6 First Practice at Sentence Structure and Immediate Constituent Analysis

WHAT WE ARE DOING THIS CHAPTER.

In this chapter we set out a range of ways of examining linguistic data, particularly sentences, in order to fit them into the models of structure developed in the previous chapter. We need to be able to sharpen our tools for revealing syntactic structure, so that we can ourselves, when confronted with any new sentence, work out what its constituent structure must be.

We saw in Chapters 1—3 how there are rather diverse meanings made in texts and in the sentences that make up texts, and these meanings are, somehow, distributed around linguistic units in a way that hearers and readers can, again somehow, recover. We have now also seen in Chapter 4 how different linguistic theories can be treated as maps which set out the territory to be explored. In this chapter, we turn in considerably more detail to the linguistic forms used for capturing meaning. Meanings are complex, and so the structures necessary to carry them can also be complex. We will therefore also use this chapter to begin becoming more sure of our abilities to recognise those parts of sentences that can carry meanings.

Along with the development of the basic parts of speech found in languages begun by the ancient Greeks, then, there has more recently been a similar refinement concerning the kinds of words and sequences that can be grouped together as phrases. This has strengthened our tools for investigating linguistic form beyond all comparison. In particular, we will see a collection of probes and tests for exploring the basic components of sentences. This collection draws primarily on the discovery procedures for grammatical structure worked out by the Bloomfieldian structuralists in their attempts to provide a secure foundation for linguistic description.

These discovery procedures and statements about distribution—i.e., in what context items occur—of elements and their combinations mark the later stages of the progression from simpler reflections about the nature of language structure towards the more systematic application of 'scientific' method. Discovery procedures can be given as a set of tests for interrogating grammatical structure. By applying tests of these kinds, linguists of the 1940s and 50s were able to take apart sentences of any language and to posit the kinds of constituents and phrases that that language possessed, largely independently of any assumptions about what those sentences might mean.

This is an extremely significant result here for two main reasons. First, it can help us decide more effectively how to break up sentences in order to identify the various bits which carry meaning—which is where we left off our practical work at the end of Chapter 3. And second, it can provide a starting point for asking questions about what meanings the various bits carry. As we shall see in a later chapter, structure is there for a reason, so being able to identify the structure

without making guesses about what those reasons are places us in a much more powerful position for revealing what language is doing and how it does it.

6.1   Breaking up a sentence into parts: tests and probes

As we try and find the significant parts of the sentences that make up a text—whether these be the 'first' elements, the elements to do with the main verb, or the Processes, Participants and Circumstances—we can find cases where the sentences are sufficiently complex that we might not be sure just what belongs to what. As we have seen and will see more in the next chapter, this is a very old problem that people— among them linguists—have been trying to understand better ever since paying serious attention to language. There have been very many significant contributions to this area over the centuries, but it is only in the last 50 years or so that the component parts of sentences have really been placed on a firm theoretical footing that allows detailed descriptions of sentences of almost any complexity. One of the results of this work is that there are now a range of reliable tests for taking sentences apart into their 'constituent parts'. To get us started, therefore, we shall simply list some of the more well known tests and then, in the chapter following, we will apply these in the analysis of a longer example.

The probes described here are tests that tell us about grammatical constituents. A grammatical constituent is a grammatical unit that is part of a bigger grammatical unit. Larger constituents are made up of smaller constituents. The largest grammatical constituent that is usually thought of is the sentence, the smallest is the morpheme. The words, whose parts of speech we saw above, are made up out of morphemes. Some words consist of just a single morpheme (e.g., 'but'), others consist of several morphemes (e.g., 'runs' consisting of a morpheme 'run' and a third-person, present tense morpheme '-s'). In this introduction we will not generally be concerned with what happens within words.

