteenth century and is called Hangul (Coulmas 1989; Sampson 1985). There were 28 graphemes which relate to the phonetic features of sounds. Nowadays, there are only 24 (four became obsolete), and they represent the sounds of the language classified as continuants, stops, vowels, and so on.
Now arrange all these systems of writing in a simple diagram:

Semasiographic
(not writing proper)

Logographic

Based on
word (non-existent)

Morphemic
(based on morpheme)

Syllabic
Segmental
Featural

Glottographic

Phonographic

Adapted from Sampson 1985:3

Semasiography, the representation of sense directly through pictures or signs, is not writing proper. Logographic systems based on a word (where a word means a combination of several morphemes) are not known so they are only a hypothetical possibility. Chinese, as you will recall, is based on morphemes which may also represent words, i.e. words of one morpheme (See Coulmas 1989 for Chinese).

The History of Writing

All writing does not come from one common ancestor. However, it is possible that segmental phonographic scripts—such as those of Urdu, Punjabi, English, Persian, and Sanskrit etc.—do have one source. Let us, however, look at the earliest forms of writing before coming to this interesting theory.

Evidence suggests that the Sumerian script is probably the oldest script known to us. Sumer flourished in the Middle East (roughly in Iraq) about 4500 to about 1750 years before the birth of Christ. Some scholars go so far as to narrow the search for the originator of writing to the city of Uruk in modern Iraq in about 3300 BC (Robinson 2007: 11). Nobody can be sure of this, but one thing is certain and that is that clay tablets, or tokens, with marks made by a wedge-shaped stick have been found from the ruins of these ancient dwellings of our ancestors. Are these ‘tokens’ for counting lists of goods? Probably yes. Certain marks were made on clay
probably to record a list of items or something connected with commerce and administration. The stick used for making these marks made wedge-like shapes. Later, European scholars, who used Latin words, called these shapes ‘cuneiform’ (cunei = wedge in Latin). Some of these shapes represented pictures initially, i.e. they were pictograms. Later, sounds also came to be represented. Here is the name of Hammurabi, the man who gave the first written code of law, to the human race:

\[ \text{ha + am + mu + ra + bi} \]
\[ \text{fish + wild bull + ear + ? + innkeeper} \]

The point is that only the sound of fish (ha) and wild bull (am), ear (mu), and innkeeper (bi) were considered, not the meaning. Thus, certain marks or shapes (graphemes) came to stand for certain sounds and a phonographic—but only partly phonographic—system or writing was born.

The Aramaic script, based upon the Phoenician script, was used in South Asia for over 1000 years. It was the official script of a number of ancient empires whose names will come later in this chapter. Some people say it was brought to ancient India by traders who came from Egypt, Turkey, and the Middle East (Robinson 2007: 172). But Aramaic was not the name of only a script; it was also the name of a language. Indeed, it was an important Semitic language (the family to which Arabic and Hebrew belong) and was probably the mother tongue of Jesus Christ. It was one of the earliest influences of the Middle East on India. But India did have its own scripts even then. These will be described later.

