9 Syntax

After a lecture on cosmology and the structure of the solar system, William James was accosted by a little old lady who told him that his view of the earth rotating round the sun was wrong.

"I've got a better theory," said the little old lady.

"And what is that, madam?" inquired James politely.

"That we live on a crust of earth which is on the back of a giant turtle."

"If your theory is correct, madam," he asked, "what does this turtle stand on?"

"You're a very clever man, Mr. James, and that's a very good question," replied the little old lady, "but I have an answer to it. And it's this: the first turtle stands on the back of a second, far larger, turtle, who stands directly under him."

"But what does this second turtle stand on?" persisted James patiently.

To this, the little old lady crowed triumphantly, "It's no use, Mr. James, it's turtles all the way down." **Adapted from Ross (1967)**

In the preceding chapter, we moved from the general categories and concepts of traditional grammar to more specific methods of describing the structure of phrases and sentences. When we concentrate on the structure and ordering of components within a sentence, we are studying the **syntax** of a language. The word 'syntax' comes originally from Greek and literally means 'a putting together' or 'arrangement'. In earlier approaches to the analysis of syntax, as we saw in chapter 8, there was an attempt to produce an accurate description of the sequence or ordering 'arrangement' of elements in the linear structure of the sentence. While this remains one of the goals of syntactic analysis, more recent work in syntax has taken a rather different approach in attempting to account for the types of 'arrangement' we observe in the structure of sentences.

Generative grammar

Inspired by the original work of Noam Chomsky, linguists have attempted to produce a particular type of grammar that has a very explicit system of rules specifying what combinations of basic elements would result in well-formed sentences. This very explicit system of rules, it was proposed, would have a lot in common with the types of rules found in mathematics. Indeed, Chomsky seems to have taken the view that the essential structure of language can be expressed in mathematical terms: "I will consider a language to be a set (finite

or infinite) of sentences" (1957: 13). This isn't how most people would describe a language, but it is a good definition to keep in mind as we try to take a close look at how the syntax (and only the syntax) of a language might be analyzed.

The mathematical perspective helps to explain the meaning of the term 'generative', which is used to describe this type of grammar. In basic algebra, if we have an expression such as $3x + 2y$, and we give x and y the value of any whole number, then that simple algebraic expression can 'generate' an endless set of values by following the simple rules of arithmetic. When $x = 5$ and $y = 10$, the result is 35. When $x = 2$ and $y = 1$, the result is 8. These results will follow directly and predictably from applying the explicit rules. The endless set of such results is 'generated' by the operation of the rules. If the sentences of a language can be seen as a comparable set, then there must be a set of explicit rules that can produce all those sentences. Such a set of explicit rules is a **generative grammar**.

Syntactic structures

A generative grammar defines the syntactic structures of a language. The grammar will generate all the well-formed syntactic structures (e.g. sentences) of the language and will not generate any ill-formed structures. This has been called the 'all and only' criterion, that is, *all* the grammatical sentences and *only* the grammatical sentences will be produced.

The grammar will have a finite (i.e. limited) number of rules, but will be capable of generating an infinite number of well-formed structures. In this way, the productivity of language (i.e. our ability to create totally novel yet grammatically accurate sentences) would be captured within the grammar.

The grammar should also be capable of revealing the basis of two other phenomena: first, how some superficially different sentences are closely related and, second, how some superficially similar sentences are in fact different.

Deep and surface structure

Two superficially different sentences are shown in these examples.

Charlie broke the window.
The window was broken by Charlie.

In traditional grammar, the first is called an active sentence, focusing on what *Charlie* did, and the second is a passive sentence, focusing on *The window* and what happened to it. The distinction between them is a difference in their **surface structure**, that is, the different syntactic forms they have as individual English sentences. However, this superficial difference in form disguises the fact that the two sentences are very closely related, even identical, at some less superficial level.

This other 'underlying' level, where the basic components (noun phrase + verb + noun phrase) shared by the two sentences can be represented, is called their **deep structure**. The deep structure is an abstract level of structural organization in which all the elements determining structural interpretation are represented. That same deep structure can be the source of many other surface structures such as *It was Charlie who broke the window* and *Was the window broken by Charlie?* In short, the grammar must be capable of showing how a single underlying abstract representation can become different surface structures.

