#### **Chomsky Normal Form (CNF)**

- Implementation: branching factor remains homogeneous
- Production Rule (Parse Tree): siblings hold same properties
- A. Length of the Production Rule body should be within 2
- B. No Variable, Terminal mixed up

#### **Prerequisite to construct CNF**

- **1.** Eliminate ∈ transitions
- 2. Eliminate Unit productions
- 3. Eliminate useless symbols
- Unit Productions: 1 Variable to 1 Variable

$$V \rightarrow X$$

- Useless Symbols:
  - Variable does not generate any Production Rule
  - Not reachable Variable(s)

### **Example:**

$$S \rightarrow AB \mid \in$$

$$A \rightarrow aAb|aAa| \in$$

$$B \rightarrow A \mid bB \mid \in$$

$$D \rightarrow Aa|bB|Eb$$

### **Example:**

 $S \rightarrow AB$ 

 $A \rightarrow aAb | aAa | \in$ 

 $B \rightarrow bB |D| \in$ 

 $E \rightarrow a$ 

# Step 1: Eliminate $\in$ Transitions

# Nullable symbols {A, B}

| Production rules                      |                     | Combination of Nullable symbols |                                       |  |
|---------------------------------------|---------------------|---------------------------------|---------------------------------------|--|
| $S \rightarrow AB$                    | $S \rightarrow AB$  | {A, B, AB, ∅}                   | $S \to AB$ $S \to A B AB \varnothing$ | $S \rightarrow AB$<br>$S \rightarrow A B AB$ |
|                                       | $A \rightarrow aAb$ |                                 | $A \rightarrow aAb$                   | $A \rightarrow aAb$                          |
| $A \rightarrow aAb aAa  \in$          |                     | {A, ∅}                          |                                       | $A \rightarrow ab$                           |
| $A \rightarrow aAb \mid aAa \mid \in$ | $A \rightarrow aAa$ |                                 | $A \rightarrow aAa$                   | $A \rightarrow aAa$                          |
|                                       |                     |                                 |                                       | $A \rightarrow aa$                           |
|                                       | $B \rightarrow bB$  |                                 | $B \rightarrow bB$                    | $B \rightarrow bB$                           |
| $B \rightarrow bB D  \in$             |                     | {B, ∅}                          |                                       | $B \rightarrow b$                            |
|                                       | $B \rightarrow D$   |                                 | $B \rightarrow D$                     | $B \rightarrow D$                            |
| $E \rightarrow a$                     |                     |                                 |                                       | $E \rightarrow a$                            |

$$\therefore S \rightarrow A|B|AB$$

 $A \rightarrow aAb|ab|aAa|aa$ 

 $B \rightarrow bB|b|D$ 

 $E \rightarrow a$ 

### **Step 2: Eliminate Unit Productions**

| $S \rightarrow A \mid B \mid AB$ | S → aAb ab aAa aa bB b D AB |
|----------------------------------|-----------------------------|
| $A \rightarrow aAb ab aAa aa$    | A → aAb ab aAa aa           |
| $B \rightarrow bB b D$           | $B \rightarrow bB b D$      |
| $E \rightarrow a$                | $E \rightarrow a$           |

### **Step 3: Eliminate Useless symbols**

- Not generating rules further {D}
- Not reachable {E}

| $S \rightarrow aAb ab aAa aa bB b D AB$ | S → aAb ab aAa aa bB b AB |
|---|---------------------------|
| A → aAb ab aAa aa                       | A → aAb ab aAa aa         |
| $B \rightarrow bB b D$                  | $B \rightarrow bB b$      |
| $E \rightarrow a$                       |                           |

# Step 4: Length of the Production rule body should be within 2

| S → aAb ab aAa aa bB b AB     | S → Xb ab Xa aa bB b AB |
|-------------------------------|-------------------------|
| $A \rightarrow aAb ab aAa aa$ | $X \rightarrow aA$      |
| $B \rightarrow bB b$          | A → Xb ab Xa aa         |
|                               | $B \rightarrow bB b$    |

### Step 5: No Variable, Terminal mixed up

| S → Xb ab Xa aa bB b AB     | $Y \rightarrow a$                   |
|-----------------------------|-------------------------------------|
| $X \rightarrow aA$          | $Z \rightarrow b$                   |
| $A \rightarrow Xb ab Xa aa$ | $S \rightarrow XZ ab XY aa ZB b AB$ |
| $B \rightarrow bB b$        | $X \rightarrow YA$                  |
| - ,                         | $A \rightarrow XZ ab XY aa$         |
|                             | $B \rightarrow ZB \mid b$           |

