CHOMSKY - NORMAL FORM

A context free grammar $G = \langle V, T, P, S \rangle$ is said to be CNF if the productions are of the forms, either

(i) $A \to BC$ or (ii) $A \to a$

where A, B, C are non terminals and 'a' is a terminal. In CNF, no. of symbols on the right side of the productions strictly limited. Every non-empty CFL without the empty string (ϵ) can be transformed into a CNF grammar.

Theorem: Every CFL L with $\epsilon \notin L$ can be generated by a CNF grammar.

Proof: We know that if 'L' is a CFL with $\epsilon \in L$ then there is a CFG $G = \langle V, T, P, S \rangle$ that generates 'L'. Without loss of generality we may assume that G has no ϵ production and no unit productions. First construct a grammar $G' = \langle V, T, P', S \rangle$ from G such that productions in P' are of the form

$A \to X_1 X_2 \ldots X_K$ or

$A \to a$

where each X_i, $i = 1, 2 \ldots k$ is a symbol in V and a in T.

To construct such a grammar, follow the following rules

Rule 1: If $A \to a$ is in P, then retain it in P'

Rule 2: If $A \to aB$ (or $A \to Ba$) in P then replace 'a' with a new non-terminal X_a and add a production of the form $X_a \to a$ in P'

Rule 3: If $A \to a_1 a_2 \ldots a_n B$ is in P with $n \geq 2$ replace each a_i with new non-terminals X_{a_i} where $i = 1, 2 \ldots n$ adding productions

$$A \to X_{a_1}, X_{a_2} \ldots X_{a_n} B \quad \text{and}$$

$$X_{a_i} \to a_i \quad \text{for } i = 1, 2 \ldots n \text{ in } P'$$

In next step, we break all productions in P' of the form $A \to X_1 X_2 \ldots X_n$ for $n > 2$ into a group of productions with ~~not~~ two non-terminals in each body by applying the following rule. Rule 4: Introduce $(n-2)$ new non-terminals $Y_1, Y_2 \ldots Y_{n-2}$. The original production $A \to X_1 X_2 \ldots X_n$ is replaced by the $(n-1)$ new productions given as

$$A \to X_1 Y_1 \quad Y_1 \to X_2 Y_2 \quad Y_2 \to X_3 Y_3 \ldots\ldots$$

$$Y_{n-3} \to X_{n-2} Y_{n-2}, \; Y_{n-2} \to X_{n-1} X_{n-1}$$

1. Find equivalent CNF

$S \to aAD \qquad A \to aB / bAB$

$B \to b \qquad D \to d$

$S \to aAD$ is not in required form

$S \to C_a D_1 \quad C_a \to a$

$D_1 \to AD$

$A \to aB$: $A \to C_a B$

$A \to bAB$: $A \to C_b D_2 \quad C_b \to b, \quad D_2 \to AB$

$B \to b$ & $D \to d$ are in required form.

So resultant CNF :

$G_2 = (\{S, A, B, D, C_a, C_b, D_1, D_2\}, \{a, b, d\}, P', S\}$

where P' consists of

$S \to C_a D_1, \quad C_a \to a \quad D_1 \to AD \quad A \to C_a B$

$C_b \to b \quad D_2 \to AB, \quad B \to b \quad D \to d.$

Q 2. Find equivalent CNF

(i) $S \to abAB \quad A \to bAB / \epsilon \quad B \to Baa | A | \epsilon$

Eliminating ϵ-productions

Since $A \to \epsilon$, $B \to \epsilon$

$S \to abAB | abA | abB | ab$

$A \to bAB | bB | bA | b$

$B \to Baa | Baa | A$

Eliminating unit productions from B, we get

$B \to Baa | aa | bAB | bB | bA | b$

→ Introducing new non terminals S_a, S_b we obtain

$S \to S_a S_b AB | S_a S_b A | S_a S_b B | S_a S_b$

$S_a \to a \quad S_b \to b$

$A \to S_b AB | S_b B | S_b A | \cancel{S_b} b$

$B \to B S_a S_a | S_a S_a | S_b AB | S_b A | S_b B | b.$

→ Break further productions so that the RHS contains exactly two non terminals.

