

## \* Overview of linguistics :-

- "Linguistics" is the scientific study of language.
  - It involves the analysis of language form, language meaning and language in context.
  - Linguists analyse the human language by observing an interplay between sound and meaning.
- We can also consider four major sub-disciplines:-

1. Historical linguistics.
2. Syntax and morphology.
3. Semantics and pragmatics.
4. Phonology.

## \* Grammers and Language:-

- Linguistics have attempted to define grammars since the inception (उद्भव) of natural languages like English, Sanskrit etc.

- 11
- Noam Chomsky gave a mathematical model of grammar in 1956, which is effective for writing computer languages.

### \* Grammar:-

A grammar  $G$  can be formally written as a 4-tuple  $(N, T, S, P)$  where:-

- $N$  or  $V_N$  is a set of variables or non-terminals symbols.
- $T$  is a set of Terminal symbols.
- $S$  is a special variable called the Start symbol,  $S \in N$ .
- $P$  is Production rules for Terminals and Non-Terminals.

→ There are four types of grammars:-

- Type 0:- Most general, with no restrictions.
- Type 1:- Context Sensitive. The Right hand side of the Production must contain at least  $\geq 1$  symbols.

- Type 2 :- Context-free. The left side of Production must contain exactly one symbol.
- Type 3 :- Regular Expressions.

Example :-

Grammar  $G_1$  -  
 $(\{S, A, B\}, \{a, b\}, S, \{S \rightarrow AB, A \rightarrow a, B \rightarrow b\})$

Here :-

- $S, A, B$  are Non terminal Symbols.
- $a, b$  are Terminal Symbols.
- $S$  is the start symbol.
- Productions,  $P$ :  $S \rightarrow AB, A \rightarrow a, B \rightarrow b$

\* Languages :- NLP is the capability of computer software to understand the natural language.

→ There are variety of languages in the world.

Each language has its own structure.

↓  
Grammar

↓  
Certain set of rules.

Certain set of Rules.

↓ determine.

what is allowable.

what is not allowable.

For example:- In English, S V O  
= = = =  
↓ ↓ ↓  
Subject Verb Object  
I eat Apple.

Other languages:- S O V  
O S V.

⇒ The set of all strings that can be derived from a grammar is said to be the language generated from that grammar.

A language generated by a grammar  $G$  is a subset formally defined by:-

$$L(G) = \{ S \Rightarrow G \}$$

$$L(G) = \{ W \mid W \in \Sigma^*, S \Rightarrow G W \}$$