Many of the tests are concerned with what you can and cannot do: this means that if you try, for example, to move some part of the sentence that is not a constituent, you will end up with something that is not grammatical English—in linguistics such sentences, or other grammatical units, that are not correct because they violate how

151

English builds sentences, are indicated by placing an asterisk in front of them; for example:

* This sentence grammatical not is.

Note that the existence of these tests is itself further evidence against the view of language as beads on a string. If language were so structured, then many of the probes given for recognising bits of structure would not work: it is only because language has structure that the probes do their job; they are responding to, or making visible, the linguistic structure. As an analogy, one can look at the tests as in some respects similar to the geologist's hammer: when a rock is hit with the hammer, then it breaks along its natural fault lines to show something of its combination. The grammatical probes and tests are like a range of different kind of hammers, each of which capable of making a distinct kind of fault line visible in the grammatical structure of sentences and clauses.

We now set out the tests with some very short examples; in the next chapter we apply them to some real sentences as found in their natural habitat—i.e., in texts.

Types of structure 'probes': subjects

The following are reliable tests for identifying grammatical Subject:

- the Subject and the finite part of the verb **agree in grammatical number**

  the dog chases the cat

  the dogs chase the cat

  Agreement is also sometimes called **concord** and in English can be either grammatical or semantic:

  'the staff is very well trained' *vs.* 'the staff are here to help'

- a **tag question** always picks out the Subject

  the cat chases the mice, doesn't it?

| the dogs |chase the cat, don't they?

- if you make a **passive construction** the Subject is always the one to disappear or to be moved to a 'by phrase'; so again in the sentence 'the dog chased the cat' when we make this into a passive construction:

    the cat is chased by *the dog*

    the cat is chased

we see that it is the Subject (the phrase concerning the dog) that has had some change occur to it.

But beware! What becomes the Subject in a passive construction is *not* always a complete or a simple Participant!! This is particularly the case in clauses that talk about 'mental' or 'verbal' events. For example, in a sentence such as:

    He saw his dog chasing a sheep

we can pick out 'his dog chasing a sheep' as a single constituent and we cannot split this constituent up easily by moving it around (try it with the tests below); this would be a Participant. But we can nevertheless make 'his dog' the Subject of a passive sentence:

    His dog was seen chasing a sheep.

This is a particular property of both textual and interpersonal meanings that we will see more of later in the course. We need to know both about the constituent structure and the grammatical functions and how those functions can be 'distributed' around the constituent structure in sometimes quite complex ways (but never 'any old how'!).

Types of structure 'probes': 'semantic constituents'

These are the parts of a sentence that answer the basic questions:

- **who?**
- **where?**
- **when?**

153

- **why?**
- **how?**
- **to whom?**

There tests are particularly suited to picking out the Participants and Circumstances of a clause.

Permutation tests

Permutation tests are tests where you try and 'move' some part of a sentence around and see what other bits of the sentence want to move too. For example,

> Fred Bloggs, author of 6 novels, wrote many books in New Hampshire.

If we try to find the 'first' element of the sentence, then we can see what can be moved where. Whenever we try and move 'Fred Bloggs' or 'author of 6 books' (or to move something in their way), then the other 'half' wants to move to:

> * Fred Bloggs, in New Hampshire, author of 6 novels, wrote many books.

> * *Author of 6 novels*, many books were written by *Fred Bloggs* in New *Hampshire*

But if we keep them together, then they are happy:

> Many books were written by *Fred Bloggs, author of 6 novels*, in New Hampshire

When we try and move something to the front of the sentence, then what will be moved is typically a full constituent: e.g.,

> *In New Hampshire,* Fred Bloggs, author of 6 novels, wrote many books.

Types of structure 'probes': syntactic constituents

Pseudo-clefts ('wh-cleft')

'Cleft' is a word that means to divide in two, or to divide. Cleft-sentences are then sentences that have been divided into two parts— and in the case of a wh-cleft they have been divided by using a wh-word such as 'what', 'when', etc. Sentences cannot be divided

arbitrarily however, they have natural places where they 'break'. Thus, if you have a sentence:

The boy kicked the ball

then you can pick out the constituents of the sentence using the **pseudo-cleft** construction:

(a) what the boy kicked was **the ball**

(b)(the one) who kicked the ball was **the boy**

Cleft-sentences serve to indicate constituents precisely because sentences have structure and so cannot be divided arbitrarily.

