

SYNTAX

Syntax is the way in which words are arranged to show relationships of meaning within (and sometimes between) sentences. The term comes from *syntaxis*, the Greek word for 'arrangement'. Most syntactic studies have focused on sentence structure, for this is where the most important grammatical relationships are expressed.

THE SENTENCE

Traditionally, grammars define a sentence in such terms as 'the complete expression of a single thought'. Modern studies avoid this emphasis, because of the difficulties involved in saying what 'thoughts' are. *An egg* can express a thought, but it would not be considered a complete sentence. *I shut the door, as it was cold* is one sentence, but it could easily be analysed as two thoughts.

Some traditional grammars give a logical definition to the sentence. The most common approach proposes that a sentence has a 'subject' (= the topic) and a 'predicate' (= what is being said about the topic). This approach works quite well for some sentences, such as *The book is on the table*, where we can argue that *the book* is what the sentence is 'about'. But in many sentences it is not so easy to make this distinction. *It's raining* is a sentence, but what is the topic? And in *Michael asked Mary for a pen*, it is difficult to decide which of Michael, Mary, or the pen is the topic – or whether we have *three* topics! Also, some modern grammars treat subjects and topics in completely different ways.

In some written languages, it is possible to arrive at a working definition of 'sentence' by referring to the punctuation one is taught to use in school. Thus, an English sentence for many people 'begins with a capital letter and ends with a full stop' (or some other mark of 'final' punctuation). The problem is that many languages (e.g. in Asia) do not make use of such features; and even in those that do, punctuation is not always a clear guide. It may be omitted (in notices and legal documents, for example); and it proves difficult to prescribe rules governing its use other than 'good practice'. People therefore often disagree about the best way to punctuate a text. In some manuals of style, it is recommended that one should not end a sentence before a coordinating conjunction (*and, or, but*). But there are often cases where an author might feel it necessary – for reasons of emphasis, perhaps – to do the opposite.

It is even more difficult to identify sentences in speech, where the units of rhythm and intonation often do not coincide with the places where full stops would occur in writing. In informal speech, in particular, constructions can lack the careful organization we associate with the written language (p. 52). It is not that conversation lacks grammar: it is simply that the

grammar is of a rather different kind, with sentences being particularly difficult to demarcate. In the following extract, it is not easy to decide whether a sentence ends at the points marked by pauses (–), or whether this is all one, loosely constructed sentence:

when they fed the pigs/ they all had to stand well back/ –
and they were allowed to take the buckets/ – but they
weren't allowed to get near the pigs/ you see/ – so they
weren't happy ...

Linguistic approaches

Despite all the difficulties, we continue to employ the notion of 'sentence', and modern syntacticians try to make sense of it. But they do not search for a satisfactory definition of 'sentence' at the outset – an enterprise that is unlikely to succeed, with over 200 such definitions on record to date. Rather, they aim to analyse the linguistic constructions that occur, recognizing the most independent of them as sentences. Thus, because the following constructions can stand on their own as utterances, and be assigned a syntactic structure, they would be recognized as sentences:

she asked for a book/
come in/
the horse ran away because the train was noisy/

The following combination of units, however, could not be called a sentence:

will the car be here at 3 o'clock/ it's raining/.

The syntax of the first unit and that of the second do not combine to produce a regular pattern. It would be just as possible to have:

it's raining / will the car be here at 3 o'clock /

or either unit without the other. Within each unit, however, several kinds of rules of syntactic order and selection are apparent. We may not say:

*will be here at 3 o'clock the car/
*will be here the car/
*car at 3 o'clock/.

Each unit in the sequence, then, is a sentence; but the combination does not produce a 'larger' syntactic unit.

