How do we identify constituents?

This chapter returns to the theme of sentence structure, introduced in Chapter 1. We saw in Chapter 4 that phrases consist of a head word and its complements, plus any optional modifiers to that head. In this chapter we discover how to identify phrases, and how to distinguish a phrase from a random string of words. The phrases which make up sentences are known as the CONSTITUENTS of a sentence. We will see how constituents are represented in tree diagrams, and start to investigate how languages differ in terms of constituency.

5.1 DISCOVERING THE STRUCTURE OF SENTENCES

Section 5.1.1 demonstrates the existence of syntactic structure, in particular by looking at ambiguous phrases and sentences. Section 5.1.2 introduces three syntactic tests for constituent structure, and Section 5.1.3 examines the ways in which linguists formally represent constituent structure.

5.1.1 Evidence of structure in sentences

One way to show that syntactic structure actually exists is to examine sentences which are syntactically ambiguous; that is, sentences which have more than one meaning. Not all ambiguity is syntactic: some is lexical, as in *Lee went down to the bank*; does this mean 'the river bank' or 'the place where money is kept'? In other cases, though, ambiguity arises because we can't tell which words group together to form a phrase. This is syntactic ambiguity. For instance, a sentence like the following appeared in a British national newspaper, causing an unforeseen breakdown in communication.

(1) Black cab drivers went on strike yesterday.

Readers wrote in to say, what did it matter what colour the drivers were? But, of course, the newspaper actually intended *black* to modify *cab*, not to modify *cab drivers*. The two different meanings reflect the fact that the phrase *black cab drivers* has two different STRUCTURES. We can indicate this by using brackets to show which words group together; different bracketings indicate different phrase structures. Example (2) illustrates these different structures: the outraged readers had

interpreted the sentence as in (2a), and the newspaper had intended (2b). (For readers unfamiliar with British culture, a 'black cab' is a particular kind of black taxi found in major cities.)

(2)   a.   [Black [**cab drivers**]] went on strike.
          b.   [[**Black cab**] drivers] went on strike.

In both (2a) and (2b), the whole phrase *black cab drivers* is a constituent of the sentence, hence is in brackets, but the differing internal brackets show that the words inside that phrase group together in different ways, depending on what *black cab drivers* actually means. In (2a), *cab drivers* forms a constituent, whereas in (2b), *black cab* forms a constituent. A constituent is a set of words that forms a phrase in a sentence. If you say aloud the distinct phrases in (2), you'll probably find that they each have a different intonation pattern; sometimes we show by our intonation which words group together to form constituents.

Occasionally we can discover which words form constituents by looking at inflections, as in the case of English possessive *-'s* (see Section 1.3.3). The affix *-'s* attaches to the end of a phrasal constituent (an NP) giving *Lee's*, *the boy's* and so on, so we can use *-'s* to discover whether or not a string of words is an NP (and therefore a constituent). This gives some results that might initially seem surprising, as in (3):

(3)   I'll be back in **[an hour or so]'s** time.

Here, *an hour or so* must be a constituent, an NP, since *-'s* can attach to the whole phrase.

The *-'s* inflection can itself be the cause of syntactic ambiguity, because we can't always tell what constituent it's attached to:

(4)   The boy and the girl's uncle stayed to dinner.

This, of course, is ambiguous as to whether just one person stayed, or two, as the variants with tag questions make clear:

(5)   a.   The boy and the girl's uncle stayed to dinner, didn't he?
          b.   The boy and the girl's uncle stayed to dinner, didn't they?

So in (4) there are two different meanings – or READINGS, to use the technical term – and, as we will see, each of these readings corresponds to a particular CONSTITUENT STRUCTURE, that is, a particular grouping of words.

The ambiguity in (4) lies in the phrase *the boy and the girl's uncle*. This whole string of words is a constituent of the sentence in both readings, but its internal structure is different in each case. We can't tell if *-'s* is suffixed to an NP *the boy and the girl*, in which case the uncle is related to both of them, or if *-'s* is just suffixed to an NP *the girl*, in which case the uncle is related to her, but not to the boy. Both options are possible, hence the ambiguity. The structures of the two alternatives are

shown in (6), where the brackets mark out the two possible constituents that *-'s* can attach to:

(6)   a.   [**The boy and the girl**]'s uncle stayed.   (one person)
        b.   The boy and [**the girl**]'s uncle stayed.   (two people)

Only in (6a) does the sequence *the boy and the girl* form a whole phrase, a constituent. This tells us that a sequence of words which forms a constituent in one environment need not necessarily do so in another environment. There is absolutely *no* rule of 'once a constituent, always a constituent'. To underline this point, compare the (a) and (b) sentences here:

(7)   a.   The students wondered how cheap textbooks could be obtained.
        b.   The students wondered how cheap textbooks could be.
        a.'   The students wondered how [**cheap textbooks**] could be obtained.
        b.'   The students wondered [**how cheap**] textbooks could be.

In (7a), there's a constituent *cheap textbooks*, as we can tell by the fact that we can refer to this phrase by the single word *they*: *The students wondered how **they** could be obtained*. The relevant structure is shown in (7a'). But *cheap textbooks* isn't a constituent in (7b). Instead, *how cheap* forms a phrase in (7b), as you can see from (7b'). Here, *textbooks* is a separate constituent, which can again be replaced by *they*: *The students wondered how cheap **they** could be*. The examples in (7) show that we can't look at a string of words out of context and decide whether or not they form a constituent. We can only find this out when the string of words appears in a sentence, and when we can manipulate the sentence in various ways to discover its constituent structure. This requires a set of tests for constituency, like the pronoun test we used here: a pronoun such as *they* replaces a whole NP constituent.

5.1.2   Some syntactic tests for constituent structure

We have used the possessive *-'s* suffix – which only attaches to NP constituents – as a morphological test for constituency. But to discover all the constituents of a sentence (and not just NPs) we also need syntactic tests. One syntactic test is seen at the end of the previous section: a constituent can often be replaced by a pronoun, but a random string of words cannot. We now go on to examine more syntactic tests.

5.1.2.1   The sentence fragment test

The first test in this section utilizes shortened answers to questions. If I ask *Who went on strike?*, a reasonable answer is *Black cab drivers*. Answers like these which are not full sentences are called SENTENCE FRAGMENTS, and they provide syntactic evidence about which words group together to form a constituent. A string of words that can be a sentence fragment must be a constituent. So here, *black cab drivers* is

confirmed as a constituent of (1). Of course, it is still ambiguous, as its internal structure is not revealed. And if I ask *Who stayed to dinner?*, the answer is *The boy and the girl's uncle*, so this whole phrase is a constituent of (4), whichever internal structure it has. Both of these particular sentence fragments remain ambiguous, because there is additional constituent structure inside each phrase which differs, depending on the interpretation.

However, the sentence fragment test can often be used to discover more about internal structure. On hearing (4), someone might try to resolve the confusion by asking *But **whose uncle** stayed to dinner?*. A typical answer would be either (8a) or (8b), depending on which reading of the sentence you have in mind:

(8) a.   The boy and the girl's.   (one person stays to dinner)
     b.   The girl's.   (two people stay to dinner)

In (8), the sentence fragment test confirms what we already discovered from (6): the whole sequence *the boy and the girl* is a constituent in the (a) reading, but in the (b) reading, *the girl* doesn't form a constituent with *the boy*. The fact that *-'s* can be attached to either possible sequence in (8) confirms that they are both able to be used as constituents.

We can also use the sentence fragment test for constituent structure to show that in (6b), the sequence *the girl's uncle* is a constituent.[1] Keep in mind the reading where two people stay to dinner. If you didn't hear the speaker too clearly, you might ask *The boy and **who** stayed to dinner?*. The answer is the sentence fragment *The girl's uncle*: this must therefore be a constituent. So we can bracket this phrase too, adding more information about the structure of (6b):

(9)   The boy and [[**the girl**]**'s uncle**] stayed.   (two people)

As (9) shows, constituents are in turn built up of smaller constituents. Thus we confirm what we already saw in Chapter 4, namely that phrases contain smaller phrases, with each phrase having its own head and dependent elements. In (9), *uncle* is the head of the phrase *the girl's uncle*, since this phrase is 'about' the uncle.

The sentence fragment test is one of the formal tests for constituent structure. Using such tests, we can discover whether two apparently similar sentences in fact have different structures. Consider the examples in (10) and (11): both contain words of exactly the same syntactic categories or word classes, and in just the same order, as (12) shows (to remind you, D is the category 'determiner').

(10) Kim wrote   that book with the blue cover.

(11) Kim bought   that book with her first wages.

(12) N   V   D   N   P   D   A   N

We might assume, then, that these sentences share a syntactic structure. However, native speakers feel instinctively that (10) and (11) are different; the sentences tend, for instance, to have a different intonation pattern. In (10), *with the blue cover* is a phrase (a PP) modifying the head noun *book* – *a book with a blue cover* is a type of book. So this PP belongs with *that book*, forming a constituent with it in (10). But in (11), the PP *with her first wages* tells us how she bought the book, and not anything about the book itself. So that PP modifies *bought* – it is an adjunct to *bought*. Crucially, the PP in (11) doesn't modify the noun *book*, and so doesn't form a constituent with it.

The sentence fragment test for constituent structure supports these intuitive feelings. In each case, when we ask a question, we get different sentence fragment answers:

(13) What did she write? [**That book with the blue cover**]

(14) What did she buy?   a.   [**That book**]
What did she buy?   b.   *That book with her first wages.

Example (13) shows that the entire sequence *that book with the blue cover* is a constituent of (10): it can be a sentence fragment. Example (14a) confirms that the sequence *that book* is a constituent of (11). And crucially, (14b) shows that the sequence *that book with her first wages* is not a constituent of (11): it can't be a sentence fragment (remember that the asterisk indicates an ungrammatical example). Contrasting grammatical and ungrammatical examples, as we have done here, is essential: you should use the ungrammatical examples to show that some sequence of words is *not* a constituent of the sentence you are working on.

- Please remember from now on that in the sentence fragment test, the question you ask should always be a *grammatical* one: the test is the answer itself.

- If a string of words from the original sentence can form a grammatical sentence fragment, it is likely that this sequence is a constituent of the original sentence.

- If the string of words is not grammatical as a sentence fragment, it most likely is not a constituent of the original sentence, though you need additional tests to confirm this.

