Manipulating grammars;

Chomsky normal form

S

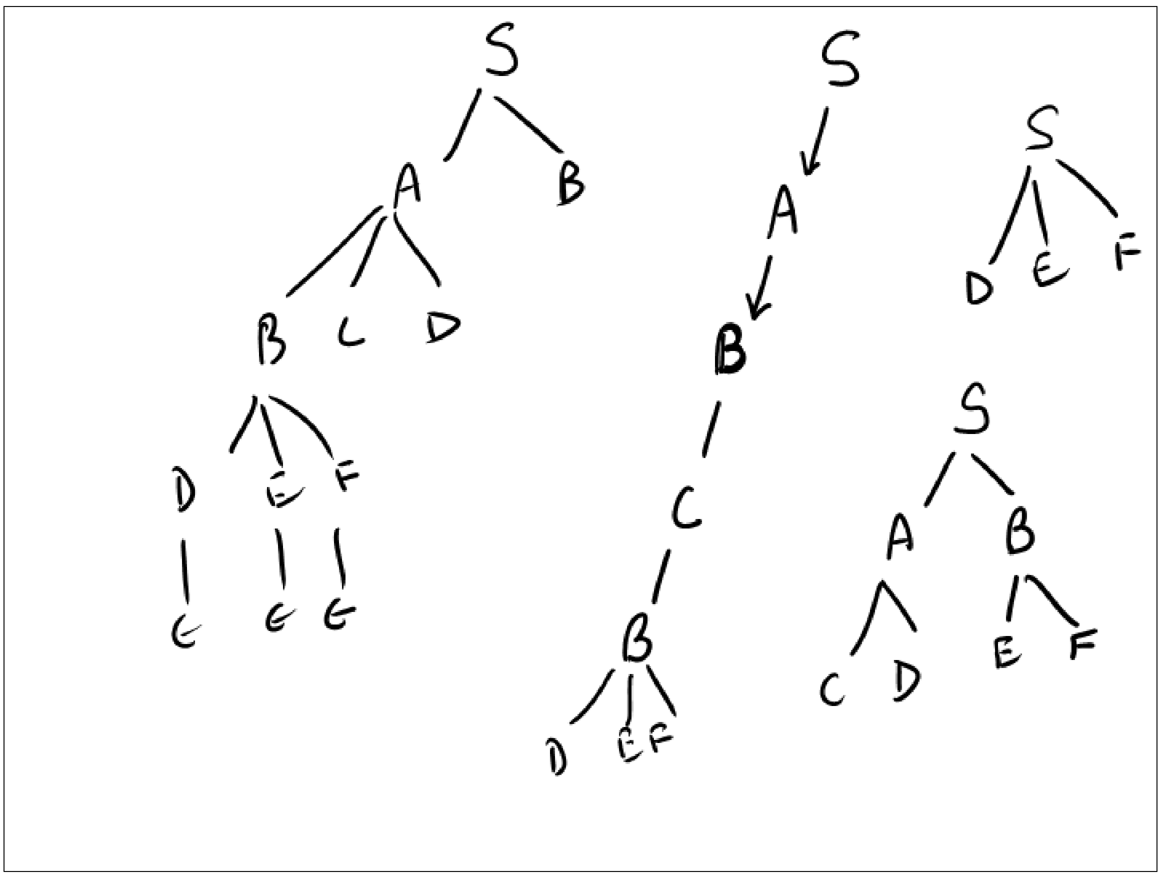
nonterminal

$$A \rightarrow aBabAc$$
$$A \rightarrow aB$$
$$A \rightarrow \epsilon$$
$$B \rightarrow aA$$
$$B \rightarrow AA\,\cancel{B}$$

Two vars on the right

Chomsky Normal Form:

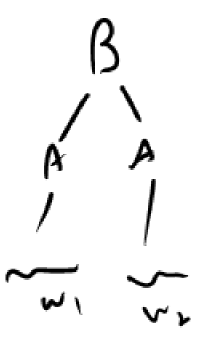
$$A \rightarrow BC$$
$$A \rightarrow a \qquad a \in \Sigma$$
$$S \rightarrow \epsilon$$