A sentence is thus the largest unit to which syntactic rules apply – 'an independent linguistic form, not included by virtue of any grammatical construction in any larger linguistic form' (L. Bloomfield, 1933, p. 170). But this approach has its exceptions, too. In particular, we have to allow for cases where sentences are permitted to omit part of their structure and thus be dependent on a previous sentence (*elliptical sentences*), as in:

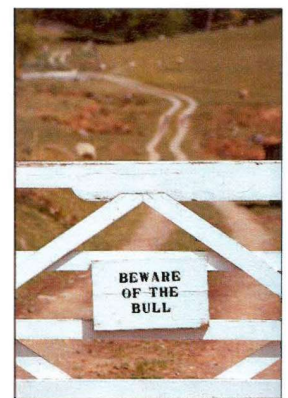
A: Where are you going?
B: To town.

Several other types of exception would be recognized in a complete grammatical description.

MINOR SENTENCE TYPES

A language contains many sentence-like units which do not conform to the regular patterns of formation. Here is a selection from English:

Yes
Gosh!
Least said, soonest mended.
How come you're early?
Oh to be free!
All aboard!
Down with racism!
No entry.
Taxi!
Good evening.
Happy birthday!
Checkmate.



A sign like this has a regular syntactic structure, but it does not use normal sentence punctuation.

ASPECTS OF SENTENCE SYNTAX

Hierarchy

Hilary couldn't open the windows.

One of the first things to do in analysing a sentence is to look for groupings within it – sets of words (or morphemes, p. 90) that hang together. In this example, we might make an initial division as follows:

Hilary / couldn't open / the windows.

Units such as *couldn't open* and *the windows* are called *phrases*. The first of these would be called a *verb phrase*, because its central word (or 'head') is a verb, *open*; the second would be called a *noun phrase*, because its head is a noun, *windows*. Other types of phrase also exist – adjective phrases, for example, such as *very nice*.

Phrases may in turn be divided into their constituent words (p. 91):

couldn't + open the + windows

And words may be divided into their constituent morphemes, if there are any:

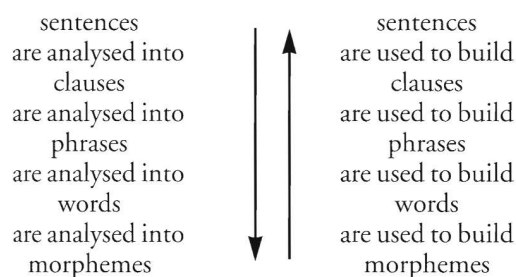
could + n't window + s

This conception of sentence structure as a hierarchy of levels, or ranks, may be extended 'upwards'. The sentence can be made larger by linking several units of the same type:

Hilary opened the windows, but David couldn't open the doors.

Here, too, we have a sentence, but now we have to recognize two major units within it – each of which has a structure closely resembling that of an independent sentence. These units are traditionally referred to as *clauses*. In the above example, the clauses have been 'coordinated' through the use of the conjunction *but*. An indefinite number of clauses can be linked within the same sentence.

A five-rank hierarchy is a widely used model of syntactic investigation:



Morphemes are the 'lower' limit of grammatical enquiry, for they have no grammatical structure. Similarly, sentences form the 'upper' limit of grammatical study, because they do not usually form a part of any larger grammatical unit.

CLAUSES

The various units that make up the structure of a clause are usually given functional labels, such as *Subject (S)*, *Verb (V)*, *Complement (C)*, *Object (O)*, and *Adverbial (A)*. A number of clause types can be identified in this way, such as:

S + V	The dog + is running.
S + V + O	The man + saw + a cow.
S + V + C	The car + is + ready.
S + V + A	A picture + lay + on the ground.
S + V + O + O	I + gave + John + a book.
S + V + O + C	He + called + John + a fool.
S + V + O + A	Mary + saw + John + yesterday.

Several approaches to grammatical analysis make use of elements of this kind, though there is considerable variation in definition and terminology. Languages also vary greatly in the way in which these elements are identified. In English, for example, word order is the main factor, with only occasional use being made of morphology (e.g. *he* (subject) *saw* (verb) *him* (object)). In Latin, word-endings provide the main clues to element function, word order being irrelevant (e.g. *puer puellam vidit* 'the boy saw the girl'). In Japanese, basic grammatical relations are marked by special particles: *ga* (subject), *o* (direct object), *ni* (indirect object), and *no* (genitive). For example,

kodomo ga tomodachi no inu ni mizu o yaru
the child friend's to dog water gives

'The child gives water to his / her friend's dog.'