Structural ambiguity

Let's say we had two distinct deep structures. One expresses the idea that 'Annie had an umbrella and she whacked a man with it.' The other expresses the idea that 'Annie whacked a man and the man happened to be carrying an umbrella.' Now, these two different versions of events can actually be expressed in the same surface structure form: *Annie whacked a man with an umbrella*. This sentence provides an example of **structural ambiguity**. It has two distinct underlying interpretations that have to be represented differently in deep structure.

The comedian Groucho Marx knew how to have fun with structural ambiguity. In the film *Animal Crackers*, he first says *I once shot an elephant in my pajamas*, then follows it with *How he got into my pajamas I'll never know*. In the non-funny interpretation, part of the underlying structure of the first sentence could be something like: 'I shot an elephant (while I was) in my pajamas.' In the other (ho, ho) interpretation, part of the underlying structure would be something like: 'I shot an elephant (which was) in my pajamas.' There are two different underlying structures with the same surface structure.

Phrases can also be structurally ambiguous, as in expressions like *small boys and girls*. The underlying interpretation can be either 'small boys and (small) girls' or 'small boys and (all) girls'. The grammar will have to be capable of showing the structural distinction between these underlying representations.

Recursion

The rules of the grammar will also need the crucial property of **recursion**. Recursive ('repeatable any number of times') rules have the capacity to be applied more than once in generating a structure. For example, we can have one prepositional phrase describing location (*on the table*) in the sentence *The gun was on the table*. We can also repeat this type of phrase, using different words (*near the window*), for as long as the sentence still makes sense (*in the bedroom*). So, in order to generate a sentence such as *The gun was on the table near the window in the bedroom*, we must be able to repeat the rule that creates a prepositional phrase over and over again.

We must also be able to put sentences inside other sentences. For example, when we produce a sentence such as *Cathy knew that Mary helped George*, we do so with the sentence *Mary helped George* inside it. And those two sentences can be generated inside another sentence such as *John believed that Cathy knew that Mary helped George*. In principle, there is no end to the recursion that would produce ever longer versions of complex sentences with this structure.

Basically, the grammar will have to capture the fact that a sentence can have another sentence inside it or that a phrase can be repeated as often as required. We should note that recursion of this type is not only a feature of grammar, but can also be an essential part of a theory of cosmic structure, as in the role of turtles in one little old lady's view of the universe.

Symbols used in syntactic description

Having reviewed some important concepts in the study of syntax, we can now look at some of the ways in which syntactic analysis is presented. We have already encountered some symbols in chapter 8 as abbreviations for syntactic categories. Examples are 'S' (= sentence), 'NP' (= noun phrase), 'N' (= noun), 'Art' (= article), and so on. There are three more symbols that are commonly used in syntactic description.

The first is in the form of an arrow →. It can be interpreted as 'consists of' or 'rewrites as'. It is typically used in the following type of rule:

NP → Art N

This is simply a shorthand way of saying that a noun phrase (NP) such as *the dog* consists of or rewrites as (→) an article (Art) *the* and a noun (N) *dog*.

The second symbol is a pair of round brackets **( )**. Whatever occurs inside these round brackets will be treated as an optional constituent. For example, we can describe an object as *the dog* or *the small dog*. We can say that both *the dog* and *the small dog* are examples of the category noun phrase (NP). When we want to use a noun phrase in English, we can include an adjective (Adj) such as *small*, but we don't have to. It's an optional constituent in a grammatically well-formed noun phrase. We can represent this observation in the following type of rule:

NP → Art (Adj) N

This shorthand notation expresses the idea that a noun phrase rewrites as (→) an article (Art) and a noun (N), with the option of including an adjective (Adj) in a specific position between them. We use the round brackets to indicate that the adjective is optional. So, we can use this notation to generate *the dog*, *the small dog*, *a cat*, *a big cat*, *the book*, *a boring book* and an endless number of other similar noun phrases.

The third symbol is in the form of curly brackets { }.These indicate that only one of the elements enclosed within the curly brackets must be selected. We use

these types of brackets when we want to indicate that there is a choice from two or more constituents. For example, we saw in chapter 8 that a noun phrase can consist of an expression such as *the dog* (article plus noun), or *it* (pronoun), or *Cathy* (proper noun). Using the abbreviations 'Pro' (for pronoun) and 'PN' (for proper noun), we can try to capture this observation about English with three separate rules, as shown on the left. However, it is more succinct to write one rule, as shown in the middle or on the right, using curly brackets and including exactly the same information.

