I

Basic concepts in grammar

The term ‘grammar’ is used in a number of different senses – the grammar of a language may be understood to be a full description of the form and meaning of the sentences of the language or else it may cover only certain, variously delimited, parts of such a description. Here we shall use it in one of these narrower senses, embracing syntax and morphology. Syntax is concerned with the way words combine to form sentences, while morphology is concerned with the form of words. We will launch without delay into a discussion of basic concepts in syntax and morphology, returning in §8 to the distinction between grammar in this sense and various other components of a full description and to the basis for dividing grammar into syntactic and morphological subcomponents. The only terms that we shall need to anticipate are ‘phonology’ and ‘semantics’: phonology deals with the sound system, with the pronunciation of words and sentences, semantics deals with meaning.

1.1 Words and lexemes

Syntax deals with combinations of words, we have said, morphology with the form of words. But again the term ‘word’ has been used in a variety of senses. For our immediate purposes it will suffice to draw just one distinction, which we can approach by considering the relation between, say, tooth and teeth: are they different words or the same word? From one point of view they are clearly different words: they are pronounced and spelt differently, they differ in meaning, and they occur in different positions in sentences (so that we could not, for example, replace tooth by teeth in This tooth is loose or teeth by tooth in These teeth are loose, and so on). Yet they are also traditionally said to be different forms of the same word. This is a more abstract sense: we abstract away the differences between them to isolate what is common to both. It will be helpful to distinguish both terminologically and notationally between these two senses. I shall use word for the less abstract concept, lexeme for the more abstract one, and I shall cite words in ordinary italics, lexemes in bold face italics. We accordingly say that tooth and teeth are different words, but forms of the same lexeme tooth.
Basic concepts in grammar

More specifically, we will say that tooth is the ‘singular’ form of teeth and that teeth is its ‘plural’ form. The words tooth and teeth are thus each analysed into two components, the abstract lexeme and what we shall call an inflectional property. These properties are relevant to both the morphological and syntactic components of the grammar (and for this reason are commonly referred to also as ‘morphosyntactic properties’). The morphology will include rules for deriving the various inflectional forms of a lexeme from the ‘lexical stem’, while the syntax will include rules specifying under what conditions a lexeme may or must carry a given inflectional property. Thus it is a fact of morphology that the plural of tooth is teeth, whereas it is a fact of syntax that if tooth enters into construction with this there must be ‘agreement’ in number, i.e. both must carry the singular inflection or both the plural. Similarly, the morphology will tell us that the ‘past participle’ of the verb see is seen, whereas the syntax will say that a past participle is required in the ‘passive’ construction, as in He was seen by the caretaker.

Not all words enter into inflectional contrasts such as we find between tooth and teeth, this and these, or see, sees, saw, seeing and seen. Usually, as with words like because, of, however, besides, this is because there is simply no inflectional property present at all – and, precisely because there is no inflectional property to abstract away, the concept of lexeme will be inapplicable in such cases. Thus because is a word that is not a form of any lexeme. In other cases we can recognise an inflectional property even though it is not independently contrastive: alms does not contrast with singular *alm, but we can still analyse it as a plural form, and conversely equipment does not contrast with plural *equipments but we can still analyse it as a singular form. In these cases we can invoke the concept of lexeme, so that equipment, for example, will be the singular form of the lexeme equipment. When we say that equipment has a singular form but no plural form we are talking about the same kind of entity as when we say that tooth has teeth as its singular form and teeth as its plural form. But it is of course contrasts like that between tooth and teeth that provide the raison d’être for the lexeme concept: if it were not for these we would have no lexeme–word distinction, tooth vs teeth, to generalise to cases like equipment vs equipment.

1.2 Constituent structure

Words are not the only units that we need in describing the structure of sentences. Although we can break a sentence down into a sequence of words, we will not go from sentence to word in a single step but will recognise units intermediate in size between sentence and word. For example, in

(1) The boss must have made a mistake

it is intuitively obvious that although a is immediately adjacent in the
1.2 Constituent structure

sequence to both made and mistake, it is more closely related to the latter than to the former: this relationship between a and mistake can then be described by saying that they go together to form a constituent of the sentence. More generally, the syntactic analysis of a sentence will assign to it a constituent structure which identifies the full hierarchy of its constituents.

A standard way of representing constituent structure diagrammatically is illustrated in (2):

(2)

```
    a
   / \ 
  /   \ /
 /     /
the  boss  must have made a mistake
```

This diagram identifies eleven constituents: the seven words, represented by the bottom row of points, and four intermediate units, the boss, must have made, a mistake, and must have made a mistake. The point from which the lines lead down to a and mistake represents the constituent a mistake, and so on. By contrast made a is not a constituent: there is no point from which the lines lead down to just this pair of words.