**Hieroglyphics and the Scripts of the Middle East**

The Egyptian system of writing, called hieroglyphics by modern scholars, was used around 3000 BC in Egypt. It is also made up of pictures, i.e. ideograms as well as phonograms. It was a very complex system which is being deciphered even now. Some of the shapes of the hieroglyphs were borrowed by the inventors of the Semitic alphabet who used them to represent speech sounds. This was the making of the alphabet as we know it to which we turn now.
<table>
<thead>
<tr>
<th>Phoenician</th>
<th>Early Greek</th>
<th>Modern Symbols</th>
<th>Hebrew</th>
<th>Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>א</td>
<td>α</td>
<td>,</td>
<td>aleph</td>
<td>א</td>
</tr>
<tr>
<td>ב</td>
<td>β</td>
<td>b</td>
<td>beth</td>
<td>ב</td>
</tr>
<tr>
<td>ג</td>
<td>γ</td>
<td>g</td>
<td>gimel</td>
<td>ג</td>
</tr>
<tr>
<td>ד</td>
<td>δ</td>
<td>d</td>
<td>daleth</td>
<td>ד</td>
</tr>
<tr>
<td>ה</td>
<td>η</td>
<td>h</td>
<td>heth</td>
<td>ה</td>
</tr>
<tr>
<td>ו</td>
<td>Ϝ</td>
<td>w</td>
<td>waw</td>
<td>ו</td>
</tr>
<tr>
<td>ז</td>
<td>ζ</td>
<td>z</td>
<td>zayin</td>
<td>ז</td>
</tr>
<tr>
<td>ח</td>
<td>χ</td>
<td>h</td>
<td>heth</td>
<td>ח</td>
</tr>
<tr>
<td>ט</td>
<td>ϴ</td>
<td>t</td>
<td>teth</td>
<td>ט</td>
</tr>
<tr>
<td>י</td>
<td>ι</td>
<td>y</td>
<td>yod</td>
<td>י</td>
</tr>
<tr>
<td>ק</td>
<td>κ</td>
<td>k</td>
<td>kaph</td>
<td>ק</td>
</tr>
<tr>
<td>ל</td>
<td>λ</td>
<td>l</td>
<td>lamed</td>
<td>ל</td>
</tr>
<tr>
<td>מ</td>
<td>μ</td>
<td>m</td>
<td>mem</td>
<td>מ</td>
</tr>
<tr>
<td>נ</td>
<td>ν</td>
<td>n</td>
<td>nun</td>
<td>נ</td>
</tr>
<tr>
<td>ס</td>
<td>σ</td>
<td>s</td>
<td>samekh</td>
<td>ס</td>
</tr>
<tr>
<td>ע</td>
<td>υ</td>
<td>ᵟ</td>
<td>ayin</td>
<td>ע</td>
</tr>
<tr>
<td>פ</td>
<td>ϕ</td>
<td>p</td>
<td>pe</td>
<td>פ</td>
</tr>
<tr>
<td>צ</td>
<td>϶</td>
<td>s</td>
<td>sade</td>
<td>צ</td>
</tr>
<tr>
<td>ק</td>
<td>ϕ</td>
<td>q</td>
<td>qoph</td>
<td>ק</td>
</tr>
<tr>
<td>ר</td>
<td>ρ</td>
<td>r</td>
<td>reš</td>
<td>ר</td>
</tr>
<tr>
<td>ש/ס</td>
<td>ϕ/σ</td>
<td>sh/s</td>
<td>shin</td>
<td>ש/ס</td>
</tr>
<tr>
<td>ת</td>
<td>ϱ</td>
<td>t</td>
<td>taw</td>
<td>ת</td>
</tr>
</tbody>
</table>

The Middle East is the cradle of the segmental phonographic script proper. Phoenician, Hebrew, and Arabic represented consonants by shapes which exist in modified forms in most of the languages of Europe and Asia. These languages had vowel sounds but did not
have any graphemes to represent them. For instance, in Arabic the consonants are written and the native speakers pronounce the vowels without actually writing them. That is why South Asian Muslims find it difficult to read the Quran printed in Arab countries. Thus, the languages which based their scripts on the Middle Eastern scripts also did not have shapes for vowels. Some of them did invent shapes for vowels but they were less than the number of vowels.

**The Scripts of Europe**

The Greeks borrowed their alphabet from the Middle East. The Romans in turn borrowed theirs from the Greeks, and all European scripts are derived from the Greek or the Roman. Arabic is, of course, of Middle Eastern origin as is Hebrew which is used in Israel. But Persian, Urdu, pre-modern Turkish, Punjabi, Siraiki, Hindko, Balochi, Brahvi, Pashto, Sindhi, Kashmiri, and other Muslim languages are also based on the Arabic script. Thus, we see that in the matter of script, European and Muslim Asian languages are related. They share a common ancestor. Remember, however, that these languages do not belong to one language family. Arabic and Hebrew are Semitic languages; Turkish is Altaic; while English, Urdu, Sindhi, Hindi, Bengali, Pashto, Persian etc. belong to the Indo-European family of languages. This script is important for us in South Asia because the subcontinent is home to so many Muslims. A number of languages used by them, notably Urdu, is written in this script. Let us now turn to the most widely used scripts in South Asia—those belonging to the Brahmi family of scripts.

