2.8 The Pumping Lemma

Theorem 2.34 (Pumping Lemma)

Let L be a regular language. Then there exists a positive integer n such that each word $x \in L$ of length $|x| \geq n$ admits a factorization of the form $x = uvw$ that satisfies all of the following properties:

(1) $|v| \geq 1$,

(2) $|uv| \leq n$,

(3) $uv^i w \in L$ for all $i \in \mathbb{N}$.

Proof.

Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA for L.
We choose $n := |Q|$, and consider a word $x \in L$ satisfying $|x| \geq n$.

$$A : q_0 \xrightarrow[x_1]{} q_1 \xrightarrow[x_2]{} \cdots \xrightarrow[x_{n-1}]{} q_{n-1} \xrightarrow[x_n]{} q_n \xrightarrow[x']{*} q' \in F,$$

where $x = x_1 x_2 \ldots x_n x'$, $x_1, x_2, \ldots, x_n \in \Sigma$ and $x' \in \Sigma^*$.

Proof of Theorem 2.34 (cont.)

Then there are r and s such that $0 \leq r < s \leq n$ and $q_r = q_s$.
Hence, x can be written as $x = uvw$, where

$$\begin{array}{rcl} |u| &=& r, \\ 1 \leq |v| &=& s - r \leq n, \\ \hat{\delta}(q_0, u) &=& q_r, \\ \hat{\delta}(q_r, v) &=& q_s = q_r, \\ \hat{\delta}(q_s, w) &=& q' \in F. \end{array}$$

It follows that

$$\begin{aligned} \hat{\delta}(q_0, uv^iw) = \hat{\delta}(q_r, v^iw) &= \hat{\delta}(q_s, v^{i-1}w) \\ &= \hat{\delta}(q_r, v^{i-1}w) = \hat{\delta}(q_s, w) = q' \in F, \end{aligned}$$

that is, $uv^iw \in L(A) = L$. □

Example 1:

Claim: $L = \{ a^m b^m \mid m \geq 1 \}$ is not regular.

Proof (indirect):

Assume that L were regular. Then L satisfies the Pumping Lemma, that is, $\exists n \in \mathbb{N}_+ \forall x \in L : |x| \geq n \leadsto \exists x = uvw :$ $|v| \geq 1, |uv| \leq n,$ and $uv^i w \in L$ for all $i \geq 0$.

Consider the word $x := a^n b^n : x \in L$ and $|x| = 2n > n$.
Hence: $\exists x = uvw$ s. t. $|v| \geq 1$, $|uv| \leq n$, and $uv^i w \in L$ for all $i \geq 0$.

$$x = a^n b^n = uvw, \text{ where } |uv| \leq n$$
$$\leadsto u = a^r, v = a^s, \text{ and } w = a^{n-s-r} b^n$$

for certain integers r, s satisfying $r \geq 0$, $s \geq 1$, $r + s \leq n$.
Thus: $uv^0 w = a^r a^{n-s-r} b^n = a^{n-s} b^n \notin L$, a contradiction!

As L does not satisfy the Pumping Lemma, it is not regular. □

Example 2:

Claim: $L = \{ 0^m \mid m \text{ is a square number} \}$ is not regular.

Proof (indirect)

Assume that L were regular.
Let n be the constant for L from the Pumping Lemma.
Consider the word $x := 0^{n^2}$: $x \in L$ and $|x| = n^2 > n$.
Hence: $\exists x = uvw$ s. t. $|v| \geq 1$, $|uv| \leq n$, and $uv^iw \in L$ for all $i \geq 0$.

Now consider the word uv^2w:

$$\begin{array}{rcl} n^2 & = & |uvw| \quad < \quad |uv^2w| \\ & = & |uvw| + |v| \quad = \quad n^2 + |v| \\ & \leq & n^2 + n \quad < \quad n^2 + 2n + 1 \\ & = & (n+1)^2, \end{array}$$

that is, $|uv^2w|$ is not a square number, and so, $uv^2w \notin L$.
This contradiction shows that L is not regular. □

Example 3:

Let $L = \{ c^m a^n b^n \mid m, n \geq 0 \} \cup \{a, b\}^*$.

