CFGs / CFLs

Standard forms of CFGs

nonterminal can derive no more than two terminals

$A \to \varepsilon$

$\times$ not included

Chomsky Normal Form

$[[\,][\,]]$

$]X]]$

Productions : $A \to BC \mid DE \mid GF$

$S \to AB \mid AC \mid SS$, $C \to SB$, $A \to [$, $B \to]$

$A \to a$ → the rule that gives a terminal should never give anything else with it

$\hookrightarrow [\,,\,] \in \Sigma$

Greibach Normal Form

generating all possible legitimate bracketing

Productions : $A \to aB_1B_2 \cdots B_k$ chars rest all nonterminals

exactly one terminal character at the left

$k \geq 0$

$S \to [B \mid [SB \mid [BS \mid [SBS$

$B \to]$

---

unit production : $A \to B$

Lemma: For any CFG $G = (N, \Sigma, P, S)$, there is a CFG G' with no ε or unit productions such that

$$L(G') = L(G) - \{\varepsilon\}.$$

$A \to B \to C \to DE \to$

$D \to \varepsilon$, $E \to \varepsilon$

$\to \varepsilon\varepsilon\varepsilon\varepsilon\varepsilon a \varepsilon\varepsilon \to$ huge collapses due to ε and unit productions

$\hat{P}$ is the smallest set of productions containing P and closed under:

a) If $A \to \alpha B \beta$ and $B \to \varepsilon$ are in $\hat{P}$, then $A \to \alpha\beta \in \hat{P}$.

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

$P \in A \to aBd$, $B \to \varepsilon$

$\hat{P}$ $A \to ad$

$A \to B \to \gamma$

Why can $\hat{P}$ be obtained from P in finitely many steps?

$G = (N, \Sigma, P, S)$

$G' = (N, \Sigma, \hat{P}, S)$

Claim: $L(G') = L(G)$ ~~$-\{\varepsilon\}$~~

Obviously, $L(G) \subseteq L(G')$

$L(G') \subseteq L(G)$ because any new production rule used from G' can be replaced by two rules from G

Any rule of $\hat{P} \setminus P$ can be substituted by a series of rules from P in any derivation

$A \to \gamma$

$\boxed{A \to B \to \gamma}$

Anything that can be generated using $\hat{P}$ can also be generated using P.

Claim: $\hat{P}$ allows us to generate $L(G)_{-\{\varepsilon\}}$ without using any ε or unit productions.

Proof: $x \neq \varepsilon$
$x \in \Sigma^*$

$S \to x$ min length → doesn't use any ε or unit productions

Suppose on the contrary an ε or unit production is used.

at least one of them $\neq \varepsilon$

$S \xrightarrow{*} \boxed{\gamma B \delta} \xrightarrow{1} \gamma\delta \xrightarrow{*} x$

ε production

This means that in the step previous to $\gamma B \delta$ could have derived $\gamma\delta$ directly according to rule (a).

This means the derivation above is not of minimum cardinality, a contradiction.

$\boxed{A \to B}$

Chomsky Normal Form

Consider any CFG $G = (N, \Sigma, P, S)$

First, express $L(G) - \{\varepsilon\}$ without any unit or ε prod.
construct $G' = (N, \Sigma, \hat{P}, S)$ to

all unit and ε prod-s can be excluded from $\hat{P}$.

Consider any terminal $a \in \Sigma$.

In every rule where a is defined, replace a with a new non terminal X_a

$C \to \underline{bafMP}$, and add a new rule

$bX_a fMP$ $\quad$ $\underline{X_a \to a}$ $\quad$ $X_b \to b$

do this for all non terminal characters.

$C \to X_b \underline{\boxed{X_a X_f MP}} \to Y_{\underline{X_a X_f MP}}$ $\quad$ $Y_{X_a X_f MP} \to X_a Y_{X_f MP}$

$Y_{X_f MP} \to X_f Y_{MP}$

$Y_{MP} \to MP$