PHRASES

Most phrases can be seen as expansions of a central element (the *head*), and these are often referred to as 'endocentric' phrases:

cars
the cars
the big cars
all the big cars
all the big cars in the garage

Phrases which cannot be analysed in this way are then called 'exocentric': *inside / the cars*.

The internal structure of an endocentric phrase is commonly described in a three-part manner:

all the big cars in the garage
PREMODIFICATION HEAD POSTMODIFICATION

COORDINATION VS SUBORDINATION

Coordination is one of two main ways of making sentences more complex; the other is known as *subordination*, or 'embedding'. The essential difference is that in the former the clauses that are linked are of equal grammatical status, whereas in the latter, one clause functions as part of another (the 'main' clause). Compare:

Coordinate clause:
The boy left on Monday and the girl left on Tuesday.

Subordinate clause:
The boy left on Monday when John rang.

The phrase *on Monday* is part of the clause, giving the time when the action took place. Similarly, the unit *when John rang* is also part of the clause, for the same reason. But *when John rang* is additionally a clause in its own right.

CONCORD

Grammatical links between words are often signalled by concord or 'agreement'. A form of one word requires a corresponding form of another, as when in English a singular noun 'agrees with' a singular verb in the present tense: *the man walks vs the men walk*.

The purpose of concord varies greatly between languages. In Latin, it is an essential means of signalling which words go together.

In the absence of fixed word-order patterns, sentences would otherwise be uninterpretable. For example, in *parvum puerum magna puella vidit* 'the tall girl saw the small boy', we know that the boy is small and the girl is tall only through the agreement of the endings, *-um vs -a*.

On the other hand, concord plays much less of a role in modern French, in cases such as *le petit garçon et la grande fille* 'the little boy and the big girl'. Because the position of adjectives is fixed (before the noun, in these cases), it would not pose any problems of intelligibility if there were no difference between the masculine and feminine forms:

le petit garçon
**la petit fille*
**le petite garçon*
la petite fille

If French allowed free word order, as in Latin, so that one could say **le garçon et la fille petit grande*, then concord would be needed to show which adjective should go with which noun – but this does not happen. The gender system is thus of limited usefulness, though it still has a role to play in certain syntactic contexts, such as cross-reference (*J'ai vu un livre et une plume. Il était nouveau.* 'I saw a book and a pen. It [i.e. the book] was new.').

IMMEDIATE CONSTITUENT DIAGRAMS

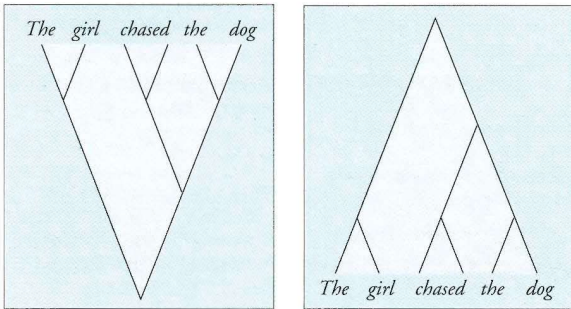
One of the most widely used techniques for displaying sentence structure is the use of *immediate constituent* (IC) analysis. This approach works through the different levels of structure within a sentence in a series of steps. At each level, a construction is divided into its major constituents, and the process continues until no further divisions can be made. For example, to make an IC analysis of the sentence *The girl chased the dog*, we carry out the following steps:

1. Identify the two major constituents, *the girl* and *chased the dog*.
2. Divide the next-biggest constituent into two, *viz. chased the dog* into *chased* and *the dog*.
3. Continue dividing constituents into two until we can go no further, *viz. the girl* and *the dog* into *the + girl*, *the + dog*, and *chased* into *chase + the -ed* ending.

