SYNTAX ANALYSIS

Regular Grammar

All regular languages can be generated by CFGs. Some nonregular languages can be generated by CFGs but not all possible languages can be generated by CFG, *e.g.*

the CFG S $\rightarrow$ aSb|ab generates the language $\{a^n b^n : n=1,2,3, \ldots\}$, which is nonregular.

Regular Grammar continued …

Semiword: A semiword is a string of terminals (may be none) concatenated with exactly one nonterminal on the right *i.e.* a semiword, in general, is of the following form

(terminal)(terminal)… (terminal)(nonterminal)

word: A word is a string of terminals. ☹ is also a word.

Theorem

If every production in a CFG is one of the following forms

1. Nonterminal → semiword
2. Nonterminal → word

then the language generated by that CFG is regular.

Regular grammar

Definition:

A CFG is said to be a **regular grammar** if it generates the regular language *i.e.* a CFG is said to be a **regular grammar** in which each production is one of the two forms

Nonterminal → semiword

Nonterminal → word

Examples

1. The CFG S → aaS|bbS|☹

   is a regular grammar. It may be observed that the above CFG generates the language of strings expressed by the RE $(aa+bb)^*$.

2. The CFG S → aA|bB, A → aS|a, B → bS|b is a regular grammar. It may be observed that the above CFG generates the language of strings expressed by RE $(aa+bb)^+$.

Transition Graph (TG) for Regular Grammar

For every regular grammar there exists a TG corresponding to the regular grammar.

Following is the method to build a TG from the given regular grammar

1. Define the states, of the required TG, equal in number to that of nonterminals of the given regular grammar. An additional state is also defined to be the final state. The initial state should correspond to the nonterminal S.

2. For every production of the given regular grammar, there are two possibilities for the transitions of the required TG

Method continued …

(i) If the production is of the form

nonterminal $\rightarrow$ semiword

then transition of the required TG would start from the state corresponding to the nonterminal on the left side of the production and would end in the state corresponding to the nonterminal on the right side of the production, labeled by string of terminals in semiword.

(ii) If the production is of the form

nonterminal $\rightarrow$ word

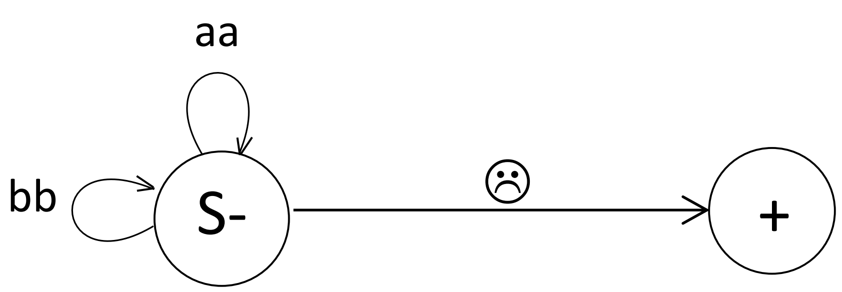
then transition of the TG would start from the state corresponding to nonterminal on the left side of the production and would end on the final state of the TG, labeled by the word. Following is an example in this regard

Example

Consider the following CFG

$S \rightarrow aaS|bbS|☹$

The TG accepting the language generated by the above CFG is given below



The corresponding RE may be $(aa+bb)^*$.

Types of Derivation.

- Replacing of a non-terminal in the current state with its corresponding production rule in the grammar in order to obtained the required string is called derivation.

- Two types of derivation.
  - Left most derivation.
  - Right most derivation.

Types of Derivation.

- Left-most Derivation.
  - The derivation in which only the left-most non-terminal in any sentential form is expanded at each step is called left most derivation.

- Right-most Derivation.
  - The derivation in which only the right-most non-terminal in any sentential form is expanded at each step is called left most derivation.

Example.

- Consider the following grammar.

$$
\begin{array}{lll}
E & \rightarrow & E + T \\
E & \rightarrow & T \\
T & \rightarrow & T * F \\
T & \rightarrow & F \\
F & \rightarrow & (E) \\
F & \rightarrow & a
\end{array}
$$

- No derive (a + a) by using both left-most and right-most derivations.

Using Right-most Derivation.

$$
\begin{aligned}
E &\rightarrow (E) \\
&\rightarrow (E + T) \\
&\rightarrow (E + F) \\
&\rightarrow (E + a) \\
&\rightarrow (T + a) \\
&\rightarrow (F + a) \\
&\rightarrow (a + a)
\end{aligned}
$$

Using Left-most Derivation.

$$
\begin{aligned}
E &\rightarrow (E) \\
&\rightarrow (E + T) \\
&\rightarrow (T + T) \\
&\rightarrow (F + T) \\
&\rightarrow (a + T) \\
&\rightarrow (a + T) \\
&\rightarrow (a + a)
\end{aligned}
$$

Example.

- Derive the string (a + a * a) by using the grammar stated in the previous slides by using both left-most and right-most derivation.