NP → Art N
NP → Pro
NP → PN

NP → { **Art N** / **Pro** / **PN** }

NP → {Art N, Pro, PN}

It is important to remember that, although there are three constituents inside these curly brackets, only one of them can be selected on any occasion.

This list of symbols and abbreviations is summarized below.

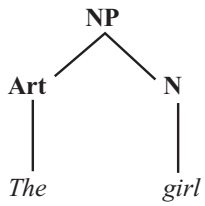
S sentence	**NP** noun phrase	**PN** proper noun
N noun	**VP** verb phrase	**Adv** adverb
V verb	**Adj** adjective	**Prep** preposition
Art article	**Pro** pronoun	**PP** prepositional phrase

*	'ungrammatical sentence'
→	'consists of' or 'rewrites as'
()	'optional constituent'
{}	'one and only one of these constituents must be selected'

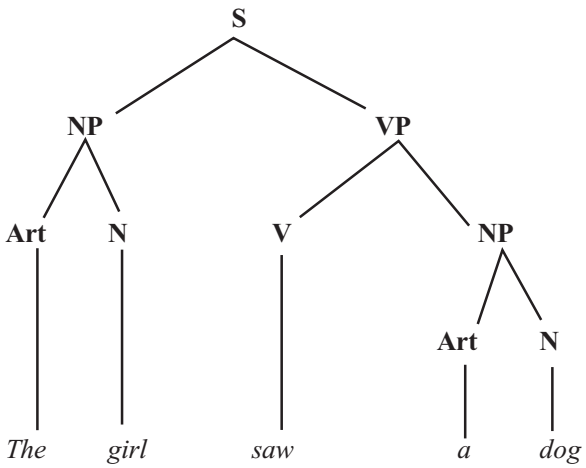
Tree diagrams

In chapter 8, we looked at ways to describe the structure of sentences by first concentrating on the linear sequence of constituents, then noting how our diagrams could capture some aspects of the hierarchical organization of those structures. To create a more explicit representation of the hierarchical organization of one structure, shown in a labeled and bracketed format on the left below, we can use a **tree diagram**, shown on the right below.

NP
[**Art** [*The*] **N** [*girl*]]



Although this kind of 'tree', with its 'branches', shown on the right, seems to grow down rather than up, it functions rather well as a diagram representing all the grammatical information found in the other analysis on the left. It also shows very explicitly that there are different levels in the analysis. That is, there is a level of analysis at which a constituent such as NP is represented and a different, lower, level at which a constituent such as N is represented. This type of hierarchical organization can be illustrated in a tree diagram for a whole sentence, beginning at the top with S.



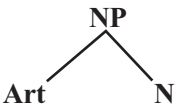
If we start at the top of this tree diagram, we begin with a sentence (S) and divide it into two constituents (NP and VP). In turn, the NP constituent is divided into two other constituents (Art and N). Finally, one word is selected that fits the label Art (*the*) and another that fits N (*girl*).

Phrase structure rules

We can think of this tree diagram format in two different ways. In one way, we can simply treat it as a static representation of the structure of the sentence shown at the bottom of the diagram. We could then propose that, for every single sentence in English, a tree diagram of this type could be drawn. An alternative view is to treat the tree diagram as a 'dynamic' format, in the sense that it represents a way of generating not only that one sentence, but a very large number of other sentences with similar structures.

This second approach is very appealing because it would enable us to generate a very large number of sentences with what look like a very small number of rules. These rules are called **phrase structure rules**. As the name suggests, these rules state that the structure of a phrase of a specific type will consist of one or more constituents in a particular order. We can use phrase structure rules to present the information of the tree diagram in another format, as we saw

when we introduced some new symbols earlier. That is, the information shown in the tree diagram on the left can be expressed in the phrase structure rule on the right.



NP → Art N

According to this rule, "a noun phrase rewrites as an article followed by a noun".

The first rule in the following set of simple (and necessarily incomplete) phrase structure rules states that "a sentence rewrites as a noun phrase and a verb phrase". The second rule states that "a noun phrase rewrites as either an article, an optional adjective and a noun or a pronoun or a proper noun".