The order of segmentation can be summarized using lines or brackets. If the first cut is symbolized by a single vertical line, the second cut by two lines, and so on, the sentence would look like this:

the ||| girl | chase ||| -ed || the ||| dog

However, a much clearer way of representing constituent structure is through the use of 'tree diagrams':

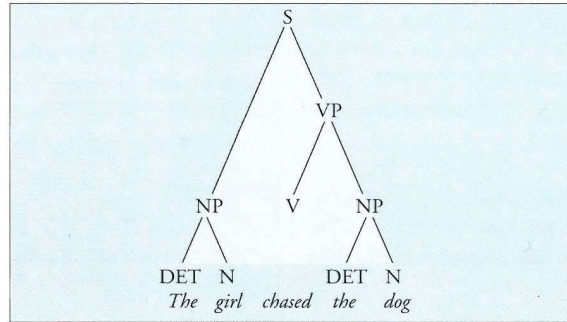


The second kind of tree diagram is in fact the normal convention in modern linguistics.

Such representations of structure are very helpful, as far as they go. But not all sentences are as easy to analyse in IC terms as this one. It is sometimes not clear where the cuts should be made (e.g. whether to divide *the three old men* into *the + three old men* or *the three old + men*, or *the three + old men*). More important, the process of segmenting individual sentences does not take us very far in understanding the grammar of a language. IC analyses do not inform us about the identity of the sentence elements they disclose, nor do they provide a means of showing how sentences relate to each other grammatically (as with statements and questions, actives and passives). To develop a deeper understanding of grammatical structure, alternative approaches must be used.

PHRASE STRUCTURE

A good way of putting more information into an analysis would be to name, or *label*, the elements that emerge each time a sentence is segmented. It would be possible to use functional labels such as 'subject' and 'predicate', but the approach that is most widely practised has developed its own terminology and abbreviations, so these will be used here. Taking the above sentence (S), the first division produces a 'noun phrase' (NP) *the girl* and a 'verb phrase' (VP) *chased the dog*. (This is a broader sense of 'verb phrase' than that used on p. 95, as it includes both the verb and the noun phrase that follows.) The second division recognizes a 'verb' (V) *chased* and another noun phrase *the dog*. The next divisions would produce combinations of 'determiner' (DET) and 'noun' (N) *the + girl*, *the + dog*. This is the 'phrase structure' of the sentence, and it can be displayed as a tree diagram:



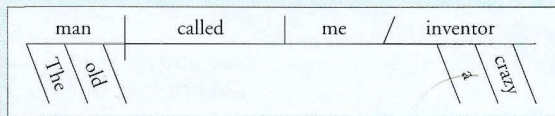
This kind of representation of the phrase structure of a sentence is known as a 'phrase marker' (or 'P-marker'). Phrase structures are also sometimes represented as labelled sets of brackets, but these are more difficult to read:

$[_S[_{NP}[_{DET}the][_Ngirl]][_{VP}[_Vchased][_NP[_{DET}the][_Ndog]]]]]$

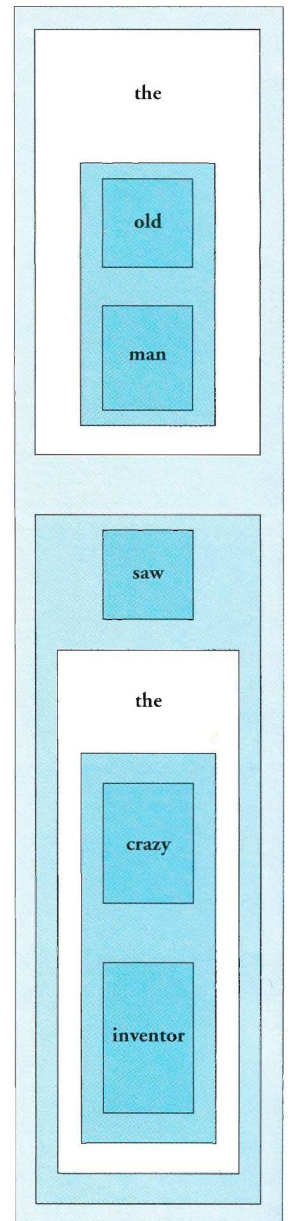
DIAGRAMMING

A frequent practice in American schools is the use of a system of vertical and slanting lines to represent the various relationships in a sentence. The representations are often called 'Reed & Kellogg' diagrams, after the authors of a 19th-century English textbook. A long vertical line marks the boundary between subject and predicate; a short vertical line divides verb and direct object; and a short slanting line marks off a complement. Other items are drawn in beneath the main parts of the sentence.