Claim: L satisfies the Pumping Lemma with the constant $k = 1$.

Proof.

Let $x \in L$, where $|x| \geq k$.

(i) $x \in \{a, b\}^*$: obvious.

(ii) $x = c^m a^n b^n$ for some $m \geq 1$:
Choose $u := \varepsilon$, $v := c$, $w := c^{m-1} a^n b^n$.
Then: $x = uvw,\ 1 \leq |v|,\ |uv| = 1 \leq k,$
$uv^i w = c^{i+m-1} a^n b^n \in L$ for all $i \geq 0$. □

Claim: L is not regular.

Proof.

L has infinite index, and hence, by Theorem 2.12, it is not regular. □

2.9 Decision Problems

The membership problem for a regular language:

INSTANCE : $x \in \Sigma^*$.
QUESTION : Is x in L?

This problem is solvable in time $|x|$ using a DFA.

The emptiness problem for a DFA (NFA):

INSTANCE : A DFA (NFA) A.
QUESTION : Is $L(A) = \emptyset$?

This is decidable in time $O(|A|^2)$, as $L(A) \neq \emptyset$ iff a final state is reachable from the initial state in the graph of A.

The finiteness problem for a DFA (NFA):

INSTANCE	:	A DFA (NFA) A.
QUESTION	:	Is $L(A)$ finite?

This is decidable in polynomial time, as $L(A)$ is infinite iff

$\exists q_0$ initial state $\exists q_1 \exists q_2$ final state: $q_0 \rightarrow^* q_1 \rightarrow^+ q_1 \rightarrow^* q_2$,

which can be checked using the graph of A.

The intersection emptiness problem for regular grammars (or NFAs):

INSTANCE	:	Two grammars G_1 and G_2.
QUESTION	:	Is $L(G_1) \cap L(G_2)$ empty?

This is decidable in quadratic time:
Construct an NFA (or a grammar) for $L(G_1) \cap L(G_2)$
and test for emptiness.

The inclusion problem for regular languages:

INSTANCE	:	Two regular grammars G_1 and G_2.
QUESTION	:	Is $L(G_1)$ a subset of $L(G_2)$?

Decidable, as $L(G_1) \subseteq L(G_2)$ iff $L(G_1) \cap \overline{L(G_2)} = \emptyset$, and a DFA for $L(G_1) \cap \overline{L(G_2)}$ can be constructed from G_1 and G_2.

If $L(G_1)$ and $L(G_2)$ are given through DFAs, then this problem is decidable in quadratic time.

The equivalence problem for regular languages:

INSTANCE	:	Two regular grammars G_1 and G_2.
QUESTION	:	Are $L(G_1)$ and $L(G_2)$ equal?

Decidable, as $L(G_1) = L(G_2)$ iff $L(G_1) \subseteq L(G_2)$ and $L(G_2) \subseteq L(G_1)$.

Chapter 3:

Context-Free Languages and Pushdown Automata

3.1. Context-Free Grammars

A phrase-structure grammar $G = (N, T, S, P)$ is called context-free, if $\ell \in N$ for each production $(\ell \to r) \in P$.

A language $L \subseteq T^*$ is called context-free, if there exists a context-free grammar G satisfying $L(G) = L$.

By $\mathrm{CFL}(\Sigma)$ we denote the class of all context-free languages over Σ, and CFL is the class of all context-free languages.

Obiously, we have $\mathrm{REG} \subseteq \mathrm{CFL}$.

Let $G = (N, T, S, P)$ be a context-free grammar, and
let $\alpha_0 \to_G \alpha_1 \to_G \ldots \to_G \alpha_n$ be a derivation in G.
Then there exist $\beta_i, \gamma_i \in (N \cup T)^*$ and $(A_i \to r_i) \in P$ such that

$$\alpha_i = \beta_i A_i \gamma_i \text{ and } \alpha_{i+1} = \beta_i r_i \gamma_i \quad (0 \leq i \leq n-1).$$

This derivation is called a left derivation if $|\beta_i| \leq |\beta_{i+1}|$ for all i,
it is called a right derivation if $|\gamma_i| \leq |\gamma_{i+1}|$ for all i,
and it is called a leftmost derivation if $\beta_i \in T^*$ for all i.
If $\alpha_i \to \alpha_{i+1}$ is a step in a leftmost derivation, this is denoted as $\alpha_i \to_{\mathrm{lm}} \alpha_{i+1}$.

