CHAPTER I

THE STRUCTURE
OF SIMPLE SENTENCES

KEY POINTS:
In this chapter you will find out about:
- the constituents of a sentence and their properties
- the syntactic status of sentences
- important relations inside the sentence

1. WORDS

Words are grouped into morphologically and syntactically defined categories.

The first intuitive answer anybody would give to the question “What are sentences made of?” is: “Words”. Sentences such as (1) below are made of words:

(1) [This] [lecture] [will] [seem] [extremely] [boring] [to] [my] [students].

What is a word?
For the purpose of the present discussion, we will adopt the view that words are free morphemes which can occur in isolation and have positional mobility, i.e. they do not have to occur in a fixed position (Lyons 1968). Consider the word students in (1). Student, but not the plural morpheme -s, can occur in isolation and also in various positions within a sentence:

(2) a. I know these students.
   b. Students do not particularly like syntax.
   c. He was talking to some students in my class.

There are, however, units such as the definite article the, which, in spite of being a free morpheme, cannot occur in isolation, nor can they enjoy positional mobility. More about the status of such elements will be said in the section dealing with functional categories.

Another important property is that of cohesiveness, i.e. nothing can be inserted within a word.

The second intuition which we all have is that certain sets of words share a number of properties; they are grouped into a small number of well-defined categories. They can be defined in terms of combinations of two categorial distinctive features: [+/- N], [+/- V]. One can distinguish [+ N(oun)] categories (nouns and adjectives) and [+ V (erb)] categories (verbs and adjectives). Each category has its own combination of [+/- N] and [+/- V] features:

---

¹ For more on the definition of words, you will benefit from reading Duțescu-Coliban (1983, Chapter 2) or Baciu (1998, Chapter 1).
Adverbs are analysed as a variety of adjectives.

\[
\begin{align*}
N &= [+ N], [- V] \\
A &= [+ N], [+ V] \\
V &= [- N], [+ V] \\
P &= [- N], [- V]
\end{align*}
\]

These categories represent classes of expressions that share a set of morphological and syntactic properties. Words like lecture and students, in spite of their different meaning, share certain morphological (i – ii) and distributional properties (iii), which qualify them as nouns:

(i) nouns can be inflected for number, having distinct singular and plural forms\(^2\). Their plural can be formed by adding –(e)s:

\[
\begin{align*}
a.& \text{ lecture} \rightarrow \text{ lectures} \\
b.& \text{ student} \rightarrow \text{ students} \\
c.& \text{ watch} \rightarrow \text{ watches}
\end{align*}
\]

(ii) they can take the Genitive morpheme ‘s:

\[
\begin{align*}
a.& \text{ the student’s} \text{ interest} \\
b.& \text{ their children’s} \text{ toys} \\
c.& \text{ her mother’s} \text{ generosity}
\end{align*}
\]

(iii) they appear in similar positions in the sentence. For example, they can be preceded by articles (6a) or by possessive pronouns (6b):

\[
\begin{align*}
a.& \text{ the lecture} \quad \text{ a lecture} \quad \text{ the student} \quad \text{ a student} \\
b.& \text{ their lecture} \quad \text{ his students}
\end{align*}
\]

\(^2\) One should point out that this is a property of the large majority of nouns in English, but that there are nouns whose plural form is irregular (child-children, man-men, woman-women, mouse-mice, louse-lice, foot-feet, goose-geese, etc.), nouns whose singular and plural form are identical (one sheep – two sheep) as well as nouns whose semantics is basically incompatible with a plural (sugar, water). This is the case of the so-called uncountable nouns.
Adjectives, just like nouns, share morphological and syntactic properties:

(i) they can take the affixes –er and –est or the degree words more/most to express the notion of degree:

(7)  

interesting  more interesting  the most interesting

nice  nicer  the nicest

(ii) they can be preceded by other degree words, such as very, too, quite, rather, that:

(8)  

very/too/quite/rather/that interesting

(iii) they occur in the structures: the + Adj + Noun and my/your/his/her/our/their + Adj+ Noun:

(9)  

the interesting lecture

their interesting lectures

Adverbs can also be modified by the degree words which co-occur with adjectives (very, quite, etc.):

(10)  

very fast

quite slowly

But they generally have a distinctive –ly ending: sad-sadly, wise-wisely, lucky-luckily, and can be differentiated from adjectives in distributional terms. An adverb cannot occur in between a Determiner and a Noun (11a) nor can it be the complement of the verb be (11b) or any copula-like verb (such as seem, look, sound, etc.) (11c):

(11)  

a.  

her slow reaction  (adjective + noun)

*her slowly reaction  (*adverb + noun)

b.  

They are very clever.  (be + adjective)

* They are very cleverly.  (*be + adverb)

c.  

She seems intelligent.  (copula-like verb + adjective)

*She seems intelligently.  (*copula-like verb + adverb)

Adverbs, but not adjectives, can modify verbs other than be or copula-like verbs (12), adjectives (13), or other adverbs (14):

(12)  

a.  

He reacted slowly.  (verb + adverb)

*He reacted slow.  (*verb + adjective)

b.  

She behaved politely.  (*verb + adjective)

(13)  

a.  

She is really nice.  (adverb modifying an adjective)

*She is real nice.  (*adjective modifying an adjective)

b.  

They are incredibly fast.  (*adjective modifying an adjective)

(14)  

a.  

John spoke really slowly.  (adverb modifying an adjective)

b.  

*John spoke real slow.  (*adjective modifying an adjective)
English verbs are associated with different distinctive inflections: -s, -ing, -(e)d, -(e)n:

(15) show shows showing showed shown

Only the infinitive of lexical verbs can be preceded by a modal (16) or by the particle to (17):

(16) should show/might leave/
    could understand (modal + lexical verb)
    *should car (*modal + noun)
    *could nice (*modal + adjective)
    *can slowly (*modal + adverb)
    *may down (*modal + preposition)

(17) to show/to leave/to understand

Verbs can also appear with an auxiliary, other than the modals:

(18)
   a. They have left for London.
   b. They will arrive here soon.
   c. This house was built by my grandfather.

Prepositions are morphologically invariant. They can be followed, for example, by the string the/a Noun (19a) and are the only category which can be preceded by right (= completely) (19b) or by straight (= directly) (19c):

(19)
   a. in [the book]/in [a book]
   b. They arrived right on time./She went right inside.
   c. They fell straight into the basket.

The brief presentation above shows that words belong to a finite set of clearly defined categories, which can be distinguished morphologically and syntactically. Each category shares a set of morphological and distributional properties. Categorial information is available in the Lexicon and plays an important part in the formation of sentences because the category to which a word belongs determines its distributional properties.

2. PHRASES

2.1. THE GENERAL FORMAT

However, if we look at a sentence like the one in (1), we will notice that it is structured into larger constituents, into higher-level groupings of words. A look at (1) shows that this "combines with" or "goes with" lecture, extremely with interesting a.s.o., i.e. that the words of a sentence merge gradually, forming larger cohesive units, called phrases. Words are the ultimate constituents of a sentence but the immediate constituents of a sentence are phrases. Words are grouped together to form phrase-level categories of various kinds. For example, a verb selects a particular complement, with which it merges, forming a phrase: {eat an apple}. The resulting phrase {eat an apple} will have the properties of the selector (the verb); it will have the status of a verb phrase.
It is intuitively obvious that within each phrase one element is dominant, it is the part that cannot be missing. For example, in boring lecture, the noun is the most important element, it is essential. We call such phrases Noun Phrases (NP). In a phrase like read a book, the verb is the essential element; this is why we call such phrases Verb Phrases (VP).

The theory that deals with the structure of phrases is X-bar theory, part of the categorial component, and which, together with the Lexicon, represents the base of a P&P grammar:

X-bar theory expresses generalizations about the phrase structure of all human languages; it restricts the combinatorial possibilities of words into larger linguistic units.

Phrases are linguistic objects larger than words. They represent projections round a head, they are endocentric. The head is the obligatory constituent of a phrase, and it determines the properties as well as the distribution of the phrase. For example, an NP will have the distributional properties of a noun, an Adjective Phrase will have the distributional properties of an adjective, a.s.o.

The general format for phrase structure is the one in (20) below, where $X^0$ stands for the head, $YP$ for a possible complement of $X^0$ (i.e. a subcategorised constituent). The role of the Specifier position will be discussed later:

(20)
$$X_P \rightarrow (\text{Specifier}) \ X^0 - YP \ (\text{Complement})$$

The format in (20) captures what is common for the different types of phrases (VP, NP, etc.), providing a universal schema for phrase structure. $X$ can be of any type: N, V, A, Adv, P. This schema is part of UG and it provides the blueprint of the composition of any phrase.

According to (20), a head can merge with a phrasal complement. Which complement a particular item can combine with is stated in the lexicon, in the entry for that specific item. The result is a larger linguistic unit. This intermediate level (or intermediate projection) is represented as $X'$ (X-bar):

(21)
$$X' \rightarrow X^0 - YP$$

For example, a verb, eat, merges with an NP, an apple. The result will be a V' unit: $V + NP \rightarrow V' \ (\text{eat an apple})$. Or a noun, book, can merge with a PP complement, of linguistics. The result will be an N' projection: book of linguistics.

