



Automata Formal Languages & Logic

Preet Kanwal

Department of Computer Science & Engineering

Automata Formal Languages & Logic

Unit 3

Preet Kanwal

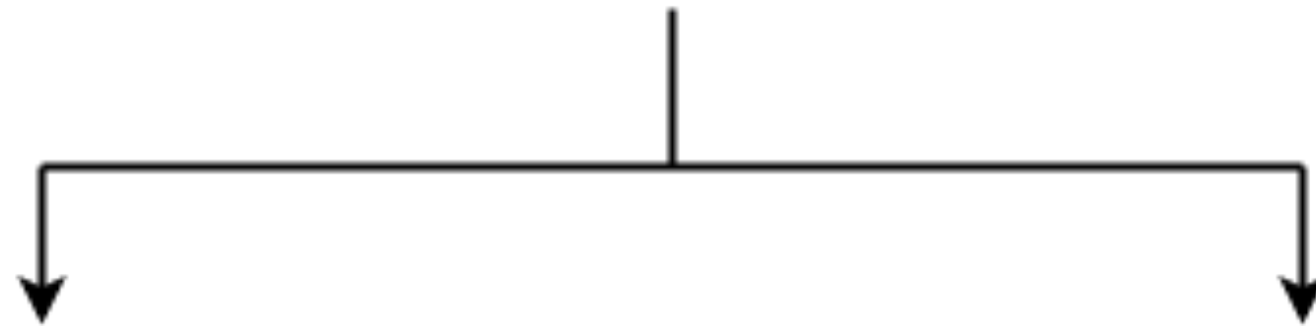
Department of Computer Science & Engineering

Normalization is performed in order to standardize the grammar.

This is because the RHS of productions have no specific format (parsing is hard!).

Idea is to make every step in derivation useful.

Normal Forms



**Chomsky Normal Form
(CNF)**

**Greibach Normal Form
(GNF)**

There are at least two relevant uses.

1. **Simplicity of proofs**

There are plenty of proofs around context-free grammars. Normal forms can be helpful there. It allows determination of:

- Membership problem –Is string w is a member of language $L(G)$?
- Emptiness problem –Is $L(G) = \emptyset$?
- Finiteness problem –Is language $L(G)$ finite?
- CNF helps prove that a language is not context-free.

There are at least two relevant uses.

2. Enables parsing

Normal forms can give us more structure to work with, resulting in easier parsing algorithms.

- **CNF is used by efficient parsing algorithm (called CYK algorithm)**
- **In a GNF grammar, any derivable string of length n can be produced with exactly n production steps.**

- **CNF restricts the number of symbols on the right side of a production to be two.**
- **It makes the parse tree for derivations using this form of the CFG a binary tree.**
- **The key advantage is that in Chomsky Normal Form, every derivation of a string of n letters has exactly $2n - 1$ steps.**
- **Thus, one can determine if a string is in the language by exhaustive search of all derivations.**

- For a given grammar, there can be more than one CNF.
- CNF produces the same language as generated by CFG.
- Every grammar in Chomsky normal form is context-free, and conversely, every context-free grammar can be transformed into an equivalent one which is in Chomsky normal form and has a size no larger than the square of the original grammar's size.

A context free grammar is in chomsky normal form (CNF) if all the productions rules satisfy one of the following condition

1) A non - terminal generating terminal

Ex: $X \rightarrow x$

2) A non - terminal generating two non terminals

Ex: $X \rightarrow XY$

3) Only Start symbol can generate λ , if λ is a part of the language

Ex: $S \rightarrow \lambda$

Step 1: Eliminate lambda productions

(Every step must be useful)

Step 2: Eliminate Unit production

(Each step must either increase the length of the sentential form or the number of terminals)

Step 3: Eliminating Useless productions and Symbols

(

There are 2 aspects :

- 1. Derivability :** Each Variable must derive a string/must end with set of terminals or lambda.
- 2. Reachability :** Each Variable must be reachable from S.

)

Remember : Lambda Units are Useless !!

Step 4: Convert the CFG to CNF as per definition.