The old man called me a crazy inventor.



The approach shows the relationships between words clearly, but it cannot handle variations in word order: both *I turned off the light* and *I turned the light off* would be diagrammed in the same way.



A little-used 'Chinese box' representation of sentence structure

RULES

Analyses of single sentences are illuminating, as far as they go, but grammarians are concerned to move beyond this point, to see whether their analyses work for other sentences in the language. To what other sentences might the above sequence of steps, and the resulting P-marker, also apply? In Noam Chomsky's approach, first outlined in *Syntactic Structures* (1957), the jump from single-sentence analysis is made by devising a set of rules that would 'generate' tree structures such as the above. The procedure can be illustrated using the following rules (but several details from the original approach are omitted for clarity):

S → NP + VP
 VP → V + NP
 NP → DET + N
 V → *chased*
 DET → *the*
 N → *girl, dog*

The first rule states that a sentence can consist of a noun phrase and a following verb phrase; the second, that a verb phrase can consist of a verb plus a following noun phrase; the third, that a noun phrase can consist of a determiner plus a noun. Each abstract category is then related to the appropriate words, thus enabling the sentence to be generated. Grammars that generate phrase structures in this way have come to be called 'phrase structure grammars' (PSGs).

If we follow these rules through, it can be seen that there is already a significant increase in the 'power' of this grammar over the single-sentence analysis used previously. If we choose *the girl* for the first NP, and *the dog* for the second, we generate *the girl chased the dog*; but if the choices are made the other way round, we generate the sentence *the dog chased the girl*. By the simple device of adding a few more words to the rules, suddenly a vast number of sentences can be generated:

V → *chased, saw, liked...*
 DET → *the, a*
 N → *girl, man, horse...*

the girl chased the horse
the man saw the girl
the horse saw the man etc.

However, if *went* were introduced into the rules, as a possible V, ungrammatical sentences would come to be generated, such as **the girl went the man*. In working out a generative grammar, therefore, a means has to be found to block the generation of this type of sentence, at the same time permitting such sentences as *the man went* to be generated. The history of generative syntax since 1957 is the study of the most efficient ways of writing rules, so as to ensure that a grammar will generate all the grammatical sentences of a language and none of the ungrammatical ones.

Transformations

This tiny fragment of a generative grammar from the 1950s suffices only to illustrate the general conception underlying the approach. 'Real' grammars of this kind contain many rules of considerable complexity and of different types. One special type of rule that was proposed in the first formulations became known as a *transformational* rule. These rules enabled the grammar to show the relationship between sentences that had the same meaning but were of different grammatical form. The link between active and passive sentences, for example, could be shown – such as *the horse chased the man* (active) and *the man was chased by the horse* (passive). The kind of formulation needed to show this is:

$$NP_1 + V + NP_2 \rightarrow NP_2 + Aux + Ven + by + NP_1$$

which is an economical way of summarizing all the changes you would have to introduce, in order to turn the first sentence into the second. If this formula were to be translated into English, four separate operations would be recognized:

- (i) The first noun phrase in the active sentence (NP₁) is placed at the end of the passive sentence.
- (ii) The second noun phrase in the active sentence (NP₂) is placed at the beginning of the passive sentence.
- (iii) The verb (V) is changed from past tense to past participle (Ven), and an auxiliary verb (Aux) is inserted before it.
- (iv) A particle *by* is inserted between the verb and the final noun phrase.

This rule will generate all regular active-passive sentences.

