2 Phrase structure grammar

This chapter deals with phase structure grammars, which play an important role in several of the theories we will encounter in later chapters.

2.1 Symbols and rewrite rules

Words can be assigned to a particular part of speech on the basis of their inflectional properties and syntactic distribution. Thus, *weil* 'because' in (1) is a conjunction, whereas *das* 'the' and *dem* 'the' are articles and therefore classed as determiners. Furthermore, *Buch* 'book' and *Mann* 'man' are nouns and *gibt* 'gives' is a verb.

(1) weil er das Buch dem Mann gibt
because he the book the man gives
'because he gives the man the book'

Using the constituency tests we introduced in Section 1.3, we can show that individual words as well as the strings *das Buch* 'the book' and *dem Mann* 'the man', form constituents. These get then assigned certain symbols. Since nouns form an important part of the phrases *das Buch* and *dem Mann*, these are referred to as *noun phrases* or NPs, for short. The pronoun *er* 'he' can occur in the same positions as full NPs and can therefore also be assigned to the category NP.

Phrase structure grammars come with rules specifying which symbols are assigned to certain kinds of words and how these are combined to create more complex units. A simple phrase structure grammar which can be used to analyze (1) is given in (2):[1,2]

(2)

NP → D N		NP → er	N → Buch
S → NP NP NP V		D → das	N → Mann
		D → dem	V → gibt

We can therefore interpret a rule such as NP → D N as meaning that a noun phrase, that is, something which is assigned the symbol NP, can consist of a determiner (D) and a noun (N).

[1] I ignore the conjunction *weil* 'because' for now. Since the exact analysis of German verb-first and verb-second clauses requires a number of additional assumptions, we will restrict ourselves to verb-final clauses in this chapter.

[2] The rule NP → er may seem odd. We could assume the rule PersPron → er instead but then would have to posit a further rule which would specify that personal pronouns can replace full NPs: NP → PersPron. The rule in (2) combines the two aforementioned rules and states that *er* 'he' can occur in positions where noun phrases can.

We can analyze the sentence in (1) using the grammar in (2) in the following way: first, we take the first word in the sentence and check if there is a rule in which this word occurs on the right-hand side of the rule. If this is the case, then we replace the word with the symbol on the left-hand side of the rule. This happens in lines 2–4, 6–7 and 9 of the derivation in (3). For instance, in line 2 *er* is replaced by NP. If there are two or more symbols which occur together on the right-hand side of a rule, then all these words are replaced with the symbol on the left. This happens in lines 5, 8 and 10. For instance, in line 5 and 8, D and N are rewritten as NP.

(3)

	words and symbols						rules that are applied
1	er	das	Buch	dem	Mann	gibt	
2	NP	das	Buch	dem	Mann	gibt	NP → er
3	NP	D	Buch	dem	Mann	gibt	D → das
4	NP	D	N	dem	Mann	gibt	N → Buch
5	NP		NP	dem	Mann	gibt	NP → D N
6	NP		NP	D	Mann	gibt	D → dem
7	NP		NP	D	N	gibt	N → Mann
8	NP		NP		NP	gibt	NP → D N
9	NP		NP		NP	V	V → gibt
10						S	S → NP NP NP V

In (3), we began with a string of words and it was shown that we can derive the structure of a sentence by applying the rules of a given phrase structure grammar. We could have applied the same steps in reverse order: starting with the sentence symbol S, we would have applied the steps 9–1 and arrived at the string of words. Selecting different rules from the grammar for rewriting symbols, we could use the grammar in (2) to get from S to the string *er dem Mann das Buch gibt* 'he the man the book gives'. We can say that this grammar licenses (or generates) a set of sentences.

The derivation in (3) can also be represented as a tree. This is shown by Figure 2.1. The

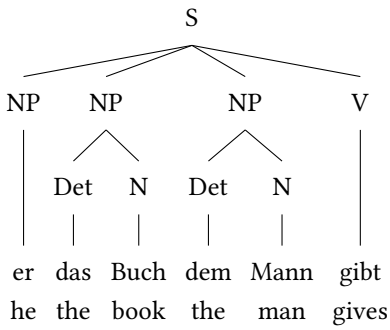

Figure 2.1: Analysis of *er das Buch dem Mann gibt* 'he the book the woman gives'

symbols in the tree are called *nodes*. We say that S immediately dominates the NP nodes

and the V node. The other nodes in the tree are also dominated, but not immediately dominated, by S. If we want to talk about the relationship between nodes, it is common to use kinship terms. In Figure 2.1, S is the mother node of the three NP nodes and the V node. The NP node and V are sisters since they have the same mother node. If a node has two daughters, then we have a binary branching structure. If there is exactly one daughter, then we have a unary branching structure. Two constituents are said to be *adjacent* if they are directly next to each other.