- Square brackets are used to show where a constituent begins and ends. Please do not put brackets round a phrase *unless* it is a constituent. You may find it useful to use a wavy underline for a string of words which is *not* a constituent, as I have done in (14b).

5.1.2.2 *The echo question test*

ECHO QUESTIONS are our next test for constituent structure. These questions are used in English when the speaker doesn't hear part of the sentence, or else is rather incredulous: e.g. *You saw **what**?*. We use a *wh*-word (*what, which, who, when, why*

and so on, and including *how*) or a *wh*-phrase (*You saw **which film?***) to replace just the part of the sentence that we want repeated, otherwise 'echoing' the speaker's words. The *wh*-word or phrase doesn't replace a random string of words, but can only stand for a constituent of the sentence:

(15)   *Kim wrote **what** with the blue cover?

(16)   Kim bought **what** with her first wages?

The grammatical echo question in (16) is fine because *what* replaces *that book*, which is a constituent of (11). But (15) is ungrammatical because *that book* is only part of a larger constituent *that book with the blue cover* in (10); crucially, *that book* itself is *not* a constituent in this case, and so can't be replaced by a *wh*-word. (We can, however, echo just a head noun on its own: for example, in (10) we can replace *book* with a *wh*-word, giving *Kim lost that **what** with the blue cover?*. The reason for this is that single words are also constituents.)

In fact, we've already used this test earlier: the echo question, *The boy and **who** stayed to dinner?*, only works when *who* replaces a constituent, such as *the girl's uncle*. So it's the question we'd ask if we were sure that two people stayed to dinner, and that one was the boy, but we weren't sure who the other person was. To summarize: in the echo question test, a *wh*-word or phrase can replace a constituent; if the resulting question is ungrammatical, though, the string of words which you've replaced is probably not a constituent.

5.1.2.3 *The cleft test*

A further test for constituent structure confirms our findings: the two sentences in (10) and (11) have different structures. In the CLEFT construction illustrated in (17), the string of words in the 'focus' position must be a constituent. So in (17), we can focus on the whole sequence *that book with the blue cover*, showing that this is a constituent:

(17)   It was [**that book with the blue cover**] that Kim wrote.

But in (18), the sequence *that book with her first wages* is not a constituent, and so can't occur in the focus position of a cleft sentence. This confirms what we saw in (14b):

(18)   *It was **that book with her first wages** that Kim bought.

Remember that we only bracket a string of words which is a constituent, so we bracket *that book with the blue cover* in (17), but not *that book with her first wages* in (18).

Our original sentence in (11) does, however, contain other word sequences which will fit into the focus position of a cleft sentence. For instance, we can focus on either

that book or *with her first wages*, showing that both these phrases are separate constituents of (11):

(19)   a.   It was [**that book**] that Kim bought with her first wages.
        b.   It was [**with her first wages**] that Kim bought that book.

Putting the information from all three tests together, we can show what we've learnt so far about the constituent structure of (10) and (11) by using brackets, as follows:

(20)   [Kim wrote [that book with the blue cover]].

(21)   [Kim bought [that book] [with her first wages]].

The whole sentence is also in brackets in each case, since both examples occur as independent sentences, and are therefore constituents – if sentence fragments are constituents, then it's not surprising that whole sentences are also constituents. Although there are other constituents in each sentence, the brackets in (20) and (21) show as much information as we have up to now.

   Of course, for our constituency tests to have real significance, we must be able to apply them to languages other than English, although not all tests apply equally well in all languages, because certain syntactic constructions may be absent. Cleft constructions occur widely. In the Irish examples below, (22a) shows the basic word order, and (22b) is a cleft construction with the noun phrase *an fear* 'the man' in the focus position:

(22)   a.   Bhí   an   fear   ag   péinteáil   cathaoir.   (Irish)
            was   the   man   PROG   paint.INFIN   chair
            'The man was painting a chair.'
        b.   Is   é   [**an   fear**]   a   bhí   ag   péinteáil   cathaoir.
            is   it   the   man   who   was   PROG   paint.INFIN   chair
            'It's *the man* who was painting a chair.'

Similarly in the next examples, from Lekeitio Basque, (23a) has basic word order, whilst (23b) is a cleft construction, with focus on the fronted noun phrase *orreri mutillari* 'that boy' (the DATIVE case marking on this NP does the work of the preposition 'to' in English, showing the boy as the recipient):

(23)   a.   Premižúa   orreri   mutillari   emon-dótze.   (Basque)
            prize   that.DATIVE   boy.DATIVE   give-AUX
            'They have given the prize to that boy.'
        b.   [**Orreri   mutillari**]   da   premižúa   emón   dotzé-na.
            that.DATIVE   boy.DATIVE   is   prize   give   AUX-that
            'It's to *that boy* that they have given the prize.'

From (23b), we can tell that *orreri mutillari* is a constituent of (23a).

5.1.2.4 *Displacement and dependency*

The constructions in Section 5.1.2 all illustrate an important property of human language: the ability to DISPLACE or MOVE a phrase from its basic position. The hallmark of such displacement is that a phrase is understood semantically as if it were in one position in the clause, but occurs physically (syntactically and audibly) in a different position in the clause. We can illustrate using the cleft examples seen earlier:

(24)   It was [$_{\text{NP}}$ that book] that Kim bought ____ with her first wages.

(25)   It was [$_{\text{PP}}$ with her first wages] that Kim bought that book ____ .

The gap in these examples shows the 'original' position of the displaced phrases. In other words, when you hear an example like (24), you understand it as if the displaced object NP *that book* were still in its normal linear position, following the verb *bought*. Importantly, the verb *bought* only has one direct object, and that syntactic fact does not change just because the object is displaced from its basic position in the usual constituent order. The same applies to (25): the displaced PP is understood as if it were in the typical adjunct position, following the direct object.

The displacement of a phrase sets up a DEPENDENCY between the displaced phrase and the 'empty' position associated with it: the displaced element provides the physical words we need, but its basic position specifies its syntactic role, for instance the role of 'direct object' in (24). The displaced element and the associated gap are of course one and the same entity – moving the object *that book* does not alter the argument structure of the verb *bought*.

It is likely that all languages have instances of displacement of one kind or another. We will see other examples as we go along.
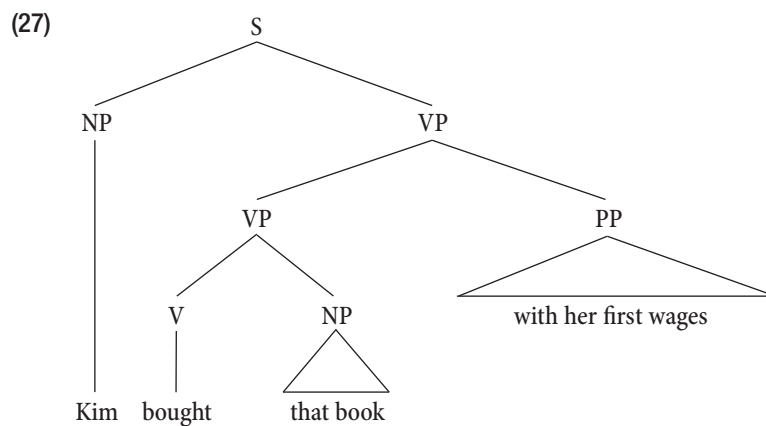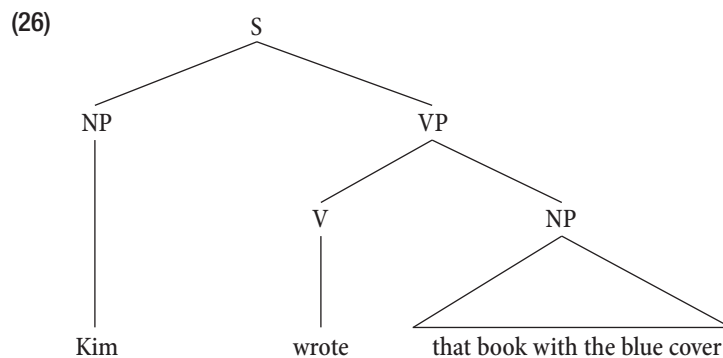
5.1.2.5 *Summary*

Each of the tests for constituent structure in Section 5.1.2 works by harnessing the intuitions of native speakers of a language. The fact that speakers share GRAMMATICALITY JUDGEMENTS – intuitions about which sentences are possible and which aren't – shows that we have an unconscious knowledge of the word groupings in a sentence. The tests for constituent structure are just particular syntactic environments which can only be filled by constituents. Whenever we put a string of words that isn't a constituent into one of these environments, the result sounds impossible to native speakers. This UNGRAMMATICALITY (the technical term for such results) tells us that in such cases, the string of words isn't a constituent.

So far in Section 5.1, I have introduced these syntactic tests for constituent structure: (1) replacement by a pronoun; (2) sentence fragments; (3) echo questions; and (4) cleft sentences. The information about constituent structure which results from our tests can be represented by using square brackets to mark off the constituents, as I have done so far, or alternatively by using tree diagrams. We turn next to this topic.

5.1.3   Introduction to constituent structure trees

In (26) and (27) I represent the structure of our two sentences in (10) and (11) by using TREE DIAGRAMS. As you can see, these are upside-down trees, with the root at the top, and branches descending from that root. The root of the tree is labelled 'S' for 'Sentence', and the clause is divided into two main branches, the subject and the predicate, as discussed in Section 3.1.1.

(26)

```
                       S
            ┌──────────┴──────────┐
           NP                     VP
            │           ┌─────────┴─────────┐
            │           V                   NP
            │           │              ╱──────────╲
          Kim         wrote      that book with the blue cover
```

(27)

```
                      S
           ┌──────────┴──────────┐
          NP                     VP
           │           ┌─────────┴─────────┐
           │          VP                   PP
           │      ┌────┴────┐         ╱──────────╲
           │      V        NP        with her first wages
           │      │       ╱──╲
         Kim   bought   that book
```

Let's now go through these tree structures. In each case there is a major split between the subject NP on the left branch and the predicate VP on the right branch. Both trees have a subject, *Kim*. The difference in structure in our two sentences lies within the VP, as we discovered from our tests on constituent structure, and this difference is reflected in the form of the two trees.

