UNIT 1 BASIC NOTIONS OF SYNTACTIC CONSTITUENCY

Structure

1.0 OBJECTIVES

The objectives of this unit are that at the end of this unit you should be able to

- begin to see what the study of syntax is all about, with reference to English,
- identify different kinds of relationships that are possible in a phrase structure tree,
- apply the various tests to identify a syntactic constituent,
- distinguish between a phrase and a clause through phrase structure rules,
- draw syntactic trees for phrases and clauses using X-bar theory.

1.1 INTRODUCTION

1.1.1 Background: Syntax and structural linguistics

A beginning definition of the term *syntax* might be that it is the study of sentences and their structure (just as morphology is the study of words and their structure). Therefore, English syntax is the study of sentences and their structure in the English language. The study of syntax has, however, acquired a special significance since the mid-1950s or so. During the four decades or so prior to the mid-1950s, the study of grammar was undertaken within a broader approach that came to be known as *Structuralism*. Structuralism analyzed human phenomena such as language, literature and culture as *structured systems*--and not merely as chaotic masses of unrelated elements. As part of this broader view, *structural linguistics* also

developed--notably beginning with the Swiss linguist Ferdinand de Saussure (who is therefore often called "the father of modern linguistics")--as the study of language viewed as a structured system with mutually inter-related elements such as phonemes, morphemes, words, and phrases. With this emphasis on the detailed study of language structure in all its complexity, morphology and syntax emerged as distinct branches of linguistic analysis (whereas traditional grammar had lumped them together simply as "grammar"). Saussure's initial ideas on language were elaborated independently in the United States of America by linguists such as Edward Sapir and Leonard Bloomfield, and in Europe by the Prague School linguists Nikolay Sergeyevitch Trubetzkoy, Roman Jakobson, and a few others; in anthropology, they were applied to the analysis of cultures notably by Claude Lévi-Strauss. Within this broader approach called *structural linguistics*, a particular school of linguistic analysis of data from actual languages emerged in the United States under the intellectual leadership of Leonard Bloomfield, which became known as *descriptive linguistics*, or the *descriptivist* approach to linguistics. Descriptive linguistics had as its goal the description of all the observable aspects of the structure of a language. No attempts at explanation were considered solid enough to be admitted within this approach. The descriptive linguist's task was to ask *what* the linguistic data were and *how* they were structured -- not to ask *why* they occurred the way they did. Descriptive linguistics was thus strongly *empiricist*, i.e., factually oriented. The analysis of words and of sentence structure was thus undertaken in a step-by-step manner, beginning with the largest unit (the word or the sentence), which was then successively broken down into smaller and smaller constituent units (much as visible matter is analysed as consisting of molecules, which in turn are subdivided into atoms, the latter again being analysable into components including various subatomic particles, etc.). This method of linguistic analysis thus came to be known as *Immediate Constituent Analysis* (or IC Analysis). Furthermore, descriptive linguistics assumed, along the lines of the approach in psychology called *behaviourism* that was then popular, that people learned and used languages through an intricate series of stimuli and responses, much as Pavlov's dog learned to salivate upon hearing the ringing of a bell associated with the arrival of food even when no food was actually brought in, or as rats learned to find their way through experimental mazes in order to get to pieces of cheese. There was thus no allowance made in descriptive linguistics for the existence of the human mind, or for any mentalistic explanation: the human mind was assumed to be simply a clean slate, a *tabula rasa*, on which linguistic data came to be imprinted through stimulus-response sequences as an infant learned a language in the course of growing into a fully articulate adult human being.

In the early 1950s, however, a descriptive linguist named Zellig Harris put forward the idea that certain systematic and non-mechanical relationships could be identified among the sentences of a given language -- an idea that was elaborated further by a gifted young student of his named Noam Chomsky, who around the mid-1950s put forward a very original and radically different view of languages and of the power of language. This new approach came to be called the *generativist* view of language, or *generative grammar*. Chomsky argued that a theory of language must have linguistic explanation as one of its primary goal, and that it must be related to properties of the human mind, since only human beings are capable of using language in a truly creative and sophisticated manner. Furthermore, certain features were to be found in every human language, precisely because the power of language, or the *faculty of language*, was innate and species-specific to all normal human beings. These features were thus integrated at a very abstract level into **Universal Grammar** (UG), which took on additional language-specific features in different languages of the world. The study of sentence structure could thus tell us deeper things about the organization of thought by the human mind. Thus, the study of sentence structure, viz., *syntax*, ceased to be simply the Immediate Constituent Analysis of sentences into their constituent phrases and the further components of these phrases. Instead, syntax came to occupy a central position in generative linguistics, and continues to do

so in ongoing work by Noam Chomsky himself and many other linguists all over the world.

1.1.2 Syntax in generative grammar

In generative grammar, syntax now has its own set of methods. Furthermore, the greater portion of these methods have been found to be applicable to the analysis of not only the structure of English and a few other well-known European languages such as French, German, Spanish, Italian, etc., but that of many non-western languages, both well-known ones such as Arabic, Swahili, Chinese, and Japanese, as well as lesser-known languages spoken in isolated geographical areas. In this and in the subsequent units of this module, our concern will be with the syntax of English sentences from the viewpoint of Universal Grammar. The syntax of a language like English, seen from this viewpoint, is a very important part of the **generative grammar** of English, which is an abstract body of rules and principles that can tell us how the words, phrases, and sentences of a language -- in the present case, English -- are built.

