Context Free Languages and Grammars

Example: Palindromes can be expressed by the following

$S \rightarrow \epsilon$

$S \rightarrow a$

$S \rightarrow b$

$S \rightarrow aSa$

$S \rightarrow bSb$

Another Example

$$E \to id$$
$$E \to E + E$$
$$E \to E * E$$

Another Example:

$S \to id = E$

$S \to$ If E Then $id = E$ Else $id = E$ EndIf

$S \to S; S$

CFG

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

- V: A finite set of variables or non-terminals.
- T: A finite set of terminals
  ($V \cap T = \emptyset$).
- P: finite set of productions. Each production is of the form $A \rightarrow \gamma$, where $A \in V$ and $\gamma \in (V \cup T)^*$.
- S: start symbol, a member of V

Simpler notation to make writing shorter:

$S \rightarrow a|b|\epsilon|aSa|bSb$

Derivations

$\alpha A\beta \Rightarrow_G \alpha\gamma\beta$, if there is a production of the form $A \rightarrow \gamma$.
We now define $\alpha \Rightarrow_G^* \beta$.
Base case: $\alpha \Rightarrow_G^* \alpha$ for all $\alpha \in (V \cup T)^*$.
Induction: If $\alpha \Rightarrow_G^* \beta$ and $\beta \Rightarrow_G \gamma$, then $\alpha \Rightarrow_G^* \gamma$.
$L(G) = \{w \in T^* \mid S \Rightarrow_G^* w\}$.
If clear from context, we drop G from $\Rightarrow_G^*$, and just write $\Rightarrow^*$.

Derivations

Consider the grammar
$S \rightarrow A1B$
$A \rightarrow 0A \mid \epsilon$
$B \rightarrow 0B \mid 1B \mid \epsilon$
The above generates the language $0^*1(0+1)^*$

Consider the derivation of 00110.
$S \Rightarrow A1B \Rightarrow A11B \Rightarrow 0A11B \Rightarrow 00A11B \Rightarrow 00A110B \Rightarrow$
$00A110 \Rightarrow 00110$
Another possible derivation is:
$S \Rightarrow A1B \Rightarrow 0A1B \Rightarrow 00A1B \Rightarrow 001B \Rightarrow 0011B \Rightarrow 00110B \Rightarrow$
00110

Sentential Forms

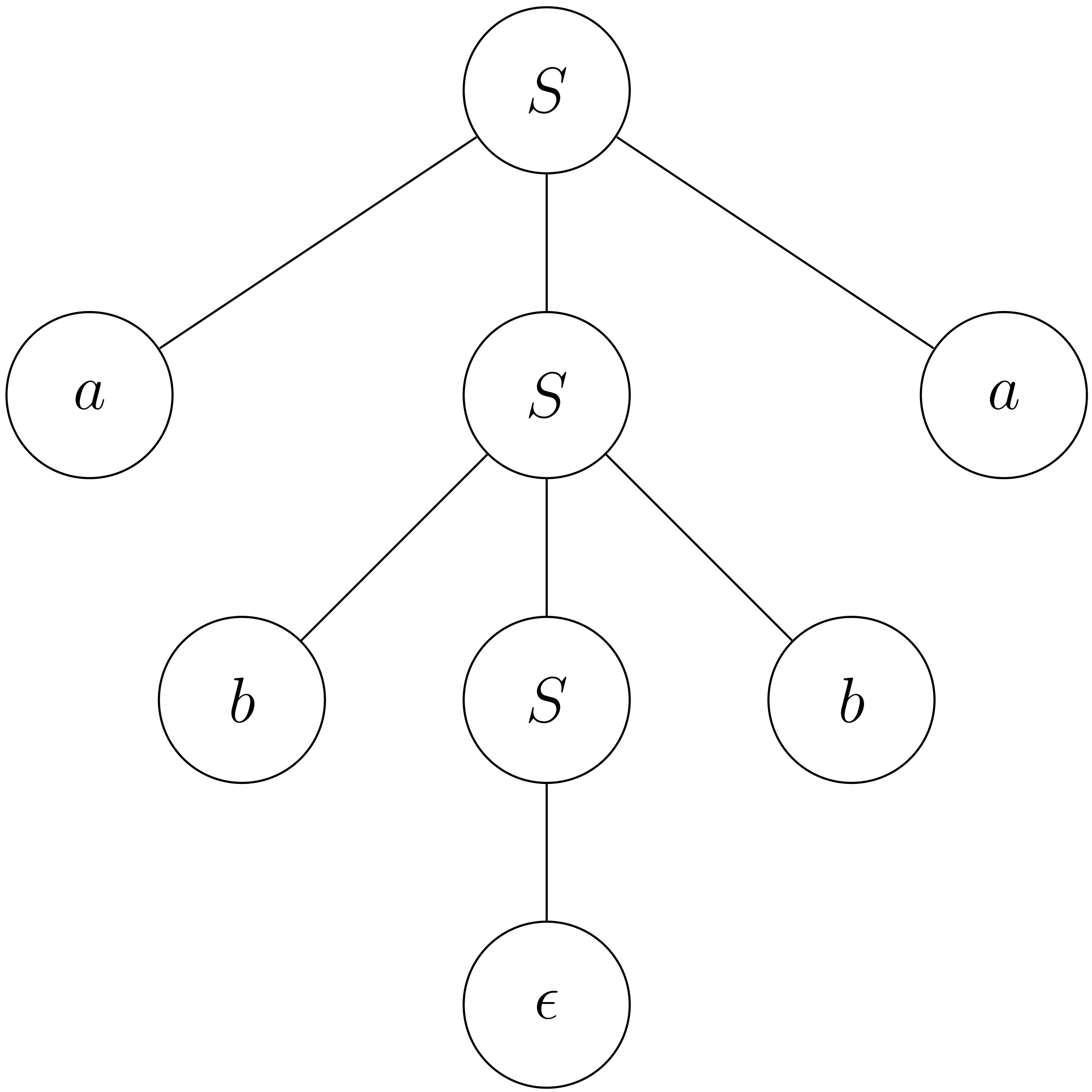
If $S \Rightarrow_G^* \alpha$, then α is called a sentential form.