This $X'$ can then merge with a Specifier, the result being a maximal projection XP:

(22)
$$XP \rightarrow \text{Specifier} - X'$$

For example, the V' eat an apple can merge with an NP in Specifier position, the children, resulting in a maximal projection VP: the children eat an apple. Similarly, the N' book of linguistics can merge with this, creating a maximal projection NP: this book of linguistics.
This means that the general format given in (20) can be actually broken into two rules, given in (21) and (22) above, and repeated for convenience in (23):

\begin{align*}
\text{(23)} & \\
a. & \text{XP} \rightarrow \text{(Specifier)} - \text{X'} \\
b. & \text{X'} \rightarrow \text{X'}^0 - \text{(YP)}
\end{align*}

The diagram that is standardly used to illustrate the hierarchical structure of phrases is the phrase structure tree or phrase marker (PM) below:

\begin{align*}
\text{(24)} & \\
\text{XP} & \\
\text{Spec} & \text{X'} \\
\text{X'}^0 & \text{YP}
\end{align*}

As expected, a tree consists of branches. A branch can be defined as the segment that links or connects two parts of a tree. For example, the segment that connects the Specifier to XP, the one that connects X' to XP, etc. are branches. The point where two branches meet or the point where a branch ends is called a node. In (24) above, the branch that links Spec to XP and the one that links X' to XP meet at XP, a node. The point where these branches end, i.e. Spec and X', are also instances of nodes. Since there is no other node above XP, we call this node a root node (25). X' has no lower branches. It is called a terminal node (or terminal element) (26):

\begin{align*}
\text{(25)} & \\
\text{XP} & = \text{root node} \\
\text{Spec} & \text{X'} \\
\text{X'}^0 & \text{YP}
\end{align*}

\begin{align*}
\text{(26)} & \\
\text{XP} & \\
\text{Spec} & \text{X'} \\
\text{X'}^0 & \\
\text{terminal element/node}
\end{align*}

As can be seen in the diagram, some nodes dominate other nodes:

\begin{align*}
\text{(27)} & \\
(i) & \text{XP dominates Spec, X'} \\
(ii) & \text{X' dominates X'}^0 \text{YP}.
\end{align*}
The domination relation is usually called a mother-daughter relationship, with the dominant node being the mother. When two nodes are immediately dominated by the same node, i.e. when they are daughters of the same mother, they are called sisters:

(28)

\[
\text{mother} \quad \quad \text{daughter} \quad \quad \text{daughter} \quad \quad \text{sisters}
\]

The tree diagram captures the fact that phrases have hierarchical organization. It can also explain why certain phrases are ambiguous. When discussing the derivational approach to language structure, in the Introduction to the present textbook, we said that the way in which we can interpret various sentences reflects the derivational steps taken in order to create that structure. For example, a sentence like (29) below was argued to have two possible interpretations precisely because the same surface sentence can be arrived at via different derivational routes, i.e. two superficially similar sentences have different derivational histories.

(29)

John touched the girl with a flower.

The sentence in (29) can be interpreted in two ways: (i) John touched the girl, and he used a flower, and (ii) John touched a girl who had a flower. The interpretations in (i) and (ii) reflect the two different derivational histories. In the case of (i), with a flower merged in the derivation with touched the girl, resulting in the Verb Phrase touched the girl with a flower. In the case of (ii), with a flower merged with the girl, the result being the linguistic object the girl with a flower.

Trees allow us to capture this difference of interpretation. The structure which corresponds to the interpretation in (i) is given in (30), whereas the one corresponding to (ii) is given in (31):

(30)

\[
\begin{align*}
\text{VP} & \quad \quad \text{Spec} \\
\text{John} & \quad \quad \text{V'} \\
\text{V'} & \quad \quad \text{PP} \\
\text{V touched} & \quad \quad \text{NP} \\
\text{with a flower} & \quad \quad \text{the girl}
\end{align*}
\]
Though we have not discussed such complex structures yet, it is not difficult to see that the bolded Prepositional Phrase with a flower occupies different positions in the two phrase markers, which account for the different readings. In (30) it is projected inside the VP, being adjoined to V'. In (31) it is projected in the complement position of the noun girl.

Let us see now how various types of phrase can be represented in accordance with the X-bar format.

2.2. Noun Phrases

As already discussed, Noun Phrases (NPs) represent projections around a nominal head, whose presence is essential for the formation and the properties of the projection. A noun may be optionally modified by a complement. Some examples of noun-complement structures are given in (32):

(32)

<table>
<thead>
<tr>
<th>Head</th>
<th>Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>book</td>
<td>of linguistics</td>
</tr>
<tr>
<td>stories</td>
<td>for children</td>
</tr>
<tr>
<td>students</td>
<td>who like linguistics</td>
</tr>
<tr>
<td>books</td>
<td>which they have never read</td>
</tr>
<tr>
<td>blouse</td>
<td>of silk</td>
</tr>
</tbody>
</table>

A noun can also merge with a Specifier. In (33) there are a few examples of Specifier-noun structures:

(33)

<table>
<thead>
<tr>
<th>Specifier</th>
<th>Noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>John’s</td>
<td>book</td>
</tr>
<tr>
<td>Their</td>
<td>teacher</td>
</tr>
<tr>
<td>Many</td>
<td>books</td>
</tr>
</tbody>
</table>

All the phrases in (34) below represent instances of NPs (whose heads have been bolded):

(34) Noam/house/a book/every book/some wine/this lecture/John’s students/every student’s problems/the student in the corner’s book/our neighbour up the street/books of my teachers/ the fact that they have few students/ our fear that she misunderstood us
As can be seen in (34), some nouns are not further determined or modified in any way (their representation would be like the one in 35), but some of them are preceded by a determiner (represented in 36) or by a possessive phrase (represented by 37), others also take a complement (represented in 38–39):

(35)

```
NP
 |  N'
  |  N^0
  house
```

(36)

```
NP
 |  Det
  |  This
  |  N'
  |  N^0
  lecture
```

(37)

```
NP
 |  PossP
  |  N'
  |  Poss
  |  N^0
  |  's
  |  every
  |  student
  |  problems
```

(38)

```
NP
 |  Det
  |  our
  |  N^0
  |  PP
  |  neighbour
  |  up the street
```

(39)

```
NP
 |  Det
  |  the
  |  N^0
  |  CP
  |  fact
  |  that they have few students
```

### 2.3. Adjective Phrases

Consider the examples in (40) which all contain an Adjective as their core, i.e. as an obligatory constituent. The head of these phrases is an adjective and they all are Adjective Phrases (APs):

(40)

a. old/young/plumpy
b. very old/rather old/so very old
c. very pleased about it/extremely angry with him/really afraid that she may leave for good/rather envious of his brother
The phrases in (40a) contain only the head:

(41)

```
  AP
     \   /  \\
A'  \  /   A
    \|/    \\
   plumpy  young
         old
```

In (40b) the adjectives are modified by an adverb, the phrase being a Specifier-head construction:

(42)

<table>
<thead>
<tr>
<th>Specifier</th>
<th>Head</th>
</tr>
</thead>
<tbody>
<tr>
<td>very</td>
<td>old</td>
</tr>
<tr>
<td>rather</td>
<td>old</td>
</tr>
<tr>
<td>so very</td>
<td>old</td>
</tr>
</tbody>
</table>

Their tree diagram is illustrated in (43):

(43)

```
  AP
    \   /  \\
 Spec A'  \  /   A
     \|/    \\
    rather old
```

In (40c) the head is modified by a Specifier, like in (40b), but it also selects a complement:

(44)

<table>
<thead>
<tr>
<th>Specifier</th>
<th>Head</th>
<th>Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>very</td>
<td>pleased</td>
<td>about it</td>
</tr>
<tr>
<td>extremely</td>
<td>angry</td>
<td>with him</td>
</tr>
<tr>
<td>really</td>
<td>afraid</td>
<td>that she may leave</td>
</tr>
<tr>
<td>rather</td>
<td>envious</td>
<td>of his brother</td>
</tr>
</tbody>
</table>

One tree diagram is provided in (45):

(45)

```
  AP
    \   /  \\
 Spec A'  \  /   A
     \|/    \\
    very pleased PP
           \   /  \\
            about it
```

In afraid that she may leave for good the complement has
sentence status (it is a CP\(^3\)):

\[(46)\]

\[\text{AP} \quad \text{Spec} \quad A' \quad A^0 \quad \text{CP} \quad \text{afraid} \quad \text{that she might leave for good}\]

When the AP precedes the noun which it modifies, it appears inside the NP, in the Specifier position of the phrase. In a very nice girl, the AP very nice modifies the noun girl. It is assumed to occupy the Specifier position, as shown in (47). In Chapter 8, which deals with the structure and the interpretation of APs in English, more will be said about the place of APs inside the NP:

\[(47)\]

\[\text{NP} \quad \text{AP} \quad \text{N'} \quad \text{very nice} \quad \text{N}^0 \quad \text{girl}\]

