1
Introduction

1.1 Some basic syntactic patterns

1.1.1 *Do human languages have a syntax?*

The syntax of a language is the system which determines the possible ways in which words can be combined to produce valid sentences. Our task is to investigate, in a methodical way, the syntactic system of French, to discover and make explicit the rules and principles which enable native speakers of French to formulate and comprehend the full range of sentences of their language, a system which we, as non-native speakers, possess only to a partial extent.

But does such a system really exist? As learners of French, we are painfully aware of aspects of the language which seem strange, illogical or unsystematic. Even in our own language, we may be baffled by apparently simple questions. If a learner of English asks us why we must say (1a) and not (1b), our best response may be simply that (1b) 'sounds funny', and the response of a native speaker of French regarding the opposite facts in (2) is likely to be the same:

(1) a That boy often watches that programme
     b *That boy watches often that programme
(2) a *Ce garçon souvent regarde cette émission
     b Ce garçon regarde souvent cette émission

Moreover, rather than admit ignorance about our own language, we may be tempted to conclude that there is nothing to explain: languages are inherently mysterious, our judgements about what we can or cannot say are based on familiarity not on rules, and so on.

This is a very defeatist conclusion. In other fields of inquiry, we do not give up just because solutions to problems are not immediately obvious. Also it is of no help to the language-learner who is anxious to avoid errors like (1b) and (2a). A more constructive response is to conclude that native speakers are not conscious of the system which underlies their language. In order to explain examples like (1) and (2) we must reconstruct this system, by analysing samples of the language and trying to fit them into a coherent general model. Only in this way can

1

we arrive at an informed answer to the general question raised above. Similarly, it is only by searching for regularities that we can identify those aspects of language which are truly idiosyncratic and which must be simply 'learnt'. Nevertheless, there are also some *a priori* reasons for supposing that human languages are fundamentally regular systems. In the following sections, we shall outline some of these reasons as a way of introducing some of the basic properties of the language system.

1.1.2 Word classes

You are unlikely to have encountered either of the sequences in (3) below, but you should be able to recognise (3a) as a valid sentence of French, and to interpret it if you know the meanings of all the words (*caribou* is a Canadian reindeer), while (3b) is gibberish:

(3)   a   Le caribou dévorait une pizza
            b   *Pizza caribou le dévorait une

Your ability to make this distinction must be based on something more abstract than mere familiarity with these examples – if you have never encountered them before, they should be equally unfamiliar, therefore equally invalid according to the 'defeatist' view envisaged in 1.1.1.

   Here, you might object that, although you have not encountered (3a) before, you have come across similar sentences, like (4a), but you have never heard anything resembling (3b) or (4b):

(4)   a   Le garçon mangeait une pomme
            b   *Pomme garçon le mangeait une

To this extent, (3a) is more familiar than (3b) in that it it conforms to a familiar pattern – it is just the particular combination of words that is novel.

   This observation leads us to a language-system of sorts; let us call it the **category-sequence** model. Somehow, we are able to perceive that *caribou* and *garçon*, *pizza* and *pomme* are words of the same type – they belong to the same **word class** (or **lexical category**), traditionally called **nouns**. Similarly, *dévorait* and *mangeait* belong to a different class (**verbs**, or rather forms of verbs) and *la* and *une* belong to a category which will be called **determiners** (sometimes called 'articles'). Thus, on encountering sentences like (4a) we can deduce that the sequence of categories in (5) constitutes a valid sentence pattern in French (and indeed English) which we can use as a template for forming novel sentences like (3a) by inserting different words of the appropriate categories:

(5)   D + N + V + D + N

2

Obviously the formula in (5) is not a complete grammar of French, but we might suppose that we could arrive at a complete grammar by pursuing the procedure sketched above – each time a new sequence of categories is encountered, it is entered in the grammar as a further formula for constructing sentences until, eventually, all the possible category-sequences which constitute valid sentences have been listed.