Example 1:

$$S \rightarrow aX \mid Yb$$
$$X \rightarrow S \mid \lambda$$
$$Y \rightarrow bY \mid b$$

Solution :

Step 1:

Eliminate λ productions

$$S \rightarrow aX \mid a \mid Yb$$
$$X \rightarrow S$$
$$Y \rightarrow bY \mid b$$

Example 1:

$$S \rightarrow aX \mid Yb$$
$$X \rightarrow S \mid \lambda$$
$$Y \rightarrow bY \mid b$$

Solution :

Step 2: Eliminate Unit Productions

$$S \rightarrow aX \mid a \mid Yb$$
$$X \rightarrow aX \mid a \mid Yb$$
$$Y \rightarrow bY \mid b$$

Example 1:

$$S \rightarrow aX \mid Yb$$
$$X \rightarrow S \mid \lambda$$
$$Y \rightarrow bY \mid b$$

Solution :

Step 3: There are no useless productions

Step 4: Conversion to CNF:

$$S \rightarrow AX \mid YB \mid a$$
$$X \rightarrow AX \mid YB \mid a$$
$$Y \rightarrow BY \mid b$$
$$A \rightarrow a$$
$$B \rightarrow b$$

Example 2:

$$S \rightarrow aSa \mid bSb \mid A \mid \lambda$$

$$A \rightarrow a \mid b \mid \lambda$$

Solution :

Step 1: Remove λ production

$$S \rightarrow aSa \mid aa \mid bSb \mid bb \mid A$$

$$A \rightarrow a \mid b$$

Example 2:

$$S \rightarrow aSa \mid bSb \mid A \mid \lambda$$

$$A \rightarrow a \mid b \mid \lambda$$

Solution :

Step 2: Remove unit production ($S \rightarrow A$)

$$S \rightarrow aSa \mid aa \mid bSb \mid bb \mid a \mid b$$

$$A \rightarrow a \mid b$$

Example 2:

$$S \rightarrow aSa \mid bSb \mid A \mid \lambda$$

$$A \rightarrow a \mid b \mid \lambda$$

Solution :

Step 3: Remove useless production(A)

$$S \rightarrow aSa \mid aa \mid bSb \mid bb \mid a \mid b$$

Now the CFG is

$$A \rightarrow a$$

$$B \rightarrow b$$

$$S \rightarrow ASA \mid BSB \mid AA \mid BB \mid a \mid b$$

Example 2:

$$S \rightarrow aSa \mid bSb \mid A \mid \lambda$$

$$A \rightarrow a \mid b \mid \lambda$$

Solution :

Step 4: To CNF

$$A \rightarrow a$$

$$B \rightarrow b$$

$$C \rightarrow AS$$

$$D \rightarrow BS$$

$$S \rightarrow CA \mid DB \mid AA \mid Bb \mid a \mid b$$

Example 3:

$S \rightarrow BAB$

$B \rightarrow bba$

$A \rightarrow Bc$

Solution :

Step 1: There are no λ production

Step 2 There are no unit productions

Step 3: There are no useless production

Example 3:

$$S \rightarrow BAB$$
$$B \rightarrow bba$$
$$A \rightarrow Bc$$

Solution :

Step 4: To CNF

$$C \rightarrow a$$
$$D \rightarrow b$$
$$E \rightarrow c$$
$$F \rightarrow BA$$
$$G \rightarrow DD$$
$$S \rightarrow FB$$
$$B \rightarrow GC$$
$$A \rightarrow BE$$

Automata Formal Languages and Logic

Unit 3 - Chomsky Normal Form



Example 4:

$$S \rightarrow Aa \mid B \mid Ca$$
$$B \rightarrow aB \mid b$$
$$C \rightarrow Db \mid D$$
$$D \rightarrow E \mid d$$
$$E \rightarrow ab$$

Solution :

Step 1: There are no λ production

Step 2: Remove unit production

$$S \rightarrow Aa \mid aB \mid b \mid Ca$$
$$B \rightarrow aB \mid b$$
$$C \rightarrow Db \mid ab \mid d$$
$$D \rightarrow ab \mid d$$
$$E \rightarrow ab$$