2.4. Verb Phrases

Just like the other types of phrases already illustrated, VPs can contain only a verb, i.e. only the head (48), or a verb and its complement (49):

\[(48)\]

sleep/dance/walk/rain/snow

\[(49)\]

buy a house/read Linguistic Inquiry/believe that this is not true

The verb, the head of the projection, selects a complement, with which it merges in the derivation, forming a head-complement structure, whose status is that of V:\n
\[(50)\]

<table>
<thead>
<tr>
<th>head</th>
<th>Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>buy</td>
<td>a house</td>
</tr>
<tr>
<td>read</td>
<td>Linguistic Inquiry</td>
</tr>
<tr>
<td>believe</td>
<td>that this is not true</td>
</tr>
<tr>
<td>insist</td>
<td>on leaving immediately</td>
</tr>
</tbody>
</table>

The Specifier position inside a VP is assumed to be occupied by the external argument of the lexical verb, i.e. by its subject:

\[(51)\]

\[\text{See the definition of CP in 5.3 and 5.4 in this chapter.}\]

46
Let us see how we can represent the Phrase marker of some of these VPs. In (52), only the head is available, in (53) the verb merged with an NP complement, in (54) the head only merged with a subject NP, whereas (55) is an example of a VP whose Specifier and Complement positions are both filled up:

(52)

```
  VP
  | V'
  | V
  sleep
```

(53)

```
  VP
  | V'
  | V
  | NP
  | buy
  Det a
  N' house
```

(54)

```
  VP
  | Spec V'
  | Dogs
  | V'
  | V
  bark
```

(55)

```
  VP
  | Spec V'
  | they
  | V'
  | V
  | NP
  | buy
  Det a
  N' house
```

What happens to those constituents of a sentence which are not obligatory constituents, i.e. which are not subcategorised for by the head of a phrase? Let us take a look at buy a house in London. Buy, the head, merges with a house, and then buy a
house, a V', merges with the PP in London. In order to capture the relationship between the verb, i.e. the head, and the two complements, we can make use of V'. Firstly, we want to capture the fact that the verb and its direct object merge, forming a larger constituent, whose status is that of a V':

\[
(56) \quad \text{V' - buy - det - N'} \quad \text{NP - house}
\]

Then, we would like to capture the fact that in London merges with this larger unit, V', resulting in a constituent which is also V'. V' can dominate two complements:

\[
(57) \quad \text{VP - V' - PP - det - N' - in London - NP - house}
\]

Notice that the lower complement is the one that is closer to the verb.

The analysis of various configurations whose head is a verb will be discussed in the chapters dealing with the structure and the properties of the VP in English\(^4\).

2.5. PREPOSITIONAL PHRASES

Some of the examples discussed so far contained Prepositional Phrases (PPs), which we did not represent in detail. PPs must contain a preposition, i.e. their head, and, optionally, a complement. We will leave out, for the moment, discussions about the Specifier position of PPs. A PP of the type on the table will have the following representation:

\[
(58) \quad \text{PP - P' - P\textsuperscript{0} - det - NP}
\]

\(^4\) See Chapters III, IV, V and VI.
In (58), the preposition on is the head and it selects an NP as its complement.

3. CONSTITUENCY TESTS

So far it has been posited, only on interpretative grounds (i.e. we assumed that some words “go together”, that they are semantically connected) that words can be expanded into phrases, which represent the immediate constituents of a sentence and which inherit the properties of their heads. No arguments besides our semantic intuition have been presented in favour of treating phrases as constituents and, consequently, no argument has been presented that sentences have a categorial constituent structure. In what follows we shall be looking at morphological and syntactic evidence in support of the assumption that phrases are constituents, i.e. that they function as a unit. The tests which are used to determine whether a group of words form a constituent are called constituency tests.

3.1. MORPHOLOGICAL EVIDENCE

When discussing Noun Phrases, we came across examples of the type John’s students or every student’s interest. In such cases, the genitive inflection ’s is attached at the end of the Noun or so it seems at first glance. Inflections usually attach to the end of words. However, the Genitive marker attaches to phrases in English. This is difficult to prove if we look only at examples like the one already mentioned. But, if one looks at cases such as the ones illustrated in (59), in which the head is followed by a complement, it will be clear that the Genitive marker attaches to the end of the phrase and not to the end of the noun:

(59)

a. [The president of the board]’s speech/*the president’s of the board speech
b. [The queen of England]’s children/*the queen’s of England children
c. [The Ambassador of Bulgaria]’s residence/*the Ambassador’s of Bulgaria residence

These examples provide evidence that phrases behave like one single syntactic unit, they represent constituents.

3.2. SYNTACTIC EVIDENCE

Most of the syntactic arguments in favour of the constituent status of phrases relates to facts about the distribution of the various strings of words. We shall mention only a few.

3.2.1. The Substitution (or Replacement) Test

If we look at sentence (60) it is easy to notice that the string this lecture can be replaced by the pronoun it or that these students can be replaced by them:

There is semantic, morphological and syntactic evidence that phrases are sentence constituents.
[This lecture] will seem extremely interesting to [these students].

These substitutions (called pronominalisations because the NPs are substituted by pronouns) show that this lecture and these students, NPs, behave like one single unit.

It is not only NPs which can be substituted. PPs (61–62), APs (63) and VPs (64) can also be substituted by various so-called pro-forms:

(61) I have been [to Antwerp] several times.

(62) They have left it on the table.

Many students consider syntax [extremely boring], but I’ve never found it SO.

(64) Most of them might [change their mind]. I hope SO.

3.2.2. Ellipsis

In English, identical VPs may be omitted within sentences, provided their content can be recovered from the "context". This phenomenon is known as ellipsis:

(65) John will leave for Canada but I won’t (leave).

He may come back any minute now, but then again he may not (come back any minute now).

Random omission of words results in ungrammaticality:

(66) * He may come back any minute now, but again he not come back any minute now.

The ungrammaticality of (66) provides further evidence that only constituents can be omitted.

3.2.3. Movement

3.2.3.1. Preposing

For emphasis reasons, some parts of sentences can be preposed, i.e. they can be moved to a position higher in the sentence, usually to sentence initial position. Obviously, such movements have a clear stylistic effect. What is relevant for the
present discussion though is the fact that only phrases (not parts of phrases) can be preposed. Let us examine (67)–(68) below:

(67)
\[\text{a. My students can't stand generative grammar.}
\text{b. Generative grammar, my students can't stand.}
\text{c. *Grammar, my students can't stand generative.}\]

(68)
\[\text{a. The children ran up the hill as fast as they could.}
\text{b. Up the hill the children ran, as fast as they could.}
\text{c. *Up the children ran the hill, as fast as they could.}\]

In (67b) an NP has been preposed, and in (68b) a PP. As the ill-formedness of (67c) and (68c) shows, when only one part of a phrase is preposed, the resulting structure will be ungrammatical. Only whole phrases can undergo preposing. Actually, this constraint is part of a larger one, which has already been mentioned: an element can Move only if it is a constituent. Preposing represents one instance of movement.

3.2.3.2. Focalised structures
Sentence constituents can be focalised by resorting to various syntactic constructions. Two such well-known focalised structures are the so-called cleft (69) and pseudo-cleft (70) sentences, which both represent instances of Move $\alpha$.

(69)
\[\text{It is generative linguistics that they don't like.}\]

(70)
\[\text{What they will have to do is try and understand the core principles.}\]

3.2.4. The coordination (or conjunction) test
Only constituents can be conjoined by coordination within a sentence. Moreover, the constituents must belong to the same class, as can be seen in the examples in (73):

(73)
\[\text{a. Students who like generative syntax AND inflectional morphology are rare. (NP AND NP)}
\text{b. The present analysis will provide arguments for the movement analysis AND against the base-generation}\]
approach. (PP AND PP)

As the examples above show, phrasal categories of the same type can be conjoined. Restrictions on coordination can only be accounted for if one assumes that sentences are made of constituents which fall into several classes and which may have different degrees of complexity.

So far, arguments that sentences are structured into word-level and phrase-level constituents have been presented, as well as morphological and syntactic evidence that phrases are constituents. But we have only taken into consideration lexical elements, such as nouns, adjectives or verbs. Nothing has been said about determiners (the or this), or auxiliaries (have or be), which evince a set of properties that clearly distinguishes them from nouns, verbs, adjectives or prepositions, and which play an equally important part in sentence formation. In what follows we will discuss the properties and the role in the formation and interpretation of sentences of categories other than the lexical ones.

4. ON LEXICAL AND FUNCTIONAL CATEGORIES

It has already been said that the Lexicon is a repository of individual items and their idiosyncratic properties (i.e. of properties which are not derivative from general principles). But so far we have only said that the Lexicon contains Nouns, Adjectives, Adverbs, Verbs or Prepositions. In this section, we shall go into further detail concerning the structure of the Lexicon.