**The Scripts of South Asia from Before the Arrival of the Perso-Arabic Scripts**

Manuscripts in Brahmi and Sanskrit languages have been found in Jaulian near Taxila. They are inscribed on burnt birch bark and are from the fourth and fifth century. Sanskrit existed in the Subcontinent from about 1000 BC. However, it was not always used for worldly purposes. A contemporary scholar of the language, Sheldon Pollock, tells us:
From around the beginning of the first millennium BCE when the earliest form of Sanskrit appeared in South Asia, until around the beginning of the first millennium CE, Sanskrit functioned as a communicative medium that was restricted both in terms of who was permitted to make use of the language and which purposes the language could subserve (Pollock 2006: 39).

The following are the graphemes of the ancient Brahmi script, the mother of most of scripts used in India and other countries today.

Source: http://www.ancient-scripts.com/brahm.html
Retrieved on 7 Jan 2010.
The Brahmi and the Kharosthi scripts, both shown above, are the oldest deciphered scripts of ancient India. An undeciphered script—if it is a script at all—is the Indus Valley script which is found in Mohenjo Daro and Harappa. The Brahmi script is written from left to right. It is the parent of most of the Indian scripts including the Devanagari script. It has also given rise to the Burmese, Thai, Lao, Khmer, and several other scripts (Bright 1966: 373–74). Kharosthi was restricted to the areas now in Pakistan and parts of Afghanistan. It was written from right to left and Ashoka’s edicts in this part of the world are in this script.

This leaves the development of the Devanagari script out. Some people think it too came across the sea from the Middle East. Most of the scholars, however, believe it is derived from the Brahmi script which has been mentioned above. The north Indian branch of the Brahmi script is called the Gupta branch, while the south Indian one is the Grantha branch. This script is used for writing Gujarati which is used both in Pakistan and India. The Devanagari script which is used for writing Hindi, Marathi, and Nepali is a huge script considering the population of the people using these languages. The scripts used for writing Hindi, Punjabi, Bengali, and Oriya in India are based on the Gupta system. The variant used for writing Punjabi is called Gurmukhi in India. The script is considered sacred by the Sikhs since it is said to have come out of the mouth (mukh) of the saintly teacher (guru). In Pakistan, it is written in the Perso-Arabic script which is also called Shahmukhi in the context of Punjabi, i.e. that which was created by the state (shah means ruler). The scripts of Tamil, Telugu, Malayalam, and Kannada are based on the Grantha system.

**THE MOST POPULAR SCRIPT OF INDIA AND NEPAL**

The Devanagari script with which students in India and Nepal are familiar but which the Hindi cinema and TV shows are showing to Pakistanis also, is based on articulatory phonetics. First come the primary vowels both short and long. These are followed by the secondary vowels, i.e. diphthongs or two vowels joined together.
Then there are velar, palatal, retroflex, dental, and labial sounds. To make the chart more organized, the unaspirated (without the little ‘h’ sound) is followed by the aspirated sound. The following chart illustrates this (The abbreviations are: Dev = Devnagari; Guj = Gujarati; Pun = Punjabi; Ben = Bengali; Ori = Oriya; Tel = Telugu; Kan = Kannada; Tam = Tamil; Mal = Malayalam; Sin = Singhalese).