Example 4:

$$S \rightarrow Aa \mid B \mid Ca$$
$$B \rightarrow aB \mid b$$
$$C \rightarrow Db \mid D$$
$$D \rightarrow E \mid d$$
$$E \rightarrow ab$$

Solution :

Step 3: Remove useless production

E is useless production and Aa is useless as there is no variable A

Step 4: To CNF

$$X \rightarrow a$$
$$Y \rightarrow b$$
$$S \rightarrow XB \mid b \mid CX$$
$$B \rightarrow XB \mid b$$
$$C \rightarrow DY \mid XY \mid d$$
$$D \rightarrow XY \mid d$$

Example 5:

$$S \rightarrow aAa \mid bBb \mid BB$$
$$A \rightarrow C$$
$$B \rightarrow S \mid A$$
$$C \rightarrow S \mid \lambda$$

Solution :

Step 1: Remove λ production

$$S \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$A \rightarrow C$$
$$B \rightarrow S \mid A$$
$$C \rightarrow S$$

Example 5:

$$S \rightarrow aAa \mid bBb \mid BB$$
$$A \rightarrow C$$
$$B \rightarrow S \mid A$$
$$C \rightarrow S \mid \lambda$$

Solution :

step 2: Remove unit productions

$$S \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$A \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$B \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$C \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$

Example 5:

$$S \rightarrow aAa \mid bBb \mid BB$$
$$A \rightarrow C$$
$$B \rightarrow S \mid A$$
$$C \rightarrow S \mid \lambda$$

Solution :

Step 3: Remove useless production

C is useless production

$$S \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$A \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$
$$B \rightarrow aAa \mid aa \mid bBb \mid bb \mid BB$$

Example 5:

$$S \rightarrow aAa \mid bBb \mid BB$$

$$A \rightarrow C$$

$$B \rightarrow S \mid A$$

$$C \rightarrow S \mid \lambda$$

Solution :

Step 4: To CNF

$$X \rightarrow a$$

$$Y \rightarrow b$$

$$P \rightarrow XA$$

$$Q \rightarrow YB$$

$$S \rightarrow PX \mid XX \mid QY \mid YY \mid BB$$

$$A \rightarrow PX \mid XX \mid QY \mid YY \mid BB$$

$$B \rightarrow PX \mid XX \mid QY \mid YY \mid BB$$

Example 6:

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow \text{num} \mid \text{id}$$

Solution :

Step 1: There are no λ production

Example 6:

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow \text{num} \mid \text{id}$$

Solution :

Step 2: Remove unit production ($E \rightarrow T$) and ($T \rightarrow F$)

$$E \rightarrow E + T \mid T * F \mid F$$

$$T \rightarrow T * F \mid \text{num} \mid \text{id}$$

$$F \rightarrow \text{num} \mid \text{id}$$

This results in $E \rightarrow F$, remove this

$$E \rightarrow E + T \mid T * F \mid \text{num} \mid \text{id}$$

$$T \rightarrow T * F \mid \text{num} \mid \text{id}$$

$$F \rightarrow \text{num} \mid \text{id}$$

Example 6:

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow \text{num} \mid \text{id}$$

Solution :

Step 3: There are no use less productions

Now the CFG is:

$$A \rightarrow +$$
$$B \rightarrow *$$
$$E \rightarrow EAT \mid TBF \mid \text{num} \mid \text{id}$$
$$T \rightarrow TBT \mid \text{num} \mid \text{id}$$
$$F \rightarrow \text{num} \mid \text{id}$$

Example 6:

$$E \rightarrow E + T \mid T$$
$$T \rightarrow T * F \mid F$$
$$F \rightarrow \text{num} \mid \text{id}$$

Solution :

Step 4 : To CFG

$$A \rightarrow +$$
$$B \rightarrow *$$
$$C \rightarrow EA$$
$$D \rightarrow TB$$
$$E \rightarrow CT \mid DF \mid \text{num} \mid \text{id}$$
$$T \rightarrow DT \mid \text{num} \mid \text{id}$$
$$F \rightarrow \text{num} \mid \text{id}$$



THANK YOU

Preet Kanwal

Department of Computer Science & Engineering

preetkanwal@pes.edu

+91 80 6666 3333 Extn 724