Thus introducing new non terminals S_c, S_d, S_e

$S \to S_a S_c B | S_a S_c | S_a S_d | S_a S_b$

$S_c \to S_b A \quad S_d \to S_b B. \quad S_e \to S_a B$

$S_a \to a \quad S_b \to b$

$A \to \cancel{S_b A} \; S_c B | S_b B | S_b A / b$

$B \to S_e S_a | S_a S_a | S_c B | S_b A | S_b B | b$

Still the production contains more than 2-NT's introduce a new NT S_f.

∴ the CNF is : $S \to S_a S_f | S_a S_d | S_a S_c | S_a S_b$

$S_c \to S_b A \quad S_d \to S_b B \quad S_e \to B S_a \quad S_f \to S_c B.$

$S_a \to a \quad S_b \to b.$

$A \to S_b A | S_b B | S_c B | b$

$B \to S_e S_a | S_a S_a | S_c B | S_b A | S_b B | b$

Convert to CNF

3. $S \to ABa \quad A \to aab \quad B \to Ac$

Introduce S_a, S_b & S_c

$\therefore S \to ABS_a \quad A \to S_aS_aS_b \quad B \to AS_c$

Introduce two NT's C & D

$C \to BS_a$ and $D \to S_cS_b$

CNF is

$S \to AC, \; A \to S_aD \quad B \to AS_c$

$C \to BS_a \quad D \to S_cS_b \mid S_a \to a \quad S_b \to b \quad S_c \to c$

4. Find equivalent CNF.

$S \to bA / aB$

$A \to bAA / aS / a$

$B \to aBB / bS / b$

Introducing $S_b \to b \quad S_a \to a$

$S \to S_bA, \; S \to S_aB$

$A \to S_bAA, \; A \to S_aS \quad A \to a$

$B \to S_aBB, \; B \to S_bS \mid B \to b$

Introducing $C \to S_bA \quad D \to S_aB$

Final set of productions

$S \to S_bA, \; S \to S_aB, \; A \to CA, \; A \to S_aS \quad A \to a$

$B \to DB, \; B \to S_bS, \; B \to b, \; S_b \to b, \; S_a \to a$

$C \to S_bA, \; D \to S_aB$

Greibach Normal Form

This normal form does not put only put restriction on the length of body of production but also put restriction on the position in which terminals and non terminals appear in body of production.

Definition: A CFG $G = \langle V, T, P, S \rangle$ is said to be in GNF if all productions are of the form $A \to aX$ where

$a \in T$, $A \in V$ and $X \in V^*$

Features:

(i) If we use GNF grammar productions into a sentential form, an i/p string of length 'n' will have derivation of exactly 'n' steps through GNF grammar.

(ii) It is used for construct PDA for a given CFG.

5. $S \to \sim S \mid [S \supset S] / p \mid q$?

- $S \to p \mid q$ in CNF
- $S \to \sim S$ replaced by $S \to AS$, $A \to \sim$
- $S \to [S \supset S]$ replaced by $S \to BSCSD$

$B \to [, \; C \to \supset, \; D \to]$

- $S \to BSCSD$ replaced by $S \to BC_1 \quad C_1 \to SCSD$

$C_1 \to SC_2 \quad C_2 \to CC_3 \quad C_3 \to SD$

So final: $S \to p/q/AS/BC_1$

$A \to \sim, \; B \to [, \; C \to \supset, \; D \to], \; C_1 \to SC_2$

$C_2 \to CC_3 \quad C_3 \to SD$ //

5. $S \rightarrow \sim S \mid [S \supset S] / p \mid q$?