Expansions and substitutions

If you have a sentence:

The king of England opened Parliament

then you can pick out constituents by trying to substitute 'smaller' but equivalent units:

(a) The king opened Parliament

(b)He opened Parliament

(c) He worked

In fact, any kind of substitution is useful: if you can replace a sequence of words by another sequence that you are more sure of, then this can be a useful indication of the kind of linguistic unit in question.

Reduction/Ellipsis test

If you have a sentence:

John won't wash the dishes

then you can pick out constituents by seeing what can be 'left out' or **ellided:**

He will _____ if you ask him

Similarly with the sentence:

John won't help me with my homework...

responding to this lets us 'leave out' even more:

He will ~~help you with your homework~~ if you ask him

that is,

He will if you ask him.

Again, English (and most languages) are rather particular about what they let you leave out and what not, so that we can use what is left out as another sign of being a constituent.

Conjunction/Co-ordination test

If you can replace a unit by that unit and another one of a *similar* kind, then you have a constituent:

The boy chased <u>the dog.</u>

The boy chased *<u>the dog and the cat</u>*.

In contrast to this, the sequence of words 'up his friends' in a sentence 'John rang up his friends' is *not* a constituent. We can see this when we try to form a conjoined phrase:

*John rang <u>up his friend</u> and up his mother.

This does not work because the 'up' here belongs to the phrasal verb 'to ring someone up' and so does not form a constituent with what follows. This is different to when it is used as a regular location Circumstance:

John climbed up the ladder and up the stairs.

Dependency test

If some words cannot be removed from a sentence or other unit **without taking others out with them** then these latter other words are **dependent on** the former and together with them make up of a larger constituent.

(a) the King of England opened Parliament

(b) the King opened Parliament

(c) * of England opened Parliament

This tells us that the part of the sentence 'of England' is *dependent* on the part of the sentence 'the King'—if the latter goes, we are left with nonsense, but if the former goes, we still have a complete and grammatical sentence.

6.2 Phrase structure

The treatment of linguistic structure took a radical turn with the publication in 1957 of Noam Chomky's *Syntactic Structures.* This slim book took on the task of providing a systematic account of linguistic structures and showed clearly that the simple probes and tests that had been developing in the preceding 20 years of Bloomfieldian structure linguistics were not up to the job of describing human language. Chomsky showed this in two steps: but we will concentrate here on only the first of these; the second (which underlies the account of *Transformational Grammar* that came to dominate linguistics and linguistic teaching throughout the 1960s and beyond) has undergone such a wide ranging series of revisions since then that it is barely recognisable in current day linguistics. The first step was the innovation of 'Phrase Structure Grammar'—a scheme for representing explicitly the kinds of structures that phrases and sentences rely on in order to do their job of representing meanings that Chomsky imported from mathematics and the study of 'formal languages'.

When we consider sequences of words, it is clear that they group together into phrases. This was the basis underlying the substitution tests given in the previous chapter.

- he   [pronoun]
- John   [proper name]
- The boy   [determiner noun]
- The good boy   [determiner adjective noun]

Observations such as these can be summarised by giving *phrase structure trees* as shown below.

```
        NP                              NP
       /  \                          /  |  \
determiner  noun          determiner adjective noun
    |         |               |        |       |
   the       boy             the      good     boy
```

In such trees, each part of the tree is called a *node*. The top of the tree is called the *root node.* Nodes that have other nodes below them are said to *dominate* those lower nodes. The higher node is called the *parent*, or *mother* node, and those below are called the *child,* or *daughter*, nodes. A node that is immediately below another node is said to be *immediately* dominated by the parent node. In these examples, all of the children nodes (for determiners, adjective and nouns) are immediately dominated by the parent node. Finally, whereas the labels of the child nodes are drawn from the familiar word classes that we have seen above, the parent node is a new kind of label, a phrase label, in this case representing a *Noun Phrase,* or *NP* for short. Phrase structure trees let us explicitly group together those parts of a sentence, or other linguistic unit, that belong together.

