Normal forms.

→ As we have seen the grammar can be simplified by reducing the ε production, removing useless symbols, unit production.

→ There is also a need to have grammar in some specific form. As you have seen in CFG at RHS of production there are any number of terminal or non terminal symbol in any combination.

→ We need to normalize such a grammar. That means we want the grammar in some specific format. That means there should be fixed no. of terminals and non terminals, in CFG.

→ Normal Forms

- Chomsky's NF
- Greibach NF

(I) Chomsky Normal Form

If language L is generated by some CFG, then there is another CFG that generates all the non – Λ words of language L, all of whose productions are of one of 2 basic forms:

Non Terminal → string of only 2 Nonterminals
Non Terminal → one terminal

→ The given CFG should be converted in the above format then we can say that the grammar is in CNF (Chomsky Normal Form).

→ Before converting the grammar into CNF it should it should be in reduced form. That means remove all the useless symbols, ε production & unit productions from it.

Thus this reduced grammar can be then converted to CNF.

eg ① Convert the following CFG into CNF.

$$\left.\begin{array}{l} S \rightarrow aSa \\ S \rightarrow bSb \\ S \rightarrow a \\ S \rightarrow b \end{array}\right\} P_1$$

Solⁿ Let $A \rightarrow a$ is added to P_1 productions

(Note that although $S \rightarrow a$ and $A \rightarrow a$ are similar still we are not replacing any a by S: This is because, S is not simply giving a. It has other productions also. & we want such a non-terminal having only one production ie. a)

Same is with B ie $B \rightarrow b$

then updated production rules after adding new we have,

$$\left.\begin{array}{l} S \rightarrow ASA \\ S \rightarrow BSB \\ A \rightarrow a \end{array}\right\} P_2$$

→ First take $S \rightarrow ASA$ for converting to CNF

$S \rightarrow A\boxed{SA}$

replace it by A_1

$$\left.\begin{array}{l} S \rightarrow AA_1 \\ A_1 \rightarrow SA \end{array}\right\} \text{both are in CNF}$$

→ Secondly consider $S \rightarrow BSB$ for converting to CNF

$S \rightarrow B\boxed{SB}$

replace it by A_2

$$\left.\begin{array}{l} S \rightarrow BA_2 \\ A_2 \rightarrow SB \end{array}\right\} \text{both are in CNF}$$

Hence we can write the rule in CNF as

$$\boxed{\begin{array}{l} S \rightarrow AA_1 \\ A_1 \rightarrow SA \\ A \rightarrow a \\ S \rightarrow BA_2 \\ A_2 \rightarrow SB \\ B \rightarrow b \\ S \rightarrow a \\ S \rightarrow b \end{array}}$$

Ans

eg ② Convert the given CFG to CNF

Consider $G = (V_N, \Sigma, P, S)$ where, SS

$V_N = \{S, A, B\}$, $\Sigma = \{a, b\}$, $P_1 = \{S \to aB \mid bA,$
$A \to bAA \mid a \mid aS$
$B \to b \mid aS \mid aBB$

Sol^n Let we add $R_1 \to a$, $R_2 \to b$ in production rule P_1. then we have,

$$\left.\begin{array}{l} S \to R_1 B \\ S \to R_2 A \\ A \to R_2 AA \\ \cancel{A \to a} \quad A \to a \\ A \to R_1 S \\ \cancel{B \to b} \quad B \to b \\ B \to R_1 S \\ B \to R_1 BB \\ R_1 \to a \\ R_2 \to b. \end{array}\right\} P_2$$

Consider those production rules which are not in CNF form. from production P_2

$A \to R_2 \boxed{AA}$
replaced by R_3

So we have

$$\left.\begin{array}{l} A \to R_2 R_3 \\ R_3 \to AA \end{array}\right\} \text{Both are in CNF.}$$

$B \to R_1 \boxed{BB}$
replaced By R_4

$$\left.\begin{array}{l} B \to R_1 R_4 \\ R_4 \to BB \end{array}\right\} \text{Both are in CNF.}$$

Finally we can write,

$S \to R_1 B$
$S \to R_2 A$
$A \to R_2 R_3$
$A \to a$
$A \to R_1 S$
$B \to b$
$B \to R_1 S$
$B \to R_1 \cancel{B} R_4$
$R_1 \to a$
$R_2 \to b$
$R_3 \to AA$
$R_4 \to BB$

Is a Chomsky's Normal Form.

Scanned by CamScanner

eg ③ Convert the following CFG to CNF:

$S \to ABA$
$A \to aA \mid \epsilon$
$B \to bB \mid \epsilon$

Soln First of all remove ϵ production. because in CNF ϵ production is not allowed.

So, $A \to \epsilon$ & $B \to \epsilon$.