As we shall get to see later, a generative grammar of a language consists of the following components: a **lexicon**, a set of rules or a **rule schema** to represent **phrase structure**, a number of **modular** sub-theories that contain **principles** obeyed by expressions in the language, a single but powerful (major) **transformational rule**, a **phonological representation** for each expression in the language (i.e., a representation in terms of the way the expression is pronounced), and a **logical form** for each expression (i.e., a representation that reflects the way in which the expression's meaning is interpreted -- i.e., the way in which one makes sense of the expression); a few minor transformational rules are also included. In what follows in this unit and in Units 2 to 6, we shall see what these various components are for the generative grammar of English.

1.2 THE IDENTIFICATION OF A SYNTACTIC CONSTITUENT

1.2.1 Introduction

Syntax is the branch of linguistics that is concerned with the study of units larger than the word: it is the study of the structure of *phrases*, *clauses*, and *sentences*. These different kinds of units larger than isolated words are different kinds of **syntactic constituents**, and are so called because they combine with each other as though they were single, indivisible units. To see this better, let us look at the simple English sentences (1a), (1b), (1c), and (1d) given below:

(1) a. Hari will speak to Veena.
b. Hari will speak to that girl.
c. This boy will speak to Veena.
d. This boy will speak to that girl.

Notice that the sequence of two words ***this boy*** in the sentences (1c) and (1d) occurs in exactly the same position as the proper noun *Hari* in the sentences (1a) and (1b). Notice also that the two-word sequence ***that girl*** in the sentences (1b) and (1d) similarly occurs in the same position as the proper noun *Veena* in the sentences (1a) and (1c). These are initial indications that the two-word sequences *this boy* and *that girl* are individual syntactic constituents. Now notice what happens if we try to replace *this boy* or *Hari* with some other two-word sequence such as *boy can* (which occurs in a sentence like (2)). The result will be patently absurd (and is marked by the asterisk * because it is absurd), as shown in the sentence (3):

(2) The boy can speak to me.
(3) *Boy can will speak to that girl.

(One can say: "This boy *can* and *will* speak to that girl". This sentence is not absurd in the way that the sentence (3) is, but notice that it is a different sentence from (3), with two extra words *This* and *and* in the correct places.) The two-word sequence *boy can* is therefore not a syntactic constituent, unlike *this boy*. We can similarly see that the two-word sequence *that girl* is a composite unit, in a way that, for example, the two-word sequence *girl will* is not. Furthermore, we can interchange the positions of *this boy* and *that girl* in a sentence like (1d) and get yet another good sentence, the sentence (4):

(4) That girl will speak to this boy.

This is an added indication that the two-word sequence *that girl* is a syntactic constituent, just like the two-word sequence *this boy* is: they can interchange positions with one another, i.e., are *distributional equivalents* of one another.

A commonly used term for a syntactic constituent is the term **phrase**. However, **clauses** and **sentences** are also special kinds of syntactic constituents. (We discuss clauses and sentences in greater detail in Unit 2.) The question now is: How do we go about determining whether a sequence of words within a given sentence is a syntactic constituent? In other words, what tests can we apply to determine whether a sequence of words is a syntactic constituent?

As it happens, there are several tests, i.e., criteria, that we can apply for a sequence of words in order to determine whether it is a syntactic constituent. We have informally applied two of these already: these are the criteria of **substitution** and **distribution**, respectively. In fact, there are five tests in all for syntactic constituency, which we shall be examining in Sections 1.2.2 to 1.2.6. We list these five tests below:

(i) substitution,
(ii) conjoining,
(iii) distribution,
(iv) replacement by anaphora,
(v) parenthetical insertion.

1.2.2 The test of substitution

A syntactic constituent has the property that it can either replace or be replaced by another syntactic constituent. Thus, to tell whether a sequence of words is a syntactic constituent, we have to try to replace it with another sequence of words that is known to be a syntactic constituent possessing certain features in common (e.g., containing a central word of the same part of speech). This kind of experiment with the sequence of words that we are interested in would be a test of *substituion*. For instance, all the sequences of words to the left of the broken line in the set (5), and even a single word such as the proper noun *Veena*, can replace one another to yield perfectly grammatical sentences:

(5)	a.	This little girl	came yesterday.
	b.	My brother	
	c.	The man who is sitting over there	
	d.	Veena	

Likewise, the sequence of words *came yesterday* in (5) can replace other similar sequences of words (or even single words), as shown in (6):

(6)	This little girl	came yesterday.
		is Veena's niece.
		ate up all the gulab-jamuns.
		stares.

Sentence (11b) is what is called a **cleft** sentence corresponding to (11a) (here, the NP *this little girl* is given special prominence, or is the **focus**). The two-word sequence *is smart* occurs as an undivided VP in (11b) as in (11a). In sentence (11c), too, the same sequence occurs as an undivided VP. Thus, the sequence *is smart* passes the test of distribution as a syntactic constituent: it occurs as an undivided whole in various distributional positions.

1.2.5 The test of replacement by anaphora

There is a special class of expressions called **anaphora** that "stand in" for, i.e., replace, certain kinds of syntactic constituents. Pronouns can occur in place of noun phrases when repetition of the noun phrase sounds tedious, for instance, and the phrase *do so* can similarly occur in place of a verb phrase. Pronouns and the phrase *do so* are thus examples of anaphora. The sentences (12b) and (13b), which contain a pronoun and the phrase *do so*, respectively, illustrate the uses of anaphora, especially when compared with (12a) and (13a):

(12) a. The boy who sits over there thinks the boy who sits over there is smart.
b. The boy who sits over there thinks he is smart.

(13) a. Monkeys eat bananas, and elephants eat bananas too.
b. Monkeys eat bananas, and elephants do so too.