Part of the value of phrase structure trees is that they make it clear how language *re-uses* certain patterns again and again. This means both that a language is easier to learn and that languages provide a force that generalizes meanings. Particular kinds of meanings are re-used in different situations, thereby providing a way of saying that situations are similar and different in certain respects. Thus, in the following sentence:

- The gnome saw the garden.

We have not one noun phrase but two: both 'the gnome' and 'the garden' have the same kind of phrase structure tree. English re-uses the noun phrase pattern when it (or rather, one of its speakers) wants to describe objects such as gnomes and gardens. Both phrases are said to be *embedded* within the sentence as a whole. We can write the structure of that sentence as something like that shown below.
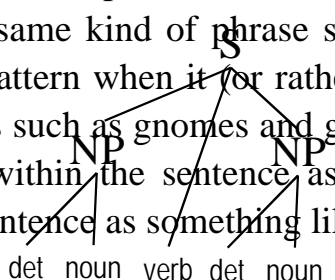
Here the entire tree is dominated by the root node 'S', standing for Sentence. This node immediately dominates three children: the two NPs and the verb. This makes it clear that we are not dealing with a simple chain consisting of:

<div align="center">determiner noun verb determiner noun</div>

but with a structure. We cannot move determiners or nouns around at will; if we move anything, then we typically must take an entire

phrase: and the phrases correspond to nodes in the tree. Thus 'the garden, the gnome saw' is a reasonable sentence (if somewhat limited in possible applications) , whereas 'the saw gnome garden the' is so-called **word salad**: it has destroyed the structural relationships and, with them, any chance of being meaningful.

We can also indicate phrases and their boundaries by using brackets which group together those bits of the sentence that belong together in phrases and exclude those belonging to other phrases:

<div align="center">((the gnome) saw (the garden))</div>

This shows grouping, but does not show the kinds of phrases involved. A way of writing the information that is completely equivalent to the tree is then to use *labelled brackets* as follows.

<div align="center">($_S$($_{NP}$ the gnome) saw ($_{NP}$ the garden))</div>

This shows both the grouping and the syntactic labels. From any tree we can write a sequence of words or word classes using labelled brackets; and from any sequence using labelled brackets we can write a tree. The two forms are interchangeable; which is used depends on how clearly the form selected shows the grouping that we want to talk about. Sometimes labelled brackets are enough; sometimes it is more useful to see the entire tree set out graphically.

There are several distinct kinds of phrases. The following is an **Adverbial Phrase** (or AdvP):

- very quickly indeed

while the following is an example of a **Prepositional Phrase**, or *PP* for short:

- in the garden

Note that if we were to **represent** this latter as a simple sequence of word classes such as:

<div align="center">preposition determiner noun</div>

then we would be missing the fact that we have seen some of this kind of structure before: it is not an accident that we again have the sequence 'determiner noun'. This type of pattern occurs both here and in the examples above; this, as we have seen, is a noun phrase. Rather than miss this detail, we can usefully describe prepositional phrases not as this simple sequence but instead in terms of the tree:
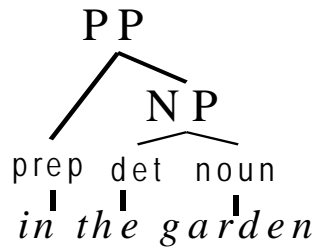
or, alternatively, as the labelled bracket expression:

($_{PP}$ preposition ($_{NP}$ determiner noun))