   In (26), I have suggested that the verb and its object together form a verb phrase (VP). The VP has two branches, V (for verb) containing just the transitive verb *wrote*, and NP, the noun phrase which is the complement of *wrote*. This sequence, *that book with the blue cover*, is shown as a triangle, which indicates that the whole

sequence forms a constituent. That doesn't mean that there is no more internal structure within that NP, just that so far, this is all we've discovered.

In (27), we again have a VP consisting of the transitive verb *bought* plus its complement, the object NP *that book*. However, we also have an ADJUNCT here, namely the PP *with her first wages*. Recall from Chapter 4 that an adjunct is a constituent which is syntactically optional, in other words not required in order to make the sentence grammatical: adjuncts are not arguments of the verb, and are therefore non-essential constituents. The structure which I've suggested for (27) reflects this by showing that if we add an adjunct to the VP, we don't get a different *kind* of phrase – it's still a verb phrase, but just one that contains more information. The structure is RECURSIVE, in that it has a VP within a larger VP.

In representing VPs in each tree, I have actually shown more structure than I gave in the brackets for each sentence in (20) and (21) – those examples did not include a set of brackets round the verb and its dependents. So we ought to check that the VP really is a constituent in each case. We can do this by using a different test for constituency: the *do so* test. A VP can be replaced by *do so* (or *did so* in the past tense) as follows:

(28)   I thought that Mel [_{VP} wrote that book with the blue cover].
        No! Kim [_{VP} *did so*].

(29)   I thought that Mel [_{VP} bought that book with her first wages]
   No! Kim [_{VP} *did so*].

The *do so* test works by replacing the entire VP with something that stands for it, and it only works if the sequence being replaced really is a constituent. In (29), I've replaced the whole larger VP *bought that book with her first wages* with *did so*. But note that (27) also contains a smaller VP, *bought that book*. If our test is to have any validity, this should also be replaceable by *did so*. And indeed it is:

(30)   I thought that Mel [_{VP} bought that book] (with some of her inheritance).
        No! Kim [_{VP} *did so*] with her first wages.

We can also use the *do so* test to confirm that the sequence *wrote that book* on its own does not form a VP constituent in (26). Once again, the contrast in grammaticality demonstrates the difference in structure between the two examples:

(31)   I thought that Mel wrote that book with the blue cover.
        No! *Kim **did so** with the blue cover.

The reason that *wrote that book* does not act like a VP here is because the sequence *that book* is itself not a constituent in this case, but rather is merely part of the larger NP *that book with the blue cover*, as we saw in Section 5.1.2. This whole NP is the object of *wrote*, so we can't take part of it and leave the rest behind. The underlining should help you see that *wrote that book* is not a constituent here:

(32)
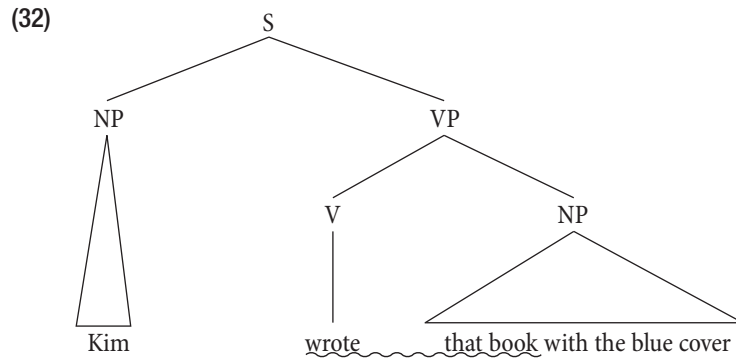
Let's look at two more examples which again contain words from exactly the same word classes and in the same order, but which again have different structures:

(33) My   brother   wrote   down   his   address.

(34) My   brother   applied   for   this   job.

(35) D   N   V   P   D   N

Tests for constituency show that (33) and (34) don't share the same syntactic structure, as you now have the chance to discover for yourself.

> ————————→ ————→ ————————→

If we want to know whether, say, the sequences *down his address / for this job* form a constituent in each case, we can try putting each sequence into the focus position of a cleft construction. Please do this before reading further, and decide what the results show. Your cleft sentences will begin '*It was …*'. Bracket the constituent that you discover.

←———————— ←———— ←————————

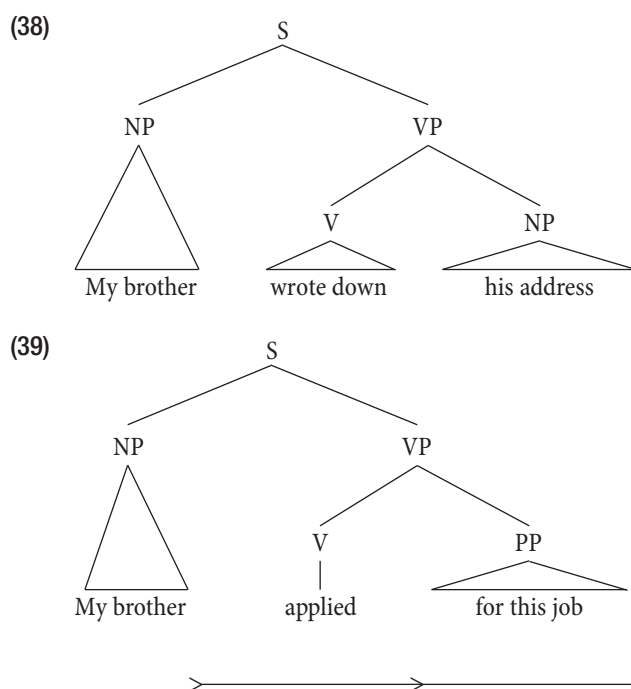The cleft constructions should be:

(36) *It was **down his address** that my brother wrote.

(37) It was [**for this job**] that my brother applied.

Only (37) is grammatical: example (36) is impossible, which indicates that *down his address* is not a constituent in (33), and therefore it can't be placed in the focus position of a cleft sentence. In (34), *for this job* is a constituent, shown by the fact

that it can be focussed in a cleft sentence. Once again, I remind you that we only bracket a string of words that is a constituent.

The tree diagrams in (38) and (39) show the structures of the two examples. The tree in (38) shows *wrote down* as a phrase, something I return to in Section 5.3.1. Note that in (38), *down* is not shown as part of the same phrase as *his address*, because we have proved by using the cleft construction that *down* doesn't form a constituent with *his address*. In (39), though, the preposition *for* forms a constituent with *this job*, as we have shown by the cleft test in (37). I should emphasize that a tree diagram simply illustrates the existence of constituents which we have already discovered by using our tests for constituent structure, and in turn, these harness our intuitions as native speakers of English.

(38)

```
                  S
         ┌────────┴────────┐
        NP                 VP
        △          ┌───────┴───────┐
    My brother      V              NP
                    △              △
                wrote down     his address
```

(39)

```
                  S
         ┌────────┴────────┐
        NP                 VP
        △          ┌───────┴───────┐
    My brother      V              PP
                    │              △
                 applied       for this job
```

The tree in (38) also shows the sequence *his address* as a constituent, although we haven't yet seen any evidence for this claim. Basing your answer on the tests for constituency that we've used so far, what evidence is there that *his address* in (38) is indeed a constituent?

First, we can use the echo question test, where the *wh*-word *what* replaces the string *his address*, as in *He wrote down **what**?*. Second, the question *What did he write down?* can also be answered with *His address* as a sentence fragment, confirming that it's a constituent. And finally, we can use the cleft test:

(40)   It was [**his address**] that my brother wrote down (not his phone number).

I will leave you to apply the same tests to show that the sequence *my brother* is also a constituent of these sentences.

Tree diagrams can be drawn to show very detailed information about the syntactic structure of a phrase or sentence, or alternatively, some of the finer details can be omitted. Linguists choose to put more or less detail into their trees depending on what information they want to convey. So, for example, the tree in (39) indicates that *for this job* is a constituent, but it doesn't show whether there are any smaller phrases within this constituent. In fact, there are. The cleft test shows that the string of words *this job* is also a constituent in this example: *It was [**this job**] that my brother applied for*. Now that we know that *this job* is a constituent in this case, we can draw a more detailed tree to represent this: (41) gives more information about the structure than (39) does.

(41)

```
                          S
                 ┌────────┴────────┐
                NP                 VP
                △          ┌───────┴───────┐
                           V               PP
                                      ┌─────┴─────┐
                                      P           NP
                                      │           △
            My brother            applied        for     this job
```
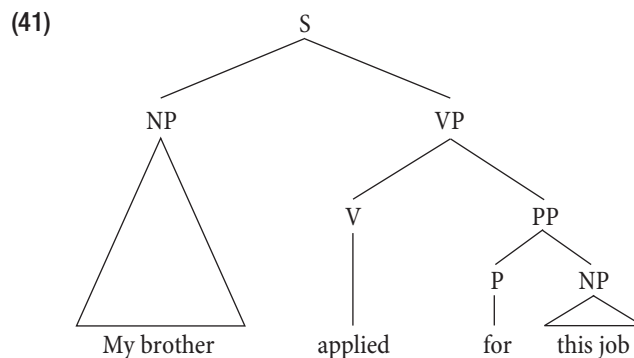
Tree (41) isn't a replacement for (39): it simply gives more information. Both trees would be used by linguists, depending on the level of detail we want to indicate. Tree (41) shows that *this job* is an NP constituent, nested inside a larger constituent *for this job*. If we are using brackets, one constituent is nested inside the other as follows: [*for* [*this job*]].

So, to summarize, exactly how much or how little structure we actually show within the tree diagram or the brackets depends on what we are trying to show. But if we are claiming that two sentences contrast in constituent structure, then the parts that differ must be shown in enough detail to make our claims clear.