Recent generative studies (Tsimpli and Ouhalla 1990, Ouhalla 1991) assume that the Lexicon consists of two compartments:

(i) a mental, conceptual compartment, which contains what has been called “substantive” elements, drawn from the universal vocabulary

(ii) a grammatical compartment, which contains functional categories, which lack descriptive content and which are determined by UG.

The conceptual compartment represents an open system, with a large number of members, called lexical categories:

Only open categories have indefinitely many members in the dictionary of a language – say several hundred at least… Conscious coining of new lexical entries by speakers is allowed only in the open categories. (Emonds 1985)

The grammatical compartment is a closed system, i.e. with a limited number of members, called functional categories, determined by UG. They correspond to what was traditionally considered the set of morphological markers, or morphological inflection.
Following a long-standing tradition, we shall include in the class of functional categories: determiners (elements such as the definite and the indefinite articles), degree (more, most, etc.), tense, agreement, aspect, mood and complementisers (that, whether, for-to). For the moment, we leave aside the problem of negation. Also, for expository reasons, we shall include tense, agreement, aspect and mood under the umbrella term of (verb) Inflection (I).

Table 2: Functional categories

<table>
<thead>
<tr>
<th>Functional categories</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>DETERMINERS (Det)</td>
<td>the, a, this, that</td>
</tr>
<tr>
<td>DEGREE (Deg)</td>
<td>more, most, -er,-est</td>
</tr>
<tr>
<td>COMPLEMENTISERS (Comp)</td>
<td>that, whether, for</td>
</tr>
</tbody>
</table>

Functional categories represent "the locus of grammatical information" (Ouhalla 1991:8) which determines the various grammatical processes that constructions can undergo. This is why they have also been labelled 'grammatical formatives'.

One important difference between lexical categories and functional categories is related to the type of meaning that they encode. Lexical categories have descriptive (or lexical) content, whereas functional categories encode some abstract grammatical meaning. Their formal or abstract meaning contributes to the interpretation of their complement (Abney 1987:65). A noun like book, for example, denotes a type of object associated with a set of prototypical properties. An article like the lacks such a descriptive content, it is a grammatical formative, with functional or grammatical meaning. When it selects the NP book as a complement it "identifies" its referential argument; the book is interpreted as denoting a unique entity that has been already identified in the discourse.

This semantic difference has important consequences for the combinatorial properties of the two types of categories. Lexical categories, which have descriptive content, will impose two types of selectional restrictions on their complements: semantic and categorial. A verb like read, for example, can merge with complements like the ones in (74) but not with complements like the ones in (75), which are semantically odd:

(74) read a book/a newspaper/a journal/ a letter/ a paper

(75) ??read a cucumber/a shoe/mosquitoes

Lexical categories can often merge with complements of

---

5 This idea goes back to Aristotle, who defines functional elements as words without meaning, and lexical ones as words with meaning.
6 For more on the meaning and role of functional categories within the P&P model, see Cornilescu (1995b), Chapter 8, section 8.
various types. Take for example a verb like believe, which can combine with an NP, as in (76) as well as with a whole clause (CP) (77):

(76)
I don’t believe [NP this story].

(77)
I believe [CP that he might have said that].

One crucial property of functional categories is that they select one unique type of complement, which is not their argument. Inflection always selects a VP as a complement, Determiner always selects an NP, Complementiser an Inflectional Phrase (IP), Degree an AP or an AdvP:

(78)

As can be seen, the relationship between a functional head and its complement is extremely restrictive. Selection takes into account only the category the complement belongs to. This type of selection has sometimes been labelled functional selection (f-selection) (Cornilescu 1995b: 227–228).

Most functional categories cannot occur independently: *the, *d, *ing.

Two more things are directly relevant for our investigation. First, one should point out that some functional categories may have an affixal nature, they may be bound morphemes. For example, Tense may be expressed in English by the affix –(e)d, Aspect by -ing and -en and Agreement by –(e)s. But there are also functional elements which are not affixal, such as the English auxiliaries, articles, and demonstratives. They are free morphemes. Whether a functional category is a free or bound morpheme is stated in its lexical entry. As will be shown later in this textbook, this property is extremely important and with far reaching consequences in the syntax of language.

Second, functional elements, be they free or bound morphemes, are assumed to project into hierarchically structured phrases, according to the blueprint in (20), just as lexical elements do. Both lexical (verbs, nouns or prepositions) and functional categories (such as Inflection) can be the head of a phrase. The theoretical implication is that inflectional formatives are syntactic heads, just like lexical categories. Any such formative which corresponds to a functional category (Tense, Aspect, Agreement, etc.) is the head of a maximal projection. In (79) Inflection – I′0, a functional category, is the head of IP, which contains a complement (VP) and a Specifier position:

(79)
Inflection, as already said, is an umbrella term for all the functional categories of the verb: Tense, Aspect, Mood, and Agreement. It will always select a VP as its complement. Its role is to determine the 'referential' properties of the event which the verb denotes, for example to locate it in time. In *He danced*, the past tense marker *-ed*, an Inflection element, locates the event of dancing at a time prior to Speech Time (*now*), in the past.

The category of Determiner (Det) is instantiated by elements whose role is to determine the referential properties of the noun with which they merge in the derivation. For example, when the definite article *the* merges with a noun, the linguistic object *the N* will be interpreted as referring to one particular entity which the speaker assumes to be known to the hearer. When the indefinite article *a* merges with a noun, the preferred reading is that the resulting phrase *a N* denotes an entity whose reference is new in the discourse. Traditionally, the class of determiners has been taken to include articles (*the, a*), demonstratives (*this, that*), possessives (*my, your, his, her*) and quantifying determiners (*all, some*). These elements form a distinct distributional class: they all select an NP as their complement, i.e. they can only merge with an NP7:

\[
\begin{array}{c}
\text{DP} \\
\text{Spec} \\
\text{D'} \\
\text{D}^0 \\
\text{NP}
\end{array}
\]

the book

The properties of functional categories and the relationship established among them and the lexical category which they select as their complement determine the structure of sentences, i.e. syntax is driven by the properties of functional elements. Sometimes, a property which may seem unrelated to syntax, such as that of being a free or a bound morpheme, may have important consequences for the syntax of a language (as we are going to see when we discuss auxiliaries or determiners). In particular, word order reflects the properties of functional categories. Since word order may vary from one language to another, one can reach the conclusion that cross-linguistic differences have their root in the different properties of functional categories in various languages. Parametric variation affects functional categories. If we want to look at differences among languages, the domain of functional categories should be the first domain of investigation.

In this subsection it has been shown that sentences are made of constituents which may be functional or lexical. Functional and lexical elements expand into endocentric projections, into phrases whose properties and

---

7 See Chapter VII for an analysis of the categorial status of these elements and the position which they occupy within the NP.
internal structure are responsible for the formation of the sentence. Remember that we started from the assumption that the structure and the interpretation of sentences are compositional, i.e. they are determined by their constituents and the relations among them. The fact that functional elements, part of the grammatical component, devoid of descriptive content, are part of the sentence provides evidence that sentence formation is guided by more than semantic information.

5. THE STATUS OF SIMPLE SENTENCES

5.1 Inflection as the head of the sentence

It has been said that sentences are made of ever-larger constituents and that sentences themselves may be constituents of complex sentences. Immediate constituents have been shown to be phrases which can combine in order to form sentences. This means that sentences themselves are also phrases of a higher degree of complexity. The obvious question is: what is the status of this larger phrase? We know that phrases are projected round heads, round elements which are essential for that particular phrase to exist and which transmit their set of properties to the whole projection. We also know, from traditional grammar, that every sentence needs a predicate. This may lead to the conclusion that the essential element in a sentence is the verb and that, consequently, the status of the sentence is that of a VP. However, even when a sentence contains both a subject and a predicate, as in (81) below, it can still be ill-formed:

(81) *John go to visit a friend.

The sentence is still incomplete. What is missing? The Inflection of the verb, which ‘locates’ the event in time and relates the two major constituents: the subject and the predicate. This suggests that without Inflection the sentence is ungrammatical. Inflection then could be taken to represent the head of a sentence and a sentence could be analysed as an IP, i.e. an endocentric projection, built round the head I. The head of the sentence can take various forms: it can be an affix attached to the verb, such as the /d/ past tense marker or it can be a modal verb. Its complement is always a VP. The Specifier position, as already discussed, has been associated with the position occupied by the subject of the sentence:

(82)
Let us see the diagrams of two sentences (one containing an auxiliary, (83) and the other one a tense morpheme (84):

(83)
John will write an extremely fascinating book.

(84)

As can be seen in the two representations above, the auxiliary and the affix are assumed to occupy the same position in the tree, I, regardless of the position in which they can be seen at S-Structure: the auxiliary precedes and the affix comes after the verb. These two types of elements represent instances of heads of a sentence. Inflection; movement operations can account for the different surface order of these two types of elements.
So far, whenever reference was made to IP, it was only said that it is an umbrella term for a bundle of functional categories associated with the VP: Tense, Agreement, Aspect, and Mood. But nothing has been said so far about its structure. In order to decide which functional categories are actually part of IP, we shall have a look at the overt morphology on the verb across languages. English is a language with impoverished morphology, so it might be difficult to decide only on the basis of the English data.

