Context Free Grammars

Example: Palindromes can be expressed by the following

$S \longrightarrow \epsilon$

$S \longrightarrow a$

$S \longrightarrow b$

$S \longrightarrow aSa$

$S \longrightarrow bSb$

CFG

$G = (V, T, P, S)$,
where

- V: Set of variables or non-terminals.

- T: Set of terminals

  $(V \cap T = \emptyset)$.

- P: finite set of productions. Each production if of form $A \rightarrow \gamma$, where $A \in V$ and $\gamma \in (V \cup T)^*$.

- S: start symbol

Derivations

$\alpha A\beta \Rightarrow_G \alpha\gamma\beta$, if there is a production of form $A \rightarrow \gamma$.

We now define $\alpha \Rightarrow^*_G \beta$.

Basis: $\alpha \Rightarrow^*_G \alpha$ for all $\alpha \in (V \cup T)^*$.

Induction: If $\alpha \Rightarrow^*_G \beta$ and $\beta \Rightarrow_G \gamma$, then $\alpha \Rightarrow^*_G \gamma$.

$L(G) = \{w \in T^* \mid S \Rightarrow^*_G w\}$.

Sentential Forms

If $S \Rightarrow_G^* \alpha$, then α is called a sentential form.

Left Most and Right Most Derivations

In Left Most Derivation, in each step of the derivation, one replaces the leftmost non-terminal in the sentential form.

In Right Most Derivation, in each step of the derivation, one replaces the rightmost non-terminal in the sentential form.

Parse Trees

Right-Linear Grammars

A CFG is called right linear if all the productions in it are of the form:

$A \to wB$, for $B \in V$ and $w \in T^*$, or
$A \to w$, for $w \in T^*$.

Theorem: Every regular language can be accepted by some right-linear grammar.

Theorem: Language accepted by a right-linear grammar is regular.

Theorem: Every regular language can be accepted by some right-linear grammar.
Suppose $A = (Q, \Sigma, \delta, q_0, F)$.
(Without loss of generality, assume that $Q \cap \Sigma = \emptyset$).
Then, let $G = (Q, \Sigma, P, q_0)$, where
i) For $q, p \in Q$, $a \in \Sigma$,
if $\delta(q, a) = p$, then we have a production in P of form $q \longrightarrow ap$.
ii) We also have productions, $q \longrightarrow \epsilon$, for each $q \in F$.
Prove by induction on length of w that $\hat{\delta}(q_0, w) = p$ iff $q_0 \Rightarrow_G^* wp$.
This would also give us that
$\hat{\delta}(q_0, w) \in F$ iff $q_0 \Rightarrow_G^* w$.

Theorem: Language accepted by a right-linear grammar is regular.
Suppose $G = (V, \Sigma, P, S)$.
(Without loss of generality, assume that $V \cap \Sigma = \emptyset$).
Assume without loss of generality that each production is of form $A \rightarrow bC$, or of form $A \rightarrow \epsilon$, where $b \in \Sigma \cup \{\epsilon\}$, $A, C \in V$.
Then, define NDFA $A = (V, \Sigma, \delta, S, F)$, as follows.
If there is a production of form $A \rightarrow aB$, then $B \in \delta(A, a)$.
$F = \{A \mid A \rightarrow \epsilon \text{ is a production in } P\}$.
Show by induction that
$A \Rightarrow_G^* wB$ iff $B \in \hat{\delta}(A, w)$.
This would also give,
$A \Rightarrow_G^* w$ iff $\hat{\delta}(A, w) \in F$.

Ambiguous Grammars

Consider
$E \longrightarrow E + E$
$E \longrightarrow E * E$
$E \longrightarrow id.$
Consider derivation of $id + id * id$.

It can be done in 2 ways:

$$E \longrightarrow E + E \longrightarrow id + E \longrightarrow id + E * E \longrightarrow id + id * E \longrightarrow id + id * id.$$

$$E \longrightarrow E * E \longrightarrow E + E * E \longrightarrow id + E * E \longrightarrow id + id * E \longrightarrow id + id * id.$$

$$S \longrightarrow S + T$$
$$S \longrightarrow T$$
$$T \longrightarrow T * id$$
$$T \longrightarrow id$$

Inherently ambiguous languages.
$L = \{a^n b^n c^m d^m \mid n, m \geq 1\} \cup \{a^n b^m c^m d^n \mid n, m \geq 1\}$.
Any grammar for above language is ambiguous.