

Simplification of CFG:

Remove ϵ productions:

- We will not be able to remove all ϵ productions, if ϵ is present in language.
- If language does not contain ϵ then we can remove all ϵ productions.

Eq: $S \rightarrow aSb \mid aAb$
 $A \rightarrow \epsilon$ epsilon

1. find out all nullable variables.

- if a variable can directly give ϵ like $A \rightarrow \epsilon$
- if a variable can generate ϵ after some no. of steps

$$A \rightarrow () \rightarrow () \rightarrow \epsilon$$

$$\text{nullable variable} = \{A\}$$

2. go to RHS of every production and wherever nullable variable is present write with it and without it.

$$\begin{array}{l} S \rightarrow aSb \mid aAb \\ A \rightarrow \epsilon \end{array} \longrightarrow \begin{array}{l} S \rightarrow aSb \mid aAb \mid ab \\ \underline{A \rightarrow \epsilon} \\ \text{eliminate it} \end{array}$$

3. now in production A is no longer present. So remove aAb

$$S \rightarrow aSb \mid ab$$

Eq: $S \rightarrow AB$
 $A \rightarrow aAA \mid \epsilon$
 $B \rightarrow bBB \mid \epsilon$

$$\text{nullable variable} = \{A, B, S\}$$

$$S \rightarrow \underline{A}B \rightarrow B \rightarrow \epsilon$$

Start Symbol is a nullable variable that means ϵ is present in language.

$$\begin{aligned} S &\rightarrow AB \mid B \mid A \mid \epsilon \\ A &\rightarrow aAA \mid aA \mid a \\ B &\rightarrow bBB \mid bB \mid b \end{aligned}$$

$L =$ Even length strings
 $\Sigma = \{a, b\}$
 $L = \{\epsilon, aa, ab, ba, bb, \dots\}$

eg:

$$\begin{aligned} S &\rightarrow AbaC \\ A &\rightarrow BC \\ B &\rightarrow b \mid \epsilon \\ C &\rightarrow D \mid \epsilon \\ D &\rightarrow d \end{aligned}$$

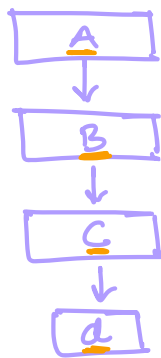
nullable variable = $\{A, B, C\}$

$$\begin{aligned} S &\rightarrow AbaC \mid baC \mid Aba \mid ba \\ A &\rightarrow BC \mid B \mid C \\ B &\rightarrow b \\ C &\rightarrow D \\ D &\rightarrow d \end{aligned}$$

In process of removing ϵ production, you might get unit production.

$$\begin{aligned} A &\rightarrow B \parallel \text{unit} \\ B &\rightarrow C \parallel \\ C &\rightarrow d \end{aligned}$$

$$\begin{aligned} A &\rightarrow d \\ B &\rightarrow d \\ C &\rightarrow d \end{aligned}$$



$$\frac{\text{Variable} \rightarrow \text{Variable}}{\text{length} \downarrow} \quad \frac{\text{Variable} \rightarrow \text{Variable}}{\text{length} \downarrow}$$

lengths is not changing.
wastage of steps
 $A \rightarrow d$

Elimination of unit productions:

eg:

$$S \rightarrow Aa \mid B$$

$$B \rightarrow A \mid bb$$

$$A \rightarrow a \mid bc \mid B$$

Write grammar without unit production.

$$S \rightarrow Aa$$

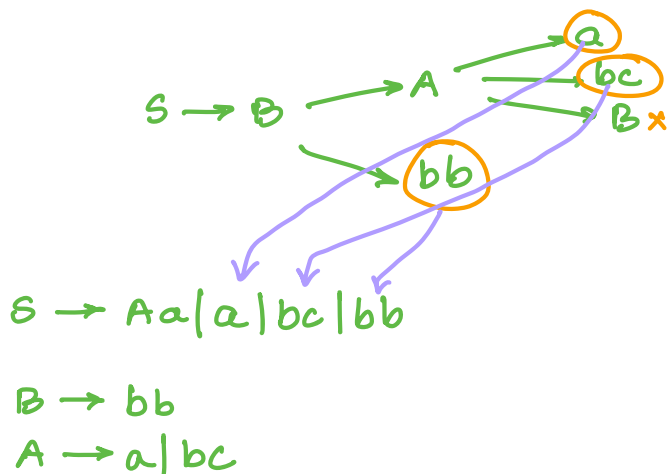
$$B \rightarrow bb$$

$$A \rightarrow a \mid bc$$

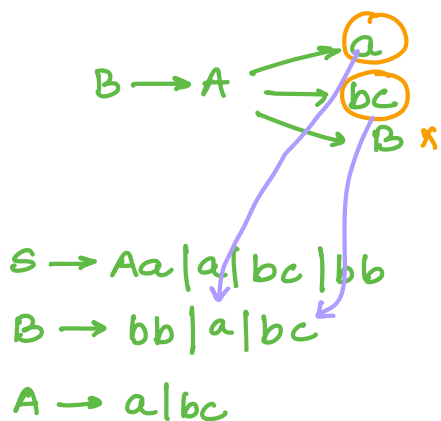
After removal of unit production, language should not be affected.