But even a language with impoverished morphology, such as English, provides evidence that the inflection on the verb encodes at least Tense information. Compare walk to walk-\textit{ed}, gleam to gleam-\textit{ed} and you will see that Tense has a distinct marker: the morpheme \textit{--ed}. Besides the past tense marker, English also marks person agreement in the present. The 3rd person singular is marked as \textit{--s}: \textit{I walk} vs. \textit{She walks}. But it is indeed difficult to decide if \textit{--s} marks only the 3rd person singular or whether it is a hybrid marker of tense and agreement, encoding the information: present tense 3rd person singular. In languages with richer morphology things are clearer though, and it is easier to see that tense markers are distinct from agreement markers. Take a look at the form of the verb for the 1st person plural in the present (86a) and the \textit{imparfait} (86b) in French:

\begin{align}
\text{(86)} & \\
\text{a. Nous dessi}nons & \quad \text{b. Nous dessi}nions \\
\text{we draw-1}^{\text{st}} \text{pl.} & \quad \text{we draw-IMPERFECT-1}^{\text{st}} \text{pl}
\end{align}

In both (a) and (b) the ending \textit{--ons} marks agreement: 1st person plural. If we break the verbal form into smaller pieces of inflection, we will see that the \textit{imparfait} form in (85b) also contains the morpheme \textit{--i}, which encodes tense information. Such data indicate that IP contains at least two distinct types of morphological information, each marked by its own morpheme. Each morpheme heads its own projection. IP should then be split into at least two projections: TP, whose head hosts morphemes that encode tense information, and AgrP, whose head hosts morphemes that encode Agreement information.

The order in which the morphemes occur provides information with respect to the place which the two projections occupy in the structure of IP. The closer a morpheme is to the lexical stem, the closer to the VP it is in the representation of the clause. In (86b) the Tense marker is closer to the verb stem. So, the order of the two projections is assumed to be the one in (87):

\begin{align}
\text{(87)} & \\
\text{AgrP} & \\
\text{Agr'} & \\
\text{Agr} & \\
\text{TP} & \\
\text{T'} & 
\end{align}
This is known as the Split-IP Hypothesis (Pollock 1989) according to which the IP is no longer seen as one single functional projection. It is split into various functional projections, each headed by a morpheme that encodes a different type of information.

There are languages which have distinct morphemes that encode Aspect information. Take, for example, the English morpheme -ing in eating, talking, sleeping, reading or the Romanian -ind in venind ‘come-ind’, dormind ‘sleep-ind’, which both have the feature [imperfective]. This leads to postulate the existence of one more functional projection in the extended functional domain of the verb, i.e. as part of IP: AspP. Evidence for the existence of a functional category Asp has been found in a number of other languages, such as Modern Greek (Rivero 1990), Scottish Gaelic (Ramchand 1993), German (Schmidt 1994), Russian (Tenny 1987), Spanish (Laka 1990), or Romanian (Avram 1999). The morpheme encoding aspect is usually the closest to the lexical verb. AspP is the first functional projection above VP:

(88)

Several studies also argue in favour of a functional projection of Mood (Rivero 1994 for various Balkan languages, Avram 1999 for English and Romanian), above AgrP:

(89)

---

* In Pollock’s analysis Negation is also part of the split IP. We will, however, not take any position with respect to the status and position of Negation in a sentence in the present textbook.

* In Avram (1999) the idea of two Mood projections is put forward: one MoodP is in the functional layer of the clause, and the other one at the border between the complementiser layer and the functional one. It is this latter position which is also assumed in Rivero (1994) for Balkan languages. In Romanian, this position hosts invariable mood elements, the subjunctive marker să and the infinitive marker a. Notice that this may lead to identifying the invariable Mood projection with FinP in the complementiser layer (see the discussion in 5.4).
Obviously, such a proliferation of functional categories, all related to the morphology of language, i.e. the place of language variation, raises at least two questions: (i) is the order of these functional projections the same cross-linguistically? and (ii) do these functional categories always project?

The inventory and the order of functional categories are assumed to be universal. Universal Grammar provides a fixed inventory of functional categories which project in the same order, cross-linguistically\(^\text{10}\) (Thráinsson 1995). What differs is the set of properties each of these categories has in each language. Being part of a universal inventory does not mean that every category must be instantiated in every language nor that it must project every time inside one and the same language. Functional projections, just like lexical ones, are subject to the Economy of Projection Principle (Speas 1994:186):

(90)

The Economy of Projection Principle: Project XP only if X has content, i.e. if there is some overt material in X or if some constituent has moved to [Spec XP] to check the abstract features in X.

Let us see what this means. Consider the maximal projection TP. If its head is filled, then one condition is met and the projection is allowed. In (90), T hosts the past tense morpheme -ed, so TP is allowed to project:

(91)

Imagine now that the T position has no overt element, only a set of abstract features. If an element moves to the Specifier of T in order to have some of its features checked against the abstract features of the head, then, according to the Economy of Projection Principle, TP is allowed to project. In (92) an NP has moved to [Spec TP] and has entered a Spec-head relation with the head. This is enough to allow TP:

(92)

\(^{10}\) For a different point of view, see Ouhalla (1991).
But if both the head and the Specifier are empty, then the projection is not licit. We are now in a position to answer the second question: functional categories do not have to always project. They project only when they meet the Economy of Projection Principle.

A look at (87) or (88) will show that the functional categories of the verb are higher than the verb. However, at S-Structure, some of the morphemes which are hosted by these projections occur in post-verbal position. For example, in English auxiliaries can move to Inflection. Imagine the auxiliary *be* needs to merge with the Tense morpheme /ed/ and the morpheme /s/ for the 3rd person singular, hosted by Tense and Agreement respectively. At D-Structure, the two morphemes precede the auxiliary, as shown in (93):

(93)

In the derivation, the auxiliary will move from its base position, under VP, first to Tense and then to Agreement, adjoining to each morpheme in turn. *Be* occupies a head position and, from this position, it will move to the head position T and then to the head position Agr. This is called head-to-head movement:

(94) a.
The auxiliary has to move to the nearest head first. It is not allowed to 'skip' it. This is the **Head-Movement Constraint**, which could be formulated as in (95):

\[(95)\]

**The Head-Movement Constraint**: A head that is moving is not allowed to bypass another head position.

In this section the structure of a split IP was examined. IP was argued to contain a number of functional projections closely related to the morphology of the verb: AspP, TP, AgrP, MoodP. The hierarchy of these projections is invariant across languages. If a category projects or not is constrained by the Economy of Projection Principle. But the properties of these categories are subject to cross-linguistic variation.

### 5.3. The Status of Embedded Clauses

The next question related to the status of the sentence is raised by the case of embedded clauses. All the sentences we have discussed so far were main clauses. Consider for example the sentences in square brackets in (95):

\[(96)\]

a. I pointed out to the little prince [that baobabs were not little bushes].

b. I did not know [whether I had made a mistake].

c. I would have preferred [for the matter to be solved between you, me and the lamp].

The bracketed sentences contain all one single predication relation but they are also introduced by a complementiser (C). Recall that C always selects an IP as its complement. This means that the embedded clauses above are CPs, and not IPs:

\[(97)\]

In certain cases, the complementiser position may be empty, i.e. there is no complementiser position. Still, one can assume that any clause is a CP. Evidence in favour of this hypothesis comes, for example, from coordination:

\[(98)\]

I know [you don’t like syntax] and [that you wish you were studying something else right now].

The two-bracketed constituents should have the same status since coordination is possible only if the two conjoining constituents belong to the same category. The second constituent has an overt complementiser, it is a CP, which means that the first one, in which the complementiser position is empty, can only be a CP.
Can we then reach the conclusion that main clauses and embedded clauses have different status or can we adopt a unifying analysis, according to which they have the same status? We are thus faced with two possible analyses:

A. main clauses are IPs, and embedded clauses are CPs
B. all clauses, main or embedded, have the same status.

Let us see if there is evidence in favour of one analysis or the other.

5.4. Sentences: IP or CP?

Evidence in favour of the existence of a position in the tree above IP in root clauses comes from sentences like the ones in (99)–(100):

(99)
  a. John should solve this puzzle.
  b. Should John solve the puzzle?

(100)
  a. "Will I get punished?" he wondered.
  b. He wondered whether he would get punished.
  c. He wondered would he get punished.
  d. *He wondered whether would he get punished.

In the interrogative sentence in (99b) or in (100a) the modal /auxiliary occupies a pre-subject position, which is typically occupied by Complementisers in embedded clauses (see 97 above):

(101)

When only the modal has been inverted with the subject (99b, 100a) or when the modal has not been inverted and the complementiser has been used (100b) the sentence is grammatical. In (102) below, the complementiser whether occupies C, but the modal has not inverted with the subject:

(102)
The ungrammaticality of (100d), where the complementiser whether and an "inverted" modal, would, are both present also suggests that the two constituents are actually competing for the same position, that of C:

(103)

This fact proves that sentences, be they main clauses or embedded ones, need a Complementiser projection. They are CPs.