Remark:

If $\alpha_0 \in N$ and $\alpha_n \in T^*$, then each left derivation $\alpha_0 \to \alpha_1 \to \cdots \to \alpha_n$ is necessarily leftmost.

Example:

(a) $G_1 := (\{A\}, \{a\}, A, \{A \to AA, A \to a\})$.

(b) $G_2 := (\{S\}, \{a, b\}, S, \{S \to aSa, S \to bSb, S \to \varepsilon\})$.

(c) $G_3 := (\{A\}, \{[,], a, \#, \uparrow\}, A, \{A \to [A\#A], A \to [A \uparrow A], A \to a\})$.

(a): The derivation $A \to AA \to Aa \to AAa \to aAa \to aaa$
is neither a left nor a right derivation.
(b): Each derivation that starts with S is a left and a right derivation.
(c): The derivation

$$A \to [A\#A] \to [A\#[A \uparrow A]] \to [a\#[A \uparrow A]] \to [a\#[a \uparrow A]]$$

is neither a left nor a right derivation, but the derivation

$$A \to [A\#A] \to [a\#A] \to [a\#[A \uparrow A]] \to [a\#[a \uparrow A]]$$

is a leftmost derivation.

The grammar $G = (N, T, S, P)$ is called unambiguous w.r.t. $A \in N$ if there exists exactly one left derivation $A \rightarrow_P^* \alpha$ for each $\alpha \in L(G, A)$.

It is called unambiguous if it is unambiguous w.r.t. all its nonterminals.

It is called ambiguous if it is not unambiguous.

A context-free language is called unambiguous if it is generated by a context-free grammar $G = (N, T, S, P)$ that is unambiguous.

A context-free language L is called inherently ambiguous if it is not generated by any unambiguous grammar.

Example (cont.):

(a) G_1 is not unambiguous, as

$$A \to AA \to AAA \to aAA \to aaA \to aaa$$

and

$$A \to AA \to aA \to aAA \to aaA \to aaa$$

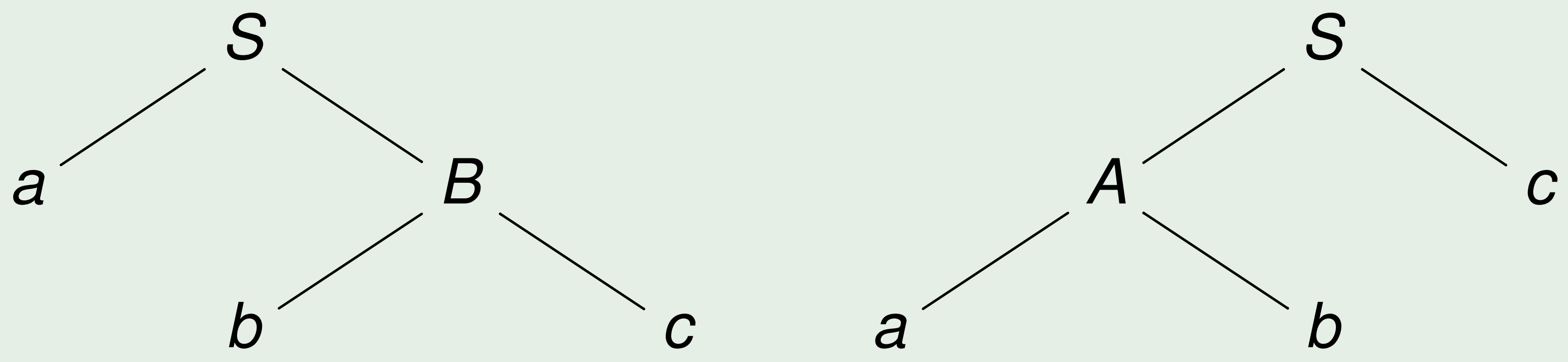
are two different left derivations for aaa.

(b) G_2 is trivially unambiguous.

(c) It can be shown that G_3 is unambiguous.