Phrase structure rules are often omitted in linguistic publications. Instead, authors opt for tree diagrams or the compact equivalent bracket notation such as (4).

(4) [$_S$ [$_{NP}$ er] [$_{NP}$ [$_D$ das] [$_N$ Buch]] [$_{NP}$ [$_D$ dem] [$_N$ Mann]] [$_V$ gibt]]
he the book the man gives

Nevertheless, it is the grammatical rules which are actually important since these represent grammatical knowledge which is independent of specific structures. In this way, we can use the grammar in (2) to parse or generate the sentence in (5), which differs from (1) in the order of objects:

(5) [weil] er dem Mann das Buch gibt
because he.NOM the.DAT man the.ACC book gives
'because he gives the man the book'

The rules for replacing determiners and nouns are simply applied in a different order than in (1). Rather than replacing the first Det with *das* 'the' and the first noun with *Buch* 'book', the first Det is replaced with *dem* 'the' and the first noun with *Mann*.

At this juncture, I should point out that the grammar in (2) is not the only possible grammar for the example sentence in (1). There is an infinite number of possible grammars which could be used to analyze these kinds of sentences (see exercise 1). Another possible grammar is given in (6):

(6)

NP → D N	NP → er	N → Buch
V → NP V	D → das	N → Mann
	D → dem	V → gibt

This grammar licenses binary branching structures as shown in Figure 2.2 on the following page.

Both the grammar in (6) and (2) are too imprecise. If we adopt additional lexical entries for *ich* 'I' and *den* 'the' (accusative) in our grammar, then we would incorrectly license the ungrammatical sentences in (7b–d):[3]

[3] With the grammar in (6), we also have the additional problem that we cannot determine when an utterance is complete since the symbol V is used for all combinations of V and NP. Therefore, we can also analyze the sentence in (i) with this grammar:

(i) a. * der Mann erwartet
the man expects

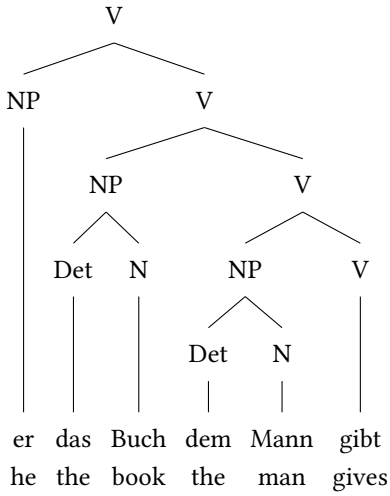

Figure 2.2: Analysis of *er das Buch dem Mann gibt* with a binary branching structure

(7) a. er das Buch dem Mann gibt
he.NOM the.ACC book the.DAT man gives
'He gives the book to the man.'

b. * ich das Buch dem Mann gibt
I.NOM the.ACC book the.DAT man gives

c. * er das Buch den Mann gibt
he.NOM the.ACC book the.ACC man gives

d. * er den Buch dem Mann gibt
he.NOM the.M book(N) the man gives

In (7b), subject-verb agreement has been violated, in other words: *ich* 'I' and *gibt* 'gives' do not fit together. (7c) is ungrammatical because the case requirements of the verb have not been satisfied: *gibt* 'gives' requires a dative object. Finally, (7d) is ungrammatical because there is a lack of agreement between the determiner and the noun. It is not possible to combine *den* 'the', which is masculine and bears accusative case, and *Buch* 'book' because *Buch* is neuter gender. For this reason, the gender properties of these two elements are not the same and the elements can therefore not be combined.

