SABANCI UNIVERSITY

Faculty of Engineering and Natural Sciences

CS 302 Automata Theory

Fall 2016

THE GREIBACH NORMAL FORM

We start with a *CFG*, $G = (V, T, R, S)$ in *CNF*, namely $V = (A_1, A_2, \ldots, A_n)$ where every production is of the form $A_i \rightarrow X\ Y$ where $X, Y \in V$ or $A_i \rightarrow a$ where $a \in T$.

We start with separating the productions of A_1 into **2** groups as follows :

$A_1 \rightarrow (A_1\ \alpha(1) \mid \ldots \mid A_1\ \alpha(k_1))$ (*all left recursive productions of* A_1) ………….*(1)*
$A_1 \rightarrow (\beta(1) \mid \ldots \mid \beta(m_1))$ (*rest of the productions*) …………………………………..*(2)*

where each $\beta(j)$ in *(2)* is either a *terminal* or starts with some A_p with $p > 1$.

As a standard way to remove left recursions (by right recursions) we replace the productions in *(1)* and *(2)* by

$A_1 \rightarrow (\beta(1)\ B_1 \mid \ldots \mid \beta(m_1)\ B_1)$ ……………………………………..*(3)*

$B_1 \rightarrow (\alpha(1)B_1 \mid \ldots \mid \alpha(k_1)B_1 \mid e)$…………………….*(4)*

To prove that (1) and (2) is equivalent to (3) and (4) simply demonstrate that every leftmost derived sentential form in which A_1 is eventually eliminated using (1) and (2) can be derived by a rightmost derivation of A_1 using (3) and (4) where B_1 is eventually eliminated.

Note that every $\beta(j)$ above is either a terminal or starts with some A_p with $p > 1$ and every $\alpha\ (j)$ consist of variable(s) in V.

Now substitute A_1 in *(3)* to all the productions of all A_p for $p > 1$ so that the right side of all these productions do not have A_1 as the starting symbol. Hence all productions of A_p for $p > 1$ either start with a terminal followed by zero or more variables or start with some A_k where $k > 1$.

Now apply the same procedure to productions of A_2 by first separating them into left recursive and the rest etc. At the ***i***th step of this algorithm we have the following productions

$A_i \rightarrow (\beta(i1)\ B_i \mid \ldots \mid \beta(im_i)\ B_i)$ ……………………………………..*(5i)*

$B_i \rightarrow (\alpha(i1)B_i \mid \ldots \mid \alpha(ik_i)B_1 \mid e)$…………………….*(6i)*

for $i = 1,2, \ldots, n$ where for notational consistency we have $\beta(1j) := \beta(j)$ and $\alpha(1j) := \alpha(j)$.

Also note that each (*formally induction may be used but it is obvious* !) $\beta(ij)$ starts either with a *terminal* or some A_p with $p > i$ and each $\alpha(ij)$ starts with some A_k. All right side of productions are in the form of a sequence of variables or a single terminal followed by a sequence of variables of zero or positive length.

Now observe that since there is no $\boldsymbol{p > n}$ all $\boldsymbol{\beta(nj)}$ ***must start*** with a ***terminal*** symbol in **T**. By back substitution of $\boldsymbol{A_n}$ productions given by *(5n)* into *(5(n-1))* for $\boldsymbol{A_n}$ we ensure that all productions of $\boldsymbol{A_{n-1}}$ start with a terminal and finally this chain of consecutive substitutions result in all productions of $\boldsymbol{A_i}$ from $\boldsymbol{i=n}$ to $\boldsymbol{i=1}$ ensures that all productions of $\boldsymbol{A_i}$ are in *GNF*. Finally we substitute these $\boldsymbol{A_i}$ productions to those $\boldsymbol{B_i}$ productions for which the corresponding $\boldsymbol{\alpha(ij)}$ starts with some $\boldsymbol{A_k}$ instead of a terminal . This completes the construction of the *GNF*.