5.1.4   Summary

We have so far used the following syntactic tests for constituency: replacement by a pronoun, the echo question test, the sentence fragment test, the cleft test, and the *do so* test for VP status. We indicate which strings of words are constituents of a sentence in two ways: either by placing square brackets round the constituents, or by using tree diagrams. Most importantly, this section shows that we must use contrasting sets of grammatical and ungrammatical examples to argue for a particular constituent structure. Our analysis is valid only if we can show that it also rules out other logically possible analyses. So, as well as using the tests to show what the constituent structure of a phrase or sentence actually is, we also use them to rule out any alternative structures.
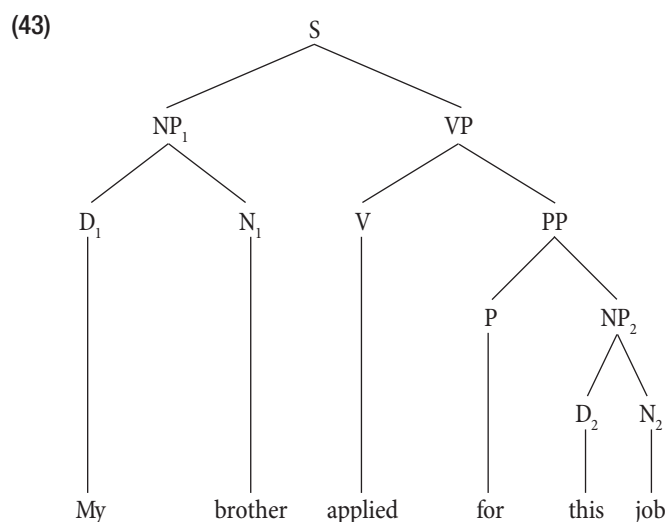
5.2   RELATIONSHIPS WITHIN THE TREE

This short section defines the technical terms used by linguists to discuss relationships between words and phrases in a tree diagram. It's common to use LABELLED BRACKETS or LABELLED TREE DIAGRAMS in which each relevant constituent has a label showing its category. Our trees include word class and phrase class labels such as V, VP, PP, P, NP, and so on, telling us that what's beneath that label is a PP, or a P, or an NP etc. This exact same information can be shown in labelled brackets. For example, the PP *for this job*, which we proved to be a constituent in (37), can be shown as follows:

(42) $[_{PP} [_P$ for$] [_{NP}$ this job$]]$

These brackets are read like this: the whole constituent is a PP, since this is the label on the outermost brackets (by convention, only the left-hand bracket is labelled). The PP comprises two main constituents, a preposition *for* and an NP *this job*: as we saw, this noun phrase fits into the focus position of a cleft sentence, so must be a constituent. Each individual lexical item (word) is in fact also a constituent, so *for*, *this* and *job* here are constituents, though I haven't labelled or bracketed the last two items here. The words are the smallest constituents of a tree.

Let's now add more information into (41) to give a fully detailed tree diagram, showing *my* and *this* as D (determiners) and *brother* and *job* as Ns.

(43)

```
                          S
                  ┌───────┴───────┐
                NP₁              VP
              ┌──┴──┐        ┌────┴────┐
             D₁     N₁       V         PP
              │      │       │      ┌──┴──┐
              │      │       │      P    NP₂
              │      │       │      │    ┌─┴─┐
              │      │       │      │   D₂  N₂
              │      │       │      │    │   │
             My   brother applied  for  this job
```

Using the tree in (43), I now introduce some of the technical terms used in syntax to describe tree structures. Recall that all the lines in the tree are known by the (reasonable!) technical term BRANCHES. Each point in the tree that has a category label or else an actual word attached to it is known as a NODE. In (43) we find PHRASAL nodes NP, VP and PP (nodes denoting the phrases in the tree), and also S. We also see in (43) the lexical nodes V, P, D and N (nodes indicating word-level elements), and the actual lexical items (words) *my*, *brother*, *applied*, *for*, *this* and *job*.

Despite the fact that I've used the label 'S', rather than calling the sentence a 'something phrase', the sentence is, of course, a phrase in its own right, and some linguists reflect this by terming the sentence 'TP', for tense phrase ('tense' in the sense of the tense of the verb). The idea is that a sentence is a phrase that denotes tense, though as we've mentioned earlier, it's not the case that *verbs* in all languages display the morphosyntactic category 'tense'. Here, I continue to use 'S', but you should be prepared for this not to be used in some frameworks if you go on to study theoretical syntax.

There are specific terms for the relationships between nodes in a tree. Each node IMMEDIATELY DOMINATES the next node below it, providing they are connected by a branch, and providing no other node intervenes. So for instance, within the PP, P immediately dominates *for*, D_2 immediately dominates *this*, and N_2 immediately dominates *job*. No other node intervenes between P and *for*, and so on. The node PP immediately dominates the two nodes P and NP_2, and NP_2 immediately dominates D_2 and N_2: again, no other nodes intervene.

A node which immediately dominates another node or set of nodes is their MOTHER: so, for example, PP is the mother of P and NP_2, and each NP is the mother of a D and an N. It won't surprise you that P and NP_2 are the DAUGHTERS of PP; and D_2 and N_2 are the daughters of NP_2, and so on. The lexical items *for*, *this* and *job* are the daughters of P, D_2 and N_2 respectively. In the same vein, the set of daughters

which share the same mother are known as SISTERS. So the nodes NP_1 and VP are sisters, and V and PP are sisters, as are D_2 and N_2, *this* and *job*, and so on.

However, the relationship between a set of nodes such as PP and the two nodes D_2 and N_2 is a different one: we say that PP DOMINATES D_2 and N_2 (though sadly we don't continue the analogy by using the term 'grandmother'). Note that PP doesn't IMMEDIATELY dominate D_2 and N_2 because the NP_2 node intervenes. But nonetheless an unbroken series of branches connects PP to D_2 and N_2: a branch first connects PP to NP_2, and then branches connect NP_2 to D_2 and N_2. When there's a path like this connecting the nodes in a tree, then the higher node is said to dominate the lower one.
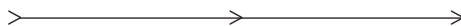
Before reading further, work out the relationships between nodes in the tree in (43) by answering the following questions:

i.   What nodes does S immediately dominate?

ii.   What nodes does S dominate?

iii. Does NP_1 dominate P and NP_2?

iv.   What nodes does VP dominate? Which are its daughters?
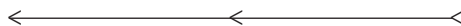
The answers are as follows:

i.   S immediately dominates NP_1 and VP.

ii.   S dominates NP_1 and VP, both Ds and Ns, V and PP, P and NP_2, and also *my*, *brother*, *applied*, *for*, *this* and *job* – in other words, all other nodes in the tree.

iii. No: NP_1 doesn't dominate P or NP_2 because there's no series of branches connecting the node NP_1 to these nodes (don't be misled by the fact that NP_1 is drawn higher up in the tree).

iv.   VP dominates V, PP, P, NP_2, D and N, as well as *applied*, *for*, *this* and *job*. Only V and PP are its daughters, because VP immediately dominates only V and PP.

Note that if a node immediately dominates a set of nodes, it automatically also dominates them. So we said, for instance, that S both immediately dominates and also dominates NP_1 and VP.

Trees also show the groupings of words into constituents:

(44)   **Definition of a constituent in a tree diagram**
       A set of elements forms a constituent in a tree diagram if and only if there is a single node that dominates just these elements, and no other items.

For instance in (43), the nodes *my* and *brother* form a constituent: they're both dominated by NP$_1$, and NP$_1$ doesn't dominate any other nodes.

>————————————→————————————→
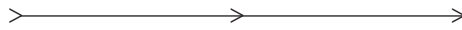
Please look again at the tree in (43) and answer these questions:

i.   Do *applied*, *for*, *this* and *job* form a constituent?

ii. Do *my*, *brother*, and *applied* form a constituent?

iii. Do *applied* and *for* form a constituent?

←————————————←————————————<
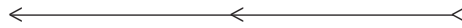
The answers are:

i.   Yes: these nodes are all dominated by VP.

ii. No: there is no single node that dominates just the elements *my*, *brother* and *applied* and no others.

iii. No: although both are dominated by VP, VP also dominates *this* and *job*.

In this section I have introduced labelled tree diagrams and discussed the terminology for the relationships between the nodes in a tree. I will make use of these terms in the following section when I discuss more complex tree diagrams.

5.3 DEVELOPING DETAILED TREE DIAGRAMS AND TESTS FOR CONSTITUENT STRUCTURE

Section 5.3.1 uses the tests for constituent structure established in Section 5.1 to work out the structure of some phrases and sentences, and also introduces a new test: ellipsis. Section 5.3.2 introduces another diagnostic for constituency: co-ordination. And Section 5.3.3 considers whether all languages have the same constituents.

5.3.1 Verb classes and constituent structure tests

5.3.1.1 *Phrasal verbs and prepositional verbs*

Before turning to some detailed tree diagrams which represent sentences of English, I first discuss the differences between two verb classes: phrasal verbs and prepositional verbs. Let's examine two more sentences which may appear superficially similar, but in fact have different constituent structures:

(45)   Those smugglers shook off their pursuers.

(46)   Those smugglers relied on the weather forecast.

There is one clear indication that these two examples are syntactically distinct. In (45), we can take the preposition *off* and place it immediately after the direct object NP *their pursuers*, to give *Those smugglers shook their pursuers off*. The verb *shake off* is a transitive verb – it must have an NP complement, i.e. a direct object. In (46), we don't have a transitive verb, but instead we have a verb *rely* and its PP complement. The verb does not have a direct object at all, and if we attempt to put the preposition *on* after the NP, the result is ungrammatical:

(47)   *Those smugglers relied the weather forecast on.

This test identifies very successfully one particular verb class: transitive PHRASAL VERBS such as *shake off* always allow the preposition to be placed after their object NP. These verbs are single lexical items comprising a V and a P: [$_V$ shake off]. We have, in fact, already met another example of a transitive phrasal verb in (33), namely *write down*, as we can confirm from the fact that we can get *My brother wrote his address down*. English has a vast number of phrasal verbs, both transitive and intransitive. Further examples of the transitive kind include *turn over*, *pull down*, *pick up*, *put out*, *switch on* and *break off*. As (48) shows, the preposition can follow the direct object (in bold) in each case. (For some of these, you may prefer to leave the preposition next to the verb. The point I am making is simply that it *may* follow the direct object.)

(48)   a.   We turned **the place** over.
            b.   They pulled **that old farm building with a thatched roof** down last
                  week.
            c.   I'd pick **that snake** up carefully.
            d.   She broke **her last engagement** off very suddenly.

This test also allows you to identify the full extent of the direct object NP, because the preposition has to be placed immediately *after* that NP (and not in the middle of it). So for instance, in (48b) the test shows that the whole of the sequence *that old farm building with a thatched roof* comprises the direct object. The preposition *down* can be placed at the end of the direct object, but not elsewhere:

(49)   *They pulled that old farm building **down** with a thatched roof last week.