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

Theorem: Every regular language can be generated by some right-linear grammar.
Suppose L is accepted by DFA $A = (Q, \Sigma, \delta, q_0, F)$.
(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 the 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 $q_0 \Rightarrow_G^* w$.

Inductive proof for: $\hat{\delta}(q_0, w) = p$ iff $q_0 \Rightarrow_G^* wp$.

If w is of length 0, then clearly, $\hat{\delta}(q_0, \epsilon) = p$ iff $p = q_0$, and $q_0 \Rightarrow_G^* p$ iff $q_0 = p$.

Suppose by induction, the claim holds for all w of length k, then we show that it holds for all w of length $k+1$.

So consider any $w = ua$, where length of u is k.

Then, $\hat{\delta}(q_0, ua) = p'$ iff $(\exists p)[\hat{\delta}(q_0, u) = p$ and $\delta(p, a) = p']$ iff $(\exists p)[q_0 \Rightarrow_G^* up$ and $p \rightarrow ap']$ iff $[q_0 \Rightarrow_G^* uap']$.

Theorem: Language generated 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 the form $A \rightarrow bC$, or of the 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 the form $A \rightarrow aB$, then $B \in \delta(A, a)$.
$F = \{A \mid A \rightarrow \epsilon$ 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 $\hat{\delta}(A, w) \cap F \neq \emptyset$
and thus,
$S \Rightarrow^*_G w$ iff $\hat{\delta}(S, w) \cap F \neq \emptyset$

To see that we can assume without loss of generality that every production in a right linear grammar is of the form $A \to bB$ or $A \to \epsilon$, where $B \in V$ and $b \in \Sigma \cup \{\epsilon\}$, we can do as follows.

If $A \to b_1 b_2 \ldots b_n B$ is a production, where $n \geq 1$, then it can be replaced by the productions:

$A \to b_1 B_1$

$B_1 \to b_2 B_2$

$\ldots$

$B_{n-1} \to b_n B$

where $B_1, B_2, \ldots, B_{n-1}$ are new non-terminals.

A production of the form $A \to b_1 b_2 \ldots b_n$, where $n \geq 1$ can be replaced by
$A \to b_1 B_1$
$B_1 \to b_2 B_2 \ \ldots$
$B_{n-1} \to b_n B_n$
$B_n \to \epsilon$
where $B_1, B_2, \ldots, B_n$ are new non-terminals.

Ambiguous Grammars

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

It can be done in 2 ways:

$$E \Rightarrow E + E \Rightarrow id + E \Rightarrow id + E * E \Rightarrow id + id * E \Rightarrow id + id * id.$$

$$E \Rightarrow E * E \Rightarrow E + E * E \Rightarrow id + E * E \Rightarrow id + id * E \Rightarrow id + id * id.$$

$$S \to S + T$$
$$S \to T$$
$$T \to T * id$$
$$T \to id$$

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.
Note that above is a context free language as shown by following grammar:

$S \to A|B$

$A \to CD$

$B \to aBd|aEd$

$C \to aCb|ab$

$D \to cCd|cd$

$E \to bEc|bc$