Definition. *A CFG, $G = (N, \Sigma, P, S)$, is in* ***Chomsky Normal Form*** *if all productions are of the form*

$$A \to BC \text{ or } A \to a$$

where $A, B, C \in N$ and $a \in \Sigma$.

Theorem. *For any CFG G, there is a CFG G' in CNF such that $L(G') = L(G) - \{\epsilon\}$.*

Definition. *A production is an* ***ϵ-production,*** *if it is of the form $V \to \epsilon$ where V is a non-terminal. A production is a* ***unit production,*** *if it is of the form $U \to V$ where U, V are non-terminals.*

Converting a CFG to Chomsky Normal Form

Given a CFG, $G = (N, \Sigma, P, S)$, construct a new grammar, $\hat{G} = (N, \Sigma, \hat{P}, S)$ by recursively adding productions to P to form $\hat{P}$, the smallest set of productions containing P that is closed under the following rules:

1. If $A \to \alpha B \beta$ and $B \to \epsilon$ are in $\hat{P}$, then $A \to \alpha\beta$ is in $\hat{P}$.

2. If $A \to B$ and $B \to \gamma$ are in $\hat{P}$, then $A \to \gamma$ is in $\hat{P}$.

After $\hat{P}$ has had enough productions added to be closed under the above rules, remove all ϵ-productions and all unit productions.

Next for each terminal $a \in \Sigma$, that occurs as part (not all) of a string on the righthand side of a production, add a new non-terminal V_a and a new production $V_a \to a$. Then replace all 'a's in those productions with V_a.

Now all productions are of the form:

$$A \to a \quad \text{or} \quad A \to B_1 B_2 B_3 \ldots B_k$$

where the B_i are all non-terminals and $k \geq 2$.

Finally do the following until all productions have righthand sides of length 2 or less. Replace productions of the form $A \to B_1 B_2 B_3 \ldots B_k$ with two new productions, namely, $A \to B_1 A_1$ and $A_1 \to B_2 B_3 \ldots B_k$ where A_1 is a new non-terminal.

Example 2.10 in Sipser (2ed).

$$S \rightarrow ASA|aB \tag{1}$$
$$A \rightarrow B|S \tag{2}$$
$$B \rightarrow b|\epsilon \tag{3}$$

$A \rightarrow B$ and $B \rightarrow b$, so add $A \rightarrow b$.

$$S \rightarrow ASA|aB \tag{4}$$
$$A \rightarrow B|S|b \tag{5}$$
$$B \rightarrow b|\epsilon \tag{6}$$

$A \rightarrow S$ and $S \rightarrow ASA$, so add $A \rightarrow ASA$.

$$S \rightarrow ASA|aB \tag{7}$$
$$A \rightarrow B|S|b|ASA \tag{8}$$
$$B \rightarrow b|\epsilon \tag{9}$$

$A \rightarrow S$ and $S \rightarrow aB$, so add $A \rightarrow aB$.

$$S \rightarrow ASA|aB \tag{10}$$
$$A \rightarrow B|S|b|ASA|aB \tag{11}$$
$$B \rightarrow b|\epsilon \tag{12}$$

$A \rightarrow B$ and $B \rightarrow \epsilon$, so add $A \rightarrow \epsilon$.

$$S \rightarrow ASA|aB \tag{13}$$
$$A \rightarrow B|S|b|ASA|aB|\epsilon \tag{14}$$
$$B \rightarrow b|\epsilon \tag{15}$$

$S \rightarrow ASA$ and $A \rightarrow \epsilon$, so add $S \rightarrow AS|SA$.

$$S \rightarrow ASA|aB|AS|SA \tag{16}$$
$$A \rightarrow B|S|b|ASA|aB|\epsilon \tag{17}$$
$$B \rightarrow b|\epsilon \tag{18}$$

$S \rightarrow aB$ and $B \rightarrow \epsilon$, so add $S \rightarrow a$.

$$S \rightarrow ASA|aB|AS|SA|a \tag{19}$$
$$A \rightarrow B|S|b|ASA|aB|\epsilon \tag{20}$$
$$B \rightarrow b|\epsilon \tag{21}$$

$A \to ASA$ and $A \to \epsilon$, so add $A \to AS|SA$.

$$S \to ASA|aB|AS|SA|a \tag{22}$$
$$A \to B|S|b|ASA|aB|\epsilon|AS|SA \tag{23}$$
$$B \to b|\epsilon \tag{24}$$

$A \to aB$ and $B \to \epsilon$, so add $A \to a$.

$$S \to ASA|aB|AS|SA|a \tag{25}$$
$$A \to B|S|b|ASA|aB|\epsilon|AA|SA|a \tag{26}$$
$$B \to b|\epsilon \tag{27}$$

Now closed so drop all ϵ and unit productions.

$$S \to ASA|aB|AS|SA|a \tag{28}$$
$$A \to b|ASA|aB|AS|SA|a \tag{29}$$
$$B \to b \tag{30}$$

Add variables for terminals in strings of two or more symbols.

$$S \to ASA|V_aB|AS|SA|a \tag{31}$$
$$A \to b|ASA|V_aB|AS|SA|a \tag{32}$$
$$B \to b \tag{33}$$
$$V_a \to a \tag{34}$$

Reduce remaining productions to no more than two variables each.

$$S \to AC|V_aB|AS|SA|a \tag{35}$$
$$A \to b|AC|V_aB|AS|SA|a \tag{36}$$
$$B \to b \tag{37}$$
$$V_a \to a \tag{38}$$
$$C \to SA \tag{39}$$