As noted, English also has phrasal verbs which are – or can be – intransitive, such as *wake up*, *sit down*, *sleep in*, *turn out* (as in *Not many people turned out*) and *break down* (as in *The car broke down*).

Some English grammarians use the term 'particle' to refer to the *over*, *down*, *up*, *off*, *out* (etc.) part of the phrasal verb, but we can tell that they are truly prepositions

by using the modifier *right*, which we saw in Section 2.6.1 to be a good test for preposition status. So for example, we get *Pull the handle **right** down*, *Break the plastic safety catch **right** off*, *There was a loud bang and I woke **right** up* and so on.

Now let's compare (46). There, we don't have a phrasal verb at all. Instead, the verb *rely* takes a PP complement, and this PP must be headed by the preposition *on*: we can only *rely **on** something*, not **rely for*, **rely off*, **rely over* or **rely out*. Verbs which select PP complements are known as PREPOSITIONAL VERBS. Their defining properties are that the PP is obligatory, and is headed by one specific preposition. Further examples of prepositional verbs include *believe in NP*, *hear from NP*, *see to NP*, *glance at NP*, *hope for NP*, *depend on NP* and *look after NP*, among many others. Quite often, the preposition has such a close relationship with the prepositional verb that not even one of the prepositional modifiers, such as *right*, *just* or *straight*, can intervene. These examples give you an idea of the variation that is found; of course, you may not agree with my judgements in each case:

(50)   We rely just/*right on our good fortune.
The politicians skated right/*just over these damaging issues.
The grandparents looked *just/*right after the children.

Prepositional verbs, then, are a rather special set. On the other hand, if a verb merely has a PP adjunct – in other words, it is modified by an optional PP – the properties are entirely different. The head preposition can be readily changed: *I jumped **on** the wall / **off** the wall / **over** the wall / **behind** the wall* (and so on), and the PP can be omitted entirely, as it is not a complement. The choice of a modifier in the PP is also much freer:

(51)   I jumped just/straight/right over the wall.
We ran just/straight/right to the end of the beach.
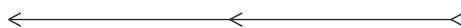The vase fell straight/right off the shelf.

The verbs illustrated in (51) are not prepositional, since the PP is an adjunct rather than a complement. We reserve the term 'prepositional verb' for a verb with an obligatory PP complement.

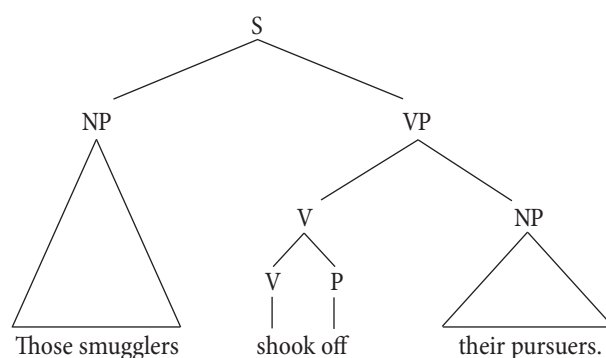### 5.3.1.2   *Tree structures for phrasal and prepositional verbs*

We already know that (45) and (46) must have different structures, because of the differing behaviour we uncovered in Section 5.3.1.1. Using constituent structure tests, we can discover which words group together. The tree diagrams in (52) and (53) give the end product of a set of tests, and I'll work back from these trees to demonstrate to you that the two distinct structures proposed here are correct. Bear in mind that a tree diagram is nothing more than a reflection of native speaker judgements about the structure of a phrase or sentence, gained mostly from applying tests for constituency.
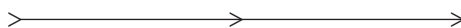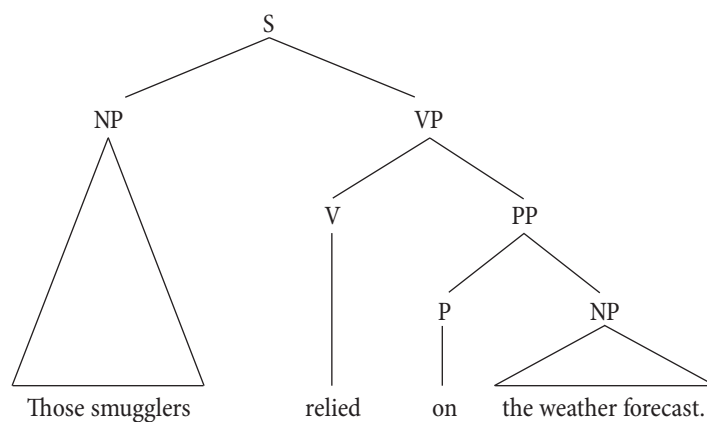
Before reading further, I suggest you draw both trees for yourself on a sheet of paper, exactly as shown below. This will give you practice with tree-drawing, and also save you having to look back at my trees as I develop the arguments for their different constituent structures.

(52)

```
                      S
              /              \
           NP                  VP
          /|\               /      \
         / | \             V         NP
        /  |  \           / \       /|\
       /   |   \         V   P     / | \
   Those smugglers    shook off   their pursuers.
```

(53)

```
                      S
              /              \
           NP                  VP
          /|\               /      \
         / | \             V         PP
        /  |  \            |        /    \
       /   |   \           |       P      NP
   Those smugglers      relied    on    the weather forecast.
```

Every time we draw a tree diagram we are making a set of claims about constituent structure – about which words group together to form the phrases of a sentence. A tree is built up from our evidence of what these phrases are, which comes from tests for constituent structure. So let's start with evidence for the two main constituents in each tree: the subject and the VP predicate.

Our tests show that *those smugglers*, the subject in each sentence, is indeed a constituent. Both subjects can be sentence fragments:

(54)　a.　　Who shook off their pursuers?　　　[_{NP} Those smugglers]
　　　　b.　　Who relied on the weather forecast?　　[_{NP} Those smugglers]

Second, both subject NPs can also appear in the focus position of a cleft sentence:

(55) a. It's [NP those smugglers] who shook off their pursuers.
      b. It's [NP those smugglers] who relied on the weather forecast.

And third, we can also replace both subject NPs with *they*, using a test for NP status introduced in Section 5.1.1. The word *they* is rather badly termed a 'pronoun'. Since it replaces a whole NP it's really a pro-NP: 'pro' means '(stands) for'. The cover term used for all pro-phrases is PROFORM: a proform takes the place of a sequence of words which form a constituent, and so any string of words that can be replaced by an appropriate proform must be a constituent.

A proform test can also prove the existence of the VP constituent, as we saw in Section 5.1.3. We use *do so* (or *did so* in the past tense) to stand for VP, therefore as a 'pro-VP':

(56) a. Those smugglers [VP shook off their pursuers], and the moonshine merchants [VP **did so**] too.
      b. Those smugglers [VP relied on the weather forecast], and these fishermen [VP **did so**] too.

Rather than repeating the whole VP, we can replace it with the proform. The *do so* test is a specific test for a VP constituent.

One of the other tests for constituent structure which was given earlier is the cleft construction. However, most dialects of English can't form a cleft using a VP constituent:

(57) *It's **shake off their pursuers** that those smugglers did.
      *It's **rely on the weather forecast** that those smugglers did.

This does not mean the cleft test is unreliable; it just means that VPs can't be focussed like this in English. In some languages, though, such as Irish, it's perfectly OK to focus VPs in the cleft construction. Given a sentence like (58), we can focus the VP to get (59), which is fully grammatical in Irish:

(58) Bhí   an   fear   ag   péinteáil   cathaoir.   (Irish)
      was   the   man   PROG   paint.INFIN   chair
      'The man was painting a chair.'

(59) Is   [VP **ag**   **péinteáil**   **cathaoir**]   a   bhí   an   fear.
      is   PROG   paint.INFIN   chair   that   was   the   man
      '*It's painting a chair that the man was.'

This shows that the string *ag péinteáil cathaoir* is a constituent of (58). The lesson here is that sometimes a test won't work in a given language, but this may be due to some language-specific quirk. VP clefts are fine in Irish but not in English. We must

make sure that our results are valid by using more than one test for constituency each time.

To confirm the existence of VPs in English I introduce another test for constituency: ELLIPSIS. Ellipsis means missing out part of the sentence, but the portion we miss out must always be a constituent:

(60) a.   Those smugglers might [$_{VP}$ shake off their pursuers], and the moonshine merchants might [$_{VP}$ ____] too.

b.   Those smugglers didn't [$_{VP}$ rely on the weather forecast], but these fishermen did [$_{VP}$ ____] for sure.

It's perfectly possible to repeat the VP from the first half of the sentence, but by omitting it as shown here, we prove that it really is a constituent.

You might have noticed, though, that some of the sequences which are constituents according to my trees in (52) and (53) cannot undergo ellipsis. First, in both trees there are sequences which are shown as NP constituents: *their pursuers* and *the weather forecast*. But if we omit these constituents from my sentences, the result is ungrammatical:

(61) a.   *Those smugglers must shake off [$_{NP}$ their pursuers], and these moonshine merchants should shake off [$_{NP}$ ____] too.

b.   *Those smugglers didn't rely on [$_{NP}$ the weather forecast], but these fishermen did rely on [$_{NP}$ ____] for sure.
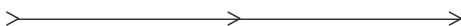
Before we examine why the examples in (61) are ungrammatical, note that I haven't yet proved that there really is an NP constituent *their pursuers / the weather forecast* in both sentences. So first, find at least two tests for each phrase to confirm that they really are constituents.
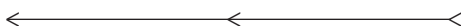
For these NPs, you could use tests as shown in (62) through (65):

(62) **Echo questions**
a.   The smugglers shook off [$_{NP}$ **who**]?
b.   The smugglers relied on [$_{NP}$ **what**]?

(63) **Sentence fragments**
a.   Who did the smugglers shake off? [$_{NP}$ **Their pursuers**]
b.   What did the smugglers rely on? [$_{NP}$ **The weather forecast**]

(64) Clefts
   a.   It was [_{NP} **their pursuers**] that the smugglers shook off.
   b.   It was [_{NP} **the weather forecast**] which the smugglers relied on.

(65) Proforms
   a.   The smugglers shook [_{NP} **them**] off.
   b.   The smugglers relied on [_{NP} **it**].

Note also that the pronoun *precedes* the preposition when we have a transitive phrasal verb, as in (65a), but *follows* the preposition when we have a prepositional verb, as in (65b). This is the way each verb class always works in English, and it is a very reliable test. For instance, we can't have **The smugglers shook off them* for a transitive phrasal verb.