Replacement by anaphora is an additional test for a syntactic constituents. Of course, not *all* syntactic constituents can satisfy this test, since there are only a few possible kinds of anaphora: thus, pronouns are NP-anaphora, the phrase *do so* is an example of VP-anaphora, and the word *such* can sometimes occur as anaphora for adjectival phrases. When we do find a sequence of words being replaced by anaphora, however, we can conclude that the sequence of words is a syntactic constituent. Note that anaphora cannot replace any sequence of words that is not a syntactic constituent. Thus, the pronoun *he* cannot replace the sequence of words *boy who sits over* alone, as the unacceptable example (14) shows:

(14) *The boy who sits over there thinks the he there is smart.

Replacement by anaphora is thus a criterion that a syntactic constituent *may* meet, provided that there are anaphora for a syntactic constituent of that particular kind (e.g., a NP, a VP, or an adjectival phrase).

1.2.6 The test of parenthetical insertion

In the everyday uses of English, certain words such as *incidentally*, *luckily*, *unfortunately*, and certain phrases like *by the way* and *of course*, ... are often inserted into sentences: they are not mandatory, but they often serve to link the sentences to other sentences before and after them in discourse, or to reveal the speaker's attitude towards what (s)he is saying in the sentence. Such expressions are often called **parenthetical expressions**. On trying to examine the positions in a sentence where such parenthetical expressions can be inserted, linguists have found that these positions tend to occur at the edges of syntactic constituents rather than within those syntactic constituents. Thus, compare the simple sentence (15a) with the acceptable sentence (15b) -- which has the parenthetical word *incidentally* inserted in it -- and the unacceptable (15c):

(15) a. My cousin is an avid mountaineer.
b. My cousin, incidentally, is an avid mountaineer.
c. *My, incidentally, cousin is an avid mountaineer.

Each of these sequences of words contain at least one **verb** -- *came*, *is*, *ate*, *stares* (this last one does not require any other supporting word) -- which is the most central word in the sequence. We therefore say that each such sequence of words is a syntactic constituent of the type **verb phrase**, indicated by the abbreviation VP. Even the single verb *stares* is an example of a VP, because it can replace other verb phrases and the result is a perfectly correct sentence of English. Similarly, in the examples (5a-c), the word sequences *this little girl*, *my brother*, and *the man who is sitting over there* each contain a **noun** -- *girl*, *brother*, *man* -- which is the most central, meaningful word in the sequence. We therefore say that *this little girl*, *my brother*, and *the man who is sitting over there* are examples of the syntactic constituent type **noun phrase**, or NP for short. The single proper noun *Veena* in example (5a) is also an instance of a NP, because it can replace any of the noun phrases shown in (5a-c) and the resulting sentence is a perfectly good sentence of English.

1.2.3 The test of conjoining

A syntactic constituent can be **conjoined** with another syntactic constituent. Thus, consider the example sentences (7a) and (7b):

(7) a. [[Veena] and [this little girl]] came yesterday.
b. Veena [[came yesterday] and [collected the letters]].

In (7a), the proper noun *Veena*, which is also a NP, is conjoined with *this little girl*, which is therefore a syntactic constituent (of the type NP). By contrast, the two-word sequence *this little* is not a syntactic constituent, since the sentence (8), with the sequence *this little* conjoined with *Veena*, is ungrammatical:

(8) *[[Veena] and [this little]] came yesterday.

Similarly, in (7b) the sequence of words *collected the letters* is a syntactic constituent, because it is conjoined with the VP *came yesterday*. However, the two-word sequence *collected the* is not a syntactic constituent at all, since, when it is conjoined with the VP *came yesterday*, the sentence (9) that results is ungrammatical:

(9) *Veena [[came yesterday] and [collected the.]]

Thus, we see that the test of conjoining can tell a syntactic constituent apart from a sequence of words that is not a syntactic constituent.

1.2.4 The test of distribution

A syntactic constituent occurs as an unbroken whole in various distributional positions. Consider the sentences (10a), (10b), and (10c):

(10) a. [This little girl] is smart.
b. I like [this little girl].
c. Veena gave the flowers [to [this little girl]].

In (10a), the three-word sequence *this little girl* occurs as the subject NP (the entity that the VP *is smart* is about). In (10b), the same sequence occurs as the direct object NP (the person towards whom my liking is directed). In (10c), it occurs as the indirect object NP (the object of the preposition *to*, and the person who is the recipient of the flowers that Veena gave). Now consider the sentences (11a), (11b), and (11c):

(11) a. This little girl [is smart].
b. It is this little girl who [is smart].
c. The one who [is smart] is this little girl.

Sentence (11b) is what is called a **cleft** sentence corresponding to (11a) (here, the NP *this little girl* is given special prominence, or is the **focus**). The two-word sequence *is smart* occurs as an undivided VP in (11b) as in (11a). In sentence (11c), too, the same sequence occurs as an undivided VP. Thus, the sequence *is smart* passes the test of distribution as a syntactic constituent: it occurs as an undivided whole in various distributional positions.

1.2.5 The test of replacement by anaphora

There is a special class of expressions called **anaphora** that "stand in" for, i.e., replace, certain kinds of syntactic constituents. Pronouns can occur in place of noun phrases when repetition of the noun phrase sounds tedious, for instance, and the phrase *do so* can similarly occur in place of a verb phrase. Pronouns and the phrase *do so* are thus examples of anaphora. The sentences (12b) and (13b), which contain a pronoun and the phrase *do so*, respectively, illustrate the uses of anaphora, especially when compared with (12a) and (13a):

(12) a. The boy who sits over there thinks the boy who sits over there is smart.
b. The boy who sits over there thinks he is smart.

(13) a. Monkeys eat bananas, and elephants eat bananas too.
b. Monkeys eat bananas, and elephants do so too.