#### CYK Algorithm: The membership problem

- To test a string is accepted or not
- J. Cocke, D. Younger, T. Kasami

# The membership problem:

- Problem:
  - · Given a context-free grammar G and a string w
    - $-\mathbf{G} = (V, \Sigma, P, S)$  where
      - » V finite set of variables
      - » ∑ (the alphabet) finite set of terminal symbols
      - » P finite set of rules
      - » S start symbol (distinguished element of V)
      - » V and ∑ are assumed to be disjoint
    - G is used to generate the string of a language
- Question:
  - Is w in L(G)?
- The Structure of the rules in a Chomsky Normal Form grammar
- Uses a "dynamic programming" or "table-filling algorithm"

- Each row corresponds to one length of substrings
  - Bottom Row Strings of length 1
  - Second from Bottom Row Strings of length 2

.

- Top Row string 'w'
- X<sub>i, i</sub> is the set of variables A such that
   A → w<sub>i</sub> is a production of G
- Compare at most n pairs of previously computed sets:

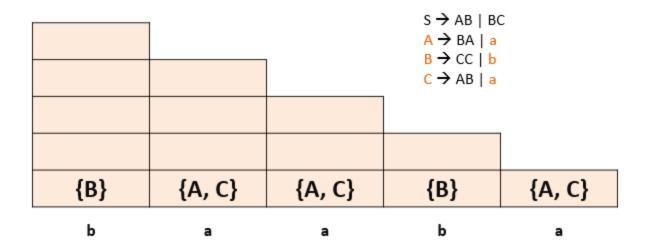
$$(X_{i,\,i}\;,\,X_{i+1,\,j}\;),\,(X_{i,\,i+1}\;,\,X_{i+2,\,j}\;)\;...\;(X_{i,\,j-1}\;,\,X_{j,\,j}\;)$$

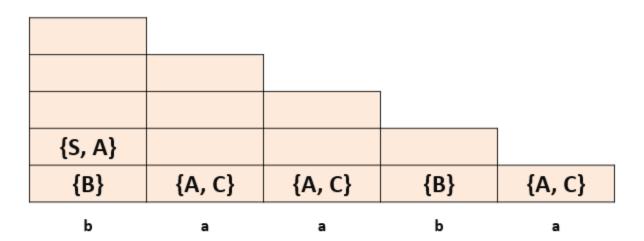
| X <sub>1,5</sub>  |                   |                   |                   |                  |
|-------------------|-------------------|-------------------|-------------------|------------------|
| X <sub>1, 4</sub> | X <sub>2,5</sub>  |                   |                   |                  |
| X <sub>1,3</sub>  | X <sub>2, 4</sub> | X <sub>3,5</sub>  |                   |                  |
| X <sub>1, 2</sub> | X <sub>2,3</sub>  | X <sub>3, 4</sub> | X <sub>4,5</sub>  |                  |
| X <sub>1, 1</sub> | X <sub>2, 2</sub> | X <sub>3,3</sub>  | X <sub>4, 4</sub> | X <sub>5,5</sub> |
| w <sub>1</sub>    | w <sub>2</sub>    | w <sub>3</sub>    | w <sub>4</sub>    | w <sub>5</sub>   |

Table for string 'w' that has length 5

### example:

- CNF grammar G
  - S → AB | BC
  - A → BA | a
  - B → CC | b
  - C → AB | a
- w is baaba
- Question Is baaba in L(G)?





• 
$$X_{1,2} = (X_{i,i}, X_{i+1,j}) = (X_{1,1}, X_{2,2})$$

# • Steps:

- Look for production rules to generate BA or BC
- There are two: S and A

$$-X_{1,2} = \{S, A\}$$

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$

|        |        |        | ı   |        |
|--------|--------|--------|-----|--------|
|        |        |        |     | ı      |
| {S, A} | {B}    |        |     |        |
| {B}    | {A, C} | {A, C} | {B} | {A, C} |
| ь      | а      | а      | ь   | а      |