S → NP VP
NP → {Art (Adj) N, Pro, PN}
VP → V NP (PP) (Adv)
PP → Prep NP

Lexical rules

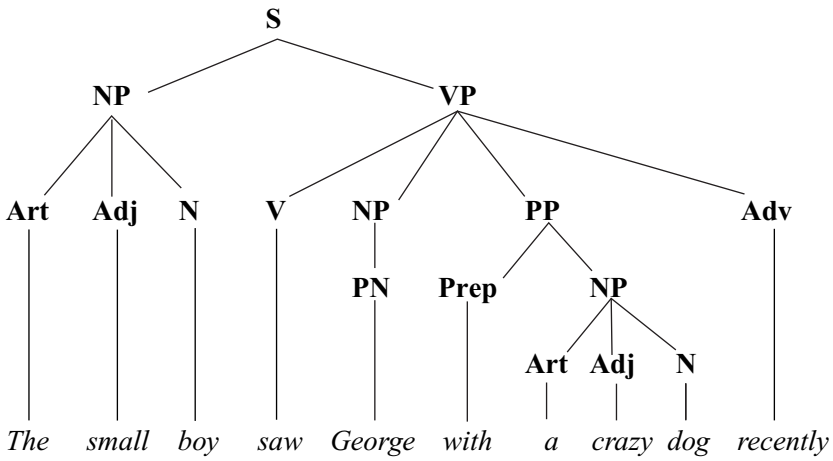
Phrase structure rules generate structures. In order to turn those structures into recognizable English, we also need **lexical rules** that specify which words can be used when we rewrite constituents such as N. The first rule in the following set states that "a proper noun rewrites as *Mary* or *George*". (It's a very small world.)

PN → {*Mary*, *George*}
N → {*girl*, *dog*, *boy*}
Art → {*a*, *the*}
Pro → {*it*, *you*}
V → {*followed*, *helped*, *saw*}
Adj → {*small*, *crazy*}
Prep → {*near*, *with*}
Adv → {*recently*, *yesterday*}

We can rely on these rules to generate the grammatical sentences 1–7 below, but not the ungrammatical sentences 8–10.

1 *The girl followed the boy.*
2 *A boy helped the dog.*
3 *You saw it yesterday.*
4 *Mary helped George. recently.*
5 *George saw a dog.*
6 *A small dog followed Mary.*
7 *The small boy saw George with a crazy dog recently.*
8 **You it saw.*
9 **Mary helped boy.*
10 **Followed the dog.*

As a way of visualizing how the phrase structure rules form the basis of these sentences, we can draw the tree diagram for sentence 7.



The very small set of phrase structure rules and lexical rules just described is a sample of what might become a more complex phrase structure grammar of English, with many more parts. However, we have still to incorporate recursion.

Back to recursion

The simple phrase structure rules listed earlier have no recursive elements. Each time we start to create an S, we only create a single S (sentence structure). We actually need to be able to include sentence structures within other sentence structures. In traditional grammar, these 'sentence structures' were described as 'clauses'. We know, for example, that *Mary helped George* is a sentence. We can put this sentence inside another sentence beginning *Cathy knew that [Mary helped George]*. And, being tediously recursive, we can put this sentence inside another sentence beginning *John believed that [Cathy knew that [Mary helped George]]*.

In these sentences, two new proper nouns and two new verbs have been used. We have to expand our earlier set of lexical rules to include PN → {*Cathy*, *John*} and V → {*believed*, *knew*}. After verbs such as *believe* and *know*, as in these examples, the word *that* introduces a 'complement phrase'.

Mary helped George.
Cathy knew that Mary helped George.
John believed that Cathy knew that Mary helped George.

Complement phrases

The word *that*, as used in these examples, is called a **complementizer** (C). The role of *that* as a complementizer is to introduce a **complement phrase** (CP). For example, in the second sentence (*Cathy knew . . .*), we can identify one CP which contains *that* plus *Mary helped George*. We already know that *Mary*

