

# CFGs / CFLs

## Standard forms of CFGs

### Chomsky Normal Form

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

$S \rightarrow AB \mid AC \mid SS$

$C \rightarrow SB, A \rightarrow [ , B \rightarrow ]$

$A \rightarrow a \rightarrow$  the rule that gives a terminal should never give anything else  
 $\rightarrow [ , ] \in \Sigma$

### Greibach Normal Form

Productions:  $A \rightarrow aB_1B_2 \dots B_k$

$S \rightarrow [B \mid [SB \mid [BS \mid [SBS$   
 $B \rightarrow ]$

exactly one terminal character at the left  
 $k \geq 0$

unit production:  $A \rightarrow B$

Lemma: For any CFG  $G = (N, \Sigma, P, S)$ , there is a CFG

$G'$  with no  $\epsilon$  or unit productions such that

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

$A \rightarrow B \rightarrow C \rightarrow DE \rightarrow$

$D \rightarrow \epsilon, E \rightarrow \epsilon$

$\epsilon \epsilon \epsilon \epsilon \epsilon a \epsilon \epsilon$

huge collapses due to  $\epsilon$  and unit productions

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

- If  $A \rightarrow \alpha B \beta$  and  $B \rightarrow \epsilon$  are in  $\hat{P}$ , then  $A \rightarrow \alpha \beta \in \hat{P}$ .
- If  $A \rightarrow B$  and  $B \rightarrow \gamma \in \hat{P}$ , then  $A \rightarrow \gamma$  is in  $\hat{P}$ .

$P \in A \rightarrow aBd, B \rightarrow \epsilon$   
 $\hat{P} \quad A \rightarrow ad$

$A \rightarrow B \rightarrow \gamma$

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

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

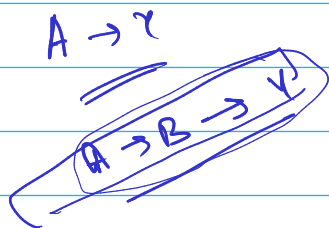
$G' = (N, \Sigma, \hat{P}, S)$

Claim:  $L(G') = L(G)$

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



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

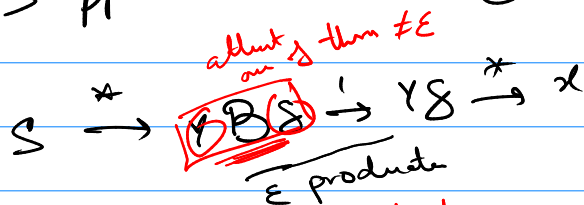
Claim:  $\hat{P}$  allows us to generate  $L(G) - \{\epsilon\}$  without using  $\epsilon$  or unit productions.

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

$S \rightarrow x$   
 min length

→ doesn't use any  $\epsilon$  or unit productions

Suppose on the contrary an  $\epsilon$  or unit production is used.



this means that what was in the step previous to  $YBS$  could have derived  $YS$  directly according to rule (a).

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

$A \rightarrow B$

## Chomsky Normal Form

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

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

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

Consider any terminal  $a \in \Sigma$ .

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

$$C \rightarrow \underline{ba_fMP} \quad , \quad \text{and add a new rule} \\ \underline{bX_a fMP} \quad \underline{X_a \rightarrow a} \quad X_b \rightarrow b$$

do this for all non terminal characters.

$$C \rightarrow \underline{(X_b X_a X_f MP)} \rightarrow \underline{Y_{X_a X_f MP}} \quad Y_{X_a X_f MP} \rightarrow X_a Y_{X_f MP} \\ Y_{X_f MP} \rightarrow X_f Y_{MP} \\ Y_{MP} \rightarrow MP$$