In subsequent development of generative grammar, many kinds of transformational rules came to be used, and the status of such rules in a grammar has proved to be controversial (§65). Recent generative grammars look very different from the model proposed in *Syntactic Structures*. But the fundamental conception of sentence organization as a single process of syntactic derivation remains influential, and it distinguishes this approach from those accounts of syntax that represent grammatical relations using a hierarchy of separate ranks (p. 95).

RULES AND 'RULES'

The 'rules' of a generative grammar are not to be identified with the prescriptive 'rules' that formed part of traditional grammar (p. 3). A prescriptive grammatical rule is a statement – such as 'You should never end a sentence with a preposition' – that tells us whether we are right or wrong to use a particular construction. Generative rules have no such implication of social correctness. They are objective descriptions of the grammatical patterns that occur.

GENERATIVE NOTATION

A major feature of generative grammar is the way special notations have been devised to enable rules to be expressed in an economical way. In particular, different types of brackets, such as (), [], and {} are given different meanings. Round brackets, for example, enclose a grammatical element that is *optional* in a sentence; that is, the sentence would be grammatical even if the element were left out. The rule

NP → DET (ADJ) N

means that a noun phrase can consist of *either* a determiner, adjective, and noun *or* simply a determiner and noun (*the old man or the man*). A grammar could, of course, list the two possibilities separately, as

NP → DET + N

NP → DET + ADJ + N

but collapsing them into a single rule, through the use of the () convention, saves a great deal of space, and represents something we all 'know' about the structure of the noun phrase.

PRINCIPLES AND PARAMETERS

Government and binding theory is an approach to generative grammar which developed in the 1980s. It takes its name from the way it focuses on the conditions which formally relate (or 'bind') certain elements of a sentence, and on the structural contexts within which these binding relationships apply ('govern').

The approach holds that the same *principles* of syntax operate in all languages, though they can differ slightly (along certain *parameters*) between languages (§14). For example it is a syntactic principle that in a noun phrase there is a chief element (the *head*), which will be the noun (*the new President*), and that other nouns may accompany it (*the President of America*). But whether the accompanying nouns occur before or after the head varies between languages: they occur after it in English, but before it in Japanese (*Amerika no Daitoryo*).

WORD ORDER

The term 'word order' is somewhat ambiguous, for it can refer both to the order of words in a phrase, and to the order of multi-word units within a sentence. Given the sentence

The cat sat on the mat

both the following involve word-order problems – but they are of very different kinds:

**cat the sat mat the on*
**sat the cat on the mat*

In linguistic description, word-order studies usually refer to the second type of problem – that is, the sequence in which grammatical elements such as Subject, Verb, and Object occur in sentences. A great deal of attention has been paid to the way in which languages vary the order of these elements, as part of typological studies (§14). Word order, it is hoped, will be a more satisfactory way of classifying languages than the older morphological method (which recognized such types as isolating and inflecting, p. 295), into which many languages do not fit neatly.

In comparing word orders across languages, it is important to appreciate that what is being compared is the 'basic' or 'favourite' pattern found in each language. For example, in English, we will encounter such sequences as:

SVO *the boy saw the man*
OVS *Jones I invited – not Smith*
VSO *govern thou my song* (Milton)
OSV *strange fits of passion have I known*
(Wordsworth)
SOV *pensive poets painful vigils keep* (Pope)

However, only the first of these is the natural, usual, 'unmarked' order in English; the others all convey special effects of an emphatic or poetic kind. The same principle must apply in studying word order in all languages, but it is often not so easy to establish which is the normal word-order pattern and which is the pattern that conveys the special effect. The mere fact of talking to a foreigner, for instance, might motivate a native speaker to change from one order to another, and it often requires great ingenuity on the part of the linguist to determine whether such stylistic changes are taking place.

Typology

Apart from cases of free word order (e.g. Latin, Quechua, Navajo, Fore), there are six logical possibilities: SVO, SOV, VSO, VOS, OSV, OVS. Of these, over 75% of the world's languages use SVO (as in English, French, Hausa, Vietnamese) or SOV (as in Japanese, Amharic, Tibetan, Korean). A further 10–15% use VSO (e.g. Welsh, Tongan, Squamish). Examples of VOS are Malagasy, Tzotzil, and Houailou.