We shall thus adopt the view that every sentence, be it a main or an embedded clause, is a CP, whose complementiser position may be filled (by a complementiser or an inverted auxiliary), or empty. Each clause must contain an IP, the complement of C:

(104)

The complementiser will select a certain type of IP, according to its subcategorisation properties. For example, that will select a finite IP (104a), whereas for will select an infinitival IP, i.e. a non-finite one (104b):

(105)

a. I guess [that you have understood my point].
b. He was waiting [for her to start the conversation].

Further evidence in favour of the view that sentences, even
main clauses, are CPs, comes from cross-linguistic data. There are many languages (among which Polish, Russian, Yiddish, Estonian, Irish, Persian, Berber) in which interrogative main clauses contain an overt complementiser (Radford 1992). Interrogation is signalled by an overt complementiser, and not by Subject-Auxiliary Inversion as in English. See, for illustration, the case of Yiddish (106) and that of Polish (107):

(106) Ts'i hot er geleient dos bux?  
'Has he read the book?' (Radford 1992:296)

(107) Czy zamykacie okna?  
'Are you closing the windows?' (Radford 1992:296)

In other languages, exclamatory main clauses may contain an overt complementiser. This is the case of Danish, German, French or Icelandic (Radford 1992). In French, for example, exclamatory main clauses are introduced by the complementiser que, which occurs in pre-subject position, i.e. in a position higher than IP:

(108) Qu'elle est belle!  
'That she is beautiful'

Imperative sentences in many languages may also be introduced by an overt complementiser. In (109) this is illustrated with an example from Spanish (Radford 1992:298):

(109) Que vengan todos!  
'Let them all come!'

Such data provide universalist arguments in favour of a Complementiser position inside all sentences. Given the fact that so many languages can have main clauses introduced by an overt complementiser, we can safely reach the conclusion that main clauses in general contain a Complementiser position, i.e. that root clauses, just like embedded clauses, are CPs.

🌟 The conclusions we can reach at this stage with respect to sentence structure are the following:
- a sentence is a hierarchically structured entity
- a sentence (be it a root clause or an embedded one) is a Complementiser Phrase
- the immediate constituents of a sentence are phrases
- the ultimate constituents of a sentence are words.

5.5. What's in a CP?

When the structure of the Inflection Phrase was discussed, it was said that IP is actually an umbrella term for various functional
categories: Tense, Agreement, Aspect, Mood, i.e. it contains functional heads which correspond to the morphological specifications on the verb (Pollock 1989). Building on the analogy with the split structure of the Inflection Phrase, linguists argued that the Complementiser Phrase can also be dissolved into a number of different functional projections which are all related to the left periphery of a clause (Rizzi 1995). The main difference between the IP and the CP systems, though, is that the properties of the former are encoded in the morphology of the verb, whereas the properties encoded by the latter are expressed by free morphemes which are not related to the VP:

Whatever “inflectional” properties C reflects, they are not encoded in the form of verbal morphology, in the general case: they are expressed on free functional morphemes (that, que, etc.) which, if anything, look nominal more than verb-like, as they often resemble demonstrative pronouns, wh elements, certain kinds of nouns (“fact”, etc.) etc. [...] the C system is fundamentally distinct from the I system, the latter but not the former being V-related in the general case. (Rizzi 1995:4)

On such a view, the complementiser layer has been defined as ‘the interface between a propositional content (expressed by IP) and the superordinate structure (a higher clause or, possibly, the articulation of discourse, if we consider a root clause). As such, we expect the C system to express at least two kinds of information, one facing the outside and the other facing the inside.’ (Rizzi 1995:3). The information ‘facing the outside’, which expresses the relations of the clause with higher elements (another clause or the discourse), states whether a sentence is declarative, interrogative, relative, etc., i.e. it specifies the Force of the sentence (Chomsky 1995). As already shown, in some languages there are overt markers which encode this type of information (see 106–109). The information ‘facing the inside’ is related to the kind of IP the complementiser requires: finite or non-finite. Compare, for example, the properties of the English complementiser that to those of the English complementiser for (already discussed in 5.3), or the properties of the Romanian complementisers că (‘that’) and ca (‘that’). While the former selects an IP whose mood is the indicative, the latter selects an IP whose mood is the subjunctive. The two types of information discussed so far make up the so-called Force-Finiteness system:

(110)

The Force-Finiteness system of C encodes two types of information:
(i) the ‘force’ type the clause belongs to (declarative, interrogative, etc.)
(ii) the type of IP (finite/non-finite) C can combine with.
Besides the Force-Finiteness system, the C layer also encodes features which are independent of selectional constraints, such as topic and focus. That is why such features will be present in the structure only when needed. Such properties are related to the information structure of a clause. Topicalised elements express old information, available from previous discourse or common background. In (111) below, the NP this dictionary expresses old information and it has been moved to the left periphery of the clause; it has been topicalised:

(111)
This dictionary, you should throw away (not buy).

According to Rizzi (1995), a Topicalised element occupies a position within a Topic Phrase (TopP), part of the CP:

(112)
```
TopP
  Spec   Top'
  Top
  YP
```

In (111) the sentence predicates about the topicalised constituent, introducing new information. It is the “comment”. This topic-comment information structure is captured at the syntactic level within the TopP: the topicalised constituent is assumed to occupy the Specifier position, while the “comment” is the YP in (112), i.e. the complement. Top connects the topicalised element and the comment, similarly to the way in which Inflection links the subject to the VP. In many languages, Top is phonetically null, but there are languages which have topic markers, such as Japanese, where the particle wa has been defined as having this role:

(113)
Watashi wa nihongo ga dai suki.
I wa Japanese ga most like
‘I like Japanese very much.’

A focalised constituent bears stress and introduces new information. Bob, in (114) is the focus. The rest of the sentence contains old information (presupposed information), which the speaker and the hearer share. The information structure is the traditional focus-presupposition one:

(114)
BOB (not John) was introduced to Sue at the party.
By analogy with the topicalised constituent, focalised constituents have been analysed as moving to the Specifier of a Focus Phrase (FocP), whose complement, YP, is the presupposition:

(115)

\[
\text{FocP} \\
\text{Spec} \quad \text{Foc’} \\
\text{Foc}^0 \quad \text{YP}
\]

Foc is, just like Top, phonetically null in many languages. But there is evidence in favour of this projection coming from languages such as Gungbe\(^\text{11}\), where the particle \(\text{wé}\) has been analysed as Foc (Aboh 1995).

Topic and focus features, then, are part of the CP, making up the Focus-Topic system. Remember that one important property which distinguishes this system from the Force-Finiteness system is that the former will be present in the structure only when needed. Also, the Force system encodes information which accounts for the way in which the CP relates to elements outside the clause, while the Finiteness system encodes information which imposes selectional restrictions on the content of the IP. When the Focus-Topic system is present, it will occupy a position in between Force and Finiteness, allowing the two to perform their selectional functions. In (116) below, the Topic and the Focus projections are in between brackets, in order to capture the fact that their presence is not compulsory:

(116)  
\[
\text{Force} \ldots (\text{Topic}) \ldots (\text{Focus}) \ldots \text{Finiteness IP}
\]

The CP, then, can be split in various projections, which account for the various types of information the C system encodes: force, finiteness, information structure (topic, focus). The Force-Finiteness system is related to the selectional properties of the complementiser, the Topic-Focus system is not, encoding a different type of information.

Rizzi (1995) proposes the following articulated structure for the CP:

(117)  
\[
\text{ForceP} \\
\text{Force’}
\]

\(^{11}\) Gungbe is a Niger-Congo language (or dialect, according to some linguists), spoken mainly in Nigeria.
In this section, a possible analysis of the structure of the Complementiser Phrase was presented. Following Rizzi (1995), the CP was argued to encode two types of information: (i) information related to the selectional properties of C and (ii) information related to the information structure of a clause. The elements which encode these various types of information head their own projections in the left periphery of the clause. This obviously led to the conclusion that any sentence contains, besides the lexical layer (NP, VP, etc.) and the functional layer (IP), a complementiser layer which contains topicalised and focalised elements, as well as elements which encode information about the ‘force’ and the finiteness properties of the clause.

6. IMPORTANT RELATIONSHIPS INSIDE THE SENTENCE

When X-bar was discussed (see section 2), we said that the elements inside a phrase are hierarchically organized, i.e. that there is a certain dominance relation between the units which make up a phrase. In what follows several important structural relations will be discussed, with a focus on their relevance for the syntax and interpretation of sentences.

6.1. C-COMMAND/M-COMMAND

The first structural relation which will be briefly presented in this section is the so-called constituent-command (c-command). In order to define this relation, let us analyse the relations between the constituents of the following sentence:

(118) John hurried Mary to the door.