Replacement by anaphora is an additional test for a syntactic constituents. Of course, not *all* syntactic constituents can satisfy this test, since there are only a few possible kinds of anaphora: thus, pronouns are NP-anaphora, the phrase *do so* is an example of VP-anaphora, and the word *such* can sometimes occur as anaphora for adjectival phrases. When we do find a sequence of words being replaced by anaphora, however, we can conclude that the sequence of words is a syntactic constituent. Note that anaphora cannot replace any sequence of words that is not a syntactic constituent. Thus, the pronoun *he* cannot replace the sequence of words *boy who sits over* alone, as the unacceptable example (14) shows:

(14) *The boy who sits over there thinks the he there is smart.

Replacement by anaphora is thus a criterion that a syntactic constituent *may* meet, provided that there are anaphora for a syntactic constituent of that particular kind (e.g., a NP, a VP, or an adjectival phrase).

1.2.6 The test of parenthetical insertion

In the everyday uses of English, certain words such as *incidentally*, *luckily*, *unfortunately*, and certain phrases like *by the way* and *of course*, ... are often inserted into sentences: they are not mandatory, but they often serve to link the sentences to other sentences before and after them in discourse, or to reveal the speaker's attitude towards what (s)he is saying in the sentence. Such expressions are often called **parenthetical expressions**. On trying to examine the positions in a sentence where such parenthetical expressions can be inserted, linguists have found that these positions tend to occur at the edges of syntactic constituents rather than within those syntactic constituents. Thus, compare the simple sentence (15a) with the acceptable sentence (15b) -- which has the parenthetical word *incidentally* inserted in it -- and the unacceptable (15c):

(15) a. My cousin is an avid mountaineer.
b. My cousin, incidentally, is an avid mountaineer.
c. *My, incidentally, cousin is an avid mountaineer.

The operation that consists of the insertion of a parenthetical expression at some point in a sentence is called **parenthetical insertion**. The test of parenthetical insertion is an additional criterion that can sometimes help us decide whether a sequence of words forms a syntactic constituent. Note that it is not necessarily true that *every* kind of syntactic constituent resists parenthetical insertion: the sentence is a special kind of syntactic constituent, for instance, and it allows parenthetical expressions to be inserted within itself. Nonetheless, if a sequence of words does resist being broken up by parenthetical insertion, we should suspect that it might be a syntactic constituent, and see whether it satisfies any of the four earlier tests that we have discussed; it is likely that at least one of the four earlier tests will be satisfied, and thus we can conclude that the sequence of words is a syntactic constituent.

1.2.7 The head of a syntactic constituent

The five tests discussed so far show that it is possible to determine whether a sequence of words is a syntactic constituent. The most general kind of syntactic constituent is the **phrase**. We have already seen examples of noun phrases and verb phrases, and adjective phrases have also been mentioned. English also has **prepositional phrases** (or PPs for short) and **adverbial phrases** (or AdvPs). Examples of these are shown in (16) and (17), respectively:

(16) a. Ram is [in the garden].
b. The mangoes fell [from the tree].

(17) a. That boy will speak [very slowly] to this girl.
b. The soldier was killed [almost instantly].

In (16a), the three-word sequence *in the garden*, which contains the preposition *in* and the noun phrase *the garden*, is a prepositional phrase; in (16b), likewise, the three-word sequence *from the tree* containing the preposition *from* and the noun phrase *the tree*, is a prepositional phrase. In (17a), the two-word sequence *very slowly*, containing the **adverb** *slowly* and the **degree modifier** *very* (not[illegible], unlike in most books on English grammar, we do not call *very* an adverb -- instead we give it the special name of "degree modifier", since that is the function it performs), is an adverbial phrase. In (17b), likewise, the two-word sequence *almost instantly*, containing the adverb *instantly* and the **quantifier** *almost* (a type of word that we shall see further in Unit 6), is an adverbial phrase.

The specific type to which a multi-word syntactic constituent can be assigned depends on the most "central" or "indispensable" element in the syntactic constituent. This element is called the **head** of the syntactic constituent. (To understand this better, one might think of the head of a phrase as being comparable to the head of a family: just as the head of the family acts as the representative for the entire family in matters such as the payment of bills, etc., likewise the head of the syntactic constituent could be thought of as representing the character of the entire constituent.) In the examples that we have seen so far, the head of the noun phrase (NP) is the noun (N); the head of the verb phrase (VP) is the verb (V); the head of the prepositional phrase (PP) is the preposition (P); the head of the adverbial phrase (AdvP) is the adverb (Adv); and so on.

1.3 PHRASE STRUCTURE

1.3.1 The phrase structure tree and its nodes

The syntactic structure of a sentence can be, and often is, represented as an inverted tree diagram, its top being the sentence (S) which then branches into its **immediate**

(which we shall learn more about in Units 2-6). A special name is therefore given to this kind of aunt-niece relationship. The node V is said to **c-command** the node P, because the first branching node immediately above--and therefore dominating--V, i.e., VP, also dominates the node P (as its grandmother node). The c-command relationship is defined as follows:

(19) **c-command**: A node Y c-commands another node Z if and only if the first branching node (say, X) dominating Y also dominates Z, and neither does Y dominate Z nor does Z dominate Y.

A c-command relationship cannot therefore occur between a direct ancestor node and any of its descendant nodes.

A rather different notion from the notion of c-command is the relationship of **command**. A node "commands" another node if and only if they are dominated by the same S node(s). In other words, a node "commands" another node if they are *clause-mates*. In this sense, the first Det node in the tree (18) "commands" all the other nodes (except the ones that dominate it) under the S node, since there is only one S node, and it dominates all the other nodes in the tree. The notion of "command" in fact historically precedes the development of the idea of "c-command". However, the notion of "c-command" has been subsequently found to be more useful in a variety of ways, which will be seen in the units to follow.