If we read the diagram from the top downwards we see that the sentence is divided first into the boss and must have made a mistake: these are said to be the immediate constituents (or ICs) of the sentence. Each of them is then broken down into its own ICs, the and boss for the first, must have made and a mistake for the second — and so on until we reach the bottom.

‘Constituent’ is a relational concept: if \( x \) is a constituent, it must be a constituent of something. For example, in (2) must have made is a constituent of the sentence — and also of must have made a mistake. Similarly with ‘immediate constituent’: must have made is an IC of must have made a mistake (but only of this). It follows that the sentence itself is not a constituent: as the maximal unit in syntax it is not part of any other unit. We will then apply the term construction to the sentence and any constituent except the minimal ones, the words. Thus with ‘constituent’ we are as it were looking upwards: \( x \) is a constituent if it is part of some element higher in the hierarchy; and with ‘construction’ we are looking downwards: \( x \) is a construction if it is analysable into, i.e. constructed from, one or more elements lower in the hierarchy.¹

¹ This allows for the special case where a construction has only one IC. For example the imperative sentence Stop! contains only one word, but we will still speak of it as a construction: it is constructed from that one word. It is for this reason that I say ‘higher/lower in the hierarchy’ rather than the more concrete ‘larger/smaller’. (See 3.3 for further discussion of this issue.)
Basic concepts in grammar

Finally we will use form as a general term covering both constituents and constructions. Thus in (2) there are eleven constituents, five constructions and twelve forms. Notice that this use of ‘form’ is consistent with that introduced in the last section, where we spoke of teeth, for example, as a form of the lexeme tooth: in He cleaned his teeth it is teeth not tooth that is a constituent of the sentence, so that teeth like his teeth, cleaned, cleaned his teeth, etc., will be a form.

A given sequence of words may be a constituent in one sentence but not in another. Thus John and Bill is a constituent of the sentence He saw John and Bill at the races but not of He saw John and Bill did too. Moreover, a single sentence may have two (or indeed more) constituent structure analyses, each corresponding to a different interpretation. Liz attacked the man with the knife, for example, is syntactically ambiguous, being analysable (approximately) as shown in (3) or (4);

(3)

![Diagram 3]

Liz attacked the man with a knife

(4)

![Diagram 4]

Liz attacked the man with a knife

Analysis (3) corresponds to the interpretation “Liz attacked the man who had a knife” – here the man with a knife forms a constituent, and serves to pick out the person whom Liz attacked. Analysis (4), by contrast, represents the structure the sentence has under the interpretation “Liz used a knife in her attack on the man” – here the man and with a knife do not go together to form a constituent, but are both ICs of the larger constituent attacked the man with a knife, with the man identifying the person attacked and with a knife giving information about the means of attack, not about the man. Sometimes such ambiguities are resolved ‘prosodically’ – the different constituent structures are distinguished by the intonation and rhythm (similarly in writing they may be resolved by punctuation). An elementary example is Liz saw John and Kim and Robin did too, which can be analysed (again approximately) as either (5) or (6):

(5)

![Diagram 5]

Liz saw John and Kim and Robin did too

4
1.3 Syntactic classes and functions

(6)

Liz saw John and Kim and Robin did too

(5) matches the interpretation “John and Kim were seen by Liz and they were also seen by Robin”, whereas (6) corresponds to “John was seen by Liz and he was also seen by Kim and Robin”. In any normal utterance of the sentence the prosodic features (or punctuation) would show clearly whether Kim was coordinated with John, as in (5), or with Robin, as in (6). But in general relatively little information about the constituent structure is derivable directly from the physical signal: constituent structure is an abstract property of sentences.

In discussing example (1) I said it was intuitively obvious that a goes with mistake to form a constituent, but clearly it will not do to proceed simply on the basis of intuition – we need to find less subjective evidence for our analysis. What kinds of evidence are relevant is a question we shall take up later – one cannot determine the constituent structure without considering other aspects of the syntax and it will therefore be better to proceed to the other main concepts, leaving till the next chapter the issue of how one chooses one analysis rather than another.

1.3 Syntactic classes and functions

The constituent structure analysis of a sentence identifies the forms and their hierarchical arrangement one within another. We must now consider how the various forms are to be further described.

In the first place, they will be assigned to syntactic classes and subclasses on the basis of various types of shared properties. Thus words will be assigned to such primary classes as noun, verb, adjective, etc., and to such subclasses as proper noun, common noun, transitive verb, etc. Forms occurring higher in the constituent hierarchy can be classified in an analogous way as noun phrases, verb phrases, clauses, interrogative clauses, and so on. If we return to our example sentence, The boss must have made a mistake, we can extend the earlier analysis by incorporating a sample of such classificatory information in the following way:
Basic concepts in grammar

The relationship between this and diagram (2) above will be clear: classificatory labels have been attached to the points representing the twelve forms, and lines have been added linking the words with the word-class labels.

