CSC-257
Theory Of Computation
(BSc CSIT, TU)

Ganesh Khatri
kh6ganesh@gmail.com

Sentential Forms

- Derivations from the start symbol produce strings that have a special role. We call these "sentential forms".
- if G = (V, T, P, S) is a CFG, then any string β is in (V U T)* such that $S \rightarrow^* \beta$ is a sentential form.
- If $S \rightarrow_{lm}^* \beta$, then β is a left sentential forms, and if $S \rightarrow_{rm}^* \beta$, then β is a right sentential form.

Ambiguity in Grammar

- A Grammar G = (V, T, P and S) is said to be ambiguous if there is a string w ε L(G) for which we can derive two or more distinct derivation tree rooted at S and yielding w.
- In other words, a grammar is ambiguous if it can produce more than one leftmost or more than one rightmost derivation for the same string in the language of the grammar
- For example : Consider a grammar

  $S \rightarrow AB \mid aaB$

  $A \rightarrow a \mid Aa$

  $B \rightarrow b$

Ambiguity in Grammar

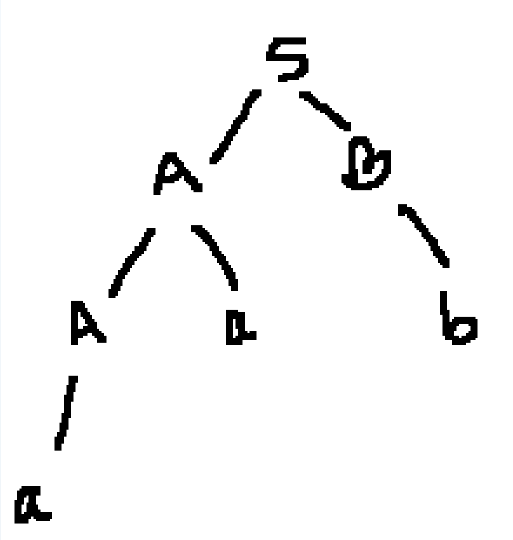
- For any string aab,
- we have two leftmost derivations as :

  S → AB

  → AaB

  → aaB

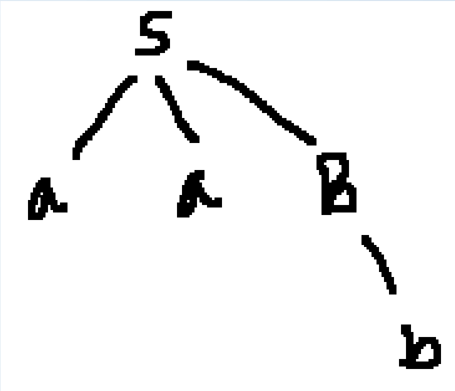
  → aab



- And also

  S → aaB

  → aab



Chomsky Normal Form

- **Theorem :** Every context free language (without Є) is generated by a CFG in which all productions are of the form A → BC or A → a, where A,B and C are variables, and a is terminal. This form is called Chomsky Normal Form.

- To prove this, we need to make a number of preliminary simplifications, which are themselves useful in various ways

1. **We must eliminate "useless symbols" :** Those variables or terminals that do not appear in any derivation of a terminal string from the start symbol.

2. **We must eliminate "Є-production" :** Those of the form A→Є for some variable A

3. **We must eliminate "unit productions" :** Those of the form A → B for variables A and B

1. Eliminating Useless Symbols

- We say a symbol x is useful for a grammar G = (V, T, P, S) if there is some derivation of the form $S \rightarrow^* \alpha x \beta \rightarrow^* w$, where w is in T*.
- Here, x may be either variable or terminal and the sentential form $\alpha x \beta$ might be the first or last in the derivation
- If x is not useful, we say it is useless.
- Thus useful symbols are those variables or terminals that appear in any derivation of a terminal string from the start symbol.
- Eliminating a useless symbol includes identifying whether or not the symbol is "generating" and "reachable"

1. Eliminating Useless Symbols

- **Generating Symbol :** We say x is generating if x $\rightarrow^*$w for some terminal string w.
- Note that every terminal is generated since w can be that terminal itself, which is derived by zero steps
- **Reachable symbol :** We say x is reachable if there is derivation S $\rightarrow^*$ αxβ for some α and β
- Thus if we eliminate the non generating symbols and then non-reachable, we shall have only the useful symbols left.

1. Eliminating Useless Symbols

- Ex : Consider a grammar defined by following productions

  $S \rightarrow aB \mid bX$

  $A \rightarrow Bad \mid bSX \mid a$

  $B \rightarrow aSB \mid bBX$

  $X \rightarrow SBd \mid aBX \mid ad$

- Here, A and X can directly generate terminal symbols. So, A and X are generating symbols. As we have the productions $A \rightarrow a$ and $X \rightarrow ad$
- Also, $S \rightarrow bX$ and X generates terminal string so S can also generate terminal string. Hence, S is also generating symbol.
- B can not produce any terminal symbol, so it is non-generating

1. Eliminating Useless Symbols

- Hence, the new grammar after removing non-generating symbols is:

  $S \rightarrow bX$

  $A \rightarrow bSX \mid a$

  $X \rightarrow ad$

- Here, A is non-reachable as there is no any derivation of the form $S \rightarrow^{*} \alpha A \beta$ in the grammar
- Thus, eliminating the non-reachable symbols, the resulting grammar is:

  $S \rightarrow bX$

  $X \rightarrow ad$

- This is the grammar with only useful symbols

1. Eliminating Useless Symbols : Exercise

- Remove useless symbol from the following grammars:

S → xyZ | XyzZ

X → Xz | xYZ

Y → yYy | XZ

Z → Zy | z

S → aC | SB

A → bSCa

B → aSB | bBC

C → aBc | ad