**Consonants**

<table>
<thead>
<tr>
<th>Velars</th>
<th>DEV</th>
<th>GUJ</th>
<th>PUN</th>
<th>BEN</th>
<th>ORI</th>
<th>TEL</th>
<th>KAN</th>
<th>TAM</th>
<th>MAL</th>
<th>SIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>k</td>
<td>क</td>
<td>ข</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>က</td>
<td>ข</td>
<td>կ</td>
<td>ค</td>
<td>က</td>
</tr>
<tr>
<td>kh</td>
<td>ख</td>
<td>ข</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>က</td>
<td>-</td>
<td>က</td>
<td>က</td>
<td>က</td>
</tr>
<tr>
<td>g</td>
<td>ग</td>
<td>ง</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>గ</td>
<td>-</td>
<td>ง</td>
<td>ง</td>
<td>ง</td>
</tr>
<tr>
<td>gh</td>
<td>घ</td>
<td>ง</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>घ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ñ</td>
<td>ण</td>
<td>ง</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>്</td>
<td>್</td>
<td>്</td>
<td>್</td>
<td>്</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Palatals</th>
<th>DEV</th>
<th>GUJ</th>
<th>PUN</th>
<th>BEN</th>
<th>ORI</th>
<th>TEL</th>
<th>KAN</th>
<th>TAM</th>
<th>MAL</th>
<th>SIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>c</td>
<td>च</td>
<td>น</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>icients</td>
<td>น</td>
<td>ញ</td>
<td>น</td>
<td>န</td>
</tr>
<tr>
<td>ch</td>
<td>छ</td>
<td>น</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>rstrip</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>j</td>
<td>ज</td>
<td>ज</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>jets</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>jh</td>
<td>झ</td>
<td>झ</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>şı</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ñ</td>
<td>ण</td>
<td>ण</td>
<td>്</td>
<td>್</td>
<td>್</td>
<td>ṇ</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
### Retroflexes

<table>
<thead>
<tr>
<th>retro.</th>
<th>DEV</th>
<th>GUJ</th>
<th>PUN</th>
<th>BEN</th>
<th>ORI</th>
<th>TEL</th>
<th>KAN</th>
<th>TAM</th>
<th>MAL</th>
<th>SIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>ट</td>
<td>ट</td>
<td>ट</td>
<td>ट</td>
<td>ट</td>
<td>ٹ</td>
<td>ट</td>
<td>ट</td>
<td>ट</td>
<td>ट</td>
</tr>
<tr>
<td>th</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
</tr>
<tr>
<td>d</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
</tr>
<tr>
<td>dh</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
<td>ठ</td>
</tr>
<tr>
<td>n</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
<td>ण</td>
</tr>
</tbody>
</table>

### Dentals

<table>
<thead>
<tr>
<th>dentals</th>
<th>DEV</th>
<th>GUJ</th>
<th>PUN</th>
<th>BEN</th>
<th>ORI</th>
<th>TEL</th>
<th>KAN</th>
<th>TAM</th>
<th>MAL</th>
<th>SIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
<td>त</td>
</tr>
<tr>
<td>th</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
<td>थ</td>
</tr>
<tr>
<td>d</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
<td>द</td>
</tr>
<tr>
<td>dh</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
<td>ध</td>
</tr>
<tr>
<td>n</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
<td>न</td>
</tr>
</tbody>
</table>

### Labials

<table>
<thead>
<tr>
<th>labials</th>
<th>DEV</th>
<th>GUJ</th>
<th>PUN</th>
<th>BEN</th>
<th>ORI</th>
<th>TEL</th>
<th>KAN</th>
<th>TAM</th>
<th>MAL</th>
<th>SIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
<td>प</td>
</tr>
<tr>
<td>ph</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
<td>फ</td>
</tr>
<tr>
<td>b</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
<td>ब</td>
</tr>
<tr>
<td>bh</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
<td>भ</td>
</tr>
<tr>
<td>m</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
<td>म</td>
</tr>
</tbody>
</table>
Figure 4 gives the origin and relationship of writing systems with each other.

<table>
<thead>
<tr>
<th>Glides (Semi-Vowels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>glides</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>y</td>
</tr>
<tr>
<td>r</td>
</tr>
<tr>
<td>l</td>
</tr>
<tr>
<td>v</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fricatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>fric.</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>s</td>
</tr>
<tr>
<td>s</td>
</tr>
<tr>
<td>s</td>
</tr>
<tr>
<td>h</td>
</tr>
</tbody>
</table>

Figure 4 gives the origin and relationship of writing systems with each other.
The invention of writing is perhaps the most significant development in the history of humanity. Had there been no writing, systematic knowledge would not have increased. That means that there would have been no science and the modern industrial revolution would not have taken place. It also means that there would have been no nuclear, chemical or biological weapons which now threaten to wipe out humanity. Whether the development of writing is good or bad is for you to decide in the light of these diverse facts.

