1

What is syntax?

1.1 SOME CONCEPTS AND MISCONCEPTIONS

1.1.1 What is the study of syntax about?

This book is about the property of human language known as syntax. 'Syntax' means 'sentence construction': how words group together to make phrases and sentences. Some people also use the term GRAMMAR to mean the same as syntax, although most linguists follow the more recent practice whereby the grammar of a language includes all of its organizing principles: information about the sound system, about the form of words, how we adjust language according to context, and so on; syntax is only one part of this grammar.

The term 'syntax' is also used to mean the *study* of the syntactic properties of languages. In this sense, it's used in the same way as we use 'stylistics' to mean the study of literary style. We're going to be studying how languages organize their syntax, so the scope of our study includes the classification of words, the order of words in phrases and sentences, the structure of phrases and sentences, and the different sentence constructions that languages use. We'll be looking at examples of sentence structure from many different languages in this book, some related to English and others not. All languages have syntax, although that syntax may look radically different from that of English. My aim is to help you understand the way syntax works in languages, and to introduce the most important syntactic concepts and technical terms which you'll need in order to see how syntax works. We'll encounter many grammatical terms, including 'noun', 'verb', 'preposition', 'relative clause', 'subject', 'nominative', 'agreement' and 'passive'. I don't expect you to know the meanings of any of these in advance. Often, terms are not formally defined when they are used for the first time, but they are illustrated so you can understand the concept, in preparation for a fuller discussion later on. More complex terms and concepts (such as 'case' and 'agreement') are discussed more than once, and a picture of their meaning is built up over several chapters.

To help you understand what the study of syntax is about, we first need to discuss some things it isn't about. When you read that 'syntax' is part of 'grammar', you may have certain impressions that differ from the aims of this book. So first, although we will be talking about grammar, this is not a DESCRIPTIVE GRAMMAR of English or any other language. Such books are certainly available, but they usually aim to catalogue the regularities and peculiarities of one language rather than looking at

the organizing principles of language in general. Second, I won't be trying to improve your 'grammar' of English. A PRESCRIPTIVE GRAMMAR (one that prescribes how the author thinks you should speak) might aim to teach you where to use *who* and *whom*; or when to say *me and Kim* and when to say *Kim and I*; it might tell you not to say *different than* or *different to*, or tell you to avoid split infinitives such as *to boldly go*. These things aren't on our agenda, because they're essentially a matter of taste – they are social, not linguistic, matters.

In fact, as a linguist, my view is that if you're a native speaker of English, no matter what your dialect, then you already know English grammar perfectly. And if you're a native speaker of a different language, then you know the grammar of that language perfectly. By this, I don't mean that you know (consciously) a few prescriptive rules, such as those mentioned in the last paragraph, but that you know (unconsciously) the much more impressive mental grammar of your own language – as do all its native speakers. Although we've all learnt this grammar, we can think of it as knowledge that we've never been taught, and it's also knowledge that we can't take out and examine. By the age of around seven, children have a fairly complete knowledge of the grammar of their native languages, and much of what happens after that age is learning more vocabulary. We can think of this as parallel to 'learning' how to walk. Children can't be taught to walk; we all do it naturally when we're ready, and we can't say how we do it. Even if we come to understand exactly what muscle movements are required, and what brain circuitry is involved, we still don't 'know' how we walk. Learning our native language is just the same: it happens without outside intervention and the resulting knowledge is inaccessible to us.

Here, you may object that you *were* taught the grammar of your native language. Perhaps you think that your parents set about teaching you it, or that you learnt it at school. But this is a misconception. All normally developing children in every culture learn their native language or languages to perfection without any formal teaching. Nothing more is required than the simple exposure to ordinary, live, human language within a society. To test whether this is true, we just need to ask if all cultures teach their children 'grammar'. Since the answer is a resounding 'no', we can be sure that all children must be capable of constructing a mental grammar of their native languages without any formal instruction. Most linguists now believe that, in order to do this, human infants are born pre-programmed to learn language, in much the same way as we are pre-programmed to walk upright. All that's needed for language to emerge is appropriate input data – hearing language (or seeing it; sign languages are full languages too) and taking part in interactions within the home and the wider society.

So if you weren't taught the grammar of your native language, what was it you were being taught when your parents tried to get you not to say things like *I ain't done nowt wrong*, or *He's more happier than what I am*, or when your school teachers tried to stop you from using a preposition to end a sentence with? (Like the sentence I just wrote.) Again, consider learning to walk. Although children learn to do this perfectly without any parental instruction, their parents might not like the way the child slouches along, or scuffs the toes of their shoes on the ground. They may tell the child to stand up straight, or to stop wearing out their shoes. It's not that the child's

way doesn't function properly, it just doesn't conform to someone's idea of what is aesthetic, or classy. In just the same way, some people have the idea that certain forms of language are more beautiful, or classier, or are simply 'correct'. But the belief that some forms of language are better than others has no linguistic basis. Since we often make social judgements about people based on their accent or dialect, we tend to transfer these judgements to their form of language. We may then think that some forms are undesirable, that some are 'good' and some 'bad'. For a linguist, however, dialectal forms of a language don't equate to 'bad grammar'.

Again, you may object here that examples of NON-STANDARD English, such as those italicized in the last paragraph, or things like *We done it well good*, are sloppy speech, or perhaps illogical. This appeal to logic and precision makes prescriptive grammar seem to be on a higher plane than if it's all down to social prejudice. So let's examine the logic argument more closely, and see if it bears scrutiny. Many speakers of English are taught that 'two negatives make a positive', so that forms like (1) 'really' mean *I did something wrong*:

(1)   I didn't do nothing wrong.

Of course, this isn't true. First, a speaker who uses a sentence like (1) doesn't *intend* it to mean *I did something wrong*. Neither would any of their addressees, however much they despise the double negative, understand (1) to mean *I did something wrong*. Second, there are languages, such as French and Breton, that use a double negative as STANDARD, not a dialectal form, as (2) illustrates:[1]

(2)   Je **ne**   mange **jamais** de viande.   (French)
       I   NEGATIVE   eat   never   of meat
       'I never eat meat.'

Example (2) shows that in Standard French the negative has two parts: in addition to the little negative word *ne* there's another negative word *jamais*, 'never'. Middle English (the English of roughly 1100 to 1500) also had a double negative. Ironically for the 'logic' argument, the variety of French that has the double negative is the most formal and prestigious variety, whereas colloquial French typically drops the initial negative word.

Another non-standard feature of certain English dialects which doesn't conform to prescriptive notions is illustrated in (3), from a northern (British) English dialect:

(3)   I aren't going with you.

Here, the logic argument runs like this: you can't say **I are not* (the star or asterisk is a convention used in linguistics to indicate an impossible sentence), so the contracted

---

1. Section 1.2.2 explains in detail how to read linguistic examples. You don't need to know any French to see the point that example (2) is making. The technique you should employ is to read the English translation, then carefully examine the second line of the example, which is the literal translation of the original language.

form *I aren't* must be wrong too. It's true that speakers who accept (3) don't ever say *I are not*. But the argument is flawed: standard English is just as illogical. Look how the statement in (4a) is turned into a question in (4b):

> **(4)**   a.   **I'm not** going with you.
>   b.   **Aren't I** going with you?

Example (4) does not conform to the usual rules of English grammar, which form questions by inverting the word order in *I can't* to give *can't I*, and *I should* to give *should I*, and so on. Given these rules, the 'logically' expected form in (4b) would be *amn't I* (and in fact this form is found in some dialects). If the standard English in (4) fails to follow the usual rules, then we can hardly criticize (3) for lack of logic. And since *aren't I* is OK, there's no logical reason for dismissing *I aren't*. The dialects that allow either *I aren't* or *amn't I* could actually be considered more logical than standard English, since they follow the general rule, while the standard dialect, in (4), has an irregularity.

It is clear, then, that socially stigmatized forms of language are potentially just as 'logical' as standard English. Speakers of non-standard dialects are, of course, following a set of mental rules, in just the same way that speakers of the most prestigious dialects are. The various dialects of a language in fact share the majority of their rules, and diverge in very few areas, but the extent of the differences tends to be exaggerated because they arouse such strong feelings. In sum, speakers of prestige dialects may feel that only their variety of English is 'grammatically correct', but these views cannot be defended on either logical or linguistic grounds.

If, by way of contrast, some speaker of English produced examples like (5), then we could justifiably claim that they were speaking ungrammatically:

> **(5)**   *I do didn't wrong anything.
>   *Do wrong didn't anything I.

Such examples completely contravene the mental rules of all dialects of English. We all agree on this, yet speakers of English haven't been taught that the sentences in (5) are bad. Our judgements must therefore be part of the shared mental grammar of English.