This model amounts to little more than a formalisation of the 'defeatist' view envisaged in 1.1.1. For example, the intuition that (2a) above 'sounds funny' in French while (2b) is fine can now be attributed to the fact that category-sequences of the type (6a) are regularly encountered in French (where *souvent* belongs to the class of **adverbs** and *ce(tte)* is a determiner) and are therefore familiar, whereas the sequence in (6b) never occurs:

(6)   a   D + N + V + Adv + D + N
        b   *D + N + Adv + V + D + N

However, this account scarcely constitutes an explanation, it merely states an observation. Moreover, the information encoded in (6) does not predict the contrast in (7), where we have added an **adjective** (*intelligent*) before the first noun:

(7)   a   *Ce garçon intelligent souvent regarde cette émission
        b   Ce garçon intelligent regarde souvent cette émission

According to the category-sequence account, we would need to be familiar with sequences of the type (8a) but not with those of type (8b):

(8)   a   D + N + A + V + Adv + D + N
        b   *D + N + A + Adv + V + D + N

This seems wrong. Intuitively, once we know that (2a) is ungrammatical and (2b) is fine, we ought to be able to deduce the position of the adverb in (7) without direct experience of sentences which conform exactly to the category-sequences in (8).

1.1.3 Phrase structure

The main problem with the category-sequence model is that it does not recognise that there are groups of words within a sentence which belong together in some sense. In the sentences considered above, it seems clear that the sequence of a determiner and a noun (e.g. *ce garçon* in (2), *une pizza* in (3)) forms a unit, in contrast to other adjacent pairs such as N + V (*garçon mangeait*) or V + D (*mangeait une*). Let us call this unit a **Noun Phrase** (**NP**). Similarly, if we place a **degree item** like *très* in before the adjective in (7b), as in (9), it seems clear that this item goes with the adjective, forming an **Adjective Phrase** (**AP**):

Introduction

(9)　　　　Ce garçon très intelligent regarde souvent cette émission

Moreover, in examples like (9) the AP appears to form part of the NP; e.g. *(très) intelligent* goes with *garçon* rather than with the following verb *regarde*.

　　By postulating **phrasal categories** like NP or AP as well as lexical categories, we can arrive at a partial solution to problems of the sort raised at the end of 1.1.2. Roughly, in French, adverbs cannot be inserted between the first NP and the following verb but they can be inserted between the verb and the NP which follows it (vice versa for English), regardless of the internal composition of the NPs (e.g. whether they contain an adjective or not):

(10)　　　　a　NP + Adv + V + NP　(English, but not French)
　　　　　　　b　NP + V + Adv + NP　(French, but not English)

There are further complications regarding the position of adverbs which will be addressed briefly in 1.4.3 and in greater detail in 7.5.

　　Having established the existence of phrasal categories, we must define their content. This can be achieved by means of **phrase-structure rules**, like those in (11), where the sequence of elements to the right of the arrow defines what the phrasal category on the left can consist of, elements in parentheses being optional:

(11)　　　　a　NP → D + N + (AP)
　　　　　　　b　AP → (Deg) + A

Similarly, having defined the internal structure of the NP, we can use this phrasal category to define the sentence pattern of the examples given so far (ignoring adverbs for the moment) as in (12), where S = sentence:

(12)　　　　S → NP + V + NP

The rules in (11)–(12) are intended as illustrations rather than accurate descriptions and will be modified as the discussion proceeds.