S
A
B
B
C
D
D
E
F
E
E
E
S
A
B
C
B
D
E
F
S
D
E
F
S
A
B
C
D
E
F

Given G, L(G) is empty or not is <u>decidable</u>

$S \Rightarrow^* \underset{\in \Sigma^*}{w}$

$X \Rightarrow^* \underset{\in \Sigma^*}{w}$



$S \to ABC \mid \underline{\underline{BA}}$

$B \to CA \mid \underline{A}\,\underline{A}$

$A \to aB \mid \underline{\underline{a}}$

$\cancel{B \to d}$

$C \to D$

$D \to C$

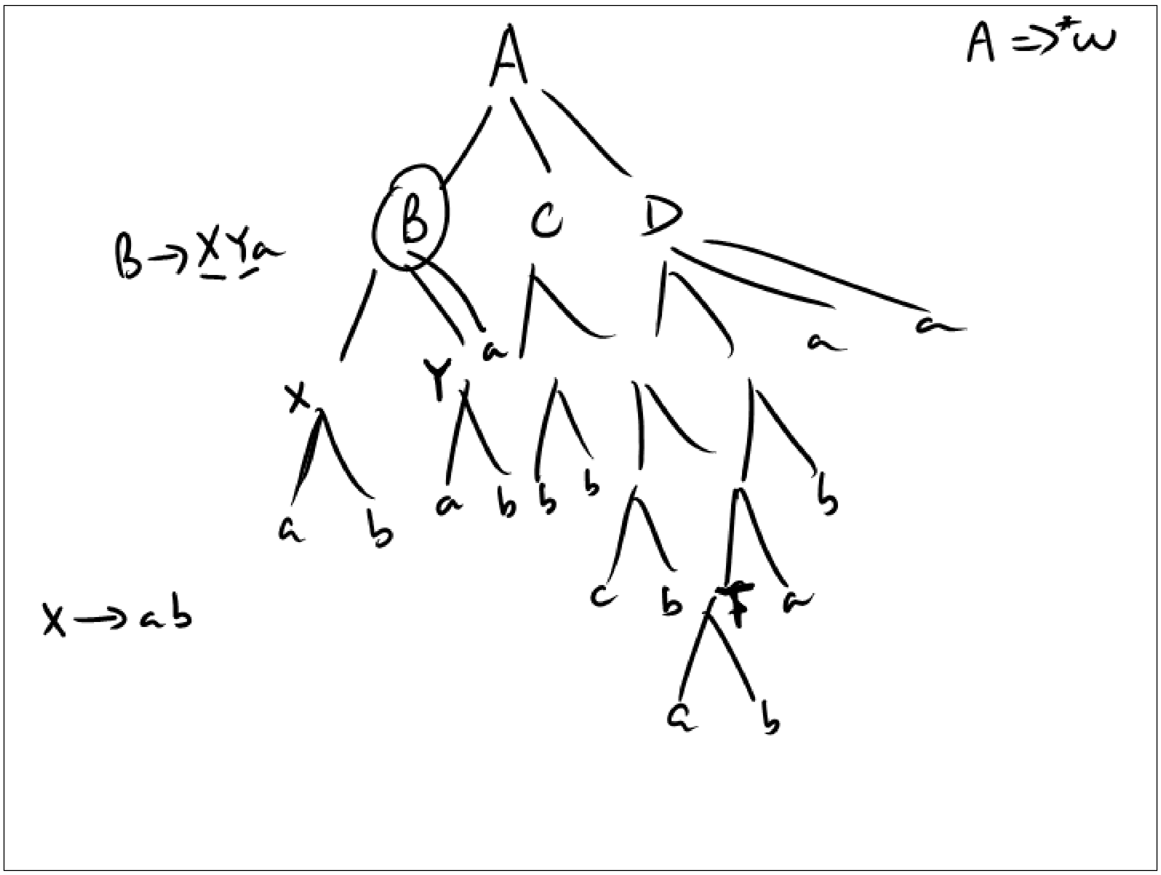
$X \to \underline{\underline{\alpha}}$

$$V' = \{ X \in V \mid X \Rightarrow^* w, \ w \in \Sigma^* \}$$

V' is the smallest subset of V such that

- If $A \rightarrow w$, $w \in \Sigma^*$, $A \in V'$
- If $A \rightarrow \beta_1 \ldots \beta_n$ where each $\beta_i \in \Sigma \cup V'$ then $A \in V'$.

```
OldV' := ∅
NewV' := { A ∈ V | A → w , w ∈ Σ* }
while ( OldV' ≠ NewV' ) do {
        OldV' := NewV'
        NewV' := NewV' ∪ { A | "A → α" ∈ P, α ∈ (Σ ∪ NewV')* }
}
V' := NewV'.
```


A =>* w
B -> XYa
X -> ab

Suppose $A \Rightarrow^* w$, I will show $\underline{\underline{A \in V'}}$

Induction on minimal # of steps to derive a word w.