• 
$$X_{2,3} = (X_{i,i}, X_{i+1,j}) = (X_{2,2}, X_{3,3})$$

- Steps:
  - Look for production rules to generate Y
  - There is one: B

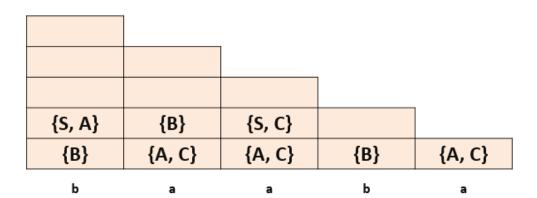
$$-X_{2,3} = \{B\}$$

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$



• 
$$X_{3,4} = (X_{i,i}, X_{i+1,j}) = (X_{3,3}, X_{4,4})$$

- Steps:
  - Look for production rules to generate Y
  - There are two: S and C

$$-X_{3,4} = \{S,C\}$$

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$

|        |        |        | l      |        |
|--------|--------|--------|--------|--------|
| {S, A} | {B}    | {S, C} | {S, A} |        |
| {B}    | {A, C} | {A, C} | {B}    | {A, C} |
| h      | a      |        | h      |        |

• 
$$X_{4,5} = (X_{i,i}, X_{i+1,j}) = (X_{4,4}, X_{5,5})$$

- Steps:
  - Look for production rules to generate Y
  - There are two: S and A

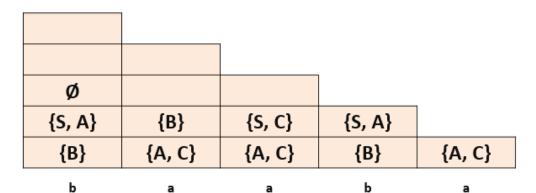
$$-X_{4,5} = \{S, A\}$$

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$



• 
$$X_{1,3} = (X_{i,i}, X_{i+1,j}) (X_{i,i+1}, X_{i+2,j})$$
  
=  $(X_{1,1}, X_{2,3}), (X_{1,2}, X_{3,3})$ 

- → {B}{B} U {S, A}{A, C}= {BB, SA, SC, AA, AC} = Y
- Steps:
  - Look for production rules to generate Y

- There are NONE: S and A
$$- X_{1,3} = \emptyset$$

$$S \rightarrow AB \mid BC$$

$$A \rightarrow BA \mid a$$

$$B \rightarrow CC \mid b$$

$$C \rightarrow AB \mid a$$

- no elements in this set (empty set)

|        |        | _      |        |        |
|--------|--------|--------|--------|--------|
|        |        |        |        |        |
| Ø      | {B}    |        |        | _      |
| {S, A} | {B}    | {S, C} | {S, A} |        |
| {B}    | {A, C} | {A, C} | {B}    | {A, C} |
|        | _      | _      | h      | _      |

• 
$$X_{2,4}$$
 =  $(X_{i,i}, X_{i+1,j})$   $(X_{i,i+1}, X_{i+2,j})$   
=  $(X_{2,2}, X_{3,4})$ ,  $(X_{2,3}, X_{4,4})$ 

#### • Steps:

- Look for production rules to generate Y

- There is one: B 
$$S \rightarrow AB \mid BC$$
  
 $A \rightarrow BA \mid a$   
-  $X_{2,4} = \{B\}$   $B \rightarrow CC \mid b$   
 $C \rightarrow AB \mid a$ 

|        |        |        | 1      |        |
|--------|--------|--------|--------|--------|
| Ø      | {B}    | {B}    |        |        |
| {S, A} | {B}    | {S, C} | {S, A} |        |
| {B}    | {A, C} | {A, C} | {B}    | {A, C} |

• 
$$X_{3,5}$$
 =  $(X_{i,i}, X_{i+1,j})$   $(X_{i,i+1}, X_{i+2,j})$   
=  $(X_{3,3}, X_{4,5})$ ,  $(X_{3,4}, X_{5,5})$ 

#### • Steps:

Look for production rules to generate Y

- There is one: B  

$$\begin{array}{ll}
S \rightarrow AB \mid BC \\
A \rightarrow BA \mid a \\
B \rightarrow CC \mid b \\
C \rightarrow AB \mid a
\end{array}$$

| {S, A, C} | ← X <sub>1,5</sub> |        |        |        |
|-----------|--------------------|--------|--------|--------|
| Ø         | {S, A, C}          |        |        |        |
| Ø         | {B}                | {B}    |        | _      |
| {S, A}    | {B}                | {S, C} | {S, A} |        |
| {B}       | {A, C}             | {A, C} | {B}    | {A, C} |
| b         | а                  | а      | ь      | а      |

We can see the S in the set  $X_{1n}$  where 'n' = 5 We can see the table the cell  $X_{15}$  = (S, A, C) then

if S ∈ X<sub>15</sub> then <u>baaba</u> ∈ L(G)

#### Homework:

- CNF grammar G
  - S → AB | BC
  - A → BA | a
  - B → CC | b
  - C → AB | a
- w is ababa
- Question Is ababa in L(G)?