So if the NPs which we've tested from (52) and (53) really are constituents, as we've shown, why can't they be omitted in (61)? The reason is that both examples contain a head word which requires these NPs to be present – the NPs are complements, and the sentences are incomplete without these complements. So the transitive phrasal verb *shake off* requires a direct object NP in (61a), and in (61b) the transitive preposition *on* also requires an NP object. Constituents which form the complement to some head, particularly a head verb or preposition, are quite generally unable to be omitted. This means that the ellipsis test can't be used to diagnose the constituent status of such phrases.

Now let's examine the sequence *on the weather forecast*, which is shown as a PP in (53). First we need to confirm the constituent status of this string of words:

(66) Sentence fragment
   What exactly did the smugglers rely on? Oh, [_{PP} **on the weather forecast**] of course!

(67) Cleft
   It was [_{PP} **on the weather forecast**] that the smugglers usually relied.

(Some speakers may not be entirely happy with (67), but the test in (66) confirms that there really is a PP.)

Again, we might expect that if it's a constituent, then the PP could be omitted, but it actually can't be:

(68) *Those smugglers actually didn't rely [_{PP} on the weather forecast], but these fishermen really did rely [_{PP} _____], for sure.

Just as with (61), the reason for the ungrammaticality of (68) is that the PP we've omitted is a complement: prepositional verbs like *rely on NP* require the PP complement to be present, so again, we can't use the ellipsis test for constituent structure in a case like this.

>————————→————————→
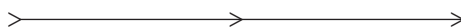
For completeness, we should use the same tests to confirm a claim made by our tree in (52): the structure proposed there says that the sequence *off their pursuers* is *not* a constituent in *Those smugglers shook off their pursuers*. Recall that to prove that two sentences such as (52) and (53) have a different structure, we must give contrasting sets of grammatical and ungrammatical examples as evidence. Please formulate the relevant sentence fragment and cleft sentence in order to demonstrate that *off their pursuers* is not a constituent in this case. Remember that in the sentence fragment test, the question you ask must be grammatical: the test is whether the answer is grammatical or not.

←————————←————————<
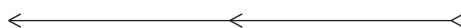
The results are:

(69)   Sentence fragment
Who did the smugglers shake off? *Off their pursuers.

(70)   Cleft
*It was off their pursuers that the smugglers shook.

(Remember that we don't put non-constituents in brackets.) Even if you were not entirely happy with (67), I expect you'll agree that (70) is far worse. These ungrammatical examples confirm that there is indeed no PP *off their pursuers* in (52).

In this section we have justified the different structures proposed for phrasal verbs and prepositional verbs, using the tests for constituency introduced in Section 5.1. We have also introduced a new test: ellipsis, or omission. If we can omit some sequence of words, then there's a good chance that it's a constituent. We also showed that if a constituent is the complement of a verb or a preposition, then we generally won't be able to omit it, because it's required by the head V or P to be present.

5.3.2   The co-ordination test for constituency

Our final test for constituency is CO-ORDINATION. Sequences of words which are constituents can be CO-ORDINATED or CONJOINED with one another, provided that they are of the same syntactic category: so we can have NP + NP, or VP + VP, for instance. For example, the sequence *their pursuers* is an NP constituent in (52) and *the weather forecast* is an NP constituent in (53), and so each can be joined together with another NP:

(71)   The smugglers shook off (both) [NP their pursuers] and [NP the revenue men].

(72)   The smugglers relied on (both) [NP the weather forecast] and [NP their years of experience].

The two NPs in brackets in these examples have been conjoined using *and*, known as a CO-ORDINATING CONJUNCTION. Other such conjunctions in English include *but*, *nor* and *or*.

You may be wondering if the transitive phrasal verb *shook off* in (71) and the preposition *on* in (72) now have two object NPs. No, they do not: when two constituents of the same category are conjoined, they simply make one larger constituent of the same category, as in (73). So there is still only one object for the transitive verb in (71) and the preposition *on* in (72), but this NP may itself contain NPs embedded within it. The node label CONJ means 'conjunction'.

(73)

```
                    NP
        _____|_____
       |           |           |
      NP         Conj.        NP
     /   \         |         /    \
 their pursuers   and   the revenue men
```
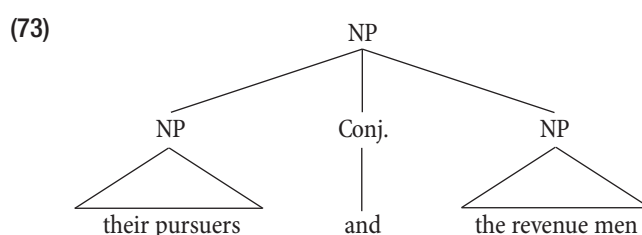
The co-ordination test can be used to confirm that a phrasal verb and a prepositional verb do have different structures, as we have proposed – look at your tree diagrams for (52) and (53). A prepositional verb contains a PP constituent, according to the tests we've seen so far. And indeed, the PP can be conjoined with another PP:

(74)   The smugglers relied [PP on the weather forecast] and (also) [PP on their years of experience].

In (75), on the other hand, we can't conjoin *off their pursuers* with *off the revenue men*, because these two strings of words are not constituents of any kind:

(75)   *The smugglers shook **off their pursuers** and **off the revenue men**.

The preposition of a phrasal verb like *shake off* isn't attached to the following NP, as we can see from the structure for phrasal verbs in (52): there is no node that dominates just the preposition and *their pursuers*, so this sequence doesn't form a constituent. Please look at your tree diagram for (52) to confirm this for yourself.

Finally, we can use the co-ordination test to discover more about the structure of one of the ambiguous sentences in Section 5.1: *The boy and the girl's uncle stayed to dinner*. You'll see that there's a conjunction *and* in this sentence. We can now show that there are two possible ways of co-ordinating constituents within the NP *the boy and the girl's uncle*, which accounts for the ambiguity. First, we can conjoin an NP *the boy* with an NP *the girl's uncle*, as in (76). This gives the reading in which two people stay to dinner:

(76)   [NP [NP the boy] and [NP the girl's uncle]]

In (76) the outermost brackets are labelled 'NP': this tells you the category of the whole phrase. Within this large NP, two smaller NPs are embedded, co-ordinated using *and*. The equivalent tree is in fact just the same as that in (73).

In the alternative reading of the phrase, where only one person stays to dinner, we conjoin an NP *the boy* with an NP *the girl*, as in (77): here, it's the uncle to both children who stays to dinner.

(77)
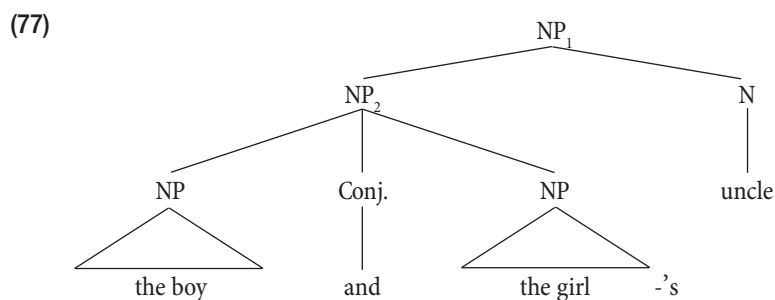
The tree in (77) says that the whole phrase is an NP (NP$_1$) which has two branches. On the left branch is NP$_2$, which immediately dominates the two conjoined NPs. Note that this whole phrase in NP$_2$ effectively replaces a single-word determiner such as *their*. On the right branch is the N *uncle*, the head noun of the *entire* phrase, NP$_1$. Again, numbering the phrases as I've done here is simply a useful way to make it clear which phrase we're referring to, when there are several phrases of the same category.

We've seen in this section that two strings of words can be conjoined if they're constituents, and (normally) of the same syntactic category. Conversely, if a sequence of words which does *not* form a constituent is conjoined with other material, then the result is always ungrammatical, just as in (75). Co-ordination can therefore be added to our set of tests for constituent structure.

5.3.3   Do all languages have the same constituents?

The answer to this question is no, they apparently don't. I illustrate this with VP. Most languages have a clear VP constituent, as can be shown, for example, using VP co-ordination. Examples (78) and (79) show conjoined VPs in Persian and in Malagasy (note that in Malagasy the subject – *Rabe*, a name – is at the end of the clause, rather than at the start):

(78) Jân   [$_{VP}$ xandid]   va   [$_{VP}$ dast   tekân   dâd].   (Persian)
          John   smiled   and   hand   sign   gave
          'John smiled and waved.'

(79) [$_{VP}$ Misotro   taoka]   sy   [$_{VP}$ mihinam-bary]   Rabe.   (Malagasy)
                 drink   alcohol and   eat-rice   Rabe
          'Rabe is drinking alcohol and eating rice.'

From this co-ordination we can conclude that there are likely to be VP constituents in both languages.

However, linguists have also used the standard tests for constituent structure to argue that some languages, such as Basque and Hungarian, have no VP constituent. In (57), we saw that VPs can't be clefted in English; nonetheless, other tests demonstrate that English clearly does have VPs:

(80) Kim says that he hasn't [VP drunk all the water]
    a.   but he has [VP __].
    b.   and he hasn't [VP __].
    c.   but he has [VP **(done) so**].

In (80a) and (80b), we see that the VP can undergo ellipsis (it can be omitted); (80c) shows that one proform for VP in English is *so* (or *done so*). Now compare the same constructions in Basque (the argument and the data here are taken from Rebuschi 1989). The putative VP (i.e. the verb and its object, the sequence being tested) is shown in bold:

(81) Peio-k   dio   **ur   guzia   edan**   du-ela.   (Basque)
      Peio-CASE   says   water   all   drunk   AUX-that
      'Peio says that he has drunk all the water.'
    a.   *eta [VP __]   du
          and   AUX
          ('and he has')
    b.   *baina   ez   du   [VP __]
          but   NEG   AUX
          ('but he hasn't')
    c.   *eta   hala   du
          and   thus   AUX
          ('and so he has')

(81a), (81b) and (81c) are all ungrammatical in Basque: there can be no ellipsis of a VP, as shown in (81a) and (81b), and nor does *hala*, 'thus', act as a VP proform. Nor, indeed, do any other standard tests indicate that Basque might have a VP constituent. So on these grounds, we can say that Basque appears to lack a VP altogether.