If we put ϵ instead of A and B.

there we get,

$S \to \underset{\epsilon}{A}BA$ i.e. ϵBA $\therefore S \to BA$

or $S \to A\underset{\epsilon}{B}A$ i.e. $A\epsilon A$ $\therefore S \to AA$

or $S \to AB\underset{\epsilon}{A}$ i.e. $AB\epsilon$ $\therefore S \to AB$

$S \to \underset{\epsilon}{A}\underset{\epsilon}{B}A$ i.e. $\epsilon\epsilon A$ $\therefore S \to A$

or $S \to \underset{\epsilon}{A}B\underset{\epsilon}{A}$ i.e. $\epsilon B \epsilon$ $\therefore S \to B$.

We can get

$$\boxed{S \to ABA \mid AB \mid BA \mid AA \mid A \mid B}$$

Similarly, $A \to a\underset{\epsilon}{A}$ i.e. $a\epsilon$

$$\boxed{\therefore A \to a \mid aA}$$

Similarly, $B \to b\underset{\epsilon}{B}$ i.e. $b\epsilon$

$$\boxed{\therefore B \to b \mid bB}$$

Now $S \to A$ & $S \to B$ is unit production getting introduced in the grammar after removal of ϵ production.

So we will remove unit production also,

$$\left\{\begin{array}{l} S \to ABA \mid AB \mid BA \mid AA \mid aA \mid a \mid bB \mid b \\ A \to aA \mid a \\ B \to bB \mid b \end{array}\right.$$

Now let us convert this grammar to CNF

Let $S \to A\boxed{BA}$
Replaced by R_1

$\left.\begin{array}{l} S \to AR_1 \\ R_1 \to BA \end{array}\right\}$ Both are in CNF form.

Let $S \to \boxed{a}A$
Replaced by R_2

$\left.\begin{array}{l} S \to R_2A \\ R_2 \to a \end{array}\right\}$ Both are in CNF form.

Let $S \to \boxed{b}B$
Replaced by R_3

$\left.\begin{array}{l} S \to R_3B \\ R_3 \to b \end{array}\right\}$ Both are in CNF form

Let $A \to \boxed{a}A$
Replaced by R_2

$A \to R_2A$
$R_2 \to a$

Let $B \to \boxed{b}B$
Replaced by R_3

$B \to R_3B$
$R_3 \to b$

Collect all the CNF(s) to integrate & we will get

$S \rightarrow AR_1 | AB | BA | AA | R_2A | R_3B | a | b$

$A \rightarrow R_2A | a$

$B \rightarrow R_3B' | b$

$R_1 \rightarrow BA$

$R_2 \rightarrow a$

$R_3 \rightarrow b$

(II) Greibach Normal form (GNF)

→ Each production in GNF is of the form.

Nonterminal → Terminal . Nonterminal

$T(T+NT)^*$

$S \rightarrow A$ ✗

eg

$S \rightarrow aA$ ✓ $\quad S \rightarrow AA$ ✗

$S \rightarrow a$ ✓ $\quad S \rightarrow Aa$ ✗

$S \rightarrow aAa$ ✓ $\quad S \rightarrow aAB$ ✓

→ To convert given CFG into GNF very interesting procedure is followed. We can use two important lemmas based on which it is easy to convert given CFG to GNF.

Lemma 1: Let $G = (V_N, \Sigma, P, S)$ be a given CFG & if there is a production $A \rightarrow Ba$ and $B \rightarrow \beta_1 | \beta_2 | \beta_3 | \dots | \beta_n$.

Then we can convert A rule to GNF as

$A \rightarrow \beta_1 a | \beta_2 a | \beta_3 a | \dots | \beta_n a$

eg.

$S \rightarrow Aa$

$A \rightarrow aA | bA | aAS | b$

Then we can convert S rule in GNF as

$$\begin{cases} S \rightarrow aAa | bAa | aASa | ba \\ A \rightarrow aA | bA | aAS | b \end{cases}$$

Note that both the rules are in GNF.

Lemma 2: Let $G = (V_N, \Sigma, P, S)$ be a given CFG and if there is a production.

$A \rightarrow A\alpha_1 | A\alpha_2 | A\alpha_3 | \dots | A\alpha_n | \beta_1 | \beta_2 | \dots | \beta_n$

Such that β_i do not start with A then equivalent grammar in Greibach Normal form can be,

$A \rightarrow \beta_1 | \beta_2 | \dots | \beta_n$

$A \rightarrow \beta_1 Z | \beta_2 Z | \dots | \beta_n Z$

$Z \rightarrow \alpha_1 | \alpha_2 | \alpha_3 | \dots | \alpha_n$

$Z \rightarrow \alpha_1 Z | \alpha_2 Z | \dots | \alpha_n Z$

eg Consider $A \rightarrow A1 | 0B | 2$

Here $\beta_1 = 0B$ & $\beta_2 = 2$ $\alpha_1 = 1$ then, attached with A $A \rightarrow A\alpha_1$

