<u>Chomsky Normal Form (CNF)</u>

- 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 $\in$ 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:

$S \rightarrow AB|\in$

$A \rightarrow aAb|aAa|\in$

$B \rightarrow A|\ bB|\in$

$D \rightarrow Aa|bB|Eb$

Example:

$S \rightarrow AB$

$A \rightarrow aAb \mid aAa \mid \in$

$B \rightarrow bB \mid D \mid \in$

$E \rightarrow a$

Step 1: Eliminate $\in$ Transitions

Nullable symbols {A, B}

Production rules		**Combination of Nullable symbols**		
$S \rightarrow AB$	$S \rightarrow AB$	$\{A, B, AB, \varnothing\}$	$S \rightarrow AB$ $S \rightarrow A \mid B \mid AB \mid \varnothing$	$S \rightarrow AB$ $S \rightarrow A \mid B \mid AB$
$A \rightarrow aAb \mid aAa \mid \in$	$A \rightarrow aAb$	$\{A, \varnothing\}$	$A \rightarrow aAb$	$A \rightarrow aAb$ $A \rightarrow ab$
	$A \rightarrow aAa$		$A \rightarrow aAa$	$A \rightarrow aAa$ $A \rightarrow aa$
$B \rightarrow bB \mid D \mid \in$	$B \rightarrow bB$	$\{B, \varnothing\}$	$B \rightarrow bB$	$B \rightarrow bB$ $B \rightarrow b$
	$B \rightarrow D$		$B \rightarrow D$	$B \rightarrow D$
$E \rightarrow a$				$E \rightarrow a$

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

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

$B \rightarrow bB \mid b \mid D$

$E \rightarrow a$

Step 2: Eliminate Unit Productions

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

Step 3: Eliminate Useless symbols

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

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

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

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

Step 5: No Variable, Terminal mixed up

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

- **<u>CYK Algorithm: The membership problem</u>**
    - To test a string is accepted or not
    - J. **C**ocke, D. **Y**ounger, T. **K**asami

The membership problem:

– Problem:
  - Given a context-free grammar **G** and a string **w**
    – **G** = (V, ∑ ,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’
- $\mathbf{X_{i,\,i}}$ is the set of variables A such that A → w_i is a production of G
- Compare at most n pairs of previously computed sets:

  $(\mathbf{X_{i,\,i}}, \mathbf{X_{i+1,\,j}}), (\mathbf{X_{i,\,i+1}}, \mathbf{X_{i+2,\,j}}) \ldots (\mathbf{X_{i,\,j-1}}, \mathbf{X_{j,\,j}})$

$X_{1,5}$				
$X_{1,4}$	$X_{2,5}$			
$X_{1,3}$	$X_{2,4}$	$X_{3,5}$		
$X_{1,2}$	$X_{2,3}$	$X_{3,4}$	$X_{4,5}$	
$X_{1,1}$	$X_{2,2}$	$X_{3,3}$	$X_{4,4}$	$X_{5,5}$
w_1	w_2	w_3	w_4	w_5

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)?

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

{S, A}				
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{1,2} = (X_{i,i}, X_{i+1,j}) = (X_{1,1}, X_{2,2})}$
- ➔ {B}{A,C} = {BA, BC}
- Steps:
  - Look for production rules to generate BA or BC
  - There are two: S and A
  - $\mathbf{X_{1,2} = \{S, A\}}$

S → AB | BC
A → BA | a
B → CC | b
C → AB | a

{S, A}	{B}			
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{2,3} = (X_{i,i}, X_{i+1,j}) = (X_{2,2}, X_{3,3})}$
- ➔ {A, C}{A,C} = {AA, AC, CA, CC} = Y
- Steps:
  - Look for production rules to generate Y
  - There is one: B
  - $\mathbf{X_{2,3}}$ **= {B}**

S ➔ AB | BC
A ➔ BA | a
B ➔ CC | b
C ➔ AB | a

{S, A}	{B}	{S, C}		
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{3,4} = (X_{i,i}, X_{i+1,j}) = (X_{3,3}, X_{4,4})}$
- ➔ {A, C}{B} = {AB, CB} = Y
- Steps:
  - Look for production rules to generate Y
  - There are two: S and C
  - $\mathbf{X_{3,4}}$ **= {S, C}**

S ➔ AB | BC
A ➔ BA | a
B ➔ CC | b
C ➔ AB | a

{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{4,5} = (X_{i,i}, X_{i+1,j}) = (X_{4,4}, X_{5,5})}$
- $\rightarrow$ {B}{A, C} = {BA, BC} = Y
- Steps:
  - Look for production rules to generate Y
  - There are two: S and A
  - $\mathbf{X_{4,5}}$ **= {S, A}**

S $\rightarrow$ AB | BC
A $\rightarrow$ BA | a
B $\rightarrow$ CC | b
C $\rightarrow$ AB | a

Ø				
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- $\mathbf{X_{1,3} = (X_{i,i}, X_{i+1,j})\ (X_{i,i+1}, X_{i+2,j})}$
  $\mathbf{= (X_{1,1}, X_{2,3}), (X_{1,2}, X_{3,3})}$
- $\rightarrow$ {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
  - $\mathbf{X_{1,3}}$ **= Ø**
  - **no elements in this set (empty set)**

S $\rightarrow$ AB | BC
A $\rightarrow$ BA | a
B $\rightarrow$ CC | b
C $\rightarrow$ AB | a

Ø	{B}			
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- **$X_{2,4}$** $= (X_{i,i}, X_{i+1,j})\ (X_{i,i+1}, X_{i+2,j})$
  $= (X_{2,2}, X_{3,4}), (X_{2,3}, X_{4,4})$
- ➔ {A, C}{S, C} **U** {B}{B}= {AS, AC, CS, CC, BB} = Y
- Steps:
  - Look for production rules to generate Y
  - There is one: B
  - $X_{2,4}$ = {B}

S ➔ AB | BC
A ➔ BA | a
B ➔ CC | b
C ➔ AB | a

Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

- **$X_{3,5}$** $= (X_{i,i}, X_{i+1,j})\ (X_{i,i+1}, X_{i+2,j})$
  $= (X_{3,3}, X_{4,5}), (X_{3,4}, X_{5,5})$
- ➔ {A,C}{S,A} **U** {S,C}{A,C}
  = {AS, AA, CS, CA, SA, SC, CA, CC} = Y
- Steps:
  - Look for production rules to generate Y
  - There is one: B
  - $X_{3,5}$ = {B}

S ➔ AB | BC
A ➔ BA | a
B ➔ CC | b
C ➔ AB | a

{S, A, C}	← $X_{1,5}$			
Ø	{S, A, C}			
Ø	{B}	{B}		
{S, A}	{B}	{S, C}	{S, A}	
{B}	{A, C}	{A, C}	{B}	{A, C}
b	a	a	b	a

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_{15} then baaba ∈ 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)?