GRAMMAR

Beulah A.
AP/CSE

Example

- The theory of formal languages is used in the field of Linguistics- to define valid sentences and give structural descriptions of sentences.

  S→<noun> <verb> <adverb>

  S→<noun><verb>

  <noun>→ Ram

  <noun> → San

  <verb>→ ran

  <verb> → ate

  <adverb> → slowly

  <adverb> → quickly

Example

- S variable to denote a sentence
- → represents a rule meaning that the word on the right side of the arrow can replace the word on the left side of the arrow.
- P collection of rules (or) productions.
- The sentences are derived from the above mentioned productions by:
  - Starting with S
  - Replacing words using the productions
  - Terminating when a string of terminals is obtained.

Example

- S→<noun><verb><adverb>
- S → Ram ate slowly

- S → <noun><verb>
- S → Sam ran

Backus-Naur Form

- Backus-Naur Form or Backus Normal Form → BNF
- BNF is formal and precise
  - BNF is a notation for context-free grammars
- BNF is essential in compiler construction
- Example

  <number> ::= <digit> | <number> <digit>
  <digit> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

Formal Grammar - Definition

A formal grammar $G = (N, \Sigma, P, S)$ consists of:

- A finite set N of non terminal symbols.
- A finite set Σ of terminal symbols that is disjoint from N.
- A finite set P of production rules where a rule is of the form
  - string in $(\Sigma \cup N)^* \rightarrow$ string in $(\Sigma \cup N)^*$
  - the left-hand side of a rule must contain at least one non terminal symbol.
- A symbol S in N that is indicated as the start symbol.

Example

- G = (N, Σ, P, S)
- N = {<sentence>,<noun>,<verb>,<adverb>}
- Σ = {Ram, Sam, ate, ran ,slowly, quickly}
- S = <sentence>
- P

Example

- The grammar G with $N = \{S, B\}$, $\Sigma = \{a, b, c\}$, P consisting of the following production rules
  - $S \longrightarrow aBSc$
  - $S \longrightarrow abc$
  - $Ba \longrightarrow aB$
  - $Bb \longrightarrow bb$

Notations

Names Beginning with	Represent Symbols In	Examples
Uppercase	N	A, B, C, Prefix
Lowercase and punctuation	Σ	a, b, c, if, then, (, ;
X, *Y*	$N \cup \Sigma$	X_i, Y_3
Other Greek letters	$(N \cup \Sigma)^*$	α, β, γ

Derivation

- If $\alpha \rightarrow \beta$ is a production in a grammar G and γ, δ are any two strings on $N \cup \Sigma$, then we say $\gamma\alpha\delta$ directly derives $\gamma\beta\delta$ in G.
  - (i.e.) $\gamma\alpha\delta \Rightarrow \gamma\beta\delta$
- This process is called *one-step derivation.*
- In particular, if $\alpha \rightarrow \beta$ is a production, then $\alpha \; G{\Rightarrow} \beta$

Derivation

- The purpose of a grammar is to derive strings in the language defined by the grammar
- $\alpha \Rightarrow \beta$, β can be derived from α in one step
- $\Rightarrow^{+}$ derived in one or more steps
- $\Rightarrow^{*}$ derived in any number of steps
- $\Rightarrow_{lm}$ leftmost derivation
  - Always substitute the leftmost non-terminal
- $\Rightarrow_{rm}$ rightmost derivation
  - Always substitute the rightmost non-terminal

Example

- $G = (\{S\}, \{0,1\}, \{S \rightarrow 0S1, S \rightarrow 01\}, S)$ then the derivation is:
- $S \Rightarrow 0S1$

  $\Rightarrow 0011$

  is a one step derivation, where S is replaced by 01.

Example

$S \rightarrow AB$

$B \rightarrow b$

$A \rightarrow aA \mid c$

- Derivation

$$
\begin{aligned}
S &\Rightarrow AB \\
&\Rightarrow aAB \\
&\Rightarrow aAb \\
&\Rightarrow aaAb \\
&\Rightarrow aacb
\end{aligned}
$$

Language

- The language generated by a grammar G, L(G) is defined as $\{w \in \Sigma^* \mid S \Rightarrow^* w\}$.
- The elements of L(G) are called sentences.
- Stated in simple way, L(G) is the set of all terminal strings derived from the start symbol S.

Language

- G_1 and G_2 are equivalent if $L(G_1) = L(G_2)$
- $A \rightarrow \alpha_1$, $A \rightarrow \alpha_2$....... $A \rightarrow \alpha_m$ said to be A-productions, rewritten as

  $A \rightarrow \alpha_1 \mid \alpha_2 \mid .....\alpha_m$

Summary

- Definition of Grammar
- Notations followed in grammar
- Different types of grammar
- Language of a grammar

Test Your Knowledge

- The entity which generate Language is termed as:
  a) Automata
  b) Tokens
  c) Grammar
  d) Data
- The minimum number of productions required to produce a language consisting of palindrome strings over $\sum=\{a,b\}$ is
  a) 3
  b) 7 c) 5 d) 6

Test Your Knowledge

- The Grammar can be defined as: $G=(V, \sum, p, S)$
  In the given definition, what does S represents?
  a) Accepting State
  b) Starting Variable
  c) Sensitive Grammar
  d) None of these

Reference

- Hopcroft J.E., Motwani R. and Ullman J.D, "Introduction to Automata Theory, Languages and Computations", Second Edition, Pearson Education, 2008