Until recently, Object-initial languages were conspicuous by their absence, and it was thought that perhaps these did not exist. But a group of OVS languages have now been found, all in the Amazon basin, mainly belonging to the Carib family, e.g. Hixkaryana, Apalai, Bacairi, Makusi. A few other languages (e.g. Jama-madi, Apurina) seem to be OSV. But there is some variability in the data that have been collected so far, with both OVS and OSV being used by some languages.

Word-order generalizations often need careful qualification. Latin, for example, is said to have a free word order, but in fact SOV is a very common pattern in that language. Modern Hebrew is SVO, but Classical Hebrew seemed to favour VSO. German prefers SVO in main clauses, but SOV in subordinate clauses. In Tagalog, the V usually comes first, but there is great variation in what follows, with both OS and SO being widely used. In Japanese, SOV is favoured, but OSV is also very common.

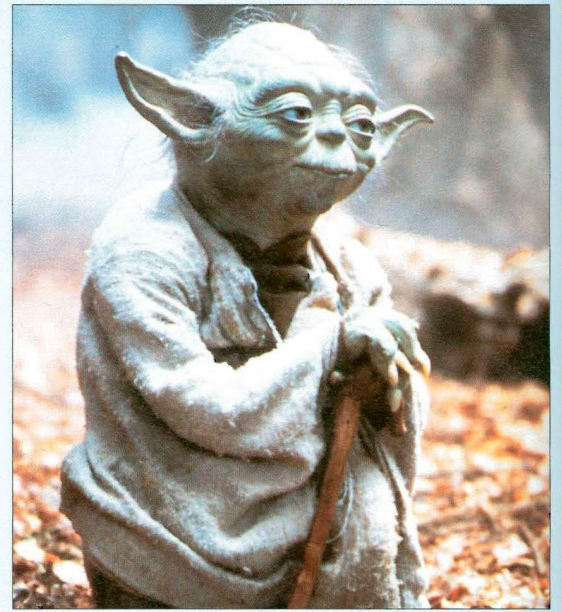
LISU

This Lolo-Burmese language seems to have free word order, yet it has no morphological cases to mark Subject and Object. A sentence Noun–Verb–Noun might therefore mean either 'N1 did V to N2' or 'N2 did V to N1'. In theory, such a language ought to be unintelligible! But in fact the speakers survive, by relying on context, the use of alternative grammatical constructions, and a modicum of common sense.

OSV IN SPACE

Sick have I become.
Strong am I with the Force.
Your father he is.
When nine hundred years
old you reach, look as
good you will not.

The rarity of OSV constructions and languages perhaps explains the impact of this strange speech style, used by the Jedi Master, Yoda, in the film *The Empire Strikes Back* (1983).



DEEP AND SURFACE STRUCTURE

In the standard approach to generative grammar, sentences are analysed in terms of two levels of organization, known as *deep structure* and *surface structure*. At the 'deep' (or 'underlying') level, a sentence structure is represented in an abstract way, displaying all the factors that govern how it should be interpreted. At the 'surface' level, there is a more concrete representation, giving the string of morphemes that closely corresponds to

what we would hear if the sentence were spoken.

This distinction was used to explain sentence ambiguities, by arguing that in such cases a single surface structure correlates with more than one deep structure. An early Chomskyan example was *Flying planes can be dangerous*, which can be related to two underlying sentences: *Planes which fly can be dangerous* and *To fly planes can be dangerous*.

The distinction was also used to relate sentences that

have different surface forms but the same underlying structure, as in the case of active and passive sentences. *Cats chase mice* and *Mice are chased by cats* were said to have different surface structures, but the same deep structure.

The interpretation and status of the two notions has altered greatly in generative theory over the years (§65), but the basic insight is one that has achieved widespread recognition in linguistics.