Most of the rules of this mental grammar are never dealt with by prescriptive or teaching grammars. So no grammar of English would ever explain that, although we can say both (6a) and (6b), we can't have questions like (7) (the gap ___ indicates an understood but 'missing' element, represented by the question word *what*):

> **(6)**   a.   They're eating eggs and chips.
>   b.   What are they eating ___?
>
> **(7)**   *What are they eating eggs and ___?

The rules that make (7) impossible are so immutable and fundamental that they hardly seem to count as a subject for discussion: native speakers never stop to

wonder why (7) is not possible. Not only are examples like (7) ungrammatical in English (i.e. they sound impossible to native speakers), they are ungrammatical in Welsh, as in (8):

(8)   *Beth   maen   nhw   yn   bwyta   wyau   a   ___ ?   (Welsh)
       what   are   they   in   eat   eggs   and   ___ ?
       *'What are they eating eggs and ___ ?'

In fact, the equivalents to (7) and (8) are generally ungrammatical in the world's languages. It seems likely, then, that many of the unconsciously 'known' rules of individual languages like English or Welsh are actually UNIVERSAL – common to all languages.

>>>>>>>>>>>>>>>>>>>>>>>>

Before reading further, note that English does have a way of expressing what (7) would mean if it were grammatical – in other words, a way of expressing the question you would ask if you wanted to know what it was that they were eating with their eggs. How is this question formed?

<<<<<<<<<<<<<<<<<<<<<<<

You could ask: *They are eating eggs and **what**?* (with heavy emphasis on the *what*).
   The fact that certain organizing rules and principles in language are universal leads many linguists to conclude that human beings have an INNATE language faculty – that is, one we are born with. We can't examine this directly, and we still know relatively little about what brain circuitry is involved, but we do know that there must be something unique to humans in this regard. All normal children learn at least one language, but no other animals have anything like language as a natural communication system, nor are they able to learn a human language, even under intense instruction. To try to understand the language faculty, we examine its output – namely the structures of natural languages. So by looking at syntax we hope to discover the common properties between languages, and maybe even ultimately to discover something about the workings of the human brain.
   As well as looking for absolutely universal principles, linguists are interested in discovering what types of construction are possible (and impossible) in the world's languages. We look for recurring patterns, and often find that amazingly similar constructions appear in unrelated languages. In the next paragraph, I give an example of this type which compares Indonesian and English. You don't have to know anything about Indonesian to get the point being made, but if the idea of looking closely at exotic languages seems too daunting at this stage, come back to the examples after you've read Section 1.2. The row of >>> arrows marks the start of a section of the text in which the reader is invited to work something out, as in the example just given; the <<< arrows mark the end of that section, and where

necessary, the exercise is followed by a suggested answer. Here, the task is simply to examine all the sentences, and try to follow the argument.

>>>>>>>>>>>>>>>>>>>>>>>>>

In English we can say either (9a) or (9b) – they alternate freely. In (9b), the word *Hasan* appears earlier in the sentence; let's say that in (9b), *Hasan* has been PROMOTED in the sentence:

| **(9)** | a. | Ali sent the letter to Hasan. | (Indonesian) |
| | b. | Ali sent Hasan the letter. | |

In Indonesian, we find the same alternation, shown in (10). If you're reading this before the discussion on the use of linguistic examples in Section 1.2, please remember to concentrate particularly on the second line of each example: the literal translation. The main 'foreign' feature in (10) is *surat itu* 'letter the' where English has the word order 'the letter'; otherwise, the word order in the two Indonesian examples is the same as that of the two English examples in (9):

(10)	a.	Ali	meng-kirim	surat	itu	kepada	Hasan.
		Ali	send	letter	the	to	Hasan
		'Ali sent the letter to Hasan.'
	b.	Ali	meng-kirim-**kan**	Hasan	surat	itu.
		Ali	send	Hasan	letter	the
		'Ali sent Hasan the letter.'

In (10b), we find an ending -*kan* on the word for 'send': this ending indicates in Indonesian that the word *Hasan* has been promoted. English has no equivalent to -*kan*.

Now look again at the English in (9). When *Hasan* is in the promoted position in (9b), we can promote it to an even higher position in the sentence, giving (11). We indicate the position that *Hasan* is understood to have moved from with the gap __. In (11), there is also a change from *sent* to *was sent*, which signals the further promotion of *Hasan*. To understand why a language would need to indicate this promotion of some part of the sentence, think about the difference in meaning between *Hasan sent the letter* and *Hasan was sent the letter*.

(11)   Hasan was sent ____ the letter by Ali.

If we start with (9a), however, where *Hasan* is not in a promoted position, then trying to promote it from there *directly* to the very highest position in the sentence would give (12): again, I show the position the word *Hasan* has moved from with the gap. But (12) is not a possible sentence of English (as indicated by the asterisk):
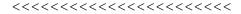
(12)   *Hasan was sent the letter to ____ by Ali.

So if the word *Hasan* is already promoted, as in (9b), then it can move again, giving (11). Otherwise, promotion of *Hasan* is impossible, as (12) shows. In fact, it seems like the promotion has to occur in stages, rather than in one single jump straight to the front of the sentence. Perhaps you're thinking, maybe it's just a question of getting rid of the *to* in (12); then it'd be fine. But if we look at (13) and (14), the Indonesian equivalents to (11) and (12), we get some strong clues that this is not the case. Note that the change from *meng-kirim* in (10) to *di-kirim* in (13) and (14) is equivalent to the change in English from *sent* to *was sent*:

(13) Hasan   di-kirim-**kan**   surat   itu   oleh   Ali.
     Hasan   was-sent   letter   the   by   Ali
     'Hasan was sent the letter by Ali.'

(14) *Hasan   di-kirim   surat   itu   (kepada)   oleh   Ali.
      Hasan   was-sent   letter   the   (to)   by   Ali
     *'Hasan was sent the letter to by Ali.'

Just as in English, one construction is fine, the other impossible. What makes the difference? In the Indonesian, we can tell that it can't be anything to do with the word for 'to' (*kepada*), because (14) is impossible with or without that word – the parentheses (...) mean that whether or not *kepada* is included makes no difference to the acceptability of the sentence. The reason only (13) is acceptable is that we have to start off with (10b) to get there – the version in which *Hasan* has already been promoted once. And we know that (13) does indeed come via (10b) because the word which means 'was sent', *di-kirim-kan*, has that ending *-kan* which shows that *Hasan* has been promoted – whereas *di-kirim* in (14) doesn't. So we could hypothesize that English probably works in the same way. Although there's nothing in English to mark the first promotion of *Hasan* in (9b), it's likely that just as in Indonesian, it's the promotion that's the distinguishing factor between the grammatical example in (11) and the ungrammatical one in (12).

   At this stage, I hope to have shown that two totally unrelated languages can display some remarkably similar syntactic behaviour. Finally, please note that although this section was rather technical, you should be able to understand it if you read it through more than once, stopping to work out each stage as you go. This tip will also be helpful throughout the book.

<<<<<<<<<<<<<<<<<<<<<<<

1.1.2   Language change

Speakers of established languages such as English often dislike changes occurring within their own language, believing that change equates with declining standards. In fact, however, the grammar of all languages changes over time, and no amount of intervention by prescriptive grammarians or language academies can prevent this. In this section, I look at some examples from the history of English, and then at more

recent changes. The examples of Middle English in (15) are from the prologue to Chaucer's *Wife of Bath's Tale*, written in the fourteenth century:

(15) a. I sey **nat** this by wyves that been wyse
'I **do not** say this for wives that are wise.'
b. But Crist ... bat **nat** every wight he sholde go selle al that he hadde
'But Christ **did not** bid every one to go (and) sell all that he had.'

The major change here is in the negation of verbs such as *say* and *bid* (Chaucer's *bat* is modern *bade*). In Chaucer's English, any verb can be negated by putting *not* directly after it: *I sey nat*; *Crist bat nat*. In modern English, we don't negate verbs directly in this way: **I say not this*, **Christ bade not* aren't possible. Instead, we use a form of *do* which doesn't add any meaning of its own, but is there purely to support *not*, as in: *I do not / don't say this*. Chaucer's English doesn't have this '*do*-support' rule, as it is sometimes known.

> >

Before reading further, think of at least five words other than forms of *do* that can be directly negated by a following *not* in modern English: find words that fit into the gap in a sentence such as: *I ____ not/n't leave.*

< < < < < < < < < < < < < < < < < < < < < < < <

This gap can be filled by *may*, *might*, *must*, *can*, *could*, *will*, *would*, *shall*, *should* as well as *dare* and *need*. By changing *leave* to *left* or *leaving* we can also add *have* and *be* to the list of words that can fit the gap, as in *I have not left*, *I am not leaving*. In modern English, only words of a certain class, a verb-like word known as an AUXILIARY, can be directly negated by *not*. Where there is no other auxiliary, *do* is used as a kind of 'dummy' auxiliary.