Example:

Let $G = (\{S, A, B\}, \{a, b, c\}, P, S)$, where
$P := \{S \to aB, S \to Ac, A \to ab, B \to bc\}$:



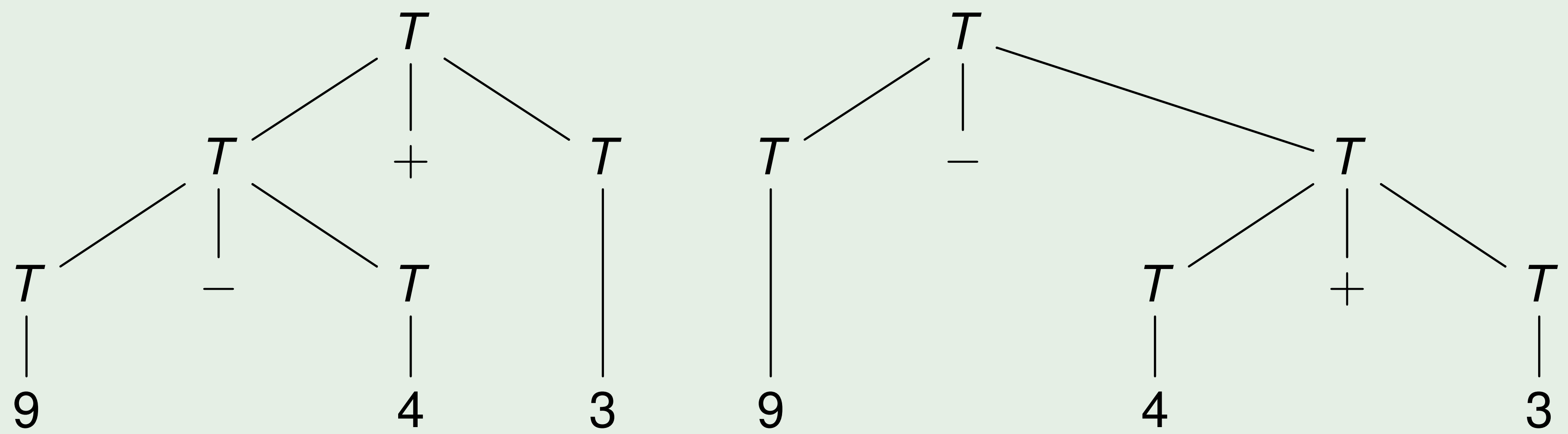
G is ambiguous.

$G' := (\{S\}, \{a, b, c\}, \{S \to abc\}, S)$ is unambiguous and
$L(G') = L(G)$.

The language $L := \{ a^i b^j c^k \mid i = j \text{ or } j = k \}$ is inherently ambiguous.

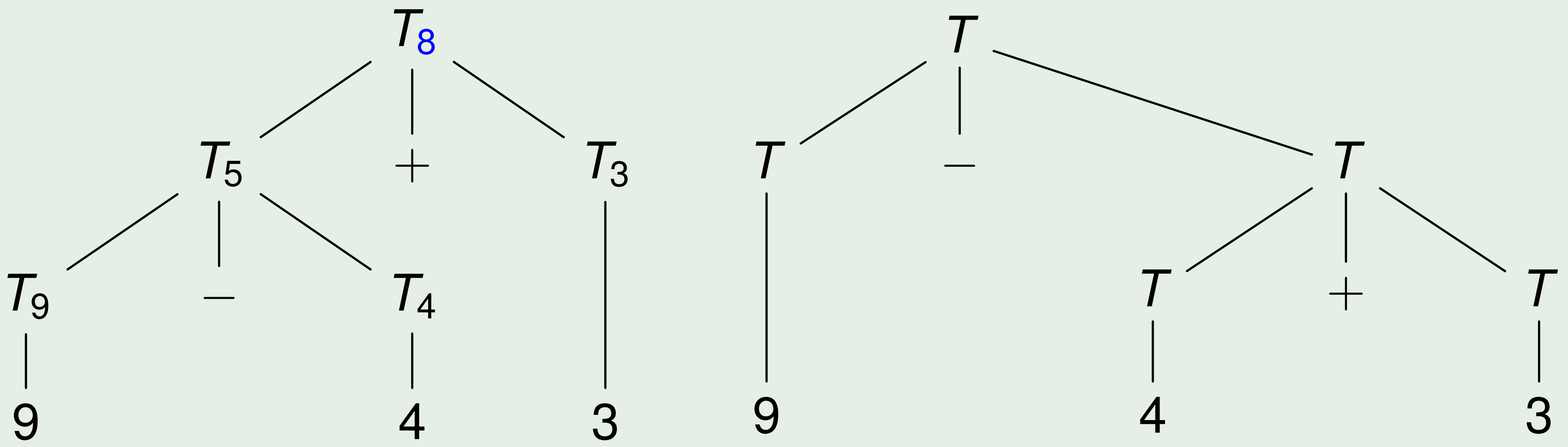
Example:

$G = (\{T\}, \{1, 2, \ldots, 9, +, -\}, \{T \to T + T \mid T - T \mid 1 \mid 2 \mid \cdots \mid 9\}, T)$:



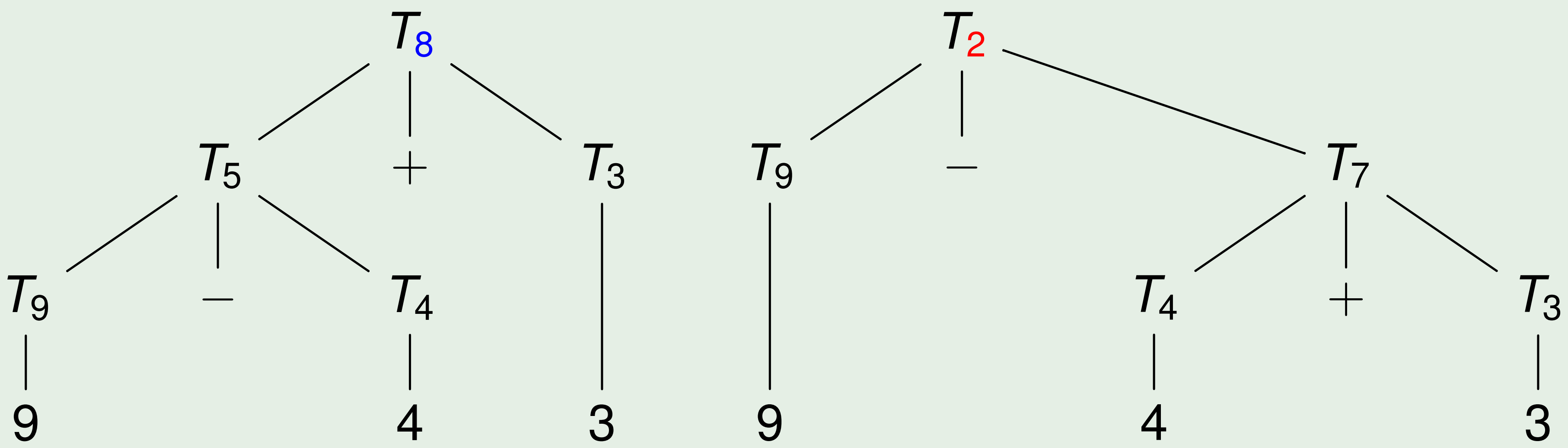
Example:

$G = (\{T\}, \{1, 2, \ldots, 9, +, -\}, \{T \to T + T \mid T - T \mid 1 \mid 2 \mid \cdots \mid 9\}, T)$:



Example:

$G = (\{T\}, \{1, 2, \ldots, 9, +, -\}, \{T \to T + T \mid T - T \mid 1 \mid 2 \mid \cdots \mid 9\}, T)$:



Syntax Trees

Let $G = (V, T, P, S)$ be a context-free grammar, and let $S = x_0 \to x_1 \to x_2 \to \ldots \to x_n = x$ be a derivation in G for a word $x \in L(G)$.