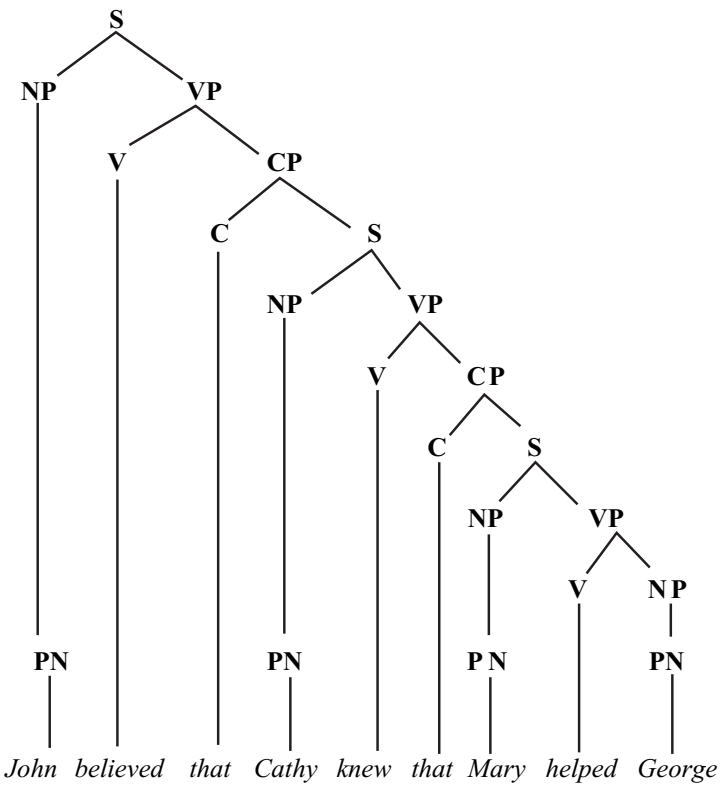
helped George is a sentence (S). So, we are now in a position to define a CP in the following way: "a complement phrase rewrites as a complementizer and a sentence", or CP → C S.

We can also see from the same sentence that the complement phrase (CP) comes after a verb (V) *knew*. This means that we are using the CP as part of a verb phrase (VP), as in *knew that Mary helped George*. So, there must be another rule that says: "a verb phrase rewrites as a verb and complement phrase", or VP → V CP.

If we now look at these two new rules in conjunction with an earlier rule, we can see how recursion is built into the grammar.

S → NP VP
VP → V CP
CP → C S

We begin with S on the left and, as we rewrite symbols, we eventually have S on the right, allowing us to go back to the beginning and go through the set of rules again (and again). This means that we can, in principle, use these rules to create an endless 'sentence' containing other 'sentences'. In practice, it allows us to draw the following tree diagram and provide a clear representation of the syntactic structure of this one fairly complex sentence.

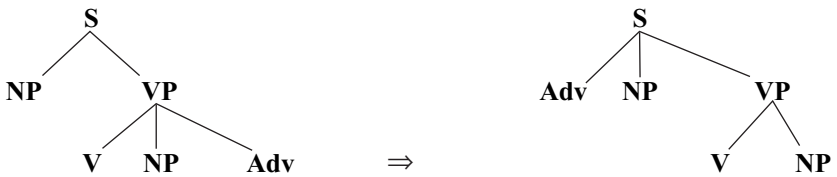


Transformational rules

The phrase structure rules (without the lexical rules) that we have described can be treated as a representation of the 'underlying' or deep structures of sentences in English. One feature of these underlying structures is that they will always generate sentences with a fixed word order. For example, if we follow the rules illustrated earlier, adverbs will always come at the end of their sentences. That would seem to work for the first sentence below, but how would we get the second sentence?

Mary saw George recently.
Recently Mary saw George.

We can think of the adverb *recently* as having been 'moved' to the beginning of the second sentence. In order to make this possible in the grammar, we need other rules that will change or move constituents in the structures derived from the phrase structure rules. These are called **transformational rules**. Essentially what they do is take a specific part of structure, like a branch of the tree, away from one part of the tree diagram and attach it to a different part. As shown below, we use the symbol ⇒ to indicate that a transformational rule is being used to derive a new structure from the basic structure.



Mary saw George recently *Recently Mary saw George*

For this particular rule, we would have to specify which type of constituent can be moved in this way, as well as from where and to where.

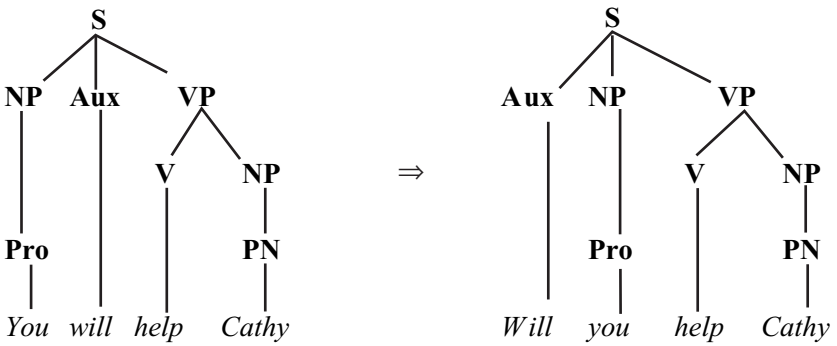
We also use a transformational rule to derive English question structures of the type illustrated in the second sentence below.