> >

Apart from its role in negation, *do*-support has another major role in modern English. Try to think of some examples of this.

< < < < < < < < < < < < < < < < < < < < < < < <

Do is also used to form 'yes/no' questions, where there's no other auxiliary. So although we can say *Might/can/will you leave?*, using one of the auxiliaries just listed, as well as *Are you leaving?* and *Have you left?*, an ordinary verb can't be used in question formation: **Left you yesterday?* Once again, Middle English did allow this construction:

(16) a. Sey you no?
'Do you say no?'

b. Why hydestow (i.e. *hidest thou*) the keyes ... ?
    'Why do you hide the keyes?'

So these are two ways in which MAIN verbs in Middle English (verbs that aren't auxiliaries) behave differently than in modern English. You may also have noticed that 'do-support' is used in modern English for emphasis too. We had an example just now: *Middle English **did** allow this construction.*

Although you may not be surprised that changes like this occurred over a period of several hundred years (with *do*-support becoming standard by around 1700), it may be less obvious that English changed in the twentieth century, and indeed, is still changing constantly. But there are plenty of syntactic changes in progress right now. At the moment, these are restricted to certain dialects or to non-standard British English, but all the examples of change discussed later are spreading, and some may eventually become standard English. First, consider TAG QUESTIONS such as those in bold in (17):

(17) a. It is a hot day, **isn't it**?
       b. I can come, **can't I**?
       c. We still lost in the end, **didn't we**?

These questions 'tagged onto' the end of a statement are formed by specific rules in standard English which match the tag to the statement. A positive statement like *It is ...* gets a negative tag, *Isn't it*. Most importantly, an auxiliary used in the statement must be used in the tag (*I can* and *can't I*) and the pronoun (such as *it, I, we*) in the statement is also in the tag. In (17c), there's no auxiliary: main verbs like *lose* can't occur in tags (**lost we*) so *do*-support occurs, as in other questions. But in some dialects of British English, a single tag question *init* is used in each of the contexts shown in (17). Example (18) illustrates. The tag *init* is a reduced form of *isn't it*, a form which, in standard English, is possible only if the statement contains *is*. In *init* dialects, however, this has become an invariant tag, so that as well as the grammatically standard *It's a hot day, init?*, we find:

(18) a. I can come, **init**?
       b. We still lost in the end, **init**?

Some other varieties of English, such as Indian English, already have an invariant *isn't it* tag. And in some languages, an invariant tag is completely standard, as in French: *n'est-ce pas* (literally, 'isn't it?') occurs whatever the form of the statement:

(19) a. Il   va   arriver   demain,   n'est-ce pas?   (French)
              he goes arrive   tomorrow   TAG
              'He will arrive tomorrow, won't he?'
       b. Nous   n'   avons   pas   de   pain,   n'est-ce pas?
              we   NEG have   NEG   of   bread   TAG
              'We haven't got any bread, have we?'

Perhaps standard British English will also have an invariant tag one day too.

A second example of ongoing change is illustrated by the differences between (20) and (21). Example (20) is standard English, but in a very common, though technically non-standard variant, (20b) is replaced by (21):

(20)   a.   less difficulty; less wheat; less boredom; less milk
        b.   fewer students; fewer sheep; fewer people; fewer difficulties

(21)   less students; less sheep; less people; less difficulties

>>>>>>>>>>>>>>>>>>>>>>>>

Look first at (20) and work out what it is that conditions the use of *less* and *fewer* in standard English. Then describe how the non-standard variety in (21) differs. If you don't have the grammatical terminology, give as accurate a description as you can of the properties involved.

<<<<<<<<<<<<<<<<<<<<<<<<

In standard English, *less* is used only with MASS or NON-COUNT nouns – words like *difficulty*, *wheat*, *boredom* and *milk*. These are inherently singular; we can't say **three boredoms*. COUNT nouns, by way of contrast, have a plural form, such as *students*, *sheep*, *people* and *difficulties*, and in standard English these occur with *fewer*. (Note that although *sheep* doesn't take the regular plural *-s*, one way we can tell that it can be a plural word is exactly by the fact that it can occur after *fewer*.) Some nouns can, in fact, be either mass or count, like *difficulty*. Example (21) reflects a widespread non-standard usage in which *less* is used before any noun, including plural count nouns.

Our final example of language change in progress comes from the non-standard use of *they*, illustrated by the attested (= real-life) examples in (22).

(22)   a.   If **any candidate** hasn't got a form, **they** need to get one from the office.
        b.   I remember **one student** who said **they** couldn't write the answers because **they**'d lost **their** one and only pen.
        c.   Do you know **which assistant** you spoke to?
              No, but **they** were tall and dark-haired.

The pronouns *they* and *their* are always plural in standard English, so can only be used to refer to a plural noun phrase, such as *the candidates*. But in (22), these pronouns refer back to a noun phrase which is singular in form: *any candidate*, *one student*, *which assistant*. This is actually not a new usage – similar uses of *they* occur even as far back as Middle English. In modern standard British English, though, still reflected in the speech of some older speakers, a singular pronoun *he* or *she* is required in each of these contexts. This gives examples like *If anyone needs to leave he should raise his hand*, whereas most speakers nowadays would say *If anyone needs to leave they should raise their hand*. Note that there's no plural intended in the

use of *they* in (22): it's used not as a plural pronoun, but rather as a gender-neutral singular pronoun. This is clear in (22a), where *any candidate* was addressed to a group of males and females; but *they* can also be used as in (22b) and (22c), where the speakers must know the actual sex of the person referred to. (I can confirm that this is the case for the response in (22c), since I was that speaker, and heard myself say this!)

Interestingly, this development seems to have occurred independently of any desire to use non-sexist language; British English has not, for example, adopted such forms as *waitperson*, often used in American English.

To summarize, I argued in Section 1.1 that all native speakers of a language share an internal grammar, although they have never been taught its rules. Evidence for this is that we largely agree about what is and what is not a possible sentence of our language, though speakers are likely to differ over their acceptance of certain non-standard or dialectal variants. What is more, languages which are unrelated share many common properties and constructions, suggesting that human beings have an innate language faculty. Finally, we saw that language changes through time, and I gave some examples of ongoing changes. I now demonstrate how to make use of examples from other languages.

1.2 USE OF LINGUISTIC EXAMPLES

1.2.1 Why not just use examples from English?

This book contains examples from a wide variety of languages, including English. At first you may find it difficult to study examples from unfamiliar languages, and perhaps you wonder why we don't just use examples from English. There are two main reasons for using foreign-language examples: to learn about the differences between languages, and to learn about the similarities between them.

First, then, languages don't all look the same, and examining just our own language and its immediate relatives doesn't show how much languages can differ. Imagine that you've met only two languages, English and German, two closely related Germanic languages from northern Europe. Example (23), from German, is a word-for-word translation of the English.

(23)	der	schöne	Wasserfall	(German)
	the	pretty	waterfall	

You might imagine that the translation of this phrase would look the same in any language: first a word for 'the', then a word for 'pretty' or 'beautiful', then a word for 'waterfall'. But this is not so. In Spanish, for instance, we'd get (24):

(24)	la	cascada	hermosa	(Spanish)
	the	waterfall	beautiful	
	'the beautiful waterfall'			

Here, the word order is different in one respect: the word for 'beautiful' follows 'waterfall'. Otherwise, the Spanish is not too different from the English: it has just the same three words, and a word for 'the' in the same position. This isn't too surprising, as Spanish is also related to English, although more distantly than German. But in certain other languages, the equivalent to 'the' comes at the end of the phrase, as in Indonesian *surat itu* 'letter the' illustrated in (10), or else there may be no word for 'the' at all, as in Japanese and Chinese, or in some languages there isn't even a direct translation of the adjective *beautiful*.

The world's languages have many interesting and important syntactic features that I'd like you to know about. English has some but not all of these features, so if we only looked at English you'd miss out on the rest. In (25), we see one example, from Spanish:

> **(25)** Es nuevo. (Spanish)
> is new
> 'It's new.'

Example (25) has no word for 'it'; it literally means 'is new' – an impossible sentence in English. Spanish typically drops the subject pronoun meaning 'it' in such examples; for this reason, it's known as a PRO-DROP language. Many languages have examples parallel to this, but confining the discussion to English would never reveal that. In yet other languages, such as Arabic and Indonesian, the three-word English sentence *It is new* translates as 'It new' (this is illustrated in Chapter 2). These simple examples show that we can't expect sentences in other languages to be word-for-word translations of English sentences. So we study other languages to discover the range of constructions and features they contain – in order to find out about LINGUISTIC DIVERSITY.