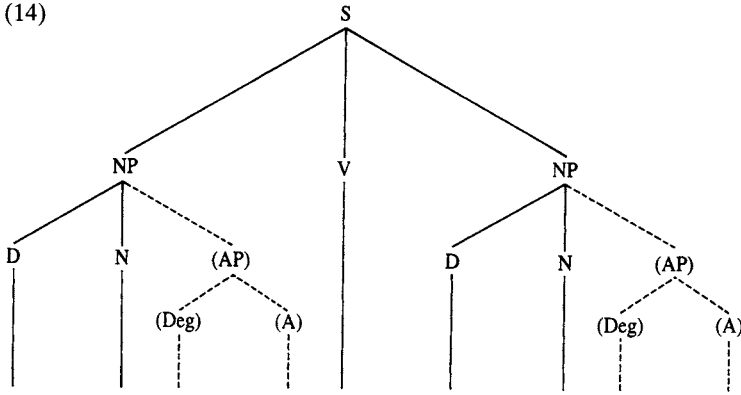　　Rule (12) defines a sentence pattern which can be represented by the **tree structure** (13):

(13)

The phrasal category NP can now be expanded by applying rule (11) to give the full structure (14), where dotted branches and parentheses indicate optional elements:

(14)

Actual sentences can be created by inserting words of the appropriate category on the branches labelled D, N, V, Deg and A.

The principal difference between the **phrase-structure** model outlined above and the category-sequence model is that the former is modular; each phrasal category is defined independently of the others and independently of the sentence as a whole.

A practical advantage of this model is that the grammar can be modified 'bit-by-bit' to accommodate new patterns. To illustrate, instead of using an NP of the form D + N + (AP) we can often use a **proper noun** (i.e. a name) like *Pierre* or a **pronoun** such as *cela*. To account for these possibilities we can simply modify rule (11a) as in (15), where the curved brace indicate alternative expansions, leaving the rest of the grammar intact:

(15)
$$\text{NP} \rightarrow \begin{Bmatrix} D + N + (AP) \\ N_{\text{PROPER}} \\ \text{Pronoun} \end{Bmatrix}$$
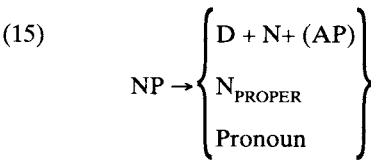
Since proper nouns and pronouns are defined as possible manifestations of the category NP, a sentence like *Pierre mangeait cela* conforms to the rule in (12). This modular approach also makes the important (and essentially correct) prediction that the possible manifestations of a particular phrasal category (e.g. NP) will be the same regardless of the position of the phrase within the sentence (e.g. before or after the verb in the above examples). Under the linear model, such similarities are purely accidental – e.g. it would be perfectly possible to describe a hypothetical language which allowed sentences of the form $N_{\text{PROPER}} + V + D + N$ and $D + N + V + N_{\text{PROPER}}$ but not, say, $N_{\text{PROPER}} + V + N_{\text{PROPER}}$. To this extent, the phrase-structure model goes some way towards explaining a fundamental property of natural languages as well as describing the facts of a particular language.

5

Introduction

> Note that rule (15) analyses items like *Pierre* and *cela* not only as single words but also as phrases (NPs). Similarly, according to rule (11b) an adjective occurring on its own constitutes an AP. This conflicts with the more usual concept of a phrase as an expression consisting of more than one word. In the framework adopted here, a more appropriate way of construing a 'phrase' is as an expression consisting of a key word (e.g. the noun in an NP, the adjective in an AP) and any elements (possibly none at all) which go with it.

1.1.4 Verb Phrases

Most linguists would dispute the analysis of the sentence presented in (12)–(13), claiming instead that the second NP goes with the verb to form a **Verb Phrase** (**VP**). Consider the following example containing an **Adverb Phrase** (**AdvP**) consisting of a manner adverb (*lentement*) optionally preceded by a degree item:
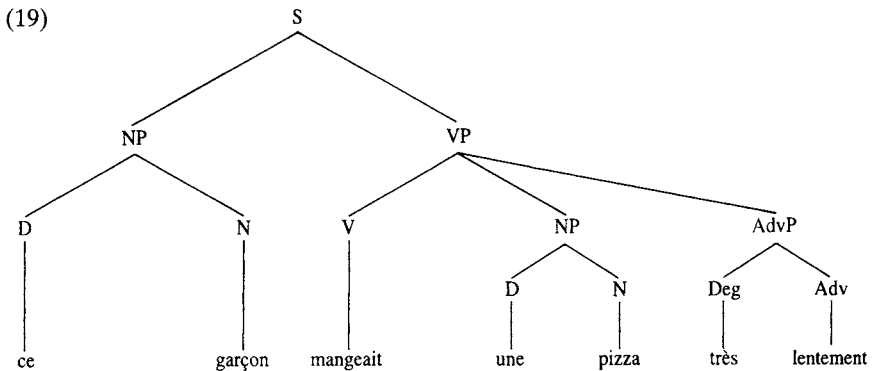
(16)   Ce garçon mangeait une pizza (très) lentement

