CHOMSKY CLASSIFICATION

Of Languages

By DIPANKAR BORUAH(dipankarborua20 twitter)

Topic of Discussion:

- Introduction.
- Chomsky Hierarchy Of Languages.
- Types Of Languages:
  - Type - 0
  - Type - 1
  - Type - 2
  - Type - 3

Introduction:

- **Noam Chomsky,** is an **American linguist**, **philosopher**, **scientist** and **social activist.**
- Chomsky hierarchy of grammars was described by **Noam Chomsky** in **1956**.
- **Grammar Definition**: It is defined by four tuples: **G = {V,T,P,S}**where
  - V = Non Terminals
  - T = Terminals
  - P = Production Rule
  - S = Start Symbol

Production Rule:

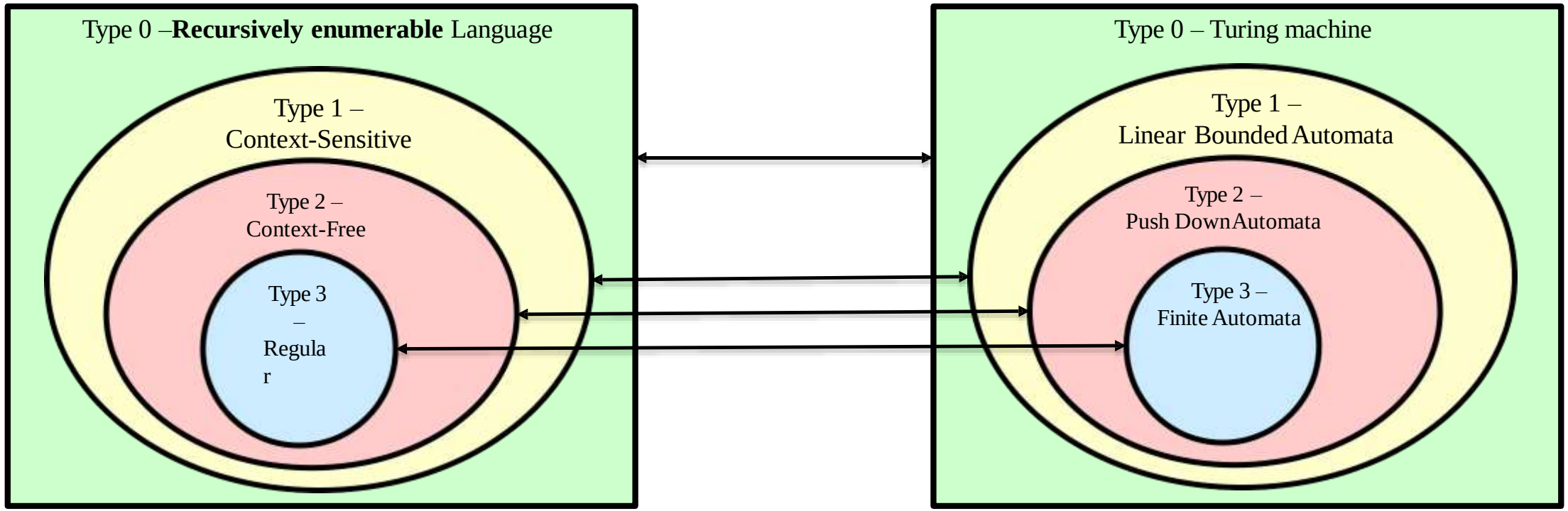
$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow b$

Chomsky Hierarchy Of Languages:

Venn Diagram of Grammar Types:



Types Of Languages:

- *Recursively enumerable* Language*(Type-0)*
- *Context-sensitive* Language*(Type-1)*
- *Context-free* Language*(Type-2)*
- *Regular* Language*(Type-3)*

Type-0:

- **Type-0 Languages** (unrestricted grammars) include all formal grammars.
- They generate exactly all languages that can be recognized by a **Turing machine.**
- These languages are also known as the **recursively enumerable languages**.
- **Type-0 grammars** are too general to describe the syntax of programming languages and natural languages.
- This grammar has rules of the form α → β (where a contains non terminal and β contains terminals or non terminals).
- Example:
- AB → A 
- AB → aB 
- S→ ^ 
- a → AB 
- ^ → a 

α = alpha β = Beta

Type-1:

- Type-1 grammar generate the context-sensitive languages.
- The languages described by these grammars are exactly all languages that can be recognized by a linear bounded automaton.
- These grammars have rules of the form $\alpha \rightarrow \beta$ with a restriction that length of $|a| \leq |\beta|$.
- Example:
  - $aAb \rightarrow bbb$ ✓
  - $aA \rightarrow bbb$ ✓
  - $aAb \rightarrow bb$ ✗

Type-2:

- Type-2 Languages generate the context-free languages.
- These languages are exactly all languages that can be recognized by a non-deterministic pushdown automaton.
- Context-free languages are the theoretical basis for the syntax of most programming languages.
- These are defined by rules of the form $A \rightarrow \alpha$ where *A* is a nonterminal and α is a string of terminals and nonterminal(there will be no context on the left and right of nonterminal).
- Example:
  - $A \rightarrow BCD$ ✓
  - $A \rightarrow aBC$ ✓
  - $a \rightarrow AbC$ ✗

Type-3:

- Type-3 Languages generate the regular languages.
- These languages are exactly all languages that can be decided by a finite state automaton
- Regular languages are commonly used to define search patterns of programming languages.
- It can be classified into two types (1)Right Linear (2)Left Linear.
- If we have repetition of non terminals on right side[$A \rightarrow xB|x$] then it is known as Right Linear.
- If we have repetition of non terminals on left side[$A \rightarrow Bx|x$] then it is known as Left Linear.($A,B \in$ non terminals and $x \in \Sigma^*$)
- Example:
- $S \rightarrow aS|b$
- $S \rightarrow aS|c$
- $S \rightarrow Sa|b$
- $A \rightarrow ba$

Σ^* = set of all terminal symbol

Reference:

- https://www.tutorialspoint.com/automata_theory/chomsky_classification_of_grammars.htm
- https://www.geeksforgeeks.org/toc-chomsky-hierarchy/
- https://www.youtube.com/watch?v=AlSB2_CehbM
- https://en.wikipedia.org/wiki/Chomsky_hierarchy

THANK YOU