The second reason for looking at examples from other languages is that linguists want to discover the common properties that languages share – their *homogeneity* or sameness. One of the most important discoveries of modern linguistics is that languages don't vary from each other at random, but are remarkably alike. Certain features occur in all languages. For instance, every language distinguishes a word class of NOUNS (words like *tree*, *liquid*, *expression* and *student*) from a word class of VERBS (words like *liquefy*, *learn*, *enjoy* and *grow*), although some languages have no other major word classes. (Chapter 2 examines word classes.) To discover this kind of information, linguists need to examine a representative sample of languages from different language families and different geographical areas.

Most linguists want to uncover the central patterns common to all languages. Although specific constructions are not universal (= common to all languages), all languages use a sub-set of the same basic tools of grammar. Each language has a wordlist or LEXICON which all its speakers share, and that wordlist always contains words from several different classes. All languages combine these words into phrases and sentences, and can manipulate the order of the phrases for various purposes – perhaps to ask questions, or to emphasize different parts of a sentence, or to show who's doing what to whom. This is syntax, and it forms the subject matter of the chapters ahead.

1.2.2 How to read linguistic examples

1.2.2.1 *The layout of examples*

Your first task as a syntactician is to learn to make use of examples from other languages. This book contains examples from over 100 different languages. Of course, I don't speak most of these – the examples come from other linguists, or from native speakers of the language (and sometimes native-speaker linguists). But I can utilize these examples because linguists set them out in a specific way for students and researchers who don't speak the language.

Examples of this special layout occur in the two Spanish illustrations in (24) and (25). Each consists of three lines. The first line is from the source language under consideration. The third line is a translation from the source language into English. You need this line to know what the original example means, but it's not the most important part of the example, because it only tells you about English – it tells you nothing about the source language. The really important line is the second one, called the GLOSS. The gloss is a literal translation of the original language. Each meaningful part of the original is translated, whether it corresponds exactly to a word in English or not. Look back at (2): French *ne* is GLOSSED (translated) simply as NEGATIVE because there's no English word that directly corresponds to it.

To see why the gloss is so important, consider (26) and (27), from Japanese and from Welsh. I have left out the gloss line. Both examples mean the same thing in the sense that they can receive the same English translation:

(26) Sensei-ga gakusei ni tegami-o kaita. (Japanese)
      'The teacher wrote a letter to the student.'

(27) Ysgrifennodd yr athro lythyr at y myfyriwr. (Welsh)
      'The teacher wrote a letter to the student.'

Let's suppose the point I'm trying to make is that sentences in Japanese, Welsh and English all have different word orders. Unless you happen to know both Japanese and Welsh, you won't be able to work this out from (26) and (27). In (28) and (29), I give the full examples, with glosses:[2]

(28) Sensei-ga   gakusei   ni   tegami-o   kaita.   (Japanese)
      teacher   student   to   letter   wrote
      'The teacher wrote a letter to the student.'

(29) Ysgrifennodd   yr   athro   lythyr   at   y   myfyriwr.   (Welsh)
      wrote   the teacher   letter   to   the   student
      'The teacher wrote a letter to the student.'

---

2. To simplify matters, I leave two small words in the Japanese unglossed: *ga* indicates that *sensei* '(the) teacher' is the subject (here, the one writing) and *o* indicates that *tegami* '(the) letter' is the object (here, the thing being written). These terms come up again later, and in Chapters 2 and 6, so don't worry if they are unfamiliar to you now.

Now we can compare the word orders of the three languages. First, the word for 'wrote' (a verb) has a different position in all three languages: at the end of the sentence in Japanese, at the beginning in Welsh, and somewhere in the middle in English – to be precise, after the phrase *the teacher*. This tells us right away that not all languages have the same sentence structure as English. Second, in both Japanese and English, the phrase '(the) teacher' is initial in the sentence, so Japanese and English have an important feature in common. In fact, at least 80 per cent of all languages would start their version of our sentence with the 'teacher' phrase. Welsh is different: its sentences start with the verb meaning 'wrote', a pattern found in perhaps 12 per cent of the world's languages. Third, both Welsh and English have the same order in the phrase 'to the student', while the Japanese in (28) has the opposite word order: *gakusei ni*, literally, 'student to'.

Using the glosses we can work out quite a lot about the word order differences – and similarities – between the three languages. Other facts about Japanese and Welsh emerge from the glosses too: for example, Japanese has no equivalent to 'the' and 'a', and Welsh has no word for 'a'.

You should now begin to see the importance of the gloss. On reaching a three-line example in the text, you should start at the bottom and work upwards, reading the translation first, then examining the gloss, then looking at the source language. Keep in mind that the English may bear little resemblance to the original source. In (28) and (29), the examples are pretty similar to the English, word-for-word (even if the word orders are different), but this certainly isn't always the case: (30) is from Rapa Nui, the Polynesian language of Easter Island.

(30)	E	tagi	ā	te	poki.	(Rapa Nui)
	NONPAST	cry	PROGRESSIVE	the	boy	
	'The boy is crying.'					

Apart from the word order differences (as in Welsh, the verb meaning 'cry' is (almost) at the beginning of the sentence), Rapa Nui has various other interesting features. In the English, *is* indicates that the crying is now, i.e. not in the past. The Rapa Nui example has no word for 'is', and instead a small word *e* indicates 'nonpast'. Second, in English the *-ing* ending on *cry* indicates an ongoing action, i.e. the boy hasn't finished crying. Rapa Nui has a separate word to indicate this: the 'progressive' word (meaning an unfinished action). Neither of these features of Rapa Nui can be discovered from the English translation, of course. So you always need to read the gloss carefully, thinking about whatever point is being made in the surrounding text. It should be clear by now that if you read only the last line of an example, you won't find out about any language other than English!

1.2.2.2 *Lexical and grammatical information*

Glosses contain both LEXICAL information, printed in normal type, and GRAMMATICAL information, printed in small capitals. Lexical information means ordinary words which are translations (or paraphrases) of the original language. In (28) and (29), the

glosses contain only lexical information. The Rapa Nui example in (30), though, has two items glossed as NONPAST and PROGRESSIVE (which indicates an ongoing action). This information concerns grammatical categories such as TENSE and ASPECT (more on these in Chapter 2). The point is that there are no separate words in English – members of the English LEXICON or vocabulary – that can translate this grammatical information, so it is glossed using the technical terms that describe its function in the source language.

All languages contain grammatical information. In (31), we show this by suggesting a precise gloss of an example from English, treating it as if it were a foreign language, and representing the grammatical information, as usual, in small capitals.

(31)　The　　　 student-s　　ask-ed　　 for　　these　　 book-s.
　　　 DEF.ART　 student-PL　ask-PAST　 for　　DEM.PL　 book-PL
　　　 'The students asked for these books.'

Taking these glosses as illustration, we can now explain the usual linguistic conventions. There are some familiar lexical items, 'student', 'ask', 'for' and 'book', but I've glossed *the* by referring to the grammatical information it represents: it's a DEFINITE ARTICLE – a word meaning 'the', as opposed to an INDEFINITE ARTICLE – a word meaning 'a'. I also glossed *these* as DEM.PL: *these* is a DEMONSTRATIVE word, a 'pointing' word from the set *this*, *that*, *these*, *those*. It's also PLURAL, therefore used before a plural word like *books*. Throughout the book, though, I will normally try where possible to use glosses you can recognize as words.

Apart from the lexical and grammatical information, the gloss also contains pieces of information separated by a hyphen (-). A hyphen preceding or following a piece of grammatical information in the gloss means that the grammatical element is attached to the word, or to another grammatical element, and can't be a separate word. Crucially, though, such grammatical elements have their own meaning. So the glosses *book*-PL and *student*-PL indicate that *books* and *students* are plural nouns; -*s* is a plural ending. And -*ed* is a past tense ending. I've also used the hyphen in the first line in (31) to indicate the boundaries in the source language between the grammatical information and the lexical items, although not all examples in this book follow this convention.

Grammatical elements attached to the beginning or end of a word, or to other pieces of grammatical information, are called AFFIXES (meaning something attached). Generally, then, a hyphen in the gloss indicates an affix, such as the plural -*s*. Grammatical affixes come in two main varieties: suffixes and prefixes. English plural -*s*, progressive -*ing* and past tense -*ed* are SUFFIXES; they're attached to the end of words. PREFIXES are attached at the beginning of words; examples from English are *un*- as in *untidy* and *re*- as in *re-seal*.