In (16) the AdvP describes the action denoted by the verb *mangeait* in much the same way as the AP *très intelligent* describes the boy in (9). A natural way of representing this relation is to assume that the AdvP is part of the VP, just as the AP *(très) intelligent* is part of the NP. However, this only makes sense if the intervening NP (*une pizza*) is also part of the VP. Thus, our earlier rule (12) expanding S can be replaced by the rules in (17)–(18):

(17)   S → NP + VP

(18)   VP → V + NP + (AdvP)

The structure of (16) can be represented as in (19):

(19)
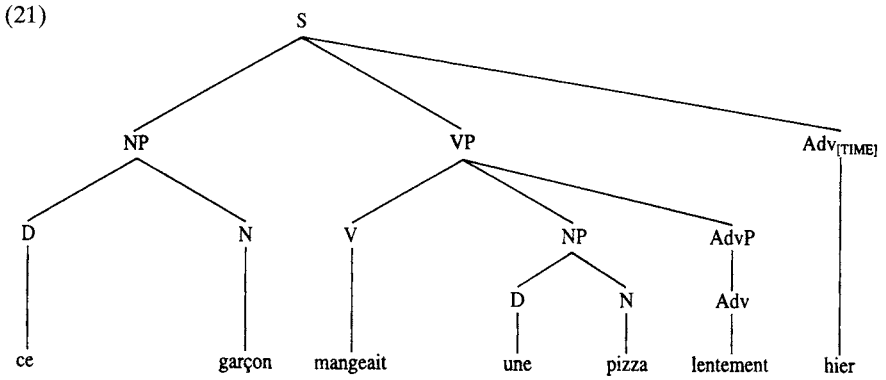

The examples in (20) provide further evidence for the category VP, as well as illustrating an important condition on syntactic structure:
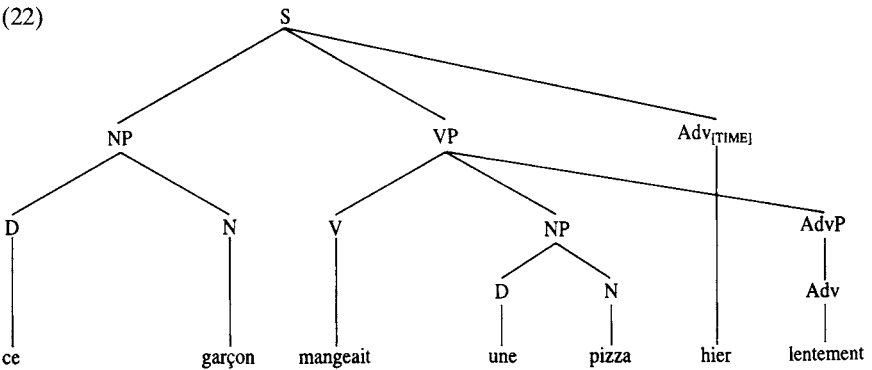
6

(20)　　　a　Ce garçon mangeait une pizza lentement hier

　　　　　b　??Ce garçon mangeait une pizza hier lentement

Time adverbs like *hier* modify the whole sentence rather than simply the verb and thus form part of the S rather than the VP, as shown in (21) for example (20a):

(21)

S
— NP [D: ce | N: garçon]
— VP [V: mangeait | NP [D: une | N: pizza] | AdvP [Adv: lentement]]
— Adv[TIME]: hier

