Context-free grammars

- a set of terminals $\Sigma$
- a set of nonterminals or variables $V$
  - syntactic categories
- a starting nonterminal $S$
  - main category being defined
- a set of productions (rewrite rules) $R$ of the form

  $$\text{variable } \rightarrow \text{string over } V \cup \Sigma$$

**Example:** $\Sigma = \{a, b\}$, $V = \{S\}$,

$$R = \{ S \rightarrow aSb, \ S \rightarrow \epsilon \}$$

Derivation step: $S$ can be replaced with $aSb$ or $\epsilon$
Derivation step

$\Rightarrow_R$ is a binary relation on $(V \cup \Sigma)^*$, defined as

\{(uAv, u\alpha v) \mid u, v \in (V \cup \Sigma)^*, (A \rightarrow \alpha) \in R\}

If $A \rightarrow \alpha$ is a rule in $R$, then

$uAv \Rightarrow u\alpha v \quad (uAv \text{ yields } u\alpha v)$

for all $u, v \in (V \cup \Sigma)^*$

$\Rightarrow^*$ is the reflexive, transitive closure of $\Rightarrow$.

$L(G) = \{w \in \Sigma^* \mid S \Rightarrow^* w\}$
Example 1

\[ \Sigma = \{a, b\}, \; V = \{S\}, \]

\[ R = \{S \rightarrow aSb, \; S \rightarrow aS, \; S \rightarrow \epsilon\} \]

\[ S \Rightarrow aSb \Rightarrow aaSb \Rightarrow aab \]

\[ S \Rightarrow aS \Rightarrow aaSb \Rightarrow aab \]

\[ L(G) = \{a^i b^j \mid i \geq j \geq 0\} \]
Example 2

\[ \Sigma = \{a, b\}, \ V = \{S\}, \]

\[ R = \{S \rightarrow aSa \mid bSb \mid a \mid b \mid \epsilon\} \]

\[ S \Rightarrow aSa \Rightarrow aba \]

\[ S \Rightarrow aSa \Rightarrow abSba \Rightarrow abba \]

\[ L(G) = \{w \mid w = w^R\} \]
Closure properties – Union

\[ G_1 = (V_1, \Sigma, R_1, S_1) \]
\[ G_2 = (V_2, \Sigma, R_2, S_2) \]

where \( V_1 \cap V_2 = \emptyset \).

Let \( S \) be a new nonterminal.

\[ G = (V, \Sigma, R, S) \]

where

\[ V = V_1 \cup V_2 \cup \{ S \} \quad \text{and} \]

\[ R = R_1 \cup R_2 \cup \{ S \rightarrow S_1 \mid S_2 \} \]

\[ L(G) = L(G_1) \cup L(G_2) \]


Closure properties – Concatenation

\[ G_1 = (V_1, \Sigma, R_1, S_1) \]
\[ G_2 = (V_2, \Sigma, R_2, S_2) \]

where \( V_1 \cap V_2 = \emptyset \).

Let \( S \) be a new nonterminal.

\[ G = (V, \Sigma, R, S) \] where

- \( V = V_1 \cup V_2 \cup \{S\} \) and
- \( R = R_1 \cup R_2 \cup \{S \to S_1S_2\} \)

\[ L(G) = L(G_1) \circ L(G_2) \]
\[ L = \{a^i b^j c^k \mid j = i + k\} \]

\[ L = L_1 L_2 \quad \text{where} \]

\[ L_1 = \{a^m b^m \mid m \geq 0\} \]

\[ L_2 = \{b^n c^n \mid n \geq 0\} \]

\[
S \rightarrow S_1 S_2 \\
S_1 \rightarrow aS_1 b \mid \epsilon \\
S_2 \rightarrow bS_2 c \mid \epsilon
\]
Closure properties – Star

\[ G_1 = (V_1, \Sigma, R_1, S_1) \]

Let \( S \) be a new nonterminal.

\[ G = (V, \Sigma, R, S) \] where

- \( V = V_1 \cup \{S\} \) and
- \( R = R_1 \cup \{S \to SS_1 \mid \epsilon\} \)

\[ L(G) = L(G_1)^* \]
Derivation tree

$\Sigma = \{a, b\}, \ V = \{S\},$

$\ R = \{S \rightarrow aSb \mid \varepsilon\}$
Derivation tree

- rooted and ordered

- nodes are labelled with nonterminals, terminals or $\epsilon$

- only nonterminal nodes have children
  leaf nodes are labelled with terminals or $\epsilon$

- if $A$ is a node and $a_1, \ldots, a_n$ (each $a_i \in V \cup \Sigma$) its children from left to right, then it must be that $A \rightarrow a_1 \ldots a_n$ is a rule in the grammar

- if $A$ is a node and $\epsilon$ is its only child, then it must be that $A \rightarrow \epsilon$ is a rule in the grammar
Ambiguity

A grammar is ambiguous if there is a string with two or more derivation trees.

\[ \Sigma = \{a, b\}, \ V = \{S\}, \]
\[ R = \{S \rightarrow SS \mid a\} \]

String \(aaa\) has two distinct derivation trees.
Ambiguity: another example

\[ \Sigma = \{a, b\}, \quad V = \{S\}, \]

\[ R = \{S \rightarrow aSb, \quad S \rightarrow aS, \quad S \rightarrow \epsilon\} \]
Ambiguity: yet another example

\[ \Sigma = \{a\}, \ V = \{S\}, \]

\[ R = \{S \rightarrow aSSa \mid a\} \]

String \(a^7\) (aaaaaaaa) has two distinct derivation trees.
For the language $a^+$,

$$\{S \rightarrow SS \mid a\} \text{ is ambiguous, but}$$

$$\{S \rightarrow aS \mid a\} \text{ is not.}$$