You will help Cathy.
Will you help Cathy?

In order to describe this process, we need to expand our phrase structure rules to include an **auxiliary verb** (Aux) such as *will* as part of the sentence. This new rule is written as: S → NP Aux VP. Although there are other forms of auxiliary verbs in English, a rudimentary lexical rule might be as follows: Aux → {*can*, *should*, *will*}.

With these components, we can specify the transformational rule that creates this basic type of English question as: NP Aux VP ⇒ Aux NP VP. We can

illustrate this change in terms of the tree on the right below being derived from the tree on the left.



Using the rules we have just described, we could also generate *Can John see it?* and *Should Mary follow the small boy?*, and many others. These would all be surface variations of a single underlying structure.

There are many more rules and concepts involved in the analysis of syntax. (We've barely scratched the surface structures.) However, having explored some of the basic issues and methods of syntactic analysis in order to talk about 'structure' in language, we must move on to consider how we might incorporate the analysis of 'meaning' in the study of language.

Study questions

1 What is the 'all and only' criterion in generative grammar?

2 In what ways are these expressions structurally ambiguous?
   (a) *The parents of the bride and groom were waiting outside.*
   (b) *We met an English history teacher.*
   (c) *Flying planes can be dangerous*.
   (d) *The students complained to everyone that they couldn't understand*.

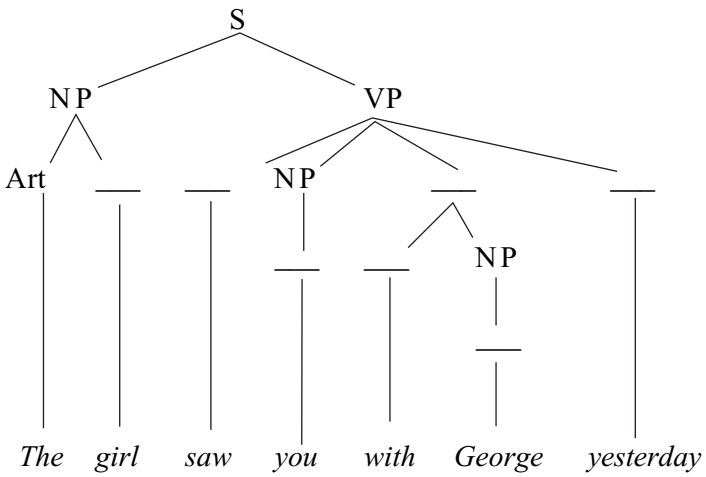
3 Do phrase structure rules represent deep structure or surface structure?

4 Which of the following expressions would be generated by this phrase structure rule: NP → Art (Adj) N?
   (a) *a house*
   (b) *the old roof*
   (c) *a big window*
   (d) *the garden*

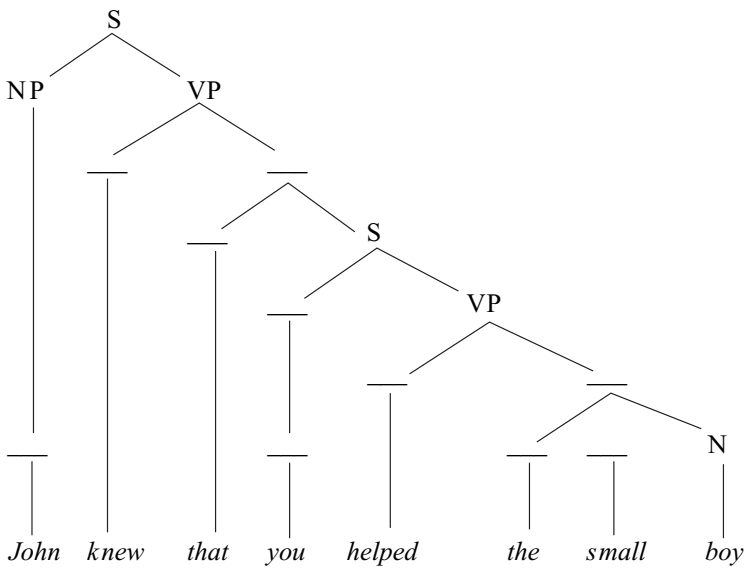
5 Which of these sentences would be generated after applying the rule: NP Aux VP ⇒ Aux NP VP?
   (a) *John will follow Mary.*
   (b) *Can George see the small dog?*
   (c) *You knew that Cathy helped the boy*.
   (d) *Should you believe that Mary saw it?*

6 Using information from the phrase structure rules presented in this chapter, complete the following tree diagrams.

(a)



(b)



Research tasks

A What is the distinction made between 'competence' and 'performance' in the study of syntax?

