Derivation

Derivation is a sequence of production rules. It is used to get input strings. During parsing, we have to take two decisions, which are as follows:

- We have to decide the non-terminal which is to be replaced.
- We have to decide the production rule by which the non-terminal will be replaced.

Two options to decide which non-terminal has to be replaced with the production rule are as follows:

- Left most derivation
- Right most derivation.

Left Most Derivation

In the leftmost derivation, the input is scanned and then replaced with the production rule from left side to right. So, we have to read that input string from left to right.

Example

Production rules:

E → E+E rule1

E → E-E rule2

E → a|b rule3

Let the input be **a-b+a**

Now, when we perform the Left Most Derivation, the result will be as follows −

E → E+E

E → E-E+E from rule2

E → a-E+E from rule3

E $\rightarrow$ a-b+E from rule3

E $\rightarrow$ a-b+a from rule3

Finally, the given string is parsed

Right Most Derivation

In Right most derivation, the input is scanned and replaced with the production rule right to left. So, we have to read the input string from right to left.

Example

Production rule:

E $\rightarrow$ E+E rule1

E $\rightarrow$ E-E rule2

E $\rightarrow$ a|b rule3

Let the input be **a-b+a**

Now, when we perform the Right Most Derivation, we get the following result −

E $\rightarrow$ E-E

E $\rightarrow$ E-E+E from rule1

E $\rightarrow$ E-E+a from rule3

E $\rightarrow$ E-b+a from rule3

E $\rightarrow$ a-b+a from rule3

A grammar G is called **ambiguous** if there exist a string **s** with two different derivations from G.

For example, Arithmetic expression grammar

E→ 0 |1|…| 9 | (E) | E*E | E+E

Is **ambiguous** because the sentence **2+3*4** has **two different derivations.**

$E \rightarrow \underline{E}*E \rightarrow \underline{E}+E*E \rightarrow 2+\underline{E}*E \rightarrow 2+3*\underline{E} \rightarrow 2+3*4$

$E \rightarrow \underline{E}+E \rightarrow 2+\underline{E} \rightarrow 2+\underline{E}*E \rightarrow 2+3*\underline{E} \rightarrow 2+3*4$

Pars trees

Here is another grammar for arithmetic expressions:

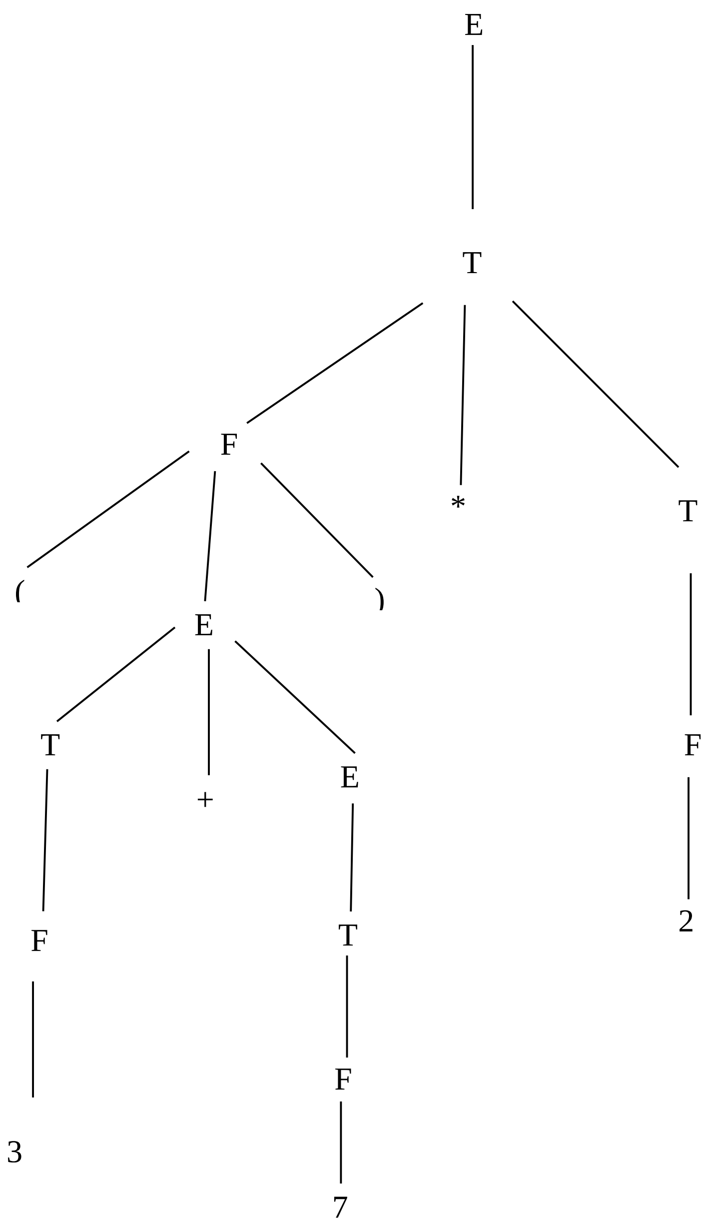
$E \rightarrow T|T+E$

$T \rightarrow F|F*T$

$F \rightarrow 0|1|\ldots|9|(E)$

This grammar is unambiguous.