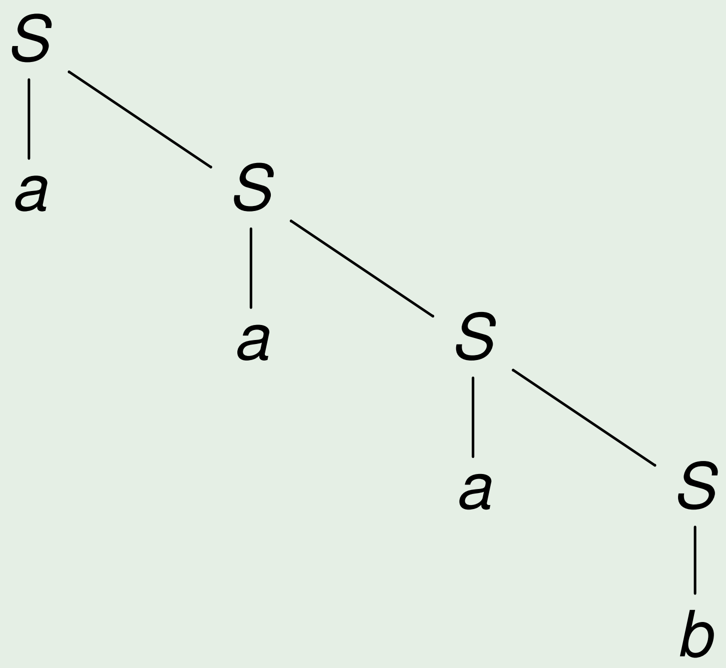
We can describe this derivation through a syntax tree:

Start: Introduce a root with label S.

Step i: If $x_{i-1} = uAv \to urv = x_i$, where $u, v \in (V \cup \Sigma)^*$ and $(A \to r) \in P$, then add $|r|$ children to the node with label A which are labelled from left to right with the symbols of r.

Example:

$G = (\{S\}, \{a, b\}, \{S \to aS, S \to b\}, S)$:

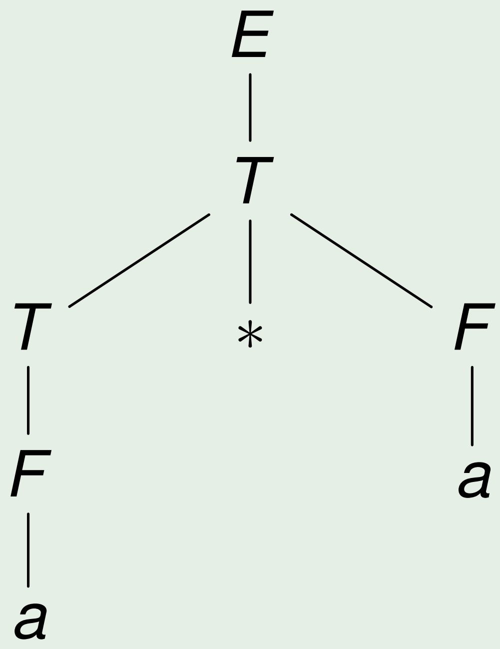


Beispiel:

Let $G = (\{E, F, T\}, \{a, *\}, E, \{E \to T, T \to F, T \to T * F, F \to a\})$:

(1.) $\underline{E} \to \underline{T} \to \underline{T} * F \to \underline{F} * F \to a * \underline{F} \to a * a$

(2.) $\underline{E} \to \underline{T} \to T * \underline{F} \to \underline{T} * a \to \underline{F} * a \to a * a$



(1) Left derivation
(2) Right derivation

Theorem 3.1

Let $G = (N, T, S, P)$ be a context-free grammar.

(a) *If $A \in N$ and $x \in (N \cup T)^*$ such that $A \to^* x$, then there exists a syntax tree with root labelled A and leaves labelled with x.*

(b) *If there exists a syntax tree with root labelled $A \in N$ and leaves labelled with $x \in T^*$, then $A \to_{\mathrm{lm}}^* x$.*

(c) *For each nonterminal $A \in N$ and each word $x \in L(G, A)$, the number of leftmost derivations of x from A coincides with the number of syntax trees with root labelled A and leaves labelled with x.*

Corollary 3.2

A context-free grammar G is unambiguous if and only if there exists a unique syntax tree with root labelled A and leaves labelled x for each $A \in N$ and each word $x \in L(G, A)$.