One notable thing about scripts is that they are part of culture, history, and identity. Thus, Muslims may not want to leave the script in which Arabic is written even if there are advantages in using some other script. Similarly, Hindus and Sikhs may not want to leave their scripts. Matters of scripts, therefore, are not straightforward questions of rationality or utility as some people wrongly assume. They are questions of politics, economics, and identity. Moreover, scripts have centuries of a groups’ literary and cultural treasures—songs, stories, riddles, aphorisms, literature, and other cultural goods. People do not want to kill them by losing the ability to read the script in which they are written. That is why they keep struggling to preserve their scripts, even if it does not appear rational to do so. Questions of changing the script are often very difficult to decide. The linguist should keep these extra-linguistic matters in mind when questions of making new scripts for unwritten languages or modifying existing scripts are disputed.
Notes for the Reader

The purpose of this introductory book was to introduce you to basic concepts and the technical terms used in linguistics. This should enable you to read books of linguistics on your own.

As for the theoretical aspects of linguistics, no attempt has been made to introduce you to theories of linguistics other than those of Chomsky. If you want to read about other theories, you can refer to Robins’ chapter on the subject (1964: 278–288). You should read Halliday’s systemic grammar (Halliday 1961), Perlmutter’s relational grammar (Blake 1990), Pike’s tagmemics (Cook 1969), Lamb’s stratificational grammar, and other such theories.

If your interest is in the relationship of language with society, you can read sociolinguistics and anthropological linguistics. If you like computers, you may be interested in computational linguistics. If you have a flair for biology, you may like to focus on neurolinguistics or animal communication etc. Those who are interested in philosophy may like to study philosophical linguistics, which also has a mathematical side to it. Linguistics is not only about language itself—though linguistic theory is—but about so many aspects of reality that everyone can enjoy it. Even political scientists can investigate how language is related to the struggle for power in the world.

A brief bibliography has been provided in the end but this is not enough. If you are interested in further research, you should look into the bibliographies of the books suggested at the end. They will lead you on into the infinite world of the linguistic sciences.
Bibliography and Further Reading


1972.


Index

A
accessibility hierarchy, 125
acquisition planning, 164, 165, 176
adverbial phrase (Adv P), 85, 87
affix hopping, 109
Akbar, 10
Allomorphs, 58, 59, 60
Allophones, 51, 52, 130
Alpha movement, 109, 110
Alveolar sounds, 28, 33, 37, 43, 45
American Indian language, 128
Anthropological linguistics, 5, 7, 22,
64, 128, 129, 140, 198
Antonymy, 63, 64
Aphasia, 5
Apical sound, 29, 47, 48
Arabic
grammars, 8, 13, 15, 19, 32, 35, 36,
73
system of writing, 8, 166, 168, 185,
186, 188, 189, 190, 192, 197
Aramaic, 183, 188
Arbitrariness of language, 12
Aristotle, 6, 12
articulation, 24, 29, 31, 33, 37, 38, 47
Ashtadhyayi (Panini), 12
Asiatic Society of Bengal, 12
Aspects of the Theory of Syntax (Noam
Chomsky), 76, 103, 199
Austin, John L., 71
Aztec, 58

B
Badeshi, 171
Balochi, 32, 186, 190
Basque language, 23
Behaviourism, 23
Bengali, 32, 190, 192, 193
Berlin, B. and Kay, P., 147, 149
Bilabial sounds, 31, 37
Bilingual education, 178
Binary branching, 100
Binary oppositions, 22, 44, 46, 131
Biological foundations of language, 5
Bloomfield, Leonard, 3, 22, 23, 128
Boas, Franz, 22, 127, 128
Bopp, Franz, 14
Bourdieu, P., 181
Brahmi, 190, 191, 192
Brahvi, 56, 186, 190
Brugmann, Karl, 14
Burushaski, 16, 186