Secondly, we shall specify the syntactic function of each constituent. Thus *the boss* functions as subject, *a mistake* as object, and with the phrase *the boss* the word *boss* functions as head, *the* as determiner, and so on. Such functional information can be incorporated into our diagrammatic representation of the structure as shown in (8), where the classificatory labels of (7) have now been abbreviated.
1.3 Syntactic classes and functions

A considerable number of technical terms have been introduced here, and during the course of the book we shall need to explain them. At this stage, however, my aim is merely to make some general points about classes and functions, to give some initial idea of what kinds of concept they are.

Functions are inherently relational – and indeed are often referred to as 'grammatical relations'. Thus when we say that the boss is subject we mean that it stands in the subject relation to the clause: it is subject of the clause. Similarly boss is head of the phrase the boss. Syntactic functions are clearly of crucial importance for the meaning. The meaning of The guard shot the intruder, where the guard functions as subject, the intruder as object, is very different from that of The intruder shot the guard, where the functions are reversed. Or again, the bottom saucepan, where bottom is modifier, saucepan head, means something quite different from the saucepan bottom, where saucepan is modifier, bottom head. The meaning of a sentence depends in part on the meanings of the vocabulary items it contains, in part on their grammatical arrangement: much of the meaning attributable to the grammatical arrangement derives from the syntactic functions.

A syntactic class or subclass, we have said, is a set of forms sharing certain properties. These shared properties are of two main types: one concerns the internal structure of the forms, the other their functional potential.

(a) With internal structure (or simply structure), we are concerned primarily with the way a form is made up of forms lower in the constituent hierarchy. For example, the forms in the garden, with considerable difficulty, behind the man I was talking to all consist of a preposition (in, with, behind) functioning as head, followed by a noun phrase functioning as complement: they are alike in their (internal) structure, and this is the major factor determining their classification as preposition phrases. Conversely, You are very careful and Be very careful differ in their internal structure (they contain different inflectional forms of be, and only the first has a subject) and this difference determines their assignment to distinct classes, declarative clause and imperative clause respectively. In this kind of classification we are considering forms as constructions, and the classificatory labels are thus commonly applied also to construction-types: we speak of Be very careful as an example of the imperative construction-type – or simply of the imperative construction. I also include under the heading of internal structure the analysis of a word into a lexeme plus its inflectional

2 This account of the differences between the two classes is intended only as a rough approximation, but the refinements needed would not affect the point being made here; see 2.1 and 11.3 for fuller discussion.
Basic concepts in grammar

properties, so that took, wrote, saw, etc., are alike in internal structure in that they consist of a lexeme plus the inflectional property past tense.

(b) With functional potential we are concerned with what functions a form can have in what kinds of construction higher in the constituent hierarchy. For example, the boss, my uncle, some of it, etc., are alike in being able to function as subject in clause structure, as in (9), object in extended verb phrase structure, as in (10), complement in preposition phrase structure, as in (11), and so on:

(9) The boss/my uncle/some of it arrived on time
(10) [He] didn’t understand the boss/my uncle/some of it
(11) [She was dissatisfied] with the boss/my uncle/some of it

This similarity of function is then a major factor justifying the classification of all three forms as noun phrases. Conversely, the reason why we assign the and of, say, to different word-classes is that they differ with respect to the functional positions they occupy in the structure of phrases: the functions as determiner in noun phrase structure (the boss), while of functions as head in preposition phrase structure (of the wine) — with noun phrases and preposition phrases likewise distinguished by their functional potential. A standard technique for investigating this kind of syntactic property involves substituting one form for another within a given sentence – a technique employed above to bring out the similarity between the boss, my uncle, and some of it. But it must be emphasised that it cannot be applied in any purely mechanical way. Two points should be borne in mind. First, the fact that we can substitute form A for form B in a given sentence will be relevant only if the syntactic function remains constant. Thus in They left the meeting we could replace the meeting by rather late to give They left rather late, but since the meeting functions here as object while rather late functions as ‘adjunct’, the substitution provides no evidence for a syntactic similarity between the meeting and rather late. The second point is that if substituting A for B results in a sentence judged to be unacceptable, this will provide evidence for a difference in functional potential only if the unacceptability is a matter of grammar. Suppose, for example, that we replace my uncle in (12) by rather sadly on the one hand, and by the saucer on the other:

(12) i My uncle laughed with delight
    ii *Rather sadly laughed with delight
    iii The saucer laughed with delight

Although (iii) might be judged unacceptable, it is not ungrammatical – it is just that the event it describes is one we regard as impossible on the basis of our knowledge of saucers. By contrast, (ii) clearly is ungrammatical, and hence, in accordance with the convention introduced in §1, is prefixed with
1.4 Paradigmatic relations and kernel clauses

an asterisk. Accordingly (12) provides no evidence for a difference in functional potential between my uncle and the saucer, but it does show my uncle and rather sadly to be different – and this kind of evidence is used in classifying my uncle as a noun phrase, rather sadly as an adverb phrase.