1.3.2 Phrase structure rules in early transformational grammar

The first kind of generative grammar to be proposed by Chomsky in 1957 was **transformational (generative) grammar**. In this kind of grammar, two basic types of syntactic rules were proposed: these were (i) **phrase structure rules,** and (ii) **transformational rules**, or simply **transformations**. The phrase structure tree in (18) is in fact constructed out of a set of phrase structure rules of the following kind:

(20) a. S → NP Pred-P
b. NP → Det N
c. Pred-P → Aux VP
d. VP → V PP
e. PP → P NP

The specific words or morphemes are then inserted into the **terminal nodes** (the smallest, indivisible nodes) by further lexical insertion rules of the following sort:

(21) a. Det → *this*
b. Det → *that*
c. N → *boy*
d. N → *girl*
e. V → *speak*
f. P → *to*

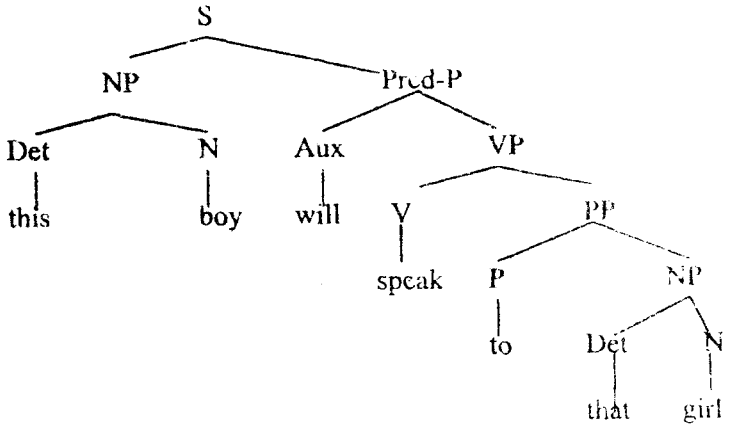
Rules of this kind are called **rewrite rules**, or **phrase structure rules**. They are called "rewrite rules" because the symbol on the left-hand side of the arrow is "rewritten", so to speak, in terms of the symbol(s) on the right-hand side. They are called "phrase structure rules" because they enable us to diagram the phrase structure of a phrase or sentence in the form of a tree.

1.3.3 From phrase structure rules to X-bar theory

The problem with phrase structure rules was found to be that they could be virtually of any kind, without any restrictions -- whereas in actual linguistic expressions there

constituents, NP and (tensed) VP; each of the constituents NP and VP then branches into its immediate syntactic constituents (e.g., V and object NP), ... and so on, until the smallest possible syntactic constituents, or **ultimate constituents**, are reached. The analysis of a linguistic expression successively into smaller and smaller immediate constituents was known prior to the days of generative grammar, and was called **Immediate Constituent Analysis**, or **IC Analysis** for short, in descriptive linguistics. Syntactic analysis more or less stopped after IC Analysis in the methodology of descriptive linguistics. In generative grammar, on the other hand, syntactic analysis of sentential structure only *begins* with IC Analysis. Because of this, it is convenient to represent the detailed syntactic structure of a sentence in the form of an inverted tree diagram, or **phrase structure tree** (earlier also called a **phrase-marker**), to which further syntactic operations can apply. Thus, the structure of a simple sentence like (1d) can be represented by the following phrase structure tree:

(18)



Let us now introduce a few new terms, to indicate the relations between the different components of this phrase structure tree. The labels for the various syntactic constituents, such as *S*, *NP*, *VP*, *Aux*, *Det*, *N*, *V* are **nodes** of the phrase structure tree: they are potential or actual joining-points for the different branches of the tree. The node S (for "Sentence") branches into the two nodes NP and Pred-P (Predicate-Phrase), which further branch into Det and N, on the one hand, and Aux and VP on the other. If we think of the nodes as being members of a family tree drawn for mothers and daughters, the node S is said to be the **mother** of the nodes NP and Predicate-Phrase (which are therefore the two **daughters** of the node S). We therefore say that the node S **immediately dominates** the nodes NP and Predicate-Phrase. Furthermore, the node S is the ancestor of the nodes NP and Predicate-Phrase and all the other nodes that NP and Predicate-Phrase branch into, and all the other nodes below those nodes; we therefore say that the node S **dominates** all the nodes that are its descendants (although it immediately dominates only its daughter nodes, NP and Predicate-Phrase). The nodes Det and N under each NP node (and therefore immediately dominated by the NP node) are **sister nodes**, since they have a common mother node, the NP node. All nodes that share a common mother node are sister nodes in relation to each other.

Notice also that the node Det occurs before the node N under each mother NP node. We therefore say that the node Det **precedes** the node N. Similarly, under the mother node VP, the daughter node V precedes its sister node PP.

What about a node that bears an aunt-niece relationship with another node? In the phrase structure tree (18), the node V is neither a direct ancestor (i.e., a mother, or grandmother, or great-grandmother, ...) nor a sister of the node P (which immediately dominates the preposition *to*). However, the node V can be said to be an aunt of the node P, since it is the sister of the node PP that is the mother of the node P. It so happens that, in the structure of English sentences, aunt-niece relationships play a vital role in the determination of links between certain kinds of expressions

(which we shall learn more about in Units 2-6). A special name is therefore given to this kind of aunt-niece relationship. The node V is said to **c-command** the node P, because the first branching node immediately above--and therefore dominating--V, i.e., VP, also dominates the node P (as its grandmother node). The c-command relationship is defined as follows:

(19) **c-command**: A node Y c-commands another node Z if and only if the first branching node (say, X) dominating Y also dominates Z, and neither does Y dominate Z nor does Z dominate Y.

