5.3 SIMPLIFICATION OF CONTEXT-FREE GRAMMARS

In a CFG G, it may not be necessary to use all the symbols in $V_N \cup \Sigma$, or all the productions in P for deriving sentences. So when we study a context-free language $L(G)$, we try to eliminate symbols and productions in G which are not useful for derivation of sentences.

Consider, for example,

$$G = (\{S, A, B, C, E\}, \{a, b, c\}, P, S)$$

where

$$P = \{S \rightarrow AB, A \rightarrow a, B \rightarrow b, B \rightarrow C, E \rightarrow c \mid \Lambda\}$$

It is easy to see that $L(G) = \{ab\}$. Let $G' = (\{S, A, B\}, \{a, b\}, P', S)$, where P' consists of $S \rightarrow AB$, $A \rightarrow a$, $B \rightarrow b$. $L(G) = L(G')$. We have eliminated the symbols C, E and c and the productions $B \rightarrow C$, $E \rightarrow c \mid \Lambda$. We note the following points regarding the symbols and productions which are eliminated:

(i) C does not derive any terminal string.
(ii) E and c do not appear in any sentential form.
(iii) $E \rightarrow \Lambda$ is a null production.
(iv) $B \rightarrow C$ simply replaces B by C.

In this section we give construction to eliminate (a) variables not deriving terminal strings, (b) symbols not appearing in any sentential form, (c) null productions, and (d) productions of the form $A \rightarrow B$.

(c) We must eli... ...variables X and Y.

6.2.1. Eliminating Use Less Symbols

Here we are going to identified those symbols, which do not play any role in the derivation of any string w in $L(G)$ (these symbols are called use less symbols) and then eliminate the identified production, which contains useless symbols, from the context free grammar.

A symbols Y in a context-free grammar is useful it and only if :

(a) $Y \overset{*}{\Rightarrow} w$, where $w \in L(G)$ and w in V_t^*, that is Y leads to a string of terminals. Here Y is said to be "generating".

(b) If there is a derivation $S \overset{*}{\Rightarrow} \alpha Y \beta \overset{*}{\Rightarrow} w$, $w \in L(G)$, for same α and β, then Y is said to be reachable.

Surely a symbol that is useful will be both generating and reachable. If we eliminate the symbols that are not generating first, and then eliminate from the remaining grammar those symbols that are not reachable. Then, after this process, context free grammar will have only useful symbols.

There for reduction of a given grammar G, involves following steps :

(a) Identified non-generating symbols in given CFG and eliminate those productions which contains non-generating symbols.

(b) Identified non-reachable symbols in grammar and eliminate those productions which contains non-reachable symbols.

Let us consider examples :

Example 6.1. Consider a CFG

$$S \rightarrow AB/a$$
$$A \rightarrow b$$

Identified and eliminate useless symbols.

Solution. Given CFG is

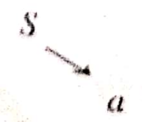
$$S \rightarrow AB/a$$
$$A \rightarrow b$$

Here by observing the given CFG, it becomes very clear that B is non-generating symbol. Since A derives b, S derives a but B is not deriving any string w in the V_t^*.

So we can eliminate $S \rightarrow AB$ from the context free grammar, now CFG becomes

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

Here A is non-reachable symbol, since it can not be reached by starting non-terminal S.



CS CamScanner

So we can eliminate $A \to b$, now grammar is

$$S \to a$$

which is reduced grammar.

Example 6.2. Remove the useless symbol from the given context free grammar :

$$S \to aB/bX$$
$$A \to BAd/bSX/a$$
$$B \to aSB/bBX$$
$$X \to SBD/aBx/ad.$$

Solution. First we choose those non-terminals which are deriving to the strings of terminals. The non-terminals A and X directly deriving to the strings of terminal a and ad respectively. Hence they are useful symbols. Since $S \to bX$ and X is useful symbol, hence S is also useful symbol. But B does not derive any string w in V_t^* so clearly B is a non-generating symbol, so eliminate those productions which contains B. Now grammar becomes :

$$S \to bX$$
$$A \to bSX/a$$
$$X \to ad$$

Again here A is non-reachable, since A can not be reached from S (starting non-terminal), it can be understand by following parse tree in Fig. 6.1.

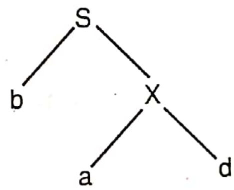


Fig. 6.1.