A prepositional phrase is therefore made up of a preposition followed by a noun phrase: or, in terms of our tree, a PP immediately dominates a preposition followed by an NP. This representation captures the fact that it is not an accident that after the preposition we can put any possible noun phrase in English, not just a sequence of determiner followed by a noun: we can substitute any NP. The tree makes the substitutions that are possible at this point in structure explicit: in fact, we can say that it is because English (and many other languages) structures a prepositional phrase like this that the substitution tests we saw above involving prepositional phases work at all.
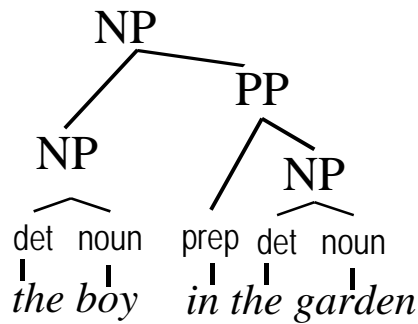
Now consider the following example:

- The boy in the garden

Again, were we to represent this just as a sequence of word classes:

determiner noun preposition determiner noun

we would miss much of the structure that is involved and end up doing more work than is necessary. This chain representation misses the fact that the first two words of the phrase and the last two words have something in common: they are both very similar to noun phrases. It also misses the fact that the phrase as a whole—the boy in the garden—can occur everywhere that a simple noun phrase can occur: that is because it *is* a kind of noun phrase. So one way of capturing these observations would again be in terms of a phrase structure; one possible phrase structure tree would be:

```
            NP
           /    \
          /      PP
         NP     /   \
        / \    /     NP
     det noun prep  / \
      |   |    |  det noun
      |   |    |   |   |
     the boy  in the garden
```

Now, this quite complicated structure shows no less than three NPs participating; whether we actually decide that such a structure is the most revealing for this phrase or not, the structure shown is one possible treatment. We will return to such issues below, where we ask just what *kinds* of structures do we want to pursue in our linguistic descriptions. This question can only be asked sensibly when we are more clear about just what 'work' the linguistic structure is meant to be doing for us. At present, the main work being asked of such structures is that they show us the constituency structure and that they group the sequences of words that we see or hear according to their natural phrases. From this perspective, a structure such as this one is certainly not a bad attempt.

This last structure also shows a further aspect of language that is absolutely crucial for how language works. It includes phrases that contain other phrases of the same kind within them: that is, the topmost NP includes other NPs within it. When we have structures of this kind, we have examples of a special kind of embedding called **recursion.** We will see that recursion is one of the most important features of linguistic structures and that without recursion human language would not be possible. This is a very long way indeed from the notion of language made up of chains of words: so much so that it often takes considerable practise to become comfortable with it. But the effort is worth it; once structure has been understood, many properties of language become easier to grasp and use.

We can now apply again the probes and tests for structure that we saw in above in order to refine our view of structure. Examining what these probes tell us, and exploring a wider range of sentences as they occur in texts, we


```
            S
          /  \
        NP  /  NP
       / | / | / \
    det noun verb det noun
```

quickly find the need for a more richly organised view of sentences
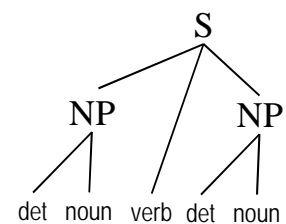
and their constituents than we have so far seen illustrated. Consider again, for example, our tree for a simple sentence given above and repeated here on the right for convenience. In fact, this tree is still giving too much of a simple chain view of language: in this case a chain consisting of a noun phrase followed by a verb followed by another noun phrase. If we apply our conjunction tests asking what parts of the sentence can be combined with 'and' so as to form bigger units of the 'same kind', then we get a several structures. We certainly get all of the constituents that we can see in the tree: e.g.,

The gnome saw the garden : The gnome and the dwarf saw the garden

The gnome saw the garden :
The gnome saw the garden and the mountain.