A c-command relationship cannot therefore occur between a direct ancestor node and any of its descendant nodes.

A rather different notion from the notion of c-command is the relationship of **command**. A node "commands" another node if and only if they are dominated by the same S node(s). In other words, a node "commands" another node if they are *clause-mates*. In this sense, the first Det node in the tree (18) "commands" all the other nodes (except the ones that dominate it) under the S node, since there is only one S node, and it dominates all the other nodes in the tree. The notion of "command" in fact historically precedes the development of the idea of "c-command". However, the notion of "c-command" has been subsequently found to be more useful in a variety of ways, which will be seen in the units to follow.

1.3.2 Phrase structure rules in early transformational grammar

The first kind of generative grammar to be proposed by Chomsky in 1957 was **transformational (generative) grammar**. In this kind of grammar, two basic types of syntactic rules were proposed: these were (i) **phrase structure rules**, and (ii) **transformational rules**, or simply **transformations**. The phrase structure tree in (18) is in fact constructed out of a set of phrase structure rules of the following kind:

(20) a. S → NP Pred-P
b. NP → Det N
c. Pred-P → Aux VP
d. VP → V PP
e. PP → P NP

The specific words or morphemes are then inserted into the **terminal nodes** (the smallest, indivisible nodes) by further lexical insertion rules of the following sort:

(21) a. Det → *this*
b. Det → *that*
c. N → *boy*
d. N → *girl*
e. V → *speak*
f. P → *to*

Rules of this kind are called **rewrite rules**, or **phrase structure rules**. They are called "rewrite rules" because the symbol on the left-hand side of the arrow is "rewritten", so to speak, in terms of the symbol(s) on the right-hand side. They are called "phrase structure rules" because they enable us to diagram the phrase structure of a phrase or sentence in the form of a tree.

1.3.3 From phrase structure rules to X-bar theory

The problem with phrase structure rules was found to be that they could be virtually anything, without any restrictions -- whereas in actual linguistic expressions there

seem to be rather well-defined constraints on what can occur on the right-hand side of a phrase structure rule, given a particular node label on the left-hand side. Thus, there is no reason why a phrase structure rule like (23) should not be possible:

(22) AP → Det V

This kind of phrase structure rule would allow us, e.g., to form an adjectival phrase like **that speak* (and hence a noun phrase such as **a that speak boy*), which is absurd: speakers of English would dismiss such a phrase (especially such a noun phrase) as nonsensical.

Secondly, the phrase structure rule (23) would allow a Verb or a Determiner to be the head of an Adjectival Phrase, which is not only counterintuitive but in fact does not conform to the linguistic facts at all. The head of an Adjectival Phrase must be an Adjective, not a Verb or Determiner or any other lexical category of word. However, so far there was no restriction on phrase structure rules to the effect that a phrase must be rewritten so that at least one lexical category of the same type (nominal, verbal, adjectival, or prepositional) appears on the right-hand side of the rule.

The notion of phrase structure rules had to be modified in the late 1960s, therefore, and Chomsky himself was one of those to take the initiative in this direction. In his important 1970 article, "Remarks on Nominalization", Chomsky proposed that phrase structure rules must be specific instants of a simple but well-defined **rule schema** (i.e., a specification of what a phrase structure rule must be like), which came to be known as **X-bar theory**. X-bar theory simply consists of the following specifications:

(23) a. XP → (Specifier), X'
b. X' → (Adjunct), X'
c. X' → X, (Complement1), (Complement2)

(The commas between the symbols indicate that the relative order of the symbols separated by commas is irrelevant: thus, there may be languages in which X' has to be rewritten as [(Complement2) (Complement1) X].) Notice that a new kind of symbol has been introduced in these rules: this is the symbol **X'** (also **X**), which is read as "X bar" and which stands for a syntactic constituent whose head is of type X (i.e., N, or V, or A, or P, ... , etc.) but which is smaller than a phrase of type XP (i.e., NP, or VP, or AP, ... , etc.). A category of the type XP is called a **phrasal category** or a **maximal projection**; a category of the type X is called a **lexical category** or a **lexical head**; therefore, a category of the type X' is called an **intermediate category** or **intermediate projection**. This rule schema therefore dictates that only phrase structure rules of the following kinds will be admitted:

(24) a. NP → [Specifier Det] N'
b. N' → N' (PP)
c. N' → AP N
d. N' → N PP
e. AP → A'
f. A' → A

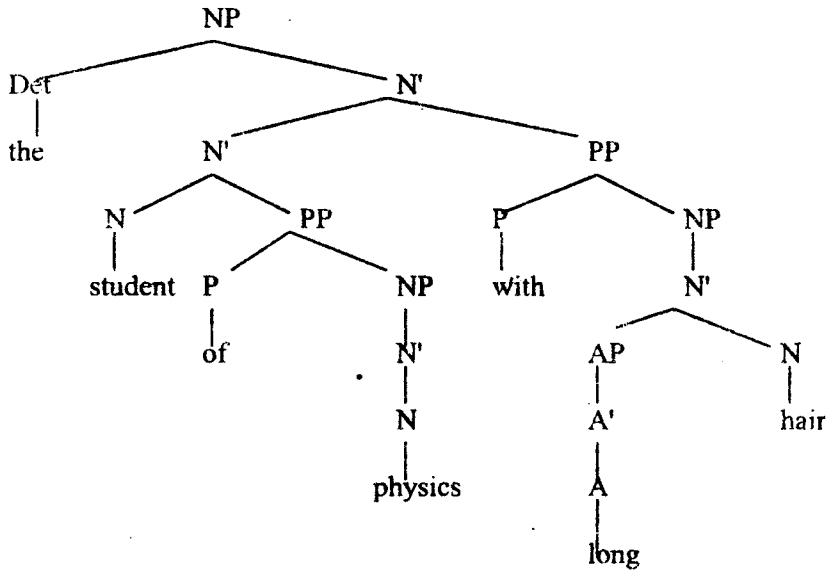
(25) a. VP → (Specifier) V'
b. V' → V NP NP

