Context Free Grammars- I

Context Free Grammars : V is the set of variables; T set of terminals, P is the set of productions, each of the form $A \to \alpha$, where $A \in V$ and $\alpha \in (V \cup T)^*$ and a start symbol S. [A, B, C, D, S will denote variables; a, b, c, d, e will denote terminals; $\alpha, \beta\gamma$ will denote elements of $(V \cup T)^*$.]

Define Derivations. For a CFG, G $L(G)$ denotes the set of strings of terminals which can be derived from S.

The main use of CFG's in practice is to specify various things in a programming language. In this setting, it is important for us to "parse" a given string $x \in T^*$: to determine whether x is in $L(G)$ and if so find a derivation. Productions of the form $A \to B$ with just one non-terminal on the right (called unit productions) and productions of the form $A \to \epsilon$ with the 0-length string on the right (called ϵ productions) are problomatic for parsing - because they are not length-increasing, they uspet the recursive procedure we usually design as a precursor to a Dyanamic Programming algorithm for parsing. So, we get rid of non-length-increasing productions at the outset.

Getting rid of ϵ productions : First determine which non-terminals can derive ϵ (Easy) - call all such non-terminals "nullable". Then for each production of the form $A \to X_1 X_2 \dots X_l$, in the given grammar G, add the (at most $2^l - 1$) productions we get by omitting a set of nullable variables among $X_1, X_2, \dots X_l$, except, we do not add the production $A \to \epsilon$. Prove that this gives us a grammar G' such that $L(G') = L(G) \setminus \{\epsilon\}$.

Getting rid of unit productions For each pair of non-terminals A, B determine whether A derives B in the given grammar G. (Easy in the abscence of ϵ productions.) Then we construct a new grammar G' as follows : for each pair of non-terminals A, B such that A derives B and for each non-unit production of the form $B \to \alpha$, we add the production $A \to \alpha$ to G'.

Chomsky Normal Form : Any CFL not containing ϵ admits a CFG with all productions of the form $A \to a$ or $A \to BC$