So eliminate $A \to bSX/a$, then reduced CFG is

$$S \to bX$$
$$X \to ad.$$

Example 6.3. Consider the following grammar and obtain an equivalent grammar containing no useless grammar symbol.

$$A \to xyz/Xyzz$$
$$X \to Xz/xYz$$
$$Y \to yYy/Xz$$
$$Z \to Zy/z.$$

Solution. Since $A \to xyz$ and $Z \to z$, hence A and Z are directly deriving to the string of terminals. Hence they are useful symbols.

Hence X and Y do not lead to a string of terminals, that is, X and Y are useless symbols. Therefore by eliminating these productions we get :

$$A \to xyz$$
$$Z \to Zy/z$$

Now since A is the starting non-terminal and right side of A does not contain Z, it means Z is not reachable. Hence after eliminating $Z \to Zy/z$, grammar becomes

$$A \to xyz.$$

which is a grammar containing no useless symbols.

CS CamScanner

Example 6.4. **Find the reduced grammar that is equivalent to the CFG given below :**

$$S \rightarrow aC/SB$$
$$A \rightarrow bSCa$$
$$B \rightarrow aSB/bBC$$
$$C \rightarrow aBC/ad.$$

Solution. Since $C \rightarrow ad$, therefore, C is generating symbol. Since $S \rightarrow aC$, therefore S is also a useful symbol according to production $A \rightarrow bSCa$, A is also generating. Right side of $B \rightarrow aSB$ and $B \rightarrow bBC$ contains B, and B is not terminating so B is not a generating symbol so we can eliminate those productions and grammar becomes

$$S \rightarrow aC$$
$$A \rightarrow bSCa$$
$$C \rightarrow ad$$

Now since S is the start symbol and right side of S does not contain A, hence A is not reachable as :

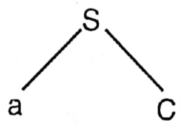


Fig. 6.2.

So by eliminating A we get

$$S \rightarrow aC$$
$$C \rightarrow ad$$

which is reduced grammar equivalent to the given grammar, containing no useless symbol.

6.2.2. Removal of Unit Production

Let us first define unit production :

A production of the form

Non-terminal $\rightarrow$ One non-terminal

that is a production of the form $A \rightarrow B$ (where A and B, both are non-terminals) is called unit production. Unit production increase the cost of derivation in a grammar. Following algorthim can be used to eliminate the unit production.

Algorithm : Removal of unit productions $\Rightarrow$

While (there exist a unit production, $A \rightarrow B$)

{

select a unit production $A \rightarrow B$, such that there exist a production $B \rightarrow \alpha$, where α is a terminal.

For (every non-unit production, $B \rightarrow \alpha$)

Add production $A \rightarrow \alpha$ to the grammar

Eliminate $A \rightarrow B$ from the grammar.

}.

Now let us apply this algorthim on the given CFG's.

CS CamScanner

Example 6.5. Consider the context free grammar G.

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

remove the unit production.

Solution. Given CFG

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

Contain three unit productions

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

Now to remove unit production $B \rightarrow C$, we see if there exists a production whose left side has C and right side contains a terminal (*i.e.* $C \rightarrow a$), but there is no such productions in G. Similar things holds for production $C \rightarrow D$. Now we try to remove unit production $D \rightarrow E$, because there is a production $E \rightarrow a$. Therefore, eliminate $D \rightarrow E$ and introduce $D \rightarrow a$, grammar becomes

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

Now we can remove $C \rightarrow D$ by using $D \rightarrow a$, we get

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

Similarly, we can remove $B \rightarrow C$ by using $C \rightarrow a$, we obtain

$$S \rightarrow AB$$
$$A \rightarrow a$$
$$B \rightarrow a/b$$

$C \to a$

$D \to a$

$E \to a$

Now it can be easily seen that productions $C \to a$, $D \to a$ $E \to a$ are useless because if we start deriving from S, these productions will never be used. Hence eliminating them gives,

$S \to AB$

$A \to a$

$B \to a/b$

which is completely reduced grammar.

Example 6.6. Consider the following unambiguous expression grammar

$I \to a/b/Ia/Ib/I0/I1$

$F \to I/(E)$

$T \to F/T*F$

$E \to T/E+T$

Identified unit production and remove them.