The structures of the two bracketed phrases in the sentence (26) are described by the phrase structure rules that we have given in (24) and (25):

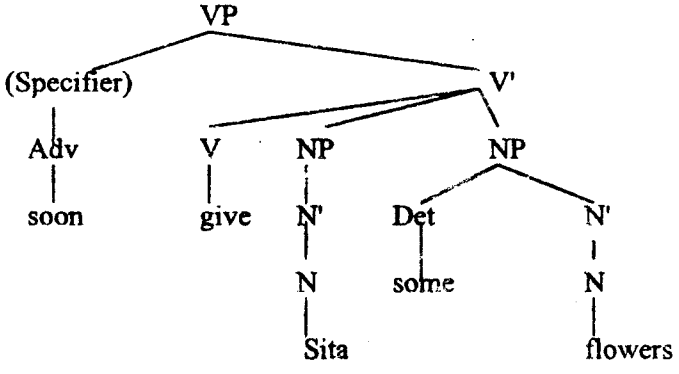
(26) [NP The student of physics with long hair] will [VP soon give Sita some flowers].

The respective structures are the following (i.e., (27a) and (27b)):

(27) a.



b.



Thus, X-bar theory allows the generation of phrase structure rules that can then be used to represent the tree structure of any phrase in English. X-bar theory gets its name from the assignment of different numbers of bars to the different category types X, X ', XP. Thus, in X-bar theory, a maximal phrasal category of the type XP is also represented as X" (X-double-bar); an intermediate category of the type X ' is of course assigned a single bar; and a lexical category of the type X is also represented as X° (X-zero or X-zero-bar).

1.4 CONCLUSION

We have thus seen that, in the study of syntax, the structure of an English sentence is arrived at step by step. The first step is to determine whether a word or sequence of words in an English phrase or sentence is a syntactic constituent: this is done by seeing whether the word or sequence of words passes one or more of a number of tests such as substitution, conjoining, distribution, replacement by anaphora, and parenthetical insertion. Once we have determined that the word or word sequence is a syntactic constituent, we can begin to formulate phrase structure rules that specify the internal structures of different kinds of syntactic constituents. However, a rule schema called X-bar theory can further specify what kinds of valid phrase structure rules can be formulated. The phrase structure of a sentence in English can then be put together on the basis of phrase structure rules in the form of a tree diagram that clearly depicts the syntactic structure of the sentence. We can then further identify

the different dominance, precedence, command, and c-command relations that hold among the different nodes of the tree diagram for the sentence. These steps constitute a method of syntactic analysis that we shall apply and find very useful for a better understanding of how the English language is built and put to use; we shall see this in greater detail in the units that follow.

1.5 LET US SUM UP

1. The study of English syntax began to receive major attention with the emergence of Structuralism in general and structuralist linguistics in particular.
2. With a shift in the focus of linguistics from Structuralist to Generative Linguistics, the study of English syntax took a new turn towards the working out of a generative grammar of English.
3. Generative linguistics has as its main goal an understanding of Universal Grammar, the body of principles and items that is common to all human languages and arises out of the innate nature of language in humankind.
4. Within generative linguistics, the study of the syntax of a language such as English seeks to describe and explain aspects of the structure of English sentences with the larger purpose of gaining a greater insight into Universal Grammar.
5. In syntax, the notion of a syntactic constituent is fundamental. Words, phrases, and clauses are all types of syntactic constituents.
6. There are at least five criteria that can be used to test whether a (randomly chosen) sequence of words in a sentence is a syntactic constituent.
7. The criteria for syntactic constituency are: (i) substitution, (ii) conjoining, (iii) distribution, (iv) replacement by anaphora, (v) parenthetical insertion.
8. Once a syntactic constituent is established, its structure can be visually depicted in a phrase structure tree diagram.
9. The details of the immediate constituents of any syntactic constituent have earlier been captured by rewrite rules or phrase structure rules.
10. The virtually unlimited power of phrase structure rules has since then come to be constrained by a rule schema known as X-bar theory.

1.6 KEY WORDS

Conjoining/Coordination: The joining or stringing together of two or more syntactic constituents by a CONJUNCTION (see below). Example: the noun phrase [[*this blue book*] *and* [*that red pen*]] has been formed by the conjoining or coordination of the two conjunct noun phrases [*this blue book*] and [*that red pen*], respectively.

Conjunction: A part of speech -- a word or pair of words that serves to conjoin or coordinate two or more syntactic constituents, e.g., ***and, but, either*** *...* ***or, neither*** *...* ***nor.***

Generative grammar: A view of grammar put forward by Noam Chomsky in the late 1950s, which holds that mere description of the structures found in any given language is inadequate and must be extended through formal explanatory mechanisms that account for the rule-governed creativity that is behind the production and understanding of those structures.

Structuralism: A major movement that took place in a number of fields of the humanities and the social sciences beginning in the earlier part of the twentieth century and continuing into the 1950s. In the field of linguistics, the Structuralist approach assumed that language consists of structured units that are most successfully described in terms of their distribution and relationships with one another.

Syntactic constituent: A syntactically significant element or word or sequence of words that behaves as a unified entity with respect to the criteria or tests of substitution, conjoining, distribution, replacement by anaphora, and parenthetical insertion. Words, phrases, and clauses are typical syntactic constituents.

X-bar syntax/theory: A rule schema that phrase structure rules have been found to conform to, in terms of maximal phrasal categories, intermediate categories, and (lexical) heads (such as N, V, A, P).