C
Chilliso, 171
Chinese, 16, 184, 187
Chomsky, Noam, ix, 4, 23, 66, 68,
72–81, 83–4, 103–4, 113, 117,
125–6
Chronos, 20
cognitive universals, 148
comparative linguistics, 13, 14, 18
competence and performance, 22, 76
complementary distribution, 130, 137,
138
complementizer phrases, (CP) 97
componential analysis, 66, 130, 132–
35, 137
computational and mathematical
linguistics, 5
connotata, 133
connotations of words, 154
consonants, 31, 37, 38, 42, 45, 189, 193
conventionalist and naturalistic debate, 13, 18
corpus planning and politics, 168
Cratylus (Plato), 12, 18, 68
Crystal, David, 3, 16, 66, 173–4
Cultural determinism, 147
Culture-specific beliefs, 8

D

De Lingua Latina (Marcus Terentius Varro), 13
dental sounds, 193
descriptive grammar, 73
Devanagari script, 6, 185, 192
diachrony, 20
diacritical marks, 30, 186
diglossia, 142, 200
diphthongs, 42, 185, 192
Domaaki, 171
dorsal sound, 29, 47–8
Durkheim, Emile, 22

E

Educational linguistics, x, 140, 175–6, 181
embedding, 82
English
  Dominant language of education, 179
  Grammar, 75
  morphology, ix, 2–4, 53, 61, 75, 137
  production of speech sounds, 24
  in South Asia, language of power, 163
English for Special Purposes (ESP), 176
English Language Teaching (ELT), viii, 176–7

ethnography of communication, 140
ethnology, 127–8
ethnic politics and language, 167
etymology, 14

F

family tree theory, 15
feral children, 10
Ferguson, Charles, 142
Foreign language teaching (FLT), 176
Formalization, 80, 84
Fortis sounds, 47–8
Frege, Gottlob, 68
French, 15, 18, 19, 22, 32, 58, 75, 80, 95, 159, 160, 181
Fricatives, 33
Functional suitability of language, 10, 11

G

Gender construction and role of language, 150–2
gender roles and language, 151–2, 154, 157
German, 13–15, 23, 40, 58, 68, 142, 176
glides (semi-vowels), 195
glottal sounds, 29
glottographic system of writing, 183–4
government and binding theory, 111
Gowro, 171
grammar
  linear model, 101, 103
  transformational generative (TG), ix, 80, 81–101
  universal, 75, 79, 109, 117, 118, 124, 126
types, 124
Grantha system of writing, 192
graphemes, 184–6, 188–9, 191
Greek, 6, 12, 14, 15, 30, 53, 62, 65, 141–2, 183, 185, 190
Grimm, Jakob, 14
Gujrati, 32, 192, 193
Gupta system of writing, 192
Gurmukhi, 192

H

Haegeman, Liliane, 111–12
Hali, Altaf Hussain, 157
Halle, Morris, 44, 47
Hammarabi, 188
*Handbook of American Indian Languages* (Franz Boas), 128
Harris, Zellig, 72
Haugen, Einar, 163
Hebrew, 16, 72, 188–90
Hermogenes, 12
hieroglyphs (Egyptian system of writing), 188
Hindi
  grammar, 73
  morphemes, 168
  writing, 192
Hindko, 186, 190
Homo sapiens, 9
homograph, 65
homonymy, 63–4
homophone, 65
human brain, 74, 76–7
Humboldt, Wilhelm von, 22–3
Hymes, Dell, 127–9, 142
Hyponymy, 63–4

I

identity and language planning, 140
implicational universals, 147
Indo-European family of language, 14, 16, 190
Indus Valley script, 192
infixes, 56
inflection, 78, 94, 96
inflectional languages, 57
ingressive air stream, 25

Institute Ferdinand de Saussure, 22
Interactionism, 129
International Phonetic Association (IPA), 30, 33, 35, 42, 48
intonation, 10
Italian, 10