Solution. In the given grammar

$F \to I$

$T \to F$

$E \to T$

are three unit production.

Unit production $F \to I$ can be removed by the help of

$I \to a/b/Ia/Ib/I0/I1$

and grammar becomes

$I \to a/b/Ia/Ib/I0/I1$

$F \to (E)/a/b/Ia/Ib/I0/I1$

$T \to F/T*F$

$E \to T/E+T$

Now we can eliminate $T \to F$ by the help of

$F \to (E)/a/b/Ia/Ib/I0/I1$

and now grammar becomes

$I \to a/b/Ia/Ib/I0/I1$

$F \to (E)/a/b/Ia/Ib/I0/I1$

$T \to T*F/(E)/a/b/Ia/Ib/I0/I1$

$E \to T/E+T$

Now let us remove $E \to T$ by the help

$T \to T*F/(E)/a/b/Ia/Ib/I0/I1$

and finally, unit production free grammar is

$$I \rightarrow a/b/Ia/Ib/I0/I1$$

$$F \rightarrow (E)/a/b/Ia/Ib/I0/I1$$

$$T \rightarrow T*F/(E)/a/b/Ia/Ib/I0/I1$$

$$E \rightarrow E+T/T*F/(E)/a/b/Ia/Ib/I0/I1$$

Example 6.7. Identified and remove the unit productions from following grammar

$$S \rightarrow A/bb$$

$$A \rightarrow B/b$$

$$B \rightarrow S/a$$

Solution. Here unit productions are

$$S \rightarrow A$$

$$A \rightarrow B$$

$$B \rightarrow S$$

We list all unit productions and sequences of unit productions.Then we will handle one unit production or sequence at a time and try to prove generating and reachable for each non-terminal involve in the unit productions, as follows :

$S \rightarrow A$	generate $S \rightarrow b$
$S \rightarrow A \rightarrow B$	generate $S \rightarrow a$
$A \rightarrow B$	generate $B \rightarrow a$
$A \rightarrow B \rightarrow S$	generate $A \rightarrow bb$
$B \rightarrow S$	generate $B \rightarrow bb$
$B \rightarrow S \rightarrow A$	generate $B \rightarrow b$

The new CFG becomes

$$S \rightarrow bb/b/a$$

$$A \rightarrow b/a/bb$$

$$B \rightarrow a/bb/b$$

which hold no unit productions.

6.2.3. Removal of ∈-productions and Nullable Non-terminal

Any context-free language in which ∈ is a word, must have some ∈-productions in its grammar since otherwise we could never derive the word ∈ from S. The statement is obvious, but it should be given some justification, mathematically, this is easy; we observe that ∈-productions are only productions that shorten the working string. If we begin with the string S and apply only non-∈-productions, we never develop a word of length zero.

We now form our attention to the eliminate the productions of the form $A \rightarrow \in$, which are called ∈-productions. Surely if ∈ is in $L(G)$, we can not eliminate all ∈-productions from G, but if ∈ is not in $L(G)$, we can eliminate all ∈-productions from G.

"In a given CFG, we call a non-terminal N Nullable if

There is a production $N \rightarrow \in$ or

There is a derivation that starts at N and leads to $\in$:

$$N \Rightarrow \ldots \Rightarrow \in"$$

To eliminate $\in$-productions from a grammar G we use the following technique :

"If $A \rightarrow \in$ is a prodoction to be eliminated then we look for all productions, whose right side contains A, and replace each occurrence of A in each of these productions to obtain the non $\in$-productions. Now these resultant non $\in$-production must be added to the grammar to keep the language generated, the same."

Let us apply this concept :

Example 6.8. Consider the following grammar

$$S \rightarrow aA$$
$$A \rightarrow b/\in$$

remove the Nullable non-terminal.

Solution. Given CFG is

$$S \rightarrow aA$$
$$A \rightarrow b/\in$$

Here $A \rightarrow \in$ is $\in$-production, so let us apply above described concept, put $\in$ in place of A, at the right side of productions and add the resulted productions to the grammar.

Here is only one productions $S \rightarrow aA$ whose right side contains A. So, replacing A by $\in$ in this production, we get $S \rightarrow a$; now add this new productions **to keep the language generated by this grammar same.** Therefore, the $\in$-free grammar, equivalent to above grammar is

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

or

$$S \rightarrow aA/a$$
$$A \rightarrow b$$

there is no $\in$-production in the above context-free grammar.