Elements of meaning such as 'ask' and 'past tense', 'un-' and 'plural' are known as MORPHEMES. As you can see, some of these represent independent words, but not all. The study of word forms is known as morphology, and though this is generally outside the scope of this book, we will often meet examples that show the interface between morphology and syntax – morphosyntax. Glosses in an example essentially

inform the reader about the morphosyntax of the words used, as well as just giving their literal meaning.

Sometimes, we recognize that a word contains more than one piece of information, that is, more than one morpheme, but these meaning elements have no discernible boundaries. For instance, if (31) had been *The students took these books home*, we would recognize that the verb *took* is past tense, just as *asked* is, but *took* is irregular, and doesn't have a past tense -*ed* suffix. We can't tell what part of *took* means 'past'. Linguists generally indicate this in the gloss using a colon (:) or a dot: thus, *took* would be glossed 'take:PAST' or 'take.PAST'. This convention means that a single source word contains more than one morpheme (such as 'take' and 'past tense') but there are no clear boundaries between these morphemes.

We also use this convention if we just don't wish to show the boundaries in a particular example, usually for the sake of keeping things clear or simple for the reader. Illustrating again with *asked*, I could show it as in (31) as *ask-ed*, with the hyphen indicating a morpheme boundary in the source word, and gloss it as ask-PAST, again showing the morpheme boundary. Alternatively, I could show *asked* in the source line, and ask.PAST in the gloss. Typically, we use this convention when we don't need to emphasize the detailed morphosyntax of the example.

1.2.2.3 *The categories of person and number*

In this section, I discuss the conventions used to represent the grammatical categories of PERSON and NUMBER, using examples from French and Kwamera (spoken in Vanuatu in the Pacific).

If you have learnt a foreign language, you will probably be used to meeting tables of verb forms like Table 1.1, from French.

Table 1.1
Present tense of French verb *parler*, 'to speak'

	Singular	Plural
1st	je parle	nous parlons
2nd	tu parles	vous parlez
3rd	il/elle parle	ils/elles parlent

Such tables, known as PARADIGMS, display the set of related forms that a particular lexical word has in a given grammatical context. The paradigm in Table 1.1 shows the set of forms that makes up the present tense of the verb *parler*, 'to speak'. Reading down the column headed Singular, the forms mean 'I speak, you (singular) speak, he/she speaks'. In the column headed Plural, the forms mean 'we speak, you (plural) speak, they (masculine/feminine) speak'.

The labels 1st (FIRST), 2nd (SECOND) and 3rd (THIRD) in the first column designate the grammatical category called PERSON. First person indicates the speaker, or a group of people that includes the speaker: so both the 'I' and 'we' forms are first person. Second person indicates the addressee(s): the 'you' forms. Third person indicates

some third party, an individual or group other than the speaker and addressee: the 'he/she/it' and 'they' forms.

The category of NUMBER refers to the distinction between SINGULAR (one person) and NON-SINGULAR (more than one person). In French, as in most European languages, number is either 'singular' or 'plural'. Note, though, that French distinguishes between *tu parles*, 'you (singular) speak', and *vous parlez*, 'you (plural) speak'. English once had this distinction too: *thou* meant 'you (singular)', equivalent to *tu*; and some varieties of modern English also have second person plural forms such as *you all* or *yous* (for instance, *yous* occurs in parts of both northeast and northwest England). Some languages divide non-singular into several categories, such as a category referring to two people (a DUAL), a category for three people (a TRIAL), and additionally a plural, used for referring to more than three people. For example, the Polynesian language Kwamera has just such a system.

Kwamera also has more PERSON distinctions than are familiar in European languages. First person in this language divides into INCLUSIVE and EXCLUSIVE forms. 'Inclusive' means 'we (as in me and you, speaker and addressee)', and 'exclusive' means 'we (speaker and other party, excluding you, the addressee)'. Imagine that a friend says 'We could go and see a film tonight'. You reply 'We? Do you mean you and me (*we inclusive*) or you and your boyfriend (*we exclusive*)?' English doesn't have different forms of 'we' to specify this information, but Kwamera does:

(32) a. **sa-ha**-akw   (Kwamera)
          1INC-PLURAL-break.up
          'We all break up.' (inclusive 'we')

      b. **ia-ha**-vehe
          1EXC-PLURAL-come
          'We came.' (exclusive 'we')

Before going any further, it's vitally important that you understand how to read the information in examples like this. The English translations in (32) contain several words – four separate words in (32a), for instance. But the Kwamera source examples each contain just *one* word, though this incorporates several distinct pieces of lexical and grammatical information. I'll explain using (32a), where there is a verb STEM, *akw*, and two prefixes attached to it – prefixes are grammatical elements which precede a stem. The *ha-* form closest to the verb stem means 'plural', and the *sa-* form on the outside means 'first person inclusive'. Together, these buy the meaning 'inclusive we'. English and Kwamera differ in a crucial way here. English has a separate pronoun *we* – it's a distinct, independent word on its own, not part of the verb. This is known as a FREE PRONOUN. But in the Kwamera, there is no separate word for 'we' at all: instead, that meaning is expressed by using grammatical elements attached to the verb itself. The forms *sa-* and *ia-* can't be separated from the verb, and don't occur on their own, and so are known as BOUND PRONOMINALS (there is more discussion of this in Section 4.3).

Note that in Kwamera, there are separate affixes representing the categories of PERSON and NUMBER, whereas in English the pronoun *we* represents both person

(1st) and number (plural) simultaneously. So the pronominal prefix *ia-* in (32b) represents not 'I' or 'we', but just first person: it becomes 'we' only when the plural prefix *ha-* follows. This means that the same 'first person exclusive' form *ia-* also translates as 'I':

(33)   **ia**-pkagkiari-mha
      1EXC-talk-NEG
      'I didn't talk.'

In future examples I gloss person and number as in Table 1.2, unless the language has some special inclusive and exclusive forms as in Kwamera. The first and second columns give the glosses and their meaning – this is grammatical information – and the third column lists the pronouns which in English are associated with this grammatical information, to help you remember.

Table 1.2
Glosses for person and number

Gloss	Meaning	English pronouns
1SG	first person singular	'I/me'
2SG	second person singular	'you [singular]'
3SG	third person singular	'he/him; she/her; it'
1PL	first person plural	'we/us'
2PL	second person plural	'you [plural]'
3PL	third person plural	'they/them'

If the gloss specifies just the person, '1', '2' or '3', but doesn't mention singular or plural, this means that the particular language being glossed does not have number distinctions in this instance.

1.2.2.4   *Writing systems and glosses*

Not all languages use the Roman alphabet (the one you're reading now). For example, Russian uses the Cyrillic alphabet, and Chinese and Japanese both use writing systems based on characters rather than an alphabet. But there usually exist conventions for writing such languages in the Roman alphabet, and this enables linguists to make use of the examples. I mostly follow the published source that my data come from, although some labels for glosses are changed to bring them into line with my practice. Additionally, I standardize glosses that are more detailed or less detailed than we need. Occasionally I simplify by not glossing some item, especially if we haven't yet met the appropriate grammatical term. I will often omit the tones in examples from languages such as Chinese: these specify pronunciation and are not vital to our discussion of syntax. Finally, note that some languages don't have a writing system at all, since they've never been written down. In this case, linguists typically give a phonetic representation of the original language. For that reason,

some of the examples don't start with capital letters; the phonetic alphabet doesn't follow the conventions of a writing system.

1.3 WHY DO LANGUAGES HAVE SYNTAX?

Speakers manipulate sentences in all sorts of ways because they're trying to convey different meanings. Syntax allows speakers to express all the meanings that they need to put across. In the simplest cases, this might mean altering the basic word order of a sentence, to emphasize or downplay a particular phrase, or to ask a question, or else grouping words together in different ways to modify the meaning. This section gives a preliminary idea of some of the typical syntactic constructions found in languages, and demonstrates that languages really do have syntactic structure.

1.3.1 Word order

In English, the WORD ORDER is pretty fixed. There are three main elements in the sentence in (34): *Kim*, the one drinking the tea; *drank*, the verb, which expresses what Kim did; and *the tea*, expressing what is being drunk. We use the term 'word order' (more accurately, as we will see later on, 'constituent order') to discuss the order in which these three main parts of a sentence occur in a language. In English, the three elements occur in the order shown in (34a). This is the normal word order, and all variants of it are impossible (therefore starred) except for (34f), which has a restricted special usage.

(34) a. Kim drank the tea.
     b. *Kim the tea drank.
     c. *Drank Kim the tea.
     d. *Drank the tea Kim.
     e. *The tea drank Kim.
     f. The tea Kim drank.