- $S \rightarrow p|q$ in CNF
- $S \rightarrow \sim S$ replaced by $S \rightarrow AS$, $A \rightarrow \sim$
- $S \rightarrow [S \supset S]$ replaced by $S \rightarrow BSCSD$

  $B \rightarrow [$, $C \rightarrow \supset$, $D \rightarrow]$
- $S \rightarrow BSCSD$ replaced By $S \rightarrow BC_1$ $C_1 \rightarrow SCSD$

  $C_1 \rightarrow SC_2$ $C_2 \rightarrow CC_3$ $C_3 \rightarrow SD$.

So final : $S \rightarrow p/q/AS/BC_1$

$A \rightarrow a$, $B \rightarrow [$, $C \rightarrow \supset$, $D \rightarrow]$, $C_1 \rightarrow$

$C_2 \rightarrow CC_3$ $C_3 \rightarrow SD$ //

GREIBACH NORMAL FORM

This normal form does not put only put restriction on the length of body of production but also put restriction on the position in which terminals and non terminals appear in body of production.

Definition: A CFG $G = \langle V, T, P, S \rangle$ is said to be in GNF if all productions are of the form $A \to aX$ where

$a \in T$, $A \in V$ and $X \in V^*$

Features:

(i) If we use GNF grammar productions into a sentential form, an I/P string of length 'n' will have derivation of exactly 'n' steps through GNF grammar.

(ii) It is used ~~to~~ to construct PDA for a given CFG

Procedure for constructing GNF

Eg: Let $G = S \to aSb \mid aA \quad A \to Aa \mid Sa \mid a$

Step 1: Rename non terminals of grammar in a sequence form like $A_1, A_2 \ldots A_n$.

→ Now $G_1 = A_1 \to aA_1b \mid aA_2 \quad A_2 \to A_2a \mid A_1a \mid a$

ie Renamed S & A as A_1 & A_2 respectively.

Step 2: Construct an equivalent grammar G_1' from G_1 in which every production body for A_i starts with a terminal or starts with a_j for some $j \geq i$. For this following lemma is applied.

Lemma 1: Let G be a CFG with rules as $A \to xBy$ with $B \to b_1 \mid b_2 \ldots b_n$ then we can replace $A \to xBy$ by

$A \to xb_1y \mid xb_2y \mid \ldots xb_ny$

Applying this lemma to our example

$A_2 \to A_1a$ should be replaced by A_1's production

ie $A_2 \to aA_1ba \mid aA_2a$

∴ we obtain G_1 as

$A_1 \to aA_1b \mid aA_2 \quad A_2 \to A_2a \mid a$ & $A_2 \to aA_1ba \mid aA_2a$

Step 3: Apply following lemma to the grammar we got from step 1 & 2.

Lemma 2: Let A_i productions in G are

$A_i \to A_ix_1 \mid \ldots \mid A_ix_m$ with

$A_i \to b_1 \mid b_2 \ldots \mid b_p$

we can introduce a variable B_i and replace first group of m; A_i productions by

$A_i \to b_1B_i \mid b_2B_i \ldots \mid b_iB_i$ and

$B_i \to x_1B_i \mid x_1 \mid x_2B_i \mid x_2 \ldots x_mB_i \mid x_m$

Back to example: Now our grammar is

$A_1 \to aA_1b \mid aA_2$; $A_2 \to A_2a \mid a \mid aA_1ba \mid aA_2a$

applying lemma 2 here!

Here $A_2 \to A_2a$ and other A_2 productions are
$A_2 \to a \mid aA_1ba \mid aA_2a$; by applying lemma 2 here we get:

$A_2 \to aB_2 \mid aA_1baB_2 \mid aA_2aB_2$ where
$B_2 \to aB_2 \mid a$

Step 4: Finally we lemma 1 once again starting from the productions for A_n then A_{n-1} and so on down to A_1. After all A_i productions have RHS that starts either with terminal or A_j for $j > i$, we get A_i productions in GNF. Likewise B_i productions can be proceed.