Example 6.9. Consider the following grammar G

$$S \rightarrow ABAC$$
$$A \rightarrow aA/\in.$$
$$B \rightarrow bB/\in$$
$$C \rightarrow c$$

remove the $\in$-production from the above grammar.

Solution. Here we have two $\in$-production to remove $A \rightarrow \in$ and $B \rightarrow \in$. To eliminate $A \rightarrow \in$ from the grammar, the non $\in$-productions to be added are obtained as follows :

List of productions whose right side contain A are :

$$S \rightarrow ABAC$$
$$A \rightarrow aA$$

So on replacing each occurrence of A by $\in$ (one by one), we get four new productions to be added to the grammar :

$$S \rightarrow ABC/BAC/BC$$
$$A \rightarrow a$$

add these productions to the grammar and eliminate $A \rightarrow \in$, to obtain the following grammar

$$S \rightarrow ABAC/ABC/BAC/BC$$
$$A \rightarrow aA/a$$
$$B \rightarrow bB/\in$$
$$C \rightarrow c$$

Now to eliminate $B \rightarrow \in$, list all productions whose right side contains B, these are :

$$S \rightarrow ABAC/BAC/ABC/BC$$
$$B \rightarrow bB$$

Replace occurrence of B in each of these productions to obtain non $\in$-productions to be added to the grammar, we get following new productions.

$$S \rightarrow AAC/AC/C$$
$$B \rightarrow b$$

Add these productions to the grammar and eliminate $B \rightarrow \in$ from the grammar, to obtain

$$S \rightarrow ABAC/BAC/ABC/BC/AAC/AC/C$$
$$A \rightarrow aA/a$$
$$B \rightarrow bB/b$$
$$C \rightarrow c$$

which is the grammar without any $\in$-production.

Example 6.10. Consider the following grammar and remove $\in$-productions

$$S \rightarrow aSa$$
$$S \rightarrow bSb/\in$$

Solution. Here $S \rightarrow \in$ is the $\in$-production, so let us replace occurrence of S by $\in$ in the following productions

$$S \rightarrow aSa$$
$$S \rightarrow bSb$$

We will get following new productions

$$S \rightarrow aa/bb$$

add these productions in the grammar and remove $S \rightarrow \in$ from the grammar

$$S \rightarrow aSa/aa/bb/bSb$$

It is $\in$-production free grammar.

CS CamScanner

Example 6.11. Consider the following grammar

$$S \rightarrow a/Xb/aYa$$
$$X \rightarrow Y/\in$$
$$Y \rightarrow b/X$$

eliminate $\in$-productions.

Solution. Here $X \rightarrow \in$ is $\in$-productions. So this problem can also solved in similar fashion and grammar without $\in$-productions will be

$$S \rightarrow a/Xb/aYa/b/aa$$
$$X \rightarrow Y$$
$$Y \rightarrow b/X.$$

Example 6.12. Design a CFG for regular expression
$r = (a + b)^* \, bb \, (a + b)^*$, which is free from $\in$-productions.

Solution. Let CFG G be required context free grammar with $\in$-productions

$$S \rightarrow XY$$
$$X \rightarrow Zb$$
$$Y \rightarrow bW$$
$$Z \rightarrow AB$$
$$W \rightarrow Z$$
$$A \rightarrow aA/bA/\in$$
$$B \rightarrow Ba/Bb/\in$$

Finally here $A \rightarrow \in$ and $B \rightarrow \in$ are $\in$-productions and A, B are nullable non-terminals.

The CFG without $\in$-production can be achieved from G in similar fashion, as above discussed. Let it be G', then G' is

$$S \rightarrow XY$$
$$X \rightarrow Zb/b$$
$$Y \rightarrow bW/b$$
$$Z \rightarrow AB/A/B$$
$$W \rightarrow Z$$
$$A \rightarrow aA/bA/a/b$$
$$B \rightarrow Ba/Bb/a/b$$

Because $\in$ was not generated by G, the new CFG G' generates exactly the same language.

6.3. CHOMSKY NORMAL FORM

Definition If a CFG has only production of the form
Non-terminals $\rightarrow$ string of exactly two non-terminals or of the form
Non-terminals $\rightarrow$ one terminal
is said to be chomsky normal form or CNF.