The phrase marker of (118), is given in (119), where, for economy reasons, the CP has not been represented:

(119) IP
As can be seen, some elements are 'superior' to others, i.e. they occupy a higher position in the tree. For example, the object NP Mary is lower in the tree than the subject NP John, the NP the door is lower in the structure than the preposition to. a.s.o. John is 'superior' to Mary or the door, but it does not contain them. Both the NP Mary and the NP the door are in the 'domain' of the NP John, but they are not contained by this NP. What does 'domain' mean? In order to establish the c-command domain of John, we go upwards until we meet a branching node, in this particular case, IP. All the elements which are contained in this projection, downwards, are within the c-command domain of John. All the elements which are c-commanded by an element represent the c-command domain of that element:

(120)

Similarly, the verb hurry c-commands the NP Mary, but it does not c-command the PP to the door. As can be seen below, the first branching node which dominates V is the lower V'. The only element contained in the domain is the NP Mary:
The direct object *Mary* cannot c-command the NP *John* because *John* is not dominated by the lower V′, which is the first branching node dominating *Mary*.

A very simple definition of c-command, then, is as follows:

(122) An element α c-commands an element β iff:
   a. α does not dominate β and β does not dominate α
   b. the first branching node above α also dominates β.

This relation is also known as strict c-command. When the first branching node is a maximal projection (an XP), we say that α m-commands β. Let us see how this works in a sentence like (123):

(123) He will publish his dissertation in January.

The VP in (123) has the representation in (124):
The verb publish c-commands the NP the book (the first branching node dominating the verb is V) but not summer; it m-commands both the NP the book and the NP summer, which are dominated by the first maximal projection dominating the verb, i.e. VP.

This relation of dominance plays an important part in the description of various phenomena, such as the reference of pronouns or anaphors.

### 6.2. Government

Another important structural relation between the constituents of a sentence is the relation of government. For example, it is easy to see that inside a VP such as eat an apple the verb selects the NP complement. In help them it is also obvious that the verb not only selects a complement, but it also assigns case to it. We say that the verb governs the NP complement, being the governor. The NP complement, its sister, is governed by the verb. It is the governee. A governer is dependent on its governor.

In a PP such as to him, P selects and assigns Case to its complement. We say that the P is the governor and the pronoun the governee. We can reach the conclusion that a head always governs its complement.

What type of element can count as a governor? A governor must be a head. In the representation of the VP eat an apple, it is obvious that the verb m-commands the NP:

(125)

\[ \begin{array}{c}
& VP \\
& \downarrow \\
V' & \\
& \downarrow \\
V^0 & \\
& \downarrow \\
NP & \\
& \downarrow \\
eat & \\
\end{array} \]

The first maximal projection dominating the V is VP, which also dominates the NP.

Consider now the VP eat an apple in the kitchen. The NP an apple is governed by the verb, but the verb cannot govern the kitchen, inside the PP. It can only govern the whole PP. PP is a maximal projection and maximal projections are barriers to government:

(126)

\[ \text{VP} \]

---

1^For the relevance of c-command/m-command in the analysis of various syntactic and semantic phenomena see, among others, Cook (1988), Cornilescu (1995b), Haegeman (1991) or Radford (1992). For the relevance of c-command in the analysis of double object constructions in English, see Chapter V in this textbook. For the relation between c-command and the referential properties of NPs, see Chapter VII, section 3.

72
The NP the kitchen is governed by the P in, which m-commands it (PP, the first maximal projection dominating the P in also dominates the NP the kitchen) and whose government is not blocked by any intervening maximal projection that contains an intervening governor, i.e. a barrier. The only intervening node is P', which is not a maximal projection.

The definition of government is summarized below:

(127)
An element $\alpha$ governs an element $\beta$ iff:
(i) $\alpha$ is a head
(ii) $\alpha$ m-commands $\beta$
(iii) no maximal projection intervenes between $\alpha$ and $\beta$.

Some typical government relations are:
(i) the relation between a V and its complement, where the V is the governor and the NP complement is the governee: eat an apple, read a book, learn syntax, take an exam, drink coffee.

(128)

(ii) the relation between a P and its complement, where the P is the governor and the NP complement is the governee: in the park, on the table, about syntax, under the bridge, with the children.

(129)

13 More recent research shows that being a maximal projection does not automatically entail the property of barrier to government. Still, for our present purpose, we will not go into further detail.
(ii) the relation between a V and a PP: read in the park, dance on the table, talk about syntax, walk under the bridge, play with the children.

6.3. GOVERNMENT AND CASE

The relationship of government plays an important part in case assignment. Case is subject to government, i.e., it can be assigned by a governor to its governee. Consider the VP help him, where help, the governor, assigns Accusative case to the governee him:

We say that the lexical verb case-marks its direct object. Whether a verb can assign Accusative case is part of its entry in the Lexicon, i.e., it is an idiosyncratic property. Some verbs can assign Accusative case, while others lack this property. This difference lies at the heart of the distinction between transitive and intransitive verbs: the former can assign Accusative case, the latter cannot.

Notice also that no barrier intervenes between the governor and the governee in (131), allowing the verb to govern into its complement and to assign case to it. The presence of intervening material would prevent the verb from governing its complement and, consequently, from assigning case to it, leading to ungrammaticality, as illustrated in (132):
* He helps today/often/in London/with money/his friends him.

As can be seen in (132), the verb can only assign case to one single complement. *He helps his friends him* is ungrammatical because help can assign case only to the closer NP, his friends, leaving the second NP, him, without case.

Similarly, inside the PP without them, the P without, the governor, assigns Accusative case to the governee them:

(133)

Case assignment under government

Just like in (131), nothing intervenes between the governor and the governee. Any intervening material would block government and, consequently, case assignment. As we are going to see in the following chapters, this adjacency requirement is very strict in English.

Because lexical verbs and prepositions can assign case, they are characterized as possible case-assigners. This property distinguishes between lexical verbs and prepositions, and nouns and adjectives. The latter cannot assign case to their complement in English.

But case is not always assigned under government. In English, Nominative case is assigned in a Spec-head agreement configuration. The [Spec IP] is the canonical subject position. The NP which occurs in this position receives Nominative case from Inflection, i.e. from the head of the IP. As can be seen in (134) below, I cannot govern the NP in [Spec IP] because in a language like English the direction of government is from left to right. Inflection assigns case from right to left:

(134)

In English, Nominative case is assigned via Spec-head agreement.

Not any type of Inflection can assign Nominative case, though. Only finite Inflection can assign Nominative case in English. This can explain why (135a), which has finite Inflection, is grammatical, whereas (135b) and (135c), which both contain non-finite Inflection (infinitival in b and gerundial in c), are ungrammatical:

(135)

Case can be assigned:
(i) under government;
(ii) in a Specifier-head configuration.
a. John attends lectures twice a week.
b. He tried ['he to attend lectures twice a week].
c. I can’t imagine ['he attending lectures twice a week].

A brief examination of the two types of case assignment discussed above will reveal that both Accusative and Nominative case are each associated with one particular configuration, or structural position. This is why they are said to illustrate structural case.

Summarizing the discussion so far, we can say that case can be assigned: (i) under government, i.e. via a relation between a head and its complement, and (ii) in a Spec-head configuration, i.e. via a relation between a head and its specifier. Case is assigned locally: a head does not assign case at a distance.

6.4. AGREEMENT

Another important relation which 'binds' the elements that make up a syntactic object is agreement. Consider the examples in (136):

(136)  
a. this textbook
b. these textbooks  
c. *this text
boks
d. *these textbooks.

In all the examples above, the head requires the dependent term to share its features. In (136a), the head textbook is [+ singular] and requires the demonstrative, which is its dependent, to have the same number feature. (136d) is ungrammatical precisely because the dependent term is [+ plural] and the head is [+ singular]. Similarly, in (136b), the head textbooks is [+ plural] and it requires its dependent term to be [+ plural]. This can explain why (136b) is grammatical, while (136c) is not. Agreement on the dependent term shows its relation of dependence on the head.

This relation is more obvious in languages with rich agreement morphology, as Romanian, where adjectives agree in case, number and gender with the nominal head:

(137)  
a. o  
a – FEM.SG.NOM/ACC  
fată  
– FEM.SG.NOM/ACC  
frumoasă  
b. unei  
– FEM.SG.GEN/DAT  
fete  
– FEM.SG.GEN/DAT  
frumoase  
c. un  
– MASC.SG.NOM/ACC  
băiat  
– MASC.SG.NOM/ACC  
iinăl  
d. unui  
– MASC.SG.GEN/DAT  
băiat  
– MASC.SG.GEN/DAT  
iinăl  

c. unui  
– MASC.SG.GEN/DAT  
băiat  
– MASC.SG.GEN/DAT  
iinăl  

c. unui  
– MASC.SG.GEN/DAT  
băiat  
– MASC.SG.GEN/DAT  
iinăl  

c. unui  
– MASC.SG.GEN/DAT  
băiat  
– MASC.SG.GEN/DAT

The subject of a sentence also 'agrees' with Inflection in person and number. In English, third person singular is marked in the present by the morpheme /z/:

(138)  
a. John attends lectures twice a week.
b. *John attend lectures twice a week.
c. Students attend lectures twice a week.

76
d.  *Students attends lectures twice a week.