$\underline{\text{If } K = 1}$ $A \Rightarrow w$ $\underline{A \rightarrow w \in P}$,

But then A gets into New V' in the first step itself.

$\underline{K > 1}$ $A \Rightarrow \underline{\beta_1} \cdots \beta_m \Rightarrow^K w$

Since $\beta_i \Rightarrow^{l_i} w_i$ for some w_i $l_i \leq K$ s.t. $w_1 \ldots w_m = w$

All β_i which are in V, are in V' (by ind hypo)

Now, $A \rightarrow \beta_1 \ldots \beta_m \in P$

and since $\{\beta_i \mid \beta_i \in V\} \subseteq V'$,

the algorithm will A to V'.

$G = (V, \Sigma, P, S)$, is $L(G) = \emptyset$?

Compute $V' = \{ X \mid X \Rightarrow^{*} \underset{w \in \Sigma^{*}}{w} \}$

Check if $S \in V'$.

If yes, $L(G) \neq \emptyset$,

o/w, $L(G) = \emptyset$.

$$S \to \epsilon$$
CNF. $$A \to BC$$
$$A \to a$$

Nontrivial	No	ϵ-productions	$A \to \epsilon$
Nontrivial	No	unit rules	$A \to B$
Easy	No	long rules	$A \to BCD, \ldots$
Easy	No	mixing nonterminals and symbols	$A \to aB$

No ϵ-productions $(A \to \epsilon)$

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

Compute the set of nullable variables

$$N = \{ X \in V \mid X \Rightarrow^* \epsilon \}$$

$OldN := \phi$
$NewN := \{ X \mid \text{"}X \to \epsilon\text{"} \in P \}$
while $(OldN \neq NewN)$ {
  $OldN := NewN$
  $NewN := \{ X \in V \mid \text{"}X \to \alpha\text{"} \in P,\ \alpha \in (NewN)^* \}$
}
$N := NewN$

$B \to ACA$

$\times$ $A \to \epsilon$

$A \to \alpha$

$B \to ACA \mid CA \mid AC \mid C$

$B \to \cancel{A}\epsilon$

$A \to \epsilon$

Transforming the grammar $G=(V,\Sigma,P,S)$

$G'=(V',\Sigma,P',S')$; $V'=V\cup\{S'\}$

Compute N .

Productions in P'

$S' \longrightarrow S$

$S' \longrightarrow \epsilon$ if $S \in N$

$m>1$

For every rule $A \longrightarrow \beta_1 \ldots \beta_m$ $\beta_i \in \Sigma \cup V$

add the rules of the form $A \rightarrow \gamma_1 \ldots \gamma_m$

where

- If $\beta_i \in \Sigma$, $\gamma_i = \beta_i$
- If $\beta_i \in N$, $(\gamma_i = \epsilon \vee \gamma_i = \beta_i)$
  If $\beta_i \in V \setminus N$, $\gamma_i = \beta_i$
- $\gamma_1 \ldots \gamma_m \neq \epsilon$

Thm: Thus resulting G' has no ϵ-productions and $L(G) = L(G')$

Removing unit rules $(A \rightarrow B)$ ...