1.7 BIBLIOGRAPHY

On the changeover from structuralist to generative linguistics: Lyons 1977: chs. 1-5.

On the goals of generative linguistics: McCawley 1988: ch. 1 (Introduction); Radford 1981: ch. 1.

On syntactic constituent structure: Radford 1981: ch. 2; Radford 1988: ch. 2.

On X-bar theory: Radford 1981: ch. 3; Radford 1988: chs. 4, 5.

Lyons, John (1977) **Chomsky**. [Series: Fontana Modern Masters.] Glasgow: Fontana/Collins.

McCawley, James D. (1988) **The Syntactic Phenomena of English**, vol. 1. Chicago: University of Chicago Press

Radford, Andrew (1981) **Transformational Syntax**. Cambridge [U. K.]: Cambridge University Press.

Radford, Andrew (1988) **Transformational Grammar: A First Course**. Cambridge [U. K.]: Cambridge University Press.

1.8 QUESTIONS & EXERCISES

1. Consider the underlined word sequences in the sentences below:

   (i) Rahul gave his brother a toy car.
   (ii) Sita was reading two books together.

   Are these sequences of words syntactic constituents according to the criteria for syntactic constituency? Which of the criteria for syntactic constituency are applicable to each of these word sequences?

2. Draw four different phrase structure trees which can be generated by a set of phrases structure rules such as (a-d) below, together with a lexicon containing the entries in (e):

a. S → NP Aux VP
b. VP → V NP
c. NP → Det (AP) N
d. AP → (DegMod) A
e. *a(n)*, Det; *apple*, N; *arrest*, V; *buy*, V; *can*, Aux; *fairly*, DegMod; *girl*, N; *handsome*, A; *lad*, N; *may*, Aux; *nice*, A; *policeman*, N; *beautiful*, A; *rather*, DegMod; *ripe*, A; *sparrow*, N; *tall*, A; *tasty*, A; *that*, Det; *the*, Det; *this*, Det; *very*, DegMod; *will*, Aux.

If you feel that you need more items in the lexicon, enter your own items along with their parts of speech (as has been done above) and insert them into your phrase structure trees.

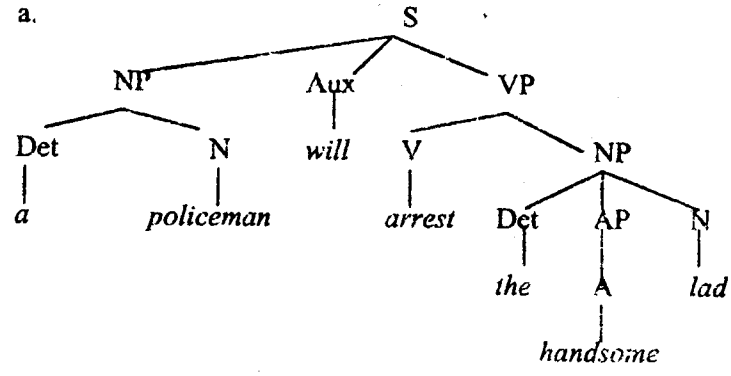
3. An earlier convention within X-bar theory has been to represent a question like *What did Ram do yesterday?* syntactically by the symbol S' ("S-bar"), where S is the symbol for the syntactic constituent Sentence, for a sentence like *Ram did a lot of work yesterday*. Does this convention conform strictly to X-bar theory? Either way, state the reasons why.

NOTES ON "QUESTIONS & EXERCISES"

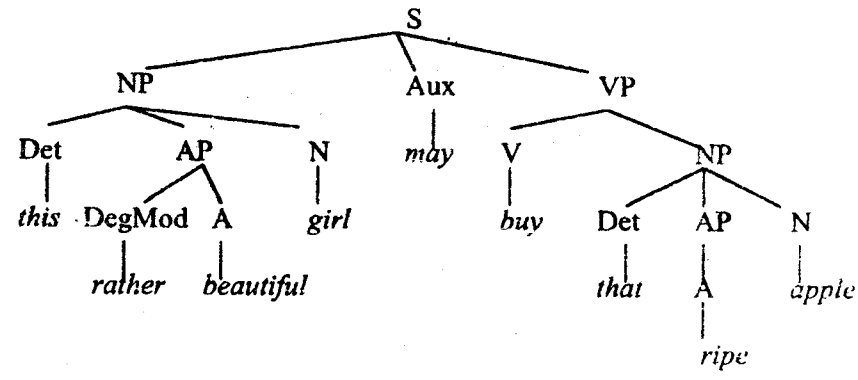
1. For the three-word sequence *a toy car* in (i), the tests of substitution, conjoining, distribution, and parenthetical insertion will be most easily applicable, and they should show that *a toy car* is a syntactic constituent. For the two-word sequence *reading two* in (ii), the tests of substitution, conjoining, distribution, and replacement by anaphora should collectively demonstrate that this sequence of words is NOT a syntactic constituent.

2. Two possible phrase structure trees generated by the phrase structure rules along with the lexicon are the following:

a.



b.



Other possible lexical items that could be inserted are: *kill*, V; *mango*, N; *snake*, N; *venomous*, A.

3. (Assuming that the answer to the first part of the question is "No":) A syntactic constituent S is not a single lexical head, unlike an A or a N or a V or a P, but a composite constituent; moreover, a question like *What did Ram do yesterday?* is better represented as a syntactic constituent "SP" rather than as S' (notice that the question word *what* and the Past-tense auxiliary verb *do* have to count as two separate syntactic constituents, one a Specifier and another a Complement), but the earlier convention did not provide for any syntactic constituent "SP" that would be larger than S'.