5.3.4   An introduction to the bar notation

This final section takes the reader somewhat beyond the simple tree structures seen so far in the chapter, and looks at a more advanced issue, though it uses exactly the same kind of argumentation. I use the co-ordination test for constituency to argue for the existence of another phrasal category which we haven't yet come across. This is a type of nominal phrase (= noun-type phrase) which is smaller than a full NP, but larger than just a head noun. I'm going to argue that a head noun together with its complement forms a constituent, which is termed N′ (pronounced N-bar).

The issue here is internal structure of the direct object NP in (82):

(82)   I admired [$_{NP}$ the director's treatment of the issues].

First, you should satisfy yourself that the whole string labelled 'NP' really is a constituent: try the cleft test *It was ...*, and the sentence fragment test, using *What did you admire?*, in order to prove my claim. I will assume that your conclusions support me. As usual, this NP can be conjoined with another full NP, as in (83):

(83)   I admired [$_{NP}$ the director's treatment of the issues] and [$_{NP}$ her sensitivity to the problems].

Each of the conjuncts (= sequences conjoined) in (83) is a full NP, as we can see from the fact that either of them could be the subject or object of a verb. Together, the two conjuncts form one large NP – the direct object of *admired* – although I haven't shown this in brackets in (83).

In (84), *admired* has a slightly different direct object. Here, I bracket the *entire* object NP, but without specifying its internal structure:

(84)   I admired [$_{NP}$ the director's treatment of the issues and sensitivity to the problems].

Inside the full NP, two smaller strings of words are conjoined. Unlike in (83), however, the conjoined strings in (84) are not full NPs.

>————————————→————————————→

Before reading any further, decide what strings of words are actually co-ordinated in (84). Don't attempt to give their category, but just say what the conjoined sequences of words are.

←————————————←————————————<
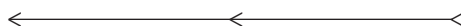
The conjuncts are *treatment of the issues* and *sensitivity to the problems*. If you didn't get this, think about what (84) means; the specifier (Section 4.1.8) *the director's* applies to both conjuncts: it's both the director's treatment of the issues and her sensitivity that you admire. So within the direct object NP we have two conjoined strings as shown in bold in (85):

(85)   I admired [$_{NP}$ the director's [$_{??}$ **treatment of the issues**] and [$_{??}$ **sensitivity to the problems**]].

If co-ordination is a reliable test for constituency, we must conclude that both of these sequences in bold are constituents. However, they're not members of any

category we've seen before, which is why I indicate the category with subscript question marks instead of the real category labels.

So, we need to work out the internal structure of an NP such as *the director's treatment of the issues*. Perhaps your first thought is that the sequence *the director's treatment* might itself be an NP in our example. But in fact this string of words is not a constituent of any kind here, as we can prove by using the proform test. The appropriate proform in English for an NP which includes a possessive form is one of the set *mine*, *yours*, *his*, *hers*, *ours*, *theirs*, as, for example, in (86):

**(86)**   I can't stand **his uncle**. Do you like **hers**?

Here, *hers* replaces an entire NP (such as *her uncle*). But it can't be used to replace the sequence *the director's treatment* in (87):

(87)   *I admired <u>hers</u> of the issues.

This means that *the director's treatment* isn't a constituent here. To make matters clear, note that I am not claiming that *the director's treatment* can never be a constituent, just that it isn't one in (82).

We do know, though, from (85), that the sequence *treatment of the issues* is a constituent in our example, because of its ability to undergo co-ordination. So far, we have had no label for the type of constituents that are co-ordinated in (85). These mystery constituents consist of a head noun *treatment* or *sensitivity*, plus a PP complement to that head: *of the issues*, *to the problems*. So a head + complement is co-ordinated with another similar sequence, and the two conjuncts share a single specifier, *the director's*: both the treatment and the sensitivity are that of the director.

The structure we propose for *the director's treatment of the issues* is shown in (88). In (88) you can see that there is a constituent, *treatment of the issues*, which is labelled N′ – this, then, is the category of the mystery constituent which was co-ordinated in (85).

(88)
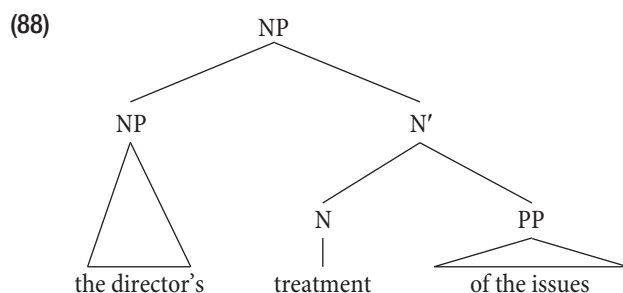
(89)   I admired the director's [$_{N'}$ treatment of the issues] and [$_{N'}$ sensitivity to the problems].

You can see that N′ is not a full NP – the whole tree is an NP, with the specifier *the director's* – but rather, N′ is a smaller nominal phrase, an intermediate category smaller than an NP but larger than an N. This N′ consists of a head noun (*treatment*) plus a PP complement to that noun.

The minimal type of nominal element is a head noun: this is sometimes referred to as N^0 ('N zero'): the bare noun. Any nominal phrase which has a determiner or a full phrase in its specifier position must be an example of the *maximal* type of nominal phrase, which is NP. One instance of such an NP is *his uncle* in (86). And *hers* in (86) is also a full NP: we can tell because it's a PROFORM for the NP *her uncle*, and therefore must itself be an NP. The intermediate category N′, on the other hand, consists of the head plus its complement, but it has no determiner or other specifier. The N′ is still a phrasal category, but not a full NP.

So the general structure of noun phrases is as shown in (90). This structure shows that an NP immediately dominates a specifier and an N′. In turn, the N′ immediately dominates the head N (the notation N^0 is often used) and its complement.

(90)

```
                    NP
                   /  \
            Specifier   N′
                       /  \
                      N    Complement
```
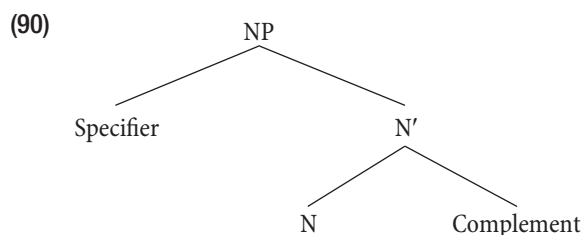
Before moving on, though, you should note that not all NPs have either a specifier or a complement – *hers* didn't in (86). This means that it is possible for an NP to consist of just its head noun, such as *cats* or *Kim*. Evidence that NPs containing only the head N are indeed full phrasal categories comes from the fact that they can co-ordinate with other NPs, as in [NP *Kim*] *and* [NP *her uncle*] *came to dinner*.

Returning to our intermediate category, we have seen one piece of evidence for the existence of an N′, from co-ordination. As I've noted, it's always best to apply more than one test for any putative constituent. And indeed, another piece of evidence for N′ comes from the fact that we can replace an N′ by a proform, *one*:

(91)   [NP This [N′ **treatment of the issues**]] is better than [NP that [N′ **one**]].

In (91), *one* stands for just the N′ *treatment of the issues*: the determiner *this* (which makes a phrase into a full NP) is excluded. If the sequence *treatment of the issues* can be replaced by a proform, then we have good evidence that it really is a constituent.

This section has presented evidence for the existence of an intermediate kind of nominal category, smaller than an NP but larger than a noun, which is called N′. The evidence was, first, that a sequence consisting of a head noun plus its complement could be co-ordinated with another similar sequence, so showing these sequences to be constituents; and second, that the sequence could be replaced by a proform *one*, again showing it to be a constituent.

5.4 SUMMARY

In Section 5.1 of this chapter, we presented evidence that sentences and phrases actually have syntactic structure. Using tests for constituent structure, we argued that various sentences which superficially appeared to be similar in fact have different structures. We represented syntactic structure using brackets or tree diagrams, which show the ways words group together to form phrases. Section 5.2 presented the standard terminology for describing the relationships between the nodes in a tree. In Section 5.3, we used the tests for constituency to work out the structure of a number of phrases and sentences, and produced labelled brackets and labelled tree diagrams to illustrate the constituent structure that we discovered. The syntactic tests for constituency used in this chapter are as follows: the sentence fragment test; echo questions; cleft sentences; proform tests; ellipsis; and co-ordination.

Checklist for Chapter 5

Although the remainder of this book does not engage further with the topic of constituent structure, you will definitely need to understand if it you're going on to further study of syntax.

- Can you outline at least four different tests for constituency, and say how they're used?

- Why do we use contrasting grammatical and ungrammatical examples in our tests?

- What kinds of constituents cannot normally be omitted under ellipsis?

- What is the difference between phrasal and prepositional verbs in English? Give some examples of each category and indicate how we can tell the difference.

FURTHER READING

Two introductory texts which concentrate on the grammar and structure of English, and which go significantly beyond what I have done in Chapters 2 through 5, are Börjars and Burridge (2010) and Lobeck (2000). Radford (1988) provides detailed (and relatively introductory) reading on constituent structure, tree diagrams and tests for constituency. At this stage you may not want to go beyond his Chapter 5. See also Aarts (2013) and Burton-Roberts (2010). For a more detailed account of how to treat the English possessive *-'s*, including proposed tree diagrams, I recommend consulting Burton-Roberts (2010). Moving on to more theoretically oriented treatments of syntax, Poole (2011) forms a good follow-up to the material discussed here.

EXERCISES

1. In the text of Chapter 5 I proposed that *one* is a proform for an N′ constituent. The examples below suggest that *one* can also be a proform for a different kind of constituent. What is this?

    (1)   Kim knows stacks of sea shanties, and Mel knows one too.

    (2)   I've seen a mockingbird, but Chris has never seen one.

    (3)   I've met several university presidents, but Lee has had dinner with one.

2. *Time flies like an arrow; fruit flies like a banana.* Explain clearly in syntactic terms how this pair of clauses work together as a nice example of a pun. Use the correct grammatical terminology plus relevant tests for constituent structure, contrasting the two clauses.

3. The two sentences in (1) and (2) below contain words from the same word classes, and in the same order, but they each have different syntactic structures.