J

Jakobson, Roman, 44
Japanese, 119, 124
Jones, William, 14

K

Kashmiri, 190
Khan, Syed Ahmad, 163
Kharosthi, 192
Kinship systems and linguistic concepts, 132, 133, 139
Kinship terms and world view, 150
Kroeber, Alfred, 128
Kurdish, 174

L

labelling, 91
labial, 47, 193
labio-dental sounds, 32
language
  acquisition device (LAD), 76, 126
  and culture, relation, 127
  death, x, 164, 171–3, 175
  faculty, 125
  and gender, 140, 150, 158
  and identity, x, 140
  learning, 5, 125–6, 164, 179, 180
  as a system, 20, 59
  and world view, 140, 146
  planning (LP), 163, 170, 176
  politics, x, 140, 162
  rights, 172, 180
  shift, 164, 165, 172, 174, 175
  varieties, 143–6, 166
**Index**

**M**

macrolinguistics, vii, 6, 181
Malinowski, 128–9, 131
Mauss, Marcel, 22
Meaning
   in philosophy, 68
medium of instruction and schooling, 68
minimalism, 100
modernity and globalization, 172
Moore, G.E., 70
morphemes
   classification, 54
   bound, 54, 56
   functional, 55
   inflectional, 54, 57–8
morphological rules, 60, 78
morphology, Ix, 2–4, 53, 61, 75, 137
morphophonology

**P**

palatal, 34–5, 37, 193
palato-alveolar sound, 34
Panini, 12
Paradigmatic relationship, 21
parole, 4, 21–2, 128
Pashto, 5, 15, 19, 32, 35, 142, 148, 149, 151, 158, 169, 186, 190
performance, 4, 22, 67, 76
Persian, 8, 13–15, 17, 19, 35–6, 57–8, 73, 75, 153, 155, 157, 163, 168, 185, 187, 190
Perso-Arabic script, 186, 192
prescriptive grammar, 73
pharyngeals, 36, 46
pharynx, 26, 36
phatic communion, 129
philology, 13, 18, 23
Philosophical Investigations (Ludwig Wittgenstein), 70
Phoenician, 182, 189
phonemes, 10, 30, 32–3, 37, 43–5, 49–52, 60, 130, 184, 186
phonetics (study of sounds), ix, 2, 3, 12, 23, 24, 30, 37, 192
phonograms, 188
phonographic script, 186, 189
phonological rules, 48, 59, 78
phonology (sound patterns), 2–4, 44, 52, 75, 130, 131, 137
phrase structure (PS) rules, 88, 93, 101–04, 106
Plato, 6, 12
pluralization morphemes, 60
politeness
  politics of language, 158
Pollock, Sheldon, 190–91
polysemy, 63–4
Portuguese, 8, 58
positron emission tomography (PET), 5
posterior sounds, 45
pragmatics, 2, 3, 70–1
Prague School of Linguistics, 21
Prefixes, 56
prepositional phrase (PP), 85, 87–8, 90–1, 105–6, 118, 120, 122
preventive linguistics, 175
projection principle, 119, 124
pronunciation of foreign languages, 30
properties, 10, 24, 66–7, 74, 79, 81, 117, 120, 124, 133
proto Indo-European (PIE) family tree, 15, 17
pulmonary (egressive) stream, 25
Punjabi, 11, 15, 32, 51, 55, 119, 149, 158, 168, 186, 187, 190, 192
received pronunciation (RP), 146
recursiveness, 82
register, 141–2
relative clauses, 124–5
retroflex sounds, 33
Reversing Language Shift (RLS), 165
Romans, 13, 190
Ronat, Mitsou, 80
Roy, Raja Ram Mohan, 163
rule and principles, 74, 76–7, 82, 117
Russell, Bertrand, 68–70, 72
Russel, Peter, 203
Ryle, G., 71
S
Sanskrit, 12–15, 17, 57, 168, 187, 190–91
Sapir, Edward, 22, 128, 146, 148
  See also Whorf, Benjamin Lee, 22, 128, 146, 147–8
Sarwar, Zakia, 176
Saussure, Ferdinand de, 3, 4, 13, 18, 19, 20, 22, 128
Schleicher, August, 15
schwa, 42
scripts, 5, 6, 170, 182–8, 190–2, 197
  See also writing, ix, x, 2, 13, 24, 74, 84, 88, 143, 145, 164–5, 173, 175, 178, 182, 183–4, 186, 188, 191–2, 197
second language teaching (SLT), 176
segmental and non-segmental features, 186
semantics (study of meanings), ix, 2, 3, 12, 62, 66, 68, 71, 75, 119, 128, 133
  component, 103
  problems, 66
  rules, 78
semasiographic systems, 183
semi vowel, 31, 35
Semitic family, 15