(138b) and (138d) above are ungrammatical precisely because the subject NP does not agree with the features of Inflection.

Again, this relation is more obvious in Romanian, where the present tense paradigm has five different forms:

(139)

\[
\begin{array}{ll}
\text{a veni ('to come')} & \text{a mînca ('to eat')} \\
1\text{sg} & 1\text{sg} \\
2\text{sg} & 2\text{sg} \\
3\text{sg} & 3\text{sg} \\
1\text{pl} & 1\text{pl} \\
2\text{pl} & 2\text{pl} \\
3\text{pl} & 3\text{pl}
\end{array}
\]

We say that verbal inflection is weak in English, because it evinces a very small number of distinct forms of agreement, and strong in Romanian, where it contains a big number of different agreement forms. As we are going to see in the next chapter, this property of Inflection can account for the differences between lexical verbs and auxiliaries in English, or between lexical verbs in English and in languages with rich inflection, such as Romanian.

Summary

This chapter has been concerned with the status and structure of root sentences. It has been shown that a sentence is a hierarchically structured entity, whose status is that of a Complementiser Phrase. Its immediate constituents are phrases and the ultimate constituents are words.

Words belong to a finite set of clearly defined categories, which can be distinguished morphologically and syntactically: nouns, verbs, prepositions, adjectives, etc. The category a word belongs to is part of the information provided by the Lexicon and it is a property which plays an important part in the formation of sentences: the category to which a word belongs determines its distribution and its notional structure determines the selection of its complements.

It has also been argued that sentences are made of both functional (Tense, Agreement, Mood, Aspect, Determiner, Degree, Complementiser) and lexical elements (nouns, verbs, adjectives, adverbs, prepositions) which expand into endocentric projections called phrases (such as NP, VP, IP, CP), according to the same X-bar format:

\[
\begin{array}{c}
\text{XP} \\
\text{Spec} \\
\text{Spec} \\
\text{X'} \\
\text{X}^0 \\
\text{YP}
\end{array}
\]

There is morphological and syntactic evidence that phrases are constituents.

The most important structural relations within a sentence are the relation of c-/m-command, the relation of government, and
that of agreement.

**C-command**
An element \( \alpha \) c-commands an element \( \beta \) iff:
\( \alpha \) does not dominate \( \beta \) and \( \beta \) does not dominate \( \alpha \)
the first branching node above \( \alpha \) also dominates \( \beta \).

**M-command**
An element \( \alpha \) m-commands an element \( \beta \) iff:
\( \alpha \) does not dominate \( \beta \) and \( \beta \) does not dominate \( \alpha \)
the first maximal projection (XP) above \( \alpha \) also dominates \( \beta \).

**Government**
An element \( \alpha \) governs an element \( \beta \) iff:
\( \alpha \) is a head
\( \alpha \) m-commands \( \beta \)
no maximal projection intervenes between \( \alpha \) and \( \beta \).

**Appendix**
In section 2 of this chapter we discussed how one could represent the structure of phrases using the so-called tree notation. Many studies, though, may use a different notation, the so-called bracketed notation or bracketed diagrams. The principles are the same. The difference between one notation and the other is only related to the way in which you “draw” the structure, i.e. it is a purely notational one.

Let us see how some of the phrases that have already been represented by using tree notation can be represented by using the bracketed notation.

**Noun Phrases**
The equivalent of the tree notation in (1a) will be (1b):

(1)  
(a)  
```
NP  
  Det N'  
    this  
      N^0  
        lecture
```
(b)  
```
[np [d this] [n lecture]]
```

As you may have already noticed, each word/constituent is placed in between square brackets. The label of each constituent is put on the left bracket, as a subscript. Let us see how this works with various types of phrases.

**Verb Phrases**

(2)  
```
VP  
  V'  
  v  
  buy  
  Det N'
```
Let us do it step by step. For your convenience, the "added" elements in each step have been bolded.

(i) bracket each part of speech and label it:
\[ [\text{v buy} [\text{D a} [\text{N house}]]] \]

(ii) bracket the NP:
\[ [\text{v buy} [\text{NP [D a] [N house]]}] \]

(iii) bracket the VP:
\[ [\text{[v buy] [NP [D a] [N house]]}] \]

---

**Adjective Phrases**

(i) bracket each part of speech of the Adjective Phrase and label it:
\[ [\text{A nice}] \]

(ii) bracket the AP:
\[ [\text{AP [A nice]}] \]

(iii) Bracket the NP now, with the AP inside it:
\[ [\text{NP [D a] [AP [A nice]] [N girl]]} \]

---

**Prepositional Phrases**

(i) bracket and label each part of the PP:
\[ [\text{P on} [\text{D the} [\text{N table}]]] \]

(ii) bracket the NP which is the complement of P:
\[ [\text{P on} [\text{NP [D the] [N table]]}] \]

(iii) bracket the PP:
\[ [\text{PP [P on] [NP [D the] [N table]]}] \]

Let us try a whole sentence now, leaving the functional projections in the verbal domain aside for the moment:

---

79
The nice girl bought a house.

(i) Bracket and label each part of speech:
[the] [A nice] [N girl] [V bought] [D a] [N house]

(ii) Bracket the NP “a house”:
[NP [D a] [N house]]

(iii)Bracket the VP “bought a house”:
[VP [V bought] [NP [D a] [N house]]]

(iv) Bracket the AP “nice”:
[AP [A nice]]

(v) Bracket the NP “the nice girl”:
[NP [D the] [AP [A nice]] [N girl]]

(vi) Bracket the sentence:
[S [NP [D the] [AP [A nice]] [N girl]] [VP [V bought] [NP [D a] [N house]]]]

**Exercises**

1) Discuss the main properties of lexical and functional categories. Provide examples.

2) Discuss and illustrate:
   (i) immediate constituent of a sentence;
   (ii) ultimate constituent of a sentence;
   (iii) phrase.

   (3) What does it mean to say that the structure and the meaning of a sentence are compositional?

   (4) Discuss the categorial status of each word in the following sentences, giving detailed reasons in support of your analysis:
   a. This boring chapter in the textbook will probably drive me mad one of these days.
   b. The book which you are reading at the moment may seem puzzling to many people.

   (5) The examples below contain nonsense words. Try to decide on the status of these nonsense words, giving arguments in support of your decision:
   I John likes to glonk in the afternoons. (2) He never glonks on Sundays. (3) He started glonking when he was 14. (4) He once glonked an out-of-work actress. (5) He has never glonked any of his classmates. (Radford 1992: 106)
   II (1) John was feeling nurgy, but happy. (2) He’s nurgier than anyone I know. (3) He’s been behaving very nurgily all week. (Radford 1992:106)
   III She put the car ung the garage. She made sure that it was right ung. (Radford 1992: 106)

   (6) Which is the head of the following phrases?
   a. a recently published book
b. They danced the whole night.
c. write a novel
d. a very romantic city
e. extremely romantic
f. rather rudely
g. Can I ask you a question?
h. I believe [that he can solve this task.]

(7)
Draw tree diagrams to represent the structure of the following phrases. Use bracketed diagrams to represent the bolded phrases:

(1) a. the house; b. the beautiful house; c. the beautiful house of brick; d. two books of my wife; e. the claim that the earth is round.
(2) a. very nice; b. so very nice; c. extremely angry with the teacher; d. quite pleased about it; e. a recent unusual book of biology; f. happy that she saw you.
(3) a. eat an apple; b. eat an apple in the kitchen; c. sleep in the bedroom; d. read a book every day; e. understand that he was lost.
(4) a. They will buy this book. b. She lived in London in the 19th. c. Should we learn generative syntax? d. John is a nice person. e. She knows that they don't like syntax. f. I wondered whether he could help us. g. They waited for the show to begin.

(8)
Discuss the syntax and semantics of the italicised pro-forms as they are used in the sentences below:

(1) I don't know whether I will pass the exam, but I certainly hope so.
(2) They say it's extremely difficult, and so it may be.
(3) You should go to Prague, if you've never been there.
(4) If the teacher is in Holland (which she is), how can we get in touch with her?

(9)
Which of the bracketed sequences in the sentences below are constituents? Explain what tests you used in order to determine constituent status. Apply at least two tests in each case:

(1) They met [near the park].
(2) [The news] took us all by surprise.
(3) The news [that they were getting married] took us all by surprise.
(4) The house [was extremely] beautiful.
(5) She has [left for London].
(6) The student looked [up the word] in the dictionary.
(7) The [student has] forgotten about the meeting.

10)
Decide whether the verb in each sentence functions as a Phrasal Verb or as a Prepositional Verb. Explain what test(s) you have used:

(1) He will pick up the ladder.
(2) He will climb up the ladder.
11) Define and illustrate with examples of your own: (i) c-command; (ii) m-command; (iii) government.

12. Does the bolded verb in each of the sentences below c-command/m-command/govern the italicised NPs and PPs?

(1) John read a newspaper on the beach.
(2) John sent a letter to Mary.
(3) John gave Mary a bunch of flowers.
(4) They were dancing in the kitchen.
(5) The girl whispered the secret to her mother.