Most of the logically possible variations are impossible in English. However, each of the word orders in (34) is attested (= found) among the world's languages, though some are much more common than others (see Chapter 6). The three most common basic word orders in languages other than English are those of (34a), (b) and (c). We saw in Section 1.2.2.1 that Japanese has the basic word order of (34b), and Welsh the basic order of (34c). Malagasy, spoken in Madagascar, has the basic order in (34d). The two word orders in (34e) and (34f) are the rarest basic word orders in the languages of the world, although they are found in the Carib language family of the Amazon basin. For example, Hixkaryana has the word order in (34e).

It is generally possible to determine the basic, neutral word order in a language, but the flexibility or rigidity of the basic word order differs widely among the world's languages. English has a fixed basic word order, while Russian has a very flexible word order, and Japanese allows many different orders provided the verb comes at the end of the sentence, as in (28). In English, some of the starred (ungrammatical)

word orders in (34) might just about be permissible in poetry, but not in the spoken language or in prose. Example (34f) may initially sound odd to you, but it can be used to FOCUS on what it was that Kim drank; the phrase *the tea* is fronted from its usual position as given in (34a), so becomes more prominent. Try adding a bit of context: Kim visits an eccentric aunt who makes tea and beer out of strange garden plants: *The tea, Kim drank ___, but the home-made nettle beer, she really hated ___.* The gaps are used to show the normal position of the fronted phrases *the tea* and *the home-made nettle beer*. In technical terms, this construction involves fronting what is known as the DIRECT OBJECT of the verbs *drank* and *hated*: the direct object (or just 'object') of *drank* is *the tea* (the 'thing drunk') and the object of *hated* is *the home-made nettle beer* (the 'thing hated'). Example (35) shows a published example of the same object-fronting construction; the context is that the writer is learning to fly a microlight aircraft:

(35)   **The last exercise, a stall while climbing**, I didn't do ___ well.
(From *Travels with Pegasus*, Christina Dodwell.
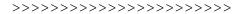Sceptre, Hodder & Stoughton, 1989)

In (35), the fronted object is shown in bold, and again its more usual position is marked by a hyphen. Object-fronting is, in fact, quite rare in English. It's known as a MARKED (= unusual) construction, while the usual basic word order as in (34a) is termed UNMARKED.

Often there are stylistic reasons for changing basic word order. The fronted phrase in (35) is rather long, and sounds clumsy in the usual object position: *I didn't do the last exercise, a stall while climbing, well*. In (36), we illustrate a different kind of word order change, which involves breaking up a rather long phrase by moving part of it to the right. The phrase in bold type moves rightwards from its basic position following the word *estimates*, shown by the gap:

(36)   Estimates ___ vary greatly **about the number of fluent speakers** (i.e. *of Esperanto*).
(From *The Cambridge Encyclopedia of Language*, David Crystal.
Cambridge University Press, 1987)

This avoids the clumsiness of a long initial phrase *estimates about the number of fluent speakers* before the short phrase *vary greatly*. (Compare the normal word order in *Estimates about this vary greatly*.) As (34) showed, English has a generally inflexible word order in the sentence, but optional MODIFYING phrases can be reordered quite easily, as is the case for the *about ...* phrase which modifies (= expands on) the word *estimates*.

>>>>>>>>>>>>>>>>>>>>>>>>>

The examples in (37) and (38) again involve rightward movement of a phrase. In both cases the moved phrase is the object of a verb – as it was in (35) – but you don't
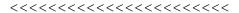
need to be able to define what an object is in order to find the phrase that's moved. Try to work out the basic word order, find the phrase that's moved, and indicate where it has moved from by using a gap, as I did earlier. Then say why you think the writer chose this construction, rather than using the basic word order of English:

(37) It may harm your defence if you do not mention when questioned something which you later rely on in court.

(From *1994 Criminal Justice and Public Order Act*)

(38) Mrs Verwoerd struggled to read without her glasses a statement appealing to Nelson Mandela to give the Afrikaner a *volkstaat*.

(From *The Guardian*, 21.8.95)

<<<<<<<<<<<<<<<<<<<<<<<<

I indicate the moved phrase by following the usual linguistic practice of enclosing it inside square brackets: [...]; the brackets signify the beginning and end of a phrase. Here are the answers to the exercise:

(39) It may harm your defence if you do not mention ___ when questioned [something which you later rely on in court].

(40) Mrs Verwoerd struggled to read _____ without her glasses [a statement appealing to Nelson Mandela to give the Afrikaner a *volkstaat*].

In (39), the basic word order would be *I mentioned something when questioned*, not *I mentioned when questioned something*. But because the bracketed object in (39) is a particularly long phrase ('heavy' is the technical term), it's allowed to shift from its normal position, and indeed sounds better that way. Similarly in (40), the basic word order has to be *She read a statement without her glasses* not *She read without her glasses a statement*. But again, the object is heavy: it's the whole phrase *a statement appealing to Nelson Mandela to give the Afrikaner a volkstaat*. So the preferred position of this heavy phrase is not its basic position, but a position to the right of the shorter phrase *without her glasses*.

   If you previously had no idea how to determine the OBJECT of a verb, look at the position of the gaps in (39) and (40). Both gaps immediately follow verbs, namely *mention* and *read*. The objects are the 'thing mentioned' and the 'thing read' here: both these phrases in some sense complete the meaning of the verb, and so are often known as the COMPLEMENTS of the verb. The normal position for a direct object in English is immediately following the verb. I discuss these technical terms in more detail in Chapter 2, but these features will help you identify objects in the next section. Please review this section before moving on if you weren't previously familiar with the grammatical term 'object'. We return to word order in Chapter 6.

1.3.2   Promotion and demotion processes

The syntactic variations in Section 1.3.1 involved simply reordering the elements of a sentence. But syntactic changes can have much more radical results than this. Section 1.1, in the discussion of Indonesian, introduced the idea of promotion processes – making a word or phrase more prominent in the sentence. There are also demotion processes, which make part of the sentence less prominent. Here I give a preliminary introduction to another construction involving promotion and demotion – the PASSIVE – in English and Japanese (more in Chapter 7).

The passive is illustrated in bold type in (41) and (42), and is an extremely common construction in both spoken and written English.

> **(41)**   The women and boys with crates converged on the boats and **their catch was counted out by the market boss**.
>
> > (From *Travels in Mauritania*, Peter Hudson. Flamingo, 1990)
>
> **(42)**   His normal work was filing girls' teeth to points, although **pointed gnashers were considered a bit old-fashioned by the girls here**.
>
> > (From *Travels with Pegasus*, Christina Dodwell.
> > Sceptre, Hodder & Stoughton, 1989)

Compare these passive constructions with the sentences in (43) and (44), which are their counterparts in meaning, but are both ACTIVE constructions:

> **(43)**   The market boss counted out **their catch**.
>
> **(44)**   The girls here considered **pointed gnashers** a bit old-fashioned.

> >

Before reading further, please try to figure out what properties differentiate the active from the passive constructions. Use the correct technical terms where you know them. Start by deciding on the grammatical role of the phrases in bold type in (43) and (44), and see what role these same phrases have in the passive constructions. What purposes do the two different construction types seem to serve? Can you describe any additional grammatical features?

< < < < < < < < < < < < < < < < < < < < < < <

First, assume that active sentences are the more basic; they are, for instance, learnt much earlier by children than are passives. Two properties of the passive occur in any language which has the construction: (i) the passive involves PROMOTION of an object phrase to a new position in the sentence, known as the SUBJECT position, and (ii) the phrase that used to be in the subject position undergoes DEMOTION. Let's go through this technical passage carefully. The phrases in bold in the active constructions in (43) and (44) are in the OBJECT position: they each immediately follow the

verb (*counted out*, *considered*) and they express what is being counted out, what is considered. In (41) and (42), the phrases *their catch* and *pointed gnashers* appear in a new, promoted position in the sentence. They have changed their grammatical function, and become the subjects of the passive sentences. How do we know that these phrases are now subjects? One major indication is that *their catch* and *pointed gnashers* appear immediately before the verb, in the normal sentence-initial position of English subjects. (We will see more tests for subjecthood in English in Chapter 2.) This advancement to subject position in (41) and (42) makes the promoted phrases more salient: it focuses attention on *their catch* and *pointed gnashers*. By contrast, the phrases that were the subjects of the active sentences in (43) and (44), namely *the market boss* and *the girls here*, are no longer subjects. In the passive sentences in (41) and (42), they have been demoted to a lower position. Demotion in this case means that they are consigned to a *by*-phrase, outside the core of the sentence. Notice that this *by*-phrase is entirely optional: we could omit it, and just have, for instance, *Their catch was counted out*. Compare that optionality with what we find in (43): both the subject *the market boss* and the object *their catch* are core elements of the sentence, and neither can be omitted. (Try this.)