As we use functional potential as a major criterion for classification, there will of course be a significant degree of correlation between our syntactic functions (subject, object, modifier, etc.) and our syntactic classes (noun phrase, noun, adjective, adverb, etc.). The correlation is, however, by no means one-to-one. In the first place, a given class may be able to fill two or more different functional positions. This we have already seen to be the case for noun phrases, which can function as subject, object, complement of a preposition – and in a number of other positions too. Similarly, an adjective phrase like very young can function either as modifier in NP structure (as in a very young child) or as ‘predicative complement’ in clause structure (The child was very young). In the second place a given functional position may be able to be filled by members of different classes. For example, although the subject of a clause is usually an NP it can also be – under certain conditions – another clause, as in That he was concealing something was becoming more and more obvious, where that he was concealing something is a (subordinate) clause functioning as subject of the larger (superordinate) clause that forms the whole sentence. Similarly, what we have just introduced as the predicative complement position can be filled by an AdjP, as in He was still very young, or by an NP, as in He was still a child. Or again, the modifier position in NP structure can be filled by an AdjP ([a man] eager for recognition), a PP ([a man] with a grudge) or a ‘relative clause’ ([a man who knew her son). (In discussing such cases I shall follow a convention whereby expressions like ‘AdjP modifier’, consisting of class label and function label, are to be interpreted as “modifier belonging to the class AdjP”, not “modifier of/within the structure of an AdjP”.)

The fact that we find a many-to-many rather than one-to-one correlation between functions and classes is of course a major reason for keeping the two kinds of category conceptually distinct. In certain specific cases the correlation is one-to-one: for example, the predicative is invariably filled by a verb phrase, and a verb phrase always functions as predicative; but for consistency we will maintain the distinction even when this is so – so that we will apply the term ‘verb phrase’ to a class of forms (such as saw, has gone, may have been) and ‘predicative’ to the function of a verb phrase in a larger construction (so that in I saw the film, say, saw is predicative and the film object).

1.4 Paradigmatic relations and kernel clauses

The linguistic relations between forms fall into two fundamentally distinct types: syntagmatic and paradigmatic. Less standard but more transparent terms are respectively ‘and-relations’ and ‘or-relations’.
Basic concepts in grammar

Thus on the one hand forms \( x, y, z, \) etc., stand in syntagmatic (‘and’) relation if \( x \) and \( y \) and \( z \) combine in the structure of some larger form. For example, in the clause John knew Max the forms John, knew and Max are related syntagmatically in that they occur together in the larger structure. On the other hand, we recognise a paradigmatic (‘or’) relation between forms \( x, y, z, \) etc., if one is substitutable for the other, if we have a choice between them. Thus the knew of John knew Max could be replaced by had known, knows, warned, had warned, etc. so that we recognise a paradigmatic relation at the predicator position between knew, had known, knows, warned, had warned, etc. Similarly the whole clause is in paradigmatic relation with Did John know Max?, John didn’t know Max, and so on.

Syntactic functions, such as subject, predicator, object, are established to handle syntagmatic relations: the specific syntagmatic relations holding between the phrases in John knew Max are that John is subject, knew predicator and Max object. In this section, our focus of attention will be on paradigmatic relations. Let us therefore pursue the above example, where we have a paradigmatic relation between knew, had known, etc. Clearly the set of forms in paradigmatic relation with knew is enormously large: it would be an unprofitable undertaking to try to list them directly. We note, however, that the differences between them involve a number of distinct variables or dimensions of contrast:

(a) knew and warned are past forms in contrast to present tense knows and warns;

(b) knew and warned are non-perfect in contrast to perfect had known, had warned;

(c) knew and warned are positive in contrast with negative didn’t know, didn’t warn;

(d) knew, knows, had known and didn’t know all contain as their final or only verb a form of know, whereas warned, warns, had warned, didn’t warn contain a form of warn;

and so on.

Where a given variable has just a small number of possible values, we shall say that we have a grammatical system, with the values being mutually exclusive terms in the system. Thus past and present are terms in one system, (a), perfect and non-perfect are terms in a second system, (b), positive and negative in a third, (c) – but we will not have a system for variable (d) contrasting verb phrases with know and warn (or admire, remember, etc.) for the differences here depend simply on which dictionary item is selected. We will return to the relationship between grammar and dictionary in §8; in the meantime we will confine our discussion of systems