B What is meant by the expression 'an embedded structure'? Were there any examples in this chapter?

C The following simplified set of phrase structure rules describes some aspects of the syntax of a language called Ewe, spoken in West Africa. Based on these rules, which of the following sentences (1–10) should have an asterisk * before them?

S → NP VP
NP → N (Art)
VP → V NP

N → {*oge*, *ika*, *amu*}
Art → *ye*
V → {*xa*, *vo*}

1 *Oge xa ika*
2 *Ye amu vo oge*
3 *Ika oge xa ye*
4 *Oge ye vo ika ye*
5 *Amu xa oge*
6 *Vo oge ika*
7 *Amu ye vo ika*
8 *Ye ika xa ye oge*
9 *Xa amu ye*
10 *Oge ye xa amu*

D Using these simple phrase structure rules for Scottish Gaelic, identify (with *) the two ungrammatical sentences below and draw tree diagrams for the two grammatical sentences.

S → V NP NP
Art → *an*
N → {*cu*, *gille*}
PN → {*Calum*, *Tearlach*}

NP → {Art N (Adj), PN}
Adj → {*beag*, *mor*}
V → {*bhuail*, *chunnaic*}

1 *Calum chunnaic an gille.*
2 *Bhuail an beag cu Tearlach.*
3 *Bhuail an gille mor an cu.*
4 *Chunnaic Tearlach an gille.*

Discussion topics/projects

I There is a principle of syntax called 'structure dependency' that is often used to show that the rules of language structure depend on hierarchical organization and not on linear position. For example, someone trying to learn English might be tempted to think that questions of the type in (ii) are formed simply by moving the second word in a statement (i) to become the first word of a question (ii).

(i) *Shaggy* ***is*** *tired.* (ii) ***Is*** *Shaggy tired?*
You ***will*** *help him.* ***Will*** *you help him?*

Using the sentences 1–4, try to decide if this is the best way to describe how all of these English questions are formed and, if it is not, try to formulate a better rule.

1 *Are the exercises in this book too easy?*
2 *Is the cat that is missing called Blackie?*
3 *Will the price of the new book you've ordered be really expensive?*
4 *Was the guy who scored the winning goal in the final playing for love or money?*

(For background reading, see chapter 4 of Fromkin *et al*., 2003.)

II We could propose that passive sentences (*George was helped by Mary*) are derived from active structures (*Mary helped George*) via a transformational

rule such as the following:

(active) NP_1 V $NP_2 \Rightarrow NP_2$ *be* V-*ed* by NP_1 (passive)

Note that the tense, past or present, of the V (e.g. *helped*) in the active structure determines the tense of *be* in the passive structure (e.g. ***was*** *helped*).

Which of the following active sentences can be transformed into passive sentences using this rule? What prevents the rule from working in the other cases?

1 *The dog chased the cat.*
2 *Snow White kissed Grumpy.*
3 *He loves them.*
4 *Betsy borrowed some money from Christopher.*
5 *The team played badly.*
6 *The bank manager laughed.*
7 *They have two children.*
8 *The duckling became a swan.*
9 *Someone mentioned that you played basketball.*
10 *The police will arrest violent demonstrators.*

(For background reading, see chapter 5 of Morenberg, 2003).

Further reading

For another short introduction to syntax, see chapter 5 of Finegan (2004) or chapter 6 of Hudson (2000). Introductory textbooks with exercises are Fabb (1994), Morenberg (2003) and Thomas (1993). Other accessible texts include Brown & Miller (1991), Burton-Roberts (1997), Miller (2002) and Tallerman (1998). On more theoretical issues, see Borsley (1995) or Green & Morgan (2001). A good overview of Chomsky's early work is Lyons (1991) and his later work is described in Radford (1997, 2004). For a more general review of his ideas, see Chomsky (2002).