look at every unit production one by one

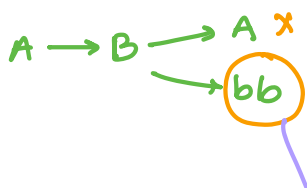
$S \rightarrow B$



$B \rightarrow A$



$A \rightarrow B$



$S \rightarrow Aa | a | bc | bb$

$B \rightarrow bb | a | bc$

$A \rightarrow a | bc | bb$

Eq:

$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow C | b$

$C \rightarrow D$

$D \rightarrow E$

$E \rightarrow a$

Grammar without unit production

$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow b$

$E \rightarrow a$

$B \rightarrow C$

$B \rightarrow C \rightarrow D \rightarrow E \rightarrow a$

$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow b | a$

$E \rightarrow a$

$C \rightarrow D$

$C \rightarrow D \rightarrow E \rightarrow a$

$S \rightarrow AB$

$A \rightarrow a$

$B \rightarrow b | a$

$E \rightarrow a$

$C \rightarrow a$

$D \rightarrow E$

$D \rightarrow E \rightarrow a$

$S \rightarrow AB$
 $A \rightarrow a$
 $B \rightarrow b|a$
 $E \rightarrow a$
 $C \rightarrow a$
 $D \rightarrow a$

C, D, E are not reachable

↳ useless symbol

Elimination of Useless Symbol:

- Variable which can derive a string of terminals is said to be useful.

$A \rightarrow \dots \rightarrow \text{terminals}$

- Variable should be reachable from start state.

useful

Eg:

$S \rightarrow AB|a$
 $A \rightarrow BC|b$
 $B \rightarrow aB|C$
 $C \rightarrow aC|B$

$T = \{a, b\}$
 $V = \{S, A, B, C\}$

- All terminals are useful symbols $\{a, b\}$

- RHS of production contains only terminals, then LHS of production will also be useful.

$S \rightarrow AB|a$
 $A \rightarrow BC|b$
 $B \rightarrow aB|C$
 $C \rightarrow aC|B$

$\{a, b, S, A\}$

- Whatever symbols have been proven to be useful, just use these symbols and find out if RHS of production is made up of only these symbols.

$B \rightarrow aB|C$ not useful
 $C \rightarrow aC|B$

B & C are useless symbols.

$B \rightarrow aB \rightarrow aaB \rightarrow aaaS \dots$

- Delete all productions having B & C on LHS & also the productions having B & C on RHS

~~$S \rightarrow AB|a$~~
 ~~$A \rightarrow BC|b$~~
 ~~$B \rightarrow aB|C$~~
 ~~$C \rightarrow aC|B$~~

$S \rightarrow a$
 $A \rightarrow b$

- Symbol which is unreachable, remove it.

~~$S \rightarrow a$~~ \rightarrow $S \rightarrow a$
 ~~$A \rightarrow b$~~

eg:

$S \rightarrow AB|AC$

$A \rightarrow aAb|bAa|a$

$B \rightarrow bbA|aaB|AB$

$C \rightarrow abCA|aDb$

$D \rightarrow bD|ac$

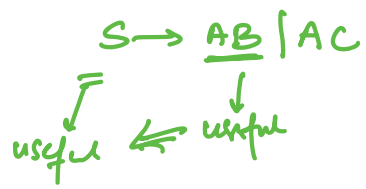
$T = \{a, b\}$

$V = \{A, B, C, D, S\}$

○ Terminals are useful $\{a, b\}$

○ Variables which are giving terminals $\{a, b, A\}$

○ $B \rightarrow \underline{bbA} | aAb | AB$ $\{a, b, A, B\}$
 \downarrow useful \leftarrow useful.



$\{a, b, A, B, S\}$

C & D are not useful.

①

$$S \rightarrow AB \mid A \cancel{C}$$

$$A \rightarrow aAb \mid bAa \mid a$$

$$B \rightarrow bbA \mid aaB \mid AB$$
~~$$C \rightarrow abCA \mid aDB$$~~
~~$$D \rightarrow bD \mid aC$$~~

②

$$S \rightarrow AB$$

$$A \rightarrow aAb \mid bAa \mid a$$

$$B \rightarrow bbA \mid aaB \mid AB$$

Simplify CFG :

1. ϵ productions
2. unit productions
3. useless symbols