R
cradle-Brown, 131
Radford, Andrew, 109, 110
Rask, Rasmus, 14
INDEX

sentence, 20–1, 55, 67, 69, 73–4, 76–7, 80–3, 85–9, 91–2, 94, 96, 97–8, 101–9, 111, 113–15, 117–21, 123–4, 150, 155, 180
sexual differences, 151
Sign, 19, 44, 51, 148, 183
Sindhi, 5, 15, 25, 27, 32, 51, 169, 186, 190, 196
Siraiki, 25, 150, 166, 186, 190
Skutnabb-Kangas, Tove, 172, 174, 179
social factors and linguistic change, 129
Society of Pakistan English Language Teachers (SPELT), 176
sociolinguistics, 5, 140, 198
Socrates, 12
sound
classification, 29
patterns, 44
physical properties, 24
South Asian language
Production of speech sounds, 24
Spanish, 32
speech defects, 5
speech sounds, production, 24
standardization, 145, 165
stops (plosives), 27, 31–2, 35, 45
Strawson, P.F., 71
strident sounds, 46
structural change (SC), 104
structural description (SD), 104, 106, 108
structuralism
structure, deep and surface, 83
structure dependency principle, 117, 118, 124
structure preserving principle, 110
style, 6, 13, 156–7
suffix, 56
Sumerian script, 187
Sweet, Henry, 30
Symbols used in writing rules, 84
synchrony versus diachrony, 20
synonymy, 63
syntactic rules, 78
Syntactic Structures (Noam Chomsky), 80, 103, 199
syntactic structures; tree diagram, 91
syntagmatic versus paradigmatic relationship, 20
syntax, 2–4, 10, 67–8, 73–6, 80, 83, 93, 100, 103, 109, 120–21, 124
T
Tarski, Alfred, 69
Teaching of English as a Foreign Language (TEFL), 176
Teaching of English as a Second Language (TESL), 177
Teeth, speech sound, role in, 27–9, 32, 45
Thanvi, Maulana Ashraf Ali, 157
Thematic (theta) roles, 122, 124
Theoretical language, 141
Tongue, speech sound, role in, 12, 26–9, 32–6, 38–9, 40, 42, 45–7
Tractatus (Ludwig Wittgenstein), 69, 70
transcription, 24, 42–3
transformational generative (TG) grammar, 80, 81, 104
transformational rules (T-rules), 103–4, 106
passive transformation, 105–6, 108
Trubetskoy, Nikolay Sergeyvich, 22
Turkish, 16, 58, 174, 190
Tyler, Sir Edward, 127
U
United Nations Educational Scientific and Cultural Organization (UNESCO), 173, 178
Urdu
grammar, 13, 75, 77–8, 95, 126
morphemes
script, 32, 185
Ushojo, 171
Uvular sounds, 36–7

V
Varro, Marcus Terentius, 13
velar sounds, 28, 35
verb phrase (VP), 85, 87–9, 90–2, 97, 103–4, 112, 116, 118–9, 126
vibration, 26
vocal cords, 25–6, 29, 31, 36, 44–5
voicing, 26
vowels
classification, 38, 47
cardinal, 39, 40–2
dipthongs, 42, 192

W
Wandt, Wilhem, 23
Western language policies, 179
Whorf, Benjamin Lee, 22, 128, 146–8
Wittgenstein, Ludwig, 69, 70–1
women’s language (WL), 156–7
writing
history, x, 187
in South Asia, 182–197

X
X-bar theory, 93, 94, 118