$A \rightarrow 0B | 2$

$A \rightarrow 0BZ | 2Z$

$Z \rightarrow 1$

$Z \rightarrow 1Z$

Then all the updated rules are in GNF

eg ① Convert given CFG to GNF where $V_N = \{S, A\}$, $\Sigma = \{0,1\}$ & $P = \{S \to AA \mid 0,\ A \to SS \mid 1\}$ & S is start symbol.

another solution is on pg-37

Sol^n Let us rename S as A_1 & A as A_2 then the given CFG becomes

$$A_1 \to A_2 A_2 \mid 0$$
$$A_2 \to A_1 A_1 \mid 1$$

Let us start with A_2.

$$A_2 \to A_1 A_1 \mid 1$$

Now replace A_1 on RHS by rule A_1

$$A_2 \to A_2 A_2 A_1 \mid 0 A_1 \mid 1$$

Acc. to lemma 2 if.

$$A \to A a_1 \mid A a_2 \mid \cdots \mid A a_n \mid \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$$

then,

$$A \to \beta_1 \mid \beta_2 \mid \cdots \mid \beta_n$$
$$A \to \beta_1 Z \mid \beta_2 Z \mid \cdots \mid \beta_n Z$$
$$Z \to a_1 \mid a_2 \mid \cdots \mid a_n$$
$$Z \to a_1 Z \mid a_2 Z \mid \cdots \mid a_n Z$$

We can map this lemma to our A_2 rule as $A = A_2$, $a_1 = A_2 A_1$, $\beta_1 = 0 A_1$ & $\beta_2 = 1$

Then we get,

$$\left.\begin{array}{l} A_2 \to 0 A_1 \mid 1 \\ A_2 \to 0 A_1 Z \mid 1 Z \\ Z \to A_2 A_1 \\ Z \to A_2 A_1 Z \end{array}\right\} \text{~~GNF form~~}$$

Now, consider production for A_1

$$A_1 \to A_2 A_2 \mid 0$$

We will replace first A_2 on RHS by its recently GNF rules, then we get.

$$A_1 \to 0 A_1 A_2 \mid 1 A_2 \mid 0 A_1 Z A_2 \mid 1 Z A_2 \mid 0$$

Thus now A_1 is also in GNF form.

Now, consider the rule for Z

$$Z \to A_2 A_1 \ \& \ Z \to A_2 A_1 Z$$

using lemma ~~1~~ replace A_2 at R.H.S.

Hence we get,

$$Z = 0 A_1 A_1 \mid 1 A_1 \mid 0 A_1 Z A_1 \mid 1 Z A_1$$
$$Z = 0 A_1 A_1 Z \mid 1 A_1 Z \mid 0 A_1 Z A_1 Z \mid 1 Z A_1 Z$$

Now let us rewrite the rules by converting back $A_1 = S$ & $A_2 = A$

$$S \to 0SA \mid 1A \mid 0SZA \mid 1ZA \mid 0$$
$$A \to 0S \mid 1 \mid 0SZ \mid 1Z$$
$$Z \to 0SS \mid 1S \mid 0SZS \mid 1ZS$$
$$Z \to 0SSZ \mid 1SZ \mid 0SZSZ \mid 1ZSZ.$$

is equivalent GNF.

eg ② Convert the given CFG to GNF.

$S \to ABA$
$A \to aA \mid \epsilon$
$B \to bB \mid \epsilon$

Sol^n We first eliminate ϵ production, unit production, useless symbols before converting CFG to its normal form.

$A \to \epsilon$ & $B \to \epsilon$.

$S \to \underset{\epsilon}{A}BA$, $\therefore S \to BA$

$S \to A\underset{\epsilon}{B}A$, $\therefore S \to AA$

$S \to AB\underset{\epsilon}{A}$, $\therefore S \to AB$

$S \to \underset{\epsilon}{A}\underset{\epsilon}{B}A$, $\therefore S \to A$

$S \to \underset{\epsilon}{A}B\underset{\epsilon}{A}$, $\therefore S \to B$

So, $\boxed{S \to ABA \mid BA \mid AB \mid AA \mid A \mid B}$

Similarly, $A \to a\underset{\epsilon}{A}$, $\therefore A \to a$.

So, $\boxed{A \to aA \mid a}$

Similarly, $B \to b\underset{\epsilon}{B}$, $\therefore B \to b$

So $\boxed{B \to bB \mid b}$

There are unit productions after removing ϵ prod^n. $S \to A$ & $S \to B$.

Let us replace A by aA or a in S.

So we get,

~~$S \to aABA \mid aAB \mid a$~~

$S \to \underbrace{aABA \mid aBA \mid aAB \mid aB \mid aAA \mid aA}_{GNF} \mid a$.