You should now be starting to have some feeling for the purpose and usual positions of different parts of the sentence. Before leaving the topic of the passive construction, note that in English (and in many other languages) it is signalled by changes in the form of the verb: compare (45a) and (45b), where the verbs are in bold type.

(45) a. Kim **broke** the vase. (active)
     b. The vase **was broken** by Kim. (passive)

The examples in (46) show the corresponding properties in Japanese. Example (46a) is the active sentence, (46b) its passive counterpart:

(46) a. Sensei   ga   John o   sikat-ta.   (active)   (Japanese)
        teacher SUBJECT John OBJECT scold-PAST
        'The teacher scolded John.'
     b. John ga   sensei   ni   sikar-**are**-ta.   (passive)
        John SUBJECT teacher by scold-PASSIVE-PAST
        'John was scolded by the teacher.'

In (46a),[3] the 'teacher' phrase *sensei ga* is the subject, and *John* is marked as the object of the 'scold' verb by the *o* marker. In (46b), *John* is promoted to subject position and the 'teacher' phrase is demoted. It appears in the equivalent of a *by*-phrase, *sensei ni* 'teacher by' – note that Japanese has a different word order from English here. The

---

3. The verb stem (the form before the affix is added) is *sikar*, but this changes to *sikat* before the past tense suffix *-ta*.

verb also has a special passive suffix, *-are*. Please make sure you understand the way these examples are structured before moving on.

Passives and other promotion and demotion constructions are discussed in detail in Chapter 7.

1.3.3   All languages have structure

All languages, whether living or dead, have syntactic structure, including, of course, sign languages (such as British Sign Language). This means that a language doesn't just consist of strings of words, but that the words group together to form phrases, and the phrases group together to form larger phrases and sentences. Linguists describe this phrases-within-phrases pattern as HIERARCHICAL STRUCTURE.

One kind of hierarchical structure is seen in EMBEDDED sentences, a construction in which sentences occur within other sentences, such as *Chris told Lee* [*Kim couldn't swim*]. This property is known as RECURSION. Here, the sentence in brackets – *Kim couldn't swim* – is the embedded sentence. It serves to tell you what it was that Chris told Lee. More examples from English are given in (47): the embedded sentences are again in square brackets.

(47)   I wonder [if Lee will arrive late].
The claim [that she doesn't like Kim] is very surprising.
[That we've no coffee left] isn't my fault.
We asked [how to get to the station].

For the first three phrases in brackets in (47), you can check that they really are sentences by removing the words *if* and *that* which introduce them: you then get perfectly good independent sentences such as *Lee will arrive late*. But this doesn't work for all embedded sentences, as is clear from *how to get to the station*; we will see much more on this in Chapter 3. Try to decide what properties this final example has that set it apart from the other embedded sentences in (47).

There are no limits to the number of embedded sentences that can be strung together. So given a sentence like *Kim couldn't swim* we can turn it into *Lee thought that Kim couldn't swim*, then *I said that Lee thought that Kim couldn't swim*, and so on. This means it's never possible to construct a 'longest sentence'.

I end this chapter with two short practical demonstrations that syntactic structure really exists, in other words that speakers of a language share the same mental representations of this structure. First, look at the examples in (48):

(48)   a.   I charged up the battery.
b.   I charged up the street.

At first glance these sentences appear to be structurally identical. Of course, you might be aware when you read them that *charge* means something different in (48a) and (48b), but otherwise, the only difference seems to be that *street* replaces *battery*. And yet the syntactic behaviour of the two sentences is entirely different. As always, the asterisks indicate ungrammatical sentences:

(49) a. I charged the battery up.
 b. *I charged the street up.

(50) a. *It was up the battery I charged (not the engine).
 b. It was up the street I charged (not the corridor).

(51) a. *I charged up Lee's battery and (then) up Kim's too.
 b. I charged up Lee's street and (then) up Kim's too.

Since native speakers of English agree about where the asterisks showing an ungrammatical sentence should be placed, we must all share an unconscious knowledge of the sentence structure of English. Even though pairs of sentences like those in (48) look the same, they in fact have different structures. Different sets of words group together to form phrases in each case, and linguists represent this using brackets:

(52) a. I [charged up] the battery.
 b. I charged [up the street].

In (52a), the brackets show that there's a phrase *charge up*. This makes sense if you think that the only thing you can do with a battery is charge it up; you can't charge it down, over, across, or anything else. So *up* belongs with *charge* in (52a). In (52b), it doesn't: instead, there's a syntactic unit *up the street*, which can be moved around the sentence for focus, as for example in (50b), *It was **up the street** I charged*. And in (52b), *up* can be replaced with a number of other words: *I charged **down** the street/**over** the street/**across** the street* and so on. The very fact that *up* forms a unit with *charge* in (52a) but with *the street* in (52b) is responsible for the patterns of (un)grammaticality in (49) to (51). We'll return in detail to questions of structure and the grouping of words to form phrases in Chapter 5.

As a second demonstration of syntactic structure, let's examine possessive *-'s* in English, as in *Lee's friend*. You might assume at first that this possessive ending simply attaches to a noun, a word such as *Lee* or *government*, as in *the government's dilemma*. But consider (53):

(53) a. I saw the woman next door's children.
 b. What was that guy who retired last month's name?
 c. The student I lent the book to's room-mate said she'd left.

Each example in (53) shows that *-'s* actually attaches to the end of a whole phrase, not to the single noun at the end of the phrase: we know that the door doesn't have children, and that the answer to (53b) couldn't be *November*. And *to* in (53c) isn't even a noun (you may find this example a little odd, because it belongs in spoken rather than written English. Try saying it aloud a few times.) Native speakers also know that you can't attach the *-'s* to the noun it logically seems to belong to: **What was that **guy's** who retired last month name?* The fact that we agree where to attach the

-'s shows once again that sentences do have structure, and that we have an intuitive knowledge of it. The phrases that -'s attaches to are shown in brackets:

(54) a. [the woman next door]'s
     b. [that guy who retired last month]'s
     c. [the student I lent the book to]'s

Demonstrations of this nature could be given from any language, because the rules of the syntax of all languages are STRUCTURE DEPENDENT. This is why no language has rules that, for example, form questions from statements by reversing the order of the words in the sentence – such a rule wouldn't depend on the structure of the sentence at all, and so can't work. When, as linguists, we try to figure out the syntactic structures of a language, we rely on the judgements of native speakers to tell us whether our example sentences are possible or impossible (the latter being starred). These GRAMMATICALITY JUDGEMENTS, along with examples that are collected from a spoken or written corpus of the language, form the data of the science of linguistics. It doesn't matter that native speakers usually can't tell us why they feel that a particular sentence is good or bad; the very fact that they have these intuitions shows up the structural differences and similarities between sentences.

FURTHER READING

An excellent general introduction to linguistics and to the views of linguists on language acquisition is Fromkin, Rodman and Hyams (2007). I can also strongly recommend Jackendoff (1993) and Pinker (1994). Baker (2001) is a recent introduction to the view of language learning most associated with the work of the linguist Noam Chomsky. Chomsky's idea that speakers possess a subconscious 'knowledge' of their native languages is explored accessibly in the early chapters of his (1986) book, *Knowledge of language*. On language change, see McMahon (1994) and Millar (2007). On person, see Siewierska (2004); on gender, see Corbett (1991); on number, see Corbett (2000); and on agreement, see Corbett (2006). Also on person, number and related issues see Whaley (1997: Chapter 10) and Comrie (1989: Chapter 9). The topics raised in Section 1.3 all appear again in later chapters: word order is in Chapter 6, promotion and demotion processes in Chapter 7, and syntactic structure in Chapters 4 and 5.

EXERCISES

1. In Chapter 1, I argued that dialectal forms of English cannot be criticized for lack of 'logic'. The table that follows lists both the standard English forms of the REFLEXIVE PRONOUNS (the 'self' forms) and the forms found in a northern dialect of British English. Which dialect has the more regular pattern for the formation of its reflexive pronouns? Why? Be as specific as you can about how the reflexives are formed in each case, using the correct technical terms if you can.

Standard dialect	Northern dialect
myself	myself
yourself	yourself
himself	hisself
herself	herself
ourselves	theirself
yourselves	ourselves
themselves	yourselves
	theirselves

I've omitted *itself*, which is the same across all dialects, and doesn't shed any light on the question. Also, the form *myself* is generally pronounced more like *meself* in this northern dialect, but I take this to be a phonetic reduction which is not relevant to the question. The northern dialect has one more reflexive pronoun than standard English. What is it, and what do you think it's used for? (If you're not a native speaker of this dialect, you may find it helpful to look back at Section 1.1.) Some English speakers have a singular form *themself*: comment on how this fits into the set of forms in standard English, and say how it is used (if you're not a speaker who has this form, see if you can imagine how it's used).