Here is the parse tree for **(3+7)*2**

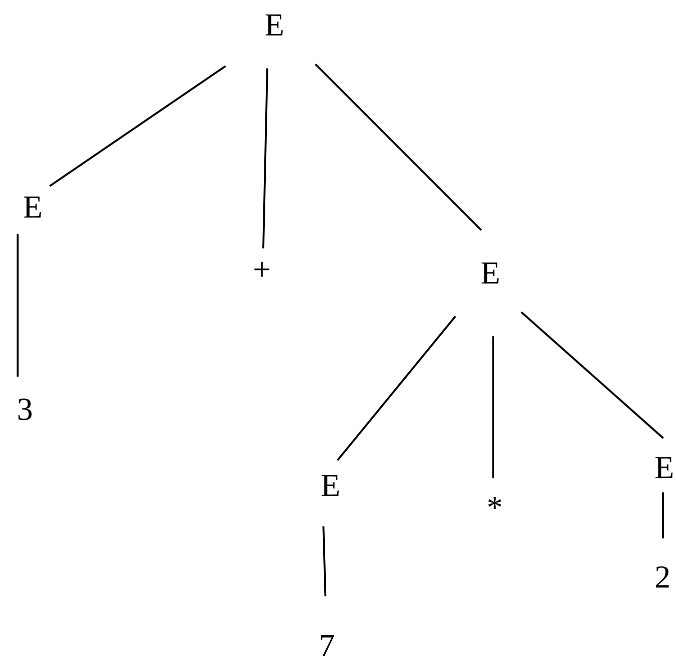


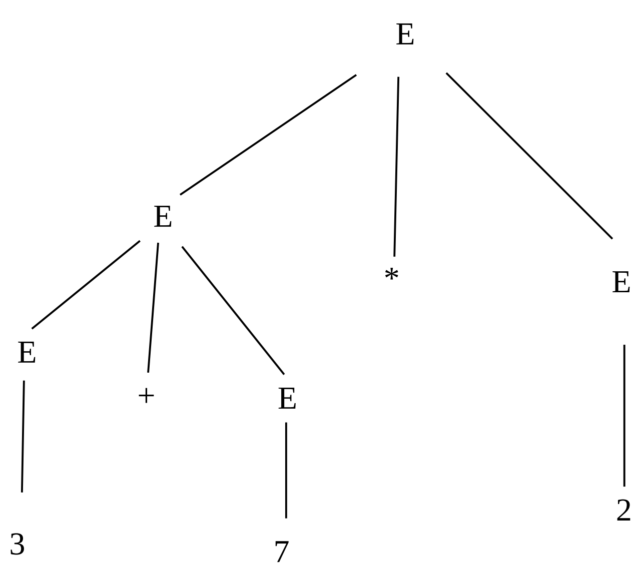
This is only parse tree for this sentence (using this grammar).

In contrast, consider the previous grammar

E→ 0 |1 |…| 9| (E)| E*E|E+E

This grammar has two different parse trees for the sentence 3+7*2.





Previously, we said a grammar is ambiguous if there exists some sentence with two different derivations.

Equivalently, a grammar is ambiguous if there exists some sentence with two different parse tree.

Examples:

Write the grammar to the regular expression

$R=(a+b)^*$

$S \longrightarrow X \mid Y$

$X \longrightarrow \varepsilon$

$Y \longrightarrow a \mid b \mid aY \mid bY$

Example:

1- $L(G)=\{0^n1^n, n \geq 0\}$

$S \rightarrow 0S1 \mid \lambda$

The derivation of (0011) is

$S \rightarrow 0S1 \rightarrow 00S11 \rightarrow 0011$

2- Let G: $S \rightarrow aSa \mid bSb \mid a \mid b \mid \lambda$

The derivation of (abaaba) is

$S \rightarrow aSa \rightarrow abSba \rightarrow abaSaba \rightarrow abaaba$

H.W 1 : $L(G)=\{0^n1^n, n \geq 1\}$

H.W 2 : $L(G)=\{a^ib^j , i \geq j\}$