So the ~~GNF~~ ~~of~~ our example question is

$A_1 \to aA_1b \mid aA_2$

$A_2 \to a \mid aA_1ba \mid aA_2a$

$A_2 \to aB_2 \mid aA_1baB_2 \mid aA_2aB_2$

$B_2 \to aB_2 \mid a$

Now apply the technique for the production of CNF to obtain GNF.

2. Convert into GNF

$A_1 \to A_2A_3$ ...(1) $A_2 \to A_3A_1 \mid b$ ...(2) $A_3 \to A_1A_2 \mid a$ ...(3)

Here 1 & 2 satisfies $A_i \to A_j\gamma$ where $j > i$

But 3 is not satisfying above condition.

...(3) $A_3 \to A_1A_2 \mid a$

Substitute $A_1 \to A_2A_3$ in this eqn

$A_3 \to A_2A_3A_2 \mid a$ — (3) again condition not satisfy

so substitute $A_2 \to A_3A_1 \mid b$ in (3); we get

$A_3 \to \underline{A_3A_1A_3A_2} \mid bA_3A_2 \mid a$.

Here recursion is present. Hence apply lemma 2

$A_3 \to bA_3A_2B \mid aB \mid bA_3A_2 \mid a$

& $B \to A_1A_3A_2B \mid A_1A_3A_2$.

Here A_3 productions are of GNF. Now apply A_3 productions on A_2

So $A_2 \to A_3A_1 \mid b$ will be ... $A_2 \to b$,

$A_2 \to bA_3A_2A_1 \mid aA_1 \mid bA_3A_2BA_1 \mid aBA_1$

Then substitute these A_2 productions in in A_1

$A_1 \to A_2A_3$ will be

$A_1 \to bA_3A_2A_1A_3 \mid bA_3A_2BA_1A_3 \mid aA_1A_3 \mid aBA_1A_3 \mid bA_3$.

Replace A_1 productions in B

$B \to bA_3A_2A_1A_3A_3A_2 \mid bA_3A_2BA_1A_3A_3A_2 \mid aA_1A_3A_3A_2$

$B \to aBA_1A_3A_3A_2 \mid bA_3A_3A_2$

$B \to bA_3A_2A_1A_3A_3A_2B \mid bA_3A_2BA_1A_3A_3A_2B$

$B \to aA_1A_3A_3A_2B \mid aBA_1A_3A_3A_2B \mid bA_3A_3A_2B$

Convert to GNF

3. $S \to BS$ $\quad A \to bc$ $\quad S \to Aa$ $\quad A \to bc$

~~$A \to b$~~ & ~~$B \to a$~~ ~~are in GNF~~.

Let us replace S, A, B by A_1, A_2, A_3 respectively

$A_1 \to A_3 A_1$ $\quad$ ~~$A_2 \to A_3 A_1$~~ $\quad A_3 \to A_2 c$

$A_1 \to A_2 a$ $\quad A_2 \to bc$

4 $S \to ASc$ $\quad S \to Ab$ $\quad A \to SA$ $\quad A \to c$

Let S & A are replaced by A_1 & A_2, the produns are

$A_1 \to A_2 A_1 c$

$A_1 \to A_2 b$ $\quad A_2 \to A_1 A_2$ $\quad A_2 \to c$

From production $A_2 \to A_1 A_2$; we replace A_1 dg

$A_2 \to A_2 A_1 c A_2 / A_2 b A_2$

Here Recursion apply Lemma 2

$A_2 \to c B_1 / c B_2$

$B_1 \to A_1 c A_2 B_1 / A_1 c A_2$

$B_2 \to b A_2 B_2 / b A_2$

Now apply A_2 productions in A_1

$A_1 \to c B_1 A_1 c / c B_2 A_1 c$

$A_1 \to B_2 A_1$