Finally, do you think that 'more logical' equals 'better', as far as languages are concerned?

2. In Chapter 1, we discussed syntactic changes in progress in English, such as the use of *init* in tag questions. Another change which seems to be gaining ground in British English concerns the COMPARISON of adjectives (words like *red*, *happy*, *difficult* in English).

Task: (i) Consider first what words can fit into the gap in (1); compile for yourself a list which you think is representative:

(1)   This is ___ than that one.

Standard English uses either a single word from a list like (a) or a phrase from a list like (b):

(a)	funnier	(b)	more difficult
	older		more dramatic
	dearer		more generous
	cleverer		more unpleasant
	pleasanter		more expensive

So the COMPARATIVE form of adjectives in standard English can consist of either ADJECTIVE + *-er* suffix, as in list (a), or *more* plus ADJECTIVE as in list (b).

(ii) Work out the standard English rule that governs the choice of these alternatives. Your answer should account for the fact that standard English does not have forms like those in (2):

(2)   *more difficulter, *more unpleasanter, *more generouser

The forms in (2) are ungrammatical in all dialects, as far as I know, but many speakers of non-standard English can fill the gap in (1) with the forms in (3):

(3)   more funnier, more dearer, more cleverer

(iii) Describe as accurately as you can the syntactic change in progress here, using the correct grammatical terms if you can.

3. The data in this exercise are from Icelandic, a Germanic language which is related quite closely to English, and are taken from Sigurðsson (2006).

   Task: All these examples illustrate a single, specific, grammatical difference between English and Icelandic. What is this? You should use the correct grammatical term, which is given earlier on in this chapter. If you find other syntactic differences between the two languages in any example, list these too. Finally, in what specific ways can you see that Icelandic is syntactically similar to English? Use the correct terminology wherever possible. **NB!** The Icelandic alphabet includes some characters that are not used in English, but this has no bearing on the answer.

   (1)   Kona   sat   á   bekk.
   woman   sat   on   bench
   'A woman sat on a bench.'

   (2)   Ég   keypti   skemmtilega   bók   í   morgun.
   I   bought   interesting   book   in   morning
   'I bought an interesting book this morning.'

   (3)   Ólafur   er   prófessor.
   Ólafur   is   professor
   'Ólafur is a professor.'

   (4)   Það   er   maður   í   garðinum.
   there   is   man   in   garden.the
   'There is a man in the garden.'

   (5)   Sá   sem   er   að   tala   er   Íslendingur.
   the.one   who   is   to   talk   is   Icelander
   'The one who is talking is an Icelander.'

4. In (1) through (8), which follow, are some more examples from Kwamera, and two closely related languages, Lenakel and Southwest Tanna, all languages of the

Republic of Vanuatu in the southwest Pacific; data are from Lindstrom and Lynch (1994) and Lynch (1998). I give the original and the gloss, and your task is to suggest a translation. You will find it useful to look back at Section 1.2.2.3, where the Kwamera examples are discussed in the text.

Hint
There is rarely just one correct way of translating each example; the important part here is to make sure you understand the role of the grammatical information in the glosses (in small capitals).

(1) t-r-uv-irapw (Kwamera)
FUTURE-3SG-move-outwards

(2) t-r-am-apri (Kwamera)
FUTURE-3SG-CONTINUOUS-sleep

Translate the 'continuous' prefix *am-* using an *-ing* form of the verb in English, as in *I was sleeping*, *I've been sleeping*, *I will be sleeping*. Both Kwamera *am-* and English *-ing* denote an ongoing action here.

(3) iak-imiki kuri u (Kwamera)
1EXC-dislike dog this

The prefix *iak-* is the form of the first person exclusive which occurs before vowels.

(4) k-rou-ánumwi (Kwamera)
1INC-DUAL-drink

Note that the first person inclusive prefix, *k-*, does not have the same form as the first person inclusive *sa-* that we saw in (32a) in the text. The reason for this is that *sa-* is used in conjunction with a plural prefix, while *k-* co-occurs with a dual prefix.

(5) K-ɨm-hal-vɨn-uas. (Lenakel)
3PL-PAST-TRIAL-go.there-together
(6) R-ɨm-avhi-in mun (Lenakel)
3SG-PAST-read-TRANS again

The suffix *-in* marks the verb as TRANSITIVE (more in Chapter 2): this means that it is understood to have an object (see Section 1.3.1), and this should affect your translation.

(7) Kɨmlu i-ɨmn-la-gɨn (Southwest Tanna)
we.two.EXC 1PL-PAST-DUAL-afraid

<table>
<tr><td>(8)</td><td>Kɨmlu</td><td>i-ɨmn-la-hai</td><td>pukah</td><td>(Southwest Tanna)</td></tr>
<tr><td></td><td>we.two.EXC</td><td>1PL-PAST-DUAL-stab</td><td>pig</td><td></td></tr>
</table>

Recall from Section 1.2.2.2 that the dot '.' in a gloss indicates that the various pieces of grammatical information can't be separated from each other: the *whole* form *kɨmlu* in (7) and (8) therefore has the meaning (glossed as) 'we.two.EXC'.

5. The following data are all from Niger-Congo languages, a massive family of African languages covering almost a quarter of the total languages in the world.

   Task: Examine each sentence and note as many grammatical differences in these examples as you can between English and each of the various source languages. Look for differences in both the syntax and the morphosyntax, and describe these differences carefully, using the correct terminology where you can. The data are taken from Watters (2000) and Payne (1985a).

   (1) è   wúru   tèe   à   bóa   (Mende)
            he   cut   stick with   knife
            'He cut the stick with a knife.'

   (2) ba-tub-aka   (Lobala)
            3PL-sing-PAST
            'They sang.'

   (3) Halima   a-na-pika   ugali   (Swahili)
            Halima   3SG-PRES-cook   porridge
            'Halima is cooking porridge.'

   (4) a. ɔ́   tẽ   kɔ́   (Kru)
                he   buy   rice
                'He bought rice.'
            b. ɔ́   sé   kɔ́   tẽ
                he   NEG   buy   rice
                'He didn't buy rice.'

6. These data are from a Melanesian language called Tinrin, spoken in part of the islands of New Caledonia, in the southwest Pacific Ocean (Osumi 1995).

   Task: (i) Figure out the function of the grammatical marker *nrâ* shown in each example in (1) through (6) in bold type, and not glossed; (ii) decide why this marker does not occur in (7), (8) and (9).

   Hints
   • Don't worry about the fact that Tinrin has another morpheme with the form *nrâ*, which occurs in (2), (4), (5), (6) and (9). This has a different meaning altogether: it's a third person singular pronoun, and has no relevance to your answer.

- Tinrin has a grammatical category 'dual', like Kwamera, discussed in this chapter. But this is not relevant to your answer.
- It will help to compare the structure of (1) through (6), which have the *nrâ* marker in question, with that of (7) through (9), which do not. What syntactic factor distinguishes these two groups of data?

(1) rru   fi   pwere   ânrâmwâ   **nrâ**   truu   truu-truare
     3.DUAL   go   to   there   ???   the.DUAL   DUAL-brother
     'The two brothers went there.'

(2) nrâ   nyôrrô   **nrâ**   wa   mwîê   mwâ
     3SG   cook   ???   the   woman   that
     'That woman cooked (something).'

(3) u   truumwêrrê   mirrî   **nrâ**   nro
     1SG   always   hungry   ???   I
     'I am always hungry.'

(4) nrâ   tewùrrù   nranri   **nrâ**   toni
     3SG   tie.up   goat   ???   Tony
     'Tony tied up the goat.'

(5) nrâ   truu   tôbwerrî-nrî   **nrâ**   magasâ
     3SG   stay   close-PASSIVE   ???   shop
     'The shop is closed.'

(6) nrâ   tôbwerrî-nrî   **nrâ**   magasâ   rugi   midi
     3SG   close-PASSIVE   ???   shop   at   noon
     'The shop is/was closed at noon.'

(7) wa   mwâ   ha   hêê-rò
     the   house   this   belonging-me
     'This house is mine.'

(8) rri   truu   hubo   ei   nrü
     3PL   stay   near   to   2SG
     'They live near you.'

(9) toni   nrâ   tuo   nrî   padrêrrê-tave
     Tony   3SG   put   3SG   side-bed
     'Tony put it beside the bed.'