    Task: (i) Using standard tests for constituency, work out what the constituents of each sentence must be. You should use at least two tests for each putative constituent. Your answers should include contrasting grammatical and ungrammatical examples which reveal the syntactic differences between (1) and (2). Use square brackets to indicate the constituents you find in each example, and remember to bracket constituents only, and not random strings of words. (ii) Next, draw labelled tree diagrams for (1) and (2), taking care that the trees correctly represent the constituent structures you discovered above.

    (1)   Kim glanced at the actor with a wig.

    (2)   Kim glanced at the actor through her binoculars.

4. The two sentences in (1) and (2) below again contain words from the same word classes, and in the same order, but again they each have different syntactic structures.

    Task: (i) Using at least three standard tests for constituency, work out what the constituents of each sentence must be. Your answers should include contrasting grammatical and ungrammatical examples which reveal the syntactic differences between (1) and (2). Can you provide any further evidence that (1) and (2) differ in structure? What sub-classes of verb does each example contain? (Don't worry if you prefer *stank* in (1); both past tense forms are acceptable!). (ii) Now draw labelled tree diagrams for (1) and (2), ensuring that the trees correctly represent the constituent structures you discovered above.

(1) The skunk stunk out my garden.

(2) The skunk slunk out my garden.

5. The data below are from Fijian, and are taken from Lynch (1998).

Task: Examine the data and explain what they show about the grammar of possession in Fijian. Next, describe carefully how the possessive construction is formed. Your answer should account for all the data.

Hint:
The prefix glossed as POSS is a possessive marker. It has three distinct forms, depending on the semantic category of the item possessed. Your answer should note all three forms. However, you are not required to specify what factors determine the appearance of any particular form.

(1)	na	tina-qu		**(2)**	na	me-na niu
	the	mother-my			the	POSS-his coconut
	'my mother'				'his coconut'	

(3)	na	ke-mu	itaba	**(4)**	na	no-mu itaba
	the	POSS-your	photo		the	POSS-your photo
	'your photo'				'your photo'	
	(i.e. a photo taken of you)				(i.e. a photo you took or have)	

(5)	na	yaca-qu	**(6)**	na	ke-mu madrai
	the	name-my		the	POSS-your bread
	'my name'			'your bread'	

(7)	na	me-qu	bia	**(8)**	na	ulu-qu
	the	POSS-my	beer		the	head-my
	'my beer'				'my head'	

(9)	na	no-qu	yaca	**(10)**	na	tama-qu
	the	POSS-my	name		the	father-my
	'my namesake'				'my father'	

6. Welsh is a VERB-INITIAL language: the finite verb or finite auxiliary appears first in the clause in unmarked (= normal) constituent order, as in (1), which has a finite auxiliary (meaning 'was') in initial position. This example also has a non-finite main verb *dweud* 'tell', lower down in the clause; this construction, then, is rather parallel to English *was telling*, which also has an auxiliary plus a non-finite main verb.

(1) Oedd fy ffrind yn dweud ei hanes wrth yr athro y bore 'ma.
was my friend PROG tell.INFIN her story to the teacher the morning here
'My friend was telling her story to the teacher this morning.'

Changes in this basic order are used to focus other constituents. Examine the data in (2) through (6) (based loosely on Jones and Thomas 1977: 289).

Task: (i) State how constituents are focussed in Welsh (being as precise as you can), and (ii) state precisely what constituent is focussed in each example, giving its category and grammatical function. (iii) A translation for (2) is already provided; suggest appropriate translations into English for the remaining examples. (iv) In both (3) and (5) there are additional grammatical changes; can you say what these are?

Hint:
In some cases your translation into English may not sound very natural. The reason for this is that languages display differences in what constituents may be focussed or otherwise manipulated, as we have seen throughout this chapter. Provide the best translations you can, whilst trying to capture the meaning of the source language.

(2)   Fy ffrind oedd yn   dweud   ei hanes wrth yr athro   y   bore   'ma.
my friend was   PROG tell.INFIN her story to   the teacher the morning here
'It was *my friend* who was telling her story to the teacher this morning.'

(3)   Ei   hanes   oedd fy   ffrind   yn   ddweud   wrth   yr   athro
her   story   was   my   friend   PROG   it.tell.INFIN to   the   teacher
y   bore   'ma.
the   morning   here

(4)   Wrth yr   athro oedd fy   ffrind yn   dweud   ei   hanes
to   the   teacher was   my   friend PROG   tell.INFIN her   story
y   bore   'ma.
the   morning   here

(5)   Dweud   ei hanes wrth yr   athro   oedd fy   ffrind   y   bore   'ma.
tell.INFIN her story   to   the teacher was   my   friend the morning here

(6)   Y   bore   'ma oedd fy ffrind yn   dweud   ei hanes wrth yr athro.
the morning here was   my friend PROG tell.INFIN her story to   the teacher

7. Examine the data in (1) through (6) from Malayalam, a Dravidian language spoken in India. These data (taken from Asher and Kumari 1997) all illustrate one particular construction which manipulates constituents in a certain way which was discussed in Chapter 5. However, I have left one crucial morpheme (part of a word) in the source language both unidentified and unglossed.

Task: (i) Identify what construction is illustrated in the data; (ii) work out *exactly* how this construction is formed in Malayalam; and (iii) work out what syntactic category of constituent (e.g. PP, NP etc.) is being manipulated in each separate Malayalam example. (To remind you, ACC is accusative case, indicating a direct object.)

(1) avan   bhaaryayooʈum   makkaḷooʈum   kuuʈe   taamasikkunnu
      he   wife.with   children.with   together.with stay.PRES
      'He stays with his wife and children.'

(2) ɲaan   raamaneyum   avanṟe   muunnaamatte   makaneyum   kaɳʈu
      I   Raman.ACC   his   third   son.ACC   see.PAST
      'I saw Raman and his third son.'

(3) avaḷ   viiʈʈilum   hoosʈʈalilum   taamasikkilla
      she   house.in   hostel.in   stay.FUT.NEG
      'She will not stay in the house or the hostel.'

(4) avan   eẓuttu   vrttiyaayum   vyaktamaayum   eẓuti
      he   letter   neatly   legibly   write.PAST
      'He wrote the letter neatly and legibly.'

(5) avan   kaappi   kuʈikkukayum   piṉṉe   vaʈa   tiṉṉukayum   ceytu
      he   coffee   drink.INFINITIVE and.then   vada   eat.INFINITIVE do.PAST
      'He drank coffee and then ate vada.'

(6) uṉṉiyum   baabuvum   vannu
      Unni   Babu   come.PAST
      'Unni and Babu came.'

8. The data in (1) and (2) (from Clamons et al. 1999) are from a Cushitic language, Oromo, spoken in Ethiopia, Kenya and Tanzania. In the examples in (1), the subject of the sentence also has a special pragmatic property: it is a TOPIC, which Clamons et al. define as 'what the sentence or discourse is about'. The topic property is marked on subjects which are topics using a topic marker -*n*. The subjects in (2) are *not* topics. Subjects in general are marked with a 'subject case' marker (SU).

Task: Discover what grammatical change occurs in the sentence when its subject is also a topic. Articulate it as clearly as you can, using the correct grammatical terminology, and generalizing so that you cover all the data with a single statement. You are looking for a property which is common to all the grammatical data in (1) versus all the grammatical data in (2), but this property is manifested slightly differently from example to example.

Hints:
- The subject of the sentence can, of course, consist of one noun phrase conjoined with another noun phrase: for instance, in (1b) the 'girl' noun phrase and the 'boy' noun phrase are co-ordinated in this way to form a subject meaning 'the girl and the boy'. The subject will then have the grammatical properties of the two conjoined phrases together.

- English does not have a special topic construction, but topics are typically associated with a particular emphatic intonation. I've indicated this by using italics in the translations in (1).

- The background information above the examples is there purely to help you see where topics are used in Oromo. You can see from (1) that topics are normally a phrase which has just been mentioned in the discourse; this contrasts with the 'out-of-the-blue' sentence that you might find at the start of a story, as in (2a), or a sentence in which the participants clearly haven't been mentioned before, as in (2b, c). Note also that the translations in (1) and (2) differ subtly, reflecting the fact that the subjects are topics in (1) but not in (2).

- It will help you to compare (1a) with (2a), (1b) with (2b), and (1c) with (2c). The additional data are there simply to give you more clues.

(1) a. (i) (In answer to: What did the girl and the boy do?)
              Intal-t-ií-n   hoolaa   bit-t-e.
              girl-FEM-SU-TOPIC   sheep   buy-FEM-PAST
              '*The girl* bought a sheep.'
     a. (ii) *Intal-t-ií-n   hoolaa   bit-e.
              girl-FEM-SU-TOPIC   sheep   buy-PAST
              ('*The girl* bought a sheep.')
     b. (In answer to: What did the girl and the boy do?)
         Intal-t-ií-n   -ifi   gurbaá-n   wal   lol-an
         girl-FEM-SU-TOPIC   and   boy.SU-TOPIC each.other   fight-3PL.PAST
         '*The girl and the boy* were fighting.'
     c. (In answer to: Where was I when the boy came?)
         Ati   -ifi   Salma-á-n   nyataa   godhu   tur-tan
         you.SG   and   Salma-SU-TOPIC food   make   were-2PL.PAST
         '*You and Salma* were cooking.'

(2) a. (Passage at the start of a story, i.e. with no previous context)
         Intala   takka-á   hoolaa   bit-e
         girl   one.FEM-SU   sheep   buy-PAST
         'A girl bought a sheep.'
     b. (In answer to: Who was fighting?)
         Intala   -afi   gurbaa   tokko-ó   wal   lol-e
         girl   -and   boy   one.MASC-SU   each.other   fight-PAST
         'Some girl and boy were fighting.'
     c. (In answer to: Who was cooking?)
         Ati   -ifi   Salma-á   nyataa   godhu   tur-e
         you.SG   and   Salma-SU   food   make   were-PAST
         'You and Salma were cooking.'
     d. (i) Intala-á   dhuf-e.
              girl-SU   come-PAST
              'The girl came.'

d.   (ii)   *Intala-á   dhuf-t-e.
                girl-SU   come-FEM-PAST
                ('The girl came.')

NOTE

1 As you can probably tell intuitively, though, the sequence *the girl's uncle* is not a constituent in (6a), where the uncle belongs to both the boy and the girl.