Similarly,

$S \to BA \mid B$.

replace B by bB or b.

$S \to \underbrace{bBA \mid bA \mid bB \mid b}_{GNF}$

Finally equivalent GNF will be:

$S \to aABA \mid aBA \mid aAB \mid aB \mid aAA \mid aA \mid a$
$S \to bBA \mid bA \mid bB \mid b$
$A \to aA \mid a$
$B \to bB \mid b$

Applications of CFG.

→ When any high level program like C or Pascal is compiled, the compiler checks the syntax of every programming statement by constructing syntax tree. And for building the syntax tree, it is necessary to write context free grammar for each statement in the program.

→ for eg If in your C prog^n the stm. is

$$x = y + z$$

Then x, y, z are identified as 'id' by lexical analyzer. It will be interpreted as.

'id' = 'id' + 'id' ;

The CFG for this will be

$S \rightarrow id = ET$

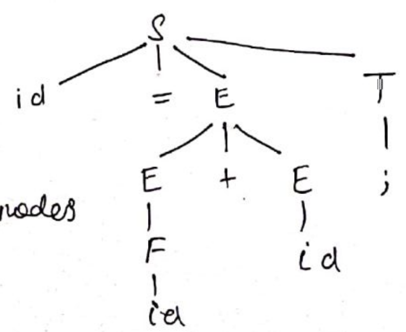
$E \rightarrow E + F$

$E \rightarrow F$

$F \rightarrow id$

$T \rightarrow ;$

Let us build the parse tree accordingly



Read out leaf nodes & you will get

id = id + id ;

Pumping lemma for Context Free Grammar

Lemma :- Let L be any Context Free Grammar language, then there is a constant n, which depends only upon L, such that there exist a string $w \in L$ & $|w| \geq n$ where $w = pqrst$ such that.

(i) $|q \& s| \geq 1$ i.e $|q| \geq 1$ & $|s| \geq 1$

(ii) $|prs| \leq n$

(iii) For all $i \geq 1$, $pq^i r s^i t$ is in L

Q1 - Show that the language $L = \{a^n b^n c^n \mid n \geq 1\}$ is not CFL.

Sol^n $L = \{abc, aabbcc, \ldots\}$

$n = 4$ as, $q_0 \xrightarrow{a} q_1 \xrightarrow{b} q_2 \xrightarrow{c} q_3$.

Choose W such that.

$|W| \geq n$ i.e $|W| \geq 4$.

So $W = aabbcc$

partition W in pqrst. Such that

(i) $|qs| \geq 1$ $\qquad \underbrace{aa}_{p}\ \underbrace{b}_{q}\ \underbrace{b}_{r}\ \underbrace{a}_{s}\ \underbrace{a}_{t}$

(ii) $|prs| \leq 4$.

(iii) If $i = 2$, then $aa(b)^2 b(a)^2 a$.

$aabbbaaa \notin L$.

So L is not CFL.

Q2 Show that Language $L = \{a^i b^j c^k \mid i<j \ \& \ i<k\}$ is not Context Free Language.

Sol^n $L = \{\overset{bc}{\underset{\wedge}{}}abbcc, \ aabbbccc, \ - - - - - \}$

$i=0 \ j=1 \ k=1$

Suppose $i=1, \ j=2, \ k=2$. $\quad$ $i=2, j=3, k=3$

$n = 3$ $\quad$ $\to q_0 \xrightarrow{b} q_1 \xrightarrow{c} q_2$ $\quad$ Base automata for smallest string of Language.

Choose word or String W such that $|W| \geq n$.

$W = abbcc$

Now partition W in pqrst such that.

(i) $|q| \geq 1 \quad |s| \geq 1$ $\quad$ $\underset{p}{a}\,\underset{q}{b}\,\underset{r}{b}\,\underset{s}{c}\,\underset{t}{c}$.

(ii) $|prs| \leq 3$

(iii) For all ~~if~~ $i=2$ $\quad a\,b^2\,b\,c^2\,c$.

$a \ bb \ b \ ccc \in L$.

Suppose $\underbrace{ab}_{p}\,\underbrace{b}_{q}\,\underbrace{c}_{r}\,\underbrace{c}_{s}\,\underbrace{\epsilon}_{t}$.

If $i=2$ $\quad ab\,b^i c\,c^i \epsilon$.

$ab \ bb \ c \ cc \ \epsilon$.

$abbbccc \in L$

Suppose.

$\underset{p}{\epsilon}\,\underbrace{a}_{q}\,\underbrace{b}_{r}\,\underbrace{b}_{s}\,\underbrace{cc}_{t}$

$\epsilon a^i b b^i cc$

for $i=2$.

$\underbrace{aa}\, b \, bb \, \underbrace{cc}$.

occurence of a = Occurence of c.

So $W \notin L$

L is not CFL.

Scanned by CamScanner