$X \Rightarrow Z \Rightarrow Y$

Compute

$$R = \{ \underset{\in V \times V}{(X, Y)} \mid X \Rightarrow^{*} Y \}$$

$A \rightarrow B$ x

$C \rightarrow AAD \mid BAD \mid ABD \mid BBD$

OldR := ϕ
NewR := $\{ (X,Y) \mid$ "$X \rightarrow Y$" $\in P \}$
while (OldR $\neq$ NewR) {
    OldR := NewR;
    NewR := NewR $\cup \{ (X,Y) \mid \exists Z \in V.\ (X,Z) \in NewR \wedge (Z,Y) \in NewR \}$
}
R := NewR;

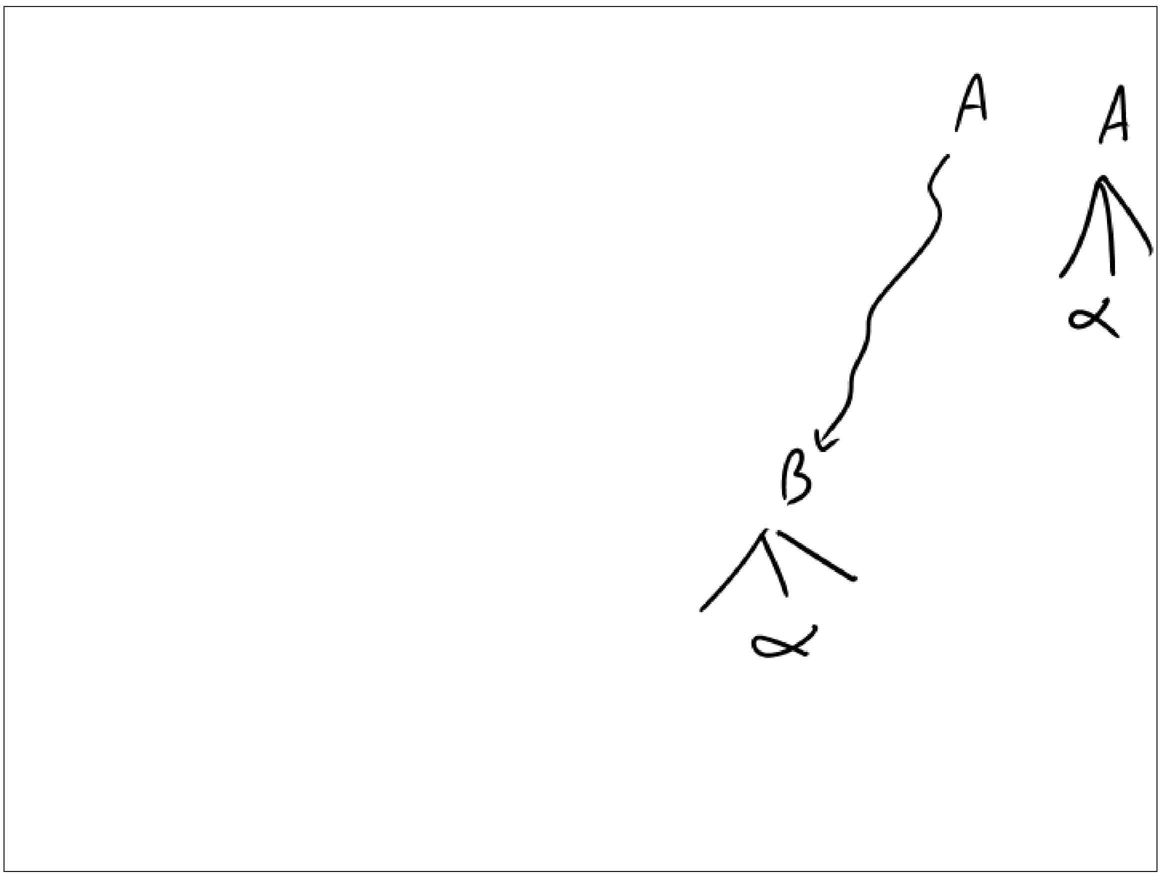
$G = (V, \Sigma, P, S)$

$G' = (V, \Sigma, P', S)$

P' : All non-unit in rules in P

$\cup$ { For every rule $B \to \alpha$ → non-unit rule. and $(A, B) \in R$, add $A \to \alpha$ }

This removes all unit-rules.



A
A
B
α
α

Removing ~~the~~ long rules

$$X \to AbBC$$

$$\hookrightarrow X \to AZ_1$$

$$Z_1 \to bZ_2$$

$$Z_2 \to BC$$

Remove mixing symbols.

$$X \longrightarrow a A$$

$$X \longrightarrow V_a A$$

$$V_a \longrightarrow a$$