The gnome saw the garden : The gnome saw and loved the garden.

But we also have sentences such as the following:

The gnome saw the garden and ran to it.

The gnome saw the garden and waited for sunrise.

According to our probes this should then mean that there is another constituent present in the sentence, a constituent identified as the sequence of elements underlined in the following:

The gnome saw the garden

We can also find evidence supporting this from the substitution probes. If we examine what constituents can be substituted for in this sentence we find combinations such as:

The dwarf saw the garden and the gnome did too.

Here the 'did' appears to be substituting just for the portion underlined above: i.e., 'saw the garden'. All of the tests in fact appear to provide evidence that there is another constituent, and that a more complete structure for the sentence would be not as given above but instead one that can be shown as:
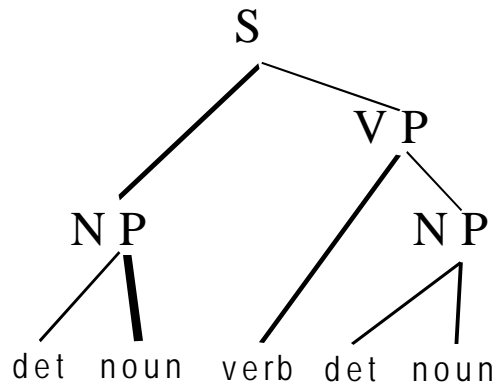
This additional node, the VP, is the **verb phrase**. We can see with this structure a return to the kind of basic division between Subject and Predicate that was developed by the ancient Greeks. Indeed, some linguists have always assumed that this division is basic and obvious; Bloomfield, for example, wrote:

> "Any English speaking person ... is sure to tell us that the immediate constituents of *Poor John ran away* are the two forms *Poor John* and *ran away*; that each of these is, in turn, a complex form." (*Language*, 1933, p161)

Whether we believe this or not probably depends on how much linguistic 'indoctrination' we have already received! But fortunately, we do not just have to believe it, i.e., take it on faith, we can instead, by following the probes and tests above, arrive at a similar conclusion ourselves—without resorting to 'belief'. There is ample evidence that there is something that is acting like a constituent in the place that we have now placed a VP node in our syntax tree. This then allows us to produce sentence structures such the following, which are necessary to cover the cases of conjoining (by means of some conjunction such as 'and') shown above:
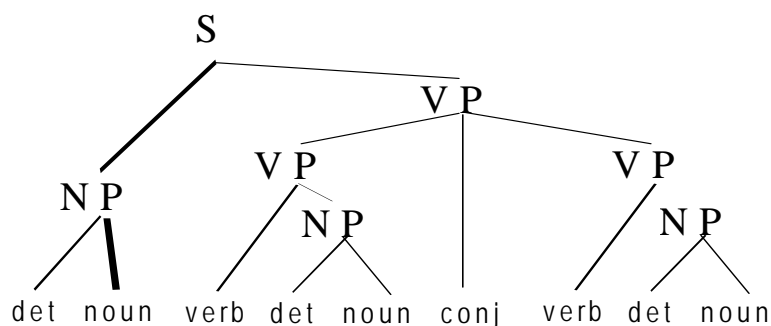
We can take this line of argument considerably further—and linguists have. In fact, there is evidence for *considerably* more structure than we have seen here; but we will leave it at that for now.

Reading and references

The tests and probes talked about here are mostly compiled from

Wagner, Karl-Heinz (and Susanne Hackmack) *Grundkurs-Skript Sprachwissenschaft,* Uni Bremen. 1996.

Wells, R. S, (1947). Immediate constituents. *Language*, 23, 81--117.

Glinz, H., (1965). *Grundbegriffe und Methoden inhaltbezogener Text- und Sprachanalyse*.

Chomsky, N. (1957) *Syntactic Structures.* Mouton.

Vater, *Einführung in die Linguistik.* (also taken from Wells in any case I think).