Given our assumption that *hier* forms part of the S whereas *lentement* belongs to the VP, we cannot draw a tree structure for (20b) without making the branches of the tree cross:

(22)

S
— NP [D: ce | N: garçon]
— VP [V: mangeait | NP [D: une | N: pizza] | AdvP [Adv: lentement]]
— Adv[TIME]: hier

Thus, example (20b) can be excluded by the following condition on syntactic structure:

The 'no-crossing constraint': *The branches of a tree cannot cross.*

Note that this account depends crucially on the assumption that *mangeait une pizza lentement* forms a phrasal category VP.

1.1.5 Some terminology

Before proceeding further, it may be helpful to introduce some technical termi-
nology and other notational devices which will be used in subsequent discussion.

   The various points in a tree structure, labelled S, NP, VP, N, V, D, etc., are called
nodes. S, NP, VP, AdvP, etc. are **phrasal nodes**, whereas N, V, Adv, D, etc. are **lexical
nodes**.

   The basic relation between nodes of a tree is that of **dominance**. A node X dom-
inates a node Y if there is a downward path from X to Y along the branches of the
tree. Thus, in structure (19), S dominates all the other nodes; VP dominates V, the
second NP, AdvP, and the nodes under them (D, N Deg and Adv), and so on. A
node X **immediately dominates** Y if X dominates Y and there is no intervening
node. Thus, in (19), S immediately dominates only the first NP and the VP while
VP immediately dominates only V, the second NP and the AdvP. Kinship terms
are sometimes used to describe such relations; e.g. if X immediately dominates Y
and Z, then X is the **mother** of Y and Z, while Y and Z are **sisters** of each other and
daughters of X. Thus, in (19) NP is the mother of D and N, D is a sister of N (and
vice versa) and D and N are daughters of NP.

   In phrasal categories, the lexical node which determines the category of the
phrase is the **head** of the phrase and, conversely, the phrase is a **projection** of the
lexical category (e.g. N is the head of NP, and NP is a projection of N – for the pur-
poses of this definition, pronouns and proper nouns introduced by the rule in (15)
can be treated as subclasses of nouns). According to the rules we have proposed, S
differs from other phrasal categories in that it does not have a clearly identifiable
head and is not therefore a projection of any category – we shall return to this
matter in 1.3.

   A sequence of words is a **constituent** if there is a node which dominates the
whole sequence but does not dominate anything else. Thus, in (19) *ce garçon* is a
constituent (dominated by NP) as is *mangeait une pizza lentement* (dominated by
VP), but *ce garçon mangeait* is not, because the only node which dominates the
whole sequence is S, but S also dominates other elements which are not part of this
sequence. Similarly, according to (19), *mangeait une pizza* is not a constituent
either (though it probably should be, see 1.4.2 below), because the lowest node
which dominates this sequence (VP) also dominates *lentement*. A constituent is
said to be an **immediate constituent** of X if the node which defines it as a constitu-
ent is immediately dominated by X; e.g. the NP *ce garçon* and the VP *mangeait une
pizza très lentement* are immediate constituents of S (because the NP and VP
nodes are immediately dominated by S), whereas the NP *une pizza* is an immedi-
ate constituent of VP.

   In example sentences, square brackets will sometimes be used to indicate

8

constituents, with the category label subscripted to the opening bracket. Thus the immediate constituents of S in (19) can be represented as in (23):

(23) $[_S [_{NP}$ ce garçon] $[_{VP}$ mangeait une pizza très lentement]]

We can use this notation to represent all the constituents in (19), as in (24):

(24) $[_S [_{NP} [_D$ ce]$[_N$ garçon]] $[_{VP} [_V$ mangeait]$[_{NP} [_D$ une]$[_N$ pizza]] $[_{AdvP} [_{Deg}$ très] $[_{Adv}$ lentement]]]]