In the following, we will consider how we would have to change our grammar to stop

b. * des Mannes er das Buch dem Mann gibt
the.GEN man.GEN he the.DAT book.DAT the man gives

The number of arguments required by a verb must be somehow represented in the grammar. In the following chapters, we will see exactly how the selection of arguments by a verb (valence) can be captured in various grammatical theories.

it from licensing the sentences in (7b–d). If we want to capture subject-verb agreement, then we have to cover the following six cases in German, as the verb has to agree with the subject in both person (1, 2, 3) and number (sg, pl):

(8)
a. Ich schlafe. (1, sg)
I sleep
b. Du schläfst. (2, sg)
you sleep
c. Er schläft. (3, sg)
he sleeps
d. Wir schlafen. (1, pl)
we sleep
e. Ihr schlaft. (2, pl)
you sleep
f. Sie schlafen. (3, pl)
they sleep

It is possible to capture these relations with grammatical rules by increasing the number of symbols we use. Instead of the rule S → NP NP NP V, we can use the following:

(9)
S → NP_1_sg NP NP V_1_sg
S → NP_2_sg NP NP V_2_sg
S → NP_3_sg NP NP V_3_sg
S → NP_1_pl NP NP V_1_pl
S → NP_2_pl NP NP V_2_pl
S → NP_3_pl NP NP V_3_pl

This would mean that we need six different symbols for noun phrases and verbs respectively, as well as six rules rather than one.

In order to account for case assignment by the verb, we can incorporate case information into the symbols in an analogous way. We would then get rules such as the following:

(10)
S → NP_1_sg_nom NP_dat NP_acc V_1_sg_nom_dat_acc
S → NP_2_sg_nom NP_dat NP_acc V_2_sg_nom_dat_acc
S → NP_3_sg_nom NP_dat NP_acc V_3_sg_nom_dat_acc
S → NP_1_pl_nom NP_dat NP_acc V_1_pl_nom_dat_acc
S → NP_2_pl_nom NP_dat NP_acc V_2_pl_nom_dat_acc
S → NP_3_pl_nom NP_dat NP_acc V_3_pl_nom_dat_acc

Since it is necessary to differentiate between noun phrases in four cases, we have a total of six symbols for NPs in the nominative and three symbols for NPs with other cases. Since verbs have to match the NPs, that is, we have to differentiate between verbs which select three arguments and those selecting only one or two (11), we have to increase the number of symbols we assume for verbs.

(11) a. Er schläft.
he sleeps
'He is sleeping.'

b. * Er schläft das Buch.
he sleeps the book

c. Er kennt das Buch.
he knows the book
'He knows the book.'

d. * Er kennt.
he knows

In the rules above, the information about the number of arguments required by a verb is included in the marking 'nom_dat_acc'.

In order to capture the determiner-noun agreement in (12), we have to incorporate information about gender (fem, mas, neu), number (sg, pl), case (nom, gen, dat, acc) and the inflectional classes (strong, weak[4]).

(12) a. der Mann, die Frau, das Buch (gender)
the man the woman the book

b. das Buch, die Bücher (number)
the book the books

c. des Buches, dem Buch (case)
the.GEN book the.DAT book

d. ein Beamter, der Beamte (inflectional class)
a civil.servant the civil.servant

Instead of the rule NP → D N, we will have to use rules such as those in (13):

(13) NP_3_sg_nom → D_fem_sg_nom_weak N_fem_sg_nom_weak
NP_3_sg_nom → D_mas_sg_nom_weak N_mas_sg_nom_weak
NP_3_sg_nom → D_neu_sg_nom_weak N_neu_sg_nom_weak
NP_3_pl_nom → D_fem_pl_nom_weak N_fem_pl_nom_weak
NP_3_pl_nom → D_mas_pl_nom_weak N_mas_pl_nom_weak
NP_3_pl_nom → D_neu_pl_nom_weak N_neu_pl_nom_weak

NP_3_sg_nom → D_fem_sg_nom_stark N_fem_sg_nom_stark
NP_3_sg_nom → D_mas_sg_nom_stark N_mas_sg_nom_stark
NP_3_sg_nom → D_neu_sg_nom_stark N_neu_sg_nom_stark
NP_3_pl_nom → D_fem_pl_nom_stark N_fem_pl_nom_stark
NP_3_pl_nom → D_mas_pl_nom_stark N_mas_pl_nom_stark
NP_3_pl_nom → D_neu_pl_nom_stark N_neu_pl_nom_stark

[4] These are inflectional classes for adjectives which are also relevant for some nouns such as *Beamter* 'civil servant', *Verwandter* 'relative', *Gesandter* 'envoy', ...). For more on adjective classes see page 23.