

Context Free Grammars

Example: Palindromes can be expressed by the following

$$S \rightarrow \epsilon$$

$$S \rightarrow a$$

$$S \rightarrow b$$

$$S \rightarrow aSa$$

$$S \rightarrow bSb$$

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 is 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:

$$A \rightarrow wB, \text{ for } B \in V \text{ and } w \in T^*, \text{ or}$$
$$A \rightarrow w, \text{ for } w \in T^*.$$

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 $q \rightarrow ap$.

ii) We also have productions, $q \rightarrow \epsilon$, 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 NFA $A = (V, \Sigma, \delta, S, F)$, 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

$$E \rightarrow E + E$$

$$E \rightarrow E * E$$

$$E \rightarrow id.$$

Consider derivation of $id + id * id$.

It can be done in 2 ways:

$$E \rightarrow E + E \rightarrow id + E \rightarrow id + E * E \rightarrow id + id * E \rightarrow id + id * id.$$

$$E \rightarrow E * E \rightarrow E + E * E \rightarrow id + E * E \rightarrow id + id * E \rightarrow id + id * id.$$

$$S \rightarrow S + T$$

$$S \rightarrow T$$

$$T \rightarrow T * id$$

$$T \rightarrow id$$

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.