The bracketted representation in (24) is equivalent to the tree structure in (19). Bracketted representations are more difficult to read than trees, but they take up less space and are useful when only a partial structural analysis is required (as in (23)). Partial analyses can also be represented in tree diagrams by using a triangle (instead of a branching structure) over the constituent which is left unanalysed; e.g. the tree in (25) is equivalent to the bracketted representation (23):

(25)
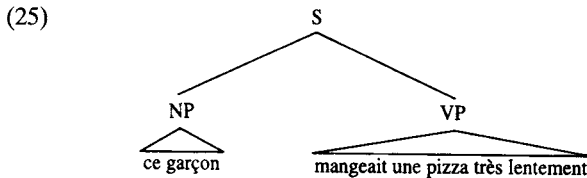
1.2 Grammatical relations

1.2.1 Subjects and objects

The division of the sentence into its immediate constituents illustrated in (23) and (25) and defined by rule (17) (S → NP + VP) corresponds closely to the traditional partition of the sentence into a **subject** and a **predicate**, as in (26):

(26)   Ce garçon | mangeait une pizza lentement
            SUBJECT | PREDICATE

The traditional concept 'predicate' can be equated with the phrasal category VP. However, we cannot simply equate 'subject' with NP, since *une pizza* in our example is an NP, but not a subject. Unfortunately, in traditional grammar the term 'subject' is sometimes used rather vaguely to encompass a range of notions which do not always coincide (e.g. 'the performer of the action' or 'the entity about which something is asserted') – more will be said about this in due course. Within the phrase-structure framework we can define the 'core' concept of a subject as the NP which is immediately dominated by S and which is the sister (to the left) of VP.

The category 'subject' is qualitatively different from phrasal categories of the sort presented so far (e.g. NP, AP, VP, ...). Whereas phrasal categories are defined

9

by their internal composition, principally the word class of the head (e.g. the noun in an NP), a subject is defined in terms of its external relation to other components of the sentence. Consequently, 'subject' will be referred to as a **relational category**.

Another relational category is the **direct object**, represented by *une pizza* in our example. In terms of the structures which have been proposed, the direct object can be defined (provisionally) as the NP which is a sister (to the right) of V and is immediately dominated by VP.

Subjects and direct objects express what linguists sometimes call **argument** relations. The term 'argument' is not widely used in traditional grammar, but is borrowed from formal logic. It denotes a useful and intuitively simple concept, which can best be explained by example. By virtue of its meaning, the verb *manger* involves two participants: an 'eater' and 'something which is eaten', represented respectively by the subject and the direct object. The expressions which identify these participants are the 'arguments' of the verb. Thus, *manger* can be classified as a verb which has two arguments. In contrast, a verb like *dormir* in (27) involves only one participant (the 'sleeper') and thus has only one argument (represented by the subject):

(27)   Le garçon dormait

An example of a verb with three arguments is *donner* in (28), which describes a relation between three participants (the 'giver', the 'gift' and the 'receiver'); the NP *son professeur* represents a further relational category, traditionally called the **indirect object**, which will be discussed in 1.2.2:

(28)   Le garçon donnait une pomme à son professeur

The arguments of a verb can be divided into two types according to their position. Direct objects (and indeed indirect objects) are classified as **internal arguments** in the sense that they occur within the VP, whereas the subject of a sentence is the **external argument** of the verb (occurring outside the VP). A more traditional term for 'internal argument' is **complement.** Thus, direct (and indirect) objects are complements of the verb.

1.2.2. Complements and modifiers

AdvPs like (*très*) *lentement* in (16) and the AP *très intelligent* in *un garçon très intelligent* represent a further relational category of **modifiers**. Unlike complements, they do not identify a participant in the relation but they describe a property of the action denoted by the verb or the entity denoted by the noun. Other phrasal categories can function as modifiers, as in (29) where the modifying expression is a **Preposition Phrase** (**PP**) consisting of a **preposition** (*à* in (29a) *avec* in (29b)) and an NP:

10