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.