CSE 431: Introduction to Theory of Computation

Converting to Chomsky Normal Form

Paul Beame

Chomsky Normal Form

- Grammar rules allowed

$A \to BC$ where $B,C \in V$ $B,C \neq S$

$A \to a$ where $a \in \Sigma$

$S \to \varepsilon$

Step 1

- Add new start symbol S_0 and rule $S_0 \rightarrow S$

$$\mathbf{S_0} \rightarrow \mathbf{S}$$
$$S \rightarrow ASA \mid aB$$
$$A \rightarrow B \mid S$$
$$B \rightarrow b \mid \varepsilon$$

Step 2

- For each a $\in\Sigma$ replace each a that appears on the RHS of a rule of size $\neq 1$ with new variable U_a and add rule $U_a \rightarrow a$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid \mathbf{a}B$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

Step 2

- For each $a \in \Sigma$ replace each a that appears on the RHS of a rule of size $\neq 1$ with new variable U_a and add rule $U_a \rightarrow a$

$S_0 \rightarrow S$

$S \rightarrow ASA \mid \mathbf{U}B$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

$\mathbf{U} \rightarrow \mathbf{a}$

Step 3

- For each rule of size >2 of the form $A \rightarrow B_1 B_2 \ldots B_k$ add new variables $T_2, \ldots, T_{k-1}$ and rules
  $A \rightarrow B_1 T_2$
  $T_2 \rightarrow B_2 T_3$
  …
  $T_{k-2} \rightarrow B_{k-2} T_{k-1}$
  $T_{k-1} \rightarrow B_{k-1} B_k$

$S_0 \rightarrow S$

$S \rightarrow \mathbf{ASA} \mid UB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

$U \rightarrow a$

Step 3

- For each rule of size >2 of the form
  $A \to B_1 B_2 \ldots B_k$ add new variables
  $T_2, \ldots, T_{k-1}$ and rules
  $A \to B_1 T_2$
  $T_2 \to B_2 T_3$
  …
  $T_{k-2} \to B_{k-2} T_{k-1}$
  $T_{k-1} \to B_{k-1} B_k$

$S_0 \to S$

$S \to A\mathbf{T} \mid UB$

$A \to B \mid S$

$B \to b \mid \varepsilon$

$U \to a$

$\mathbf{T} \to \mathbf{SA}$

Step 4

- Define set $\mathcal{E}$ by
  - For each rule of the form $A \rightarrow \varepsilon$ add A to $\mathcal{E}$
  - Repeat until done:
    If $A \rightarrow BC$ or $A \rightarrow B$ where $B, C \in \mathcal{E}$ then add A to $\mathcal{E}$

$S_0 \rightarrow S$

$S \rightarrow AT \mid UB$

$A \rightarrow B \mid S$

$B \rightarrow b \mid \varepsilon$

$U \rightarrow a$

$T \rightarrow SA$

$\mathcal{E} = \{B, A\}$

Step 4'

- For each $B \in \mathcal{E}$
  For each rule $A \rightarrow BC$
  add the rule $A \rightarrow C$
- For each $C \in \mathcal{E}$
  For each rule $A \rightarrow BC$
  add the rule $A \rightarrow B$
- Remove all $A \rightarrow \varepsilon$ rules
- If $S_0 \in \mathcal{E}$ then add $S_0 \rightarrow \varepsilon$

$S_0 \rightarrow S$
$S \rightarrow AT \mid UB$
$A \rightarrow B \mid S$
$B \rightarrow b \mid \varepsilon$
$U \rightarrow a$
$T \rightarrow SA$

$\mathcal{E}=\{B,A\}$

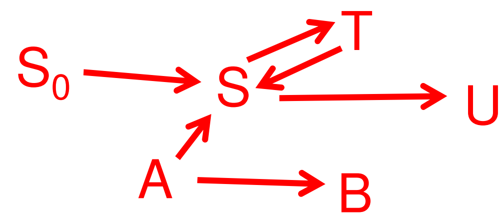
Step 4'

- For each $B \in \mathcal{E}$ For each rule $A \rightarrow BC$ add the rule $A \rightarrow C$
- For each $C \in \mathcal{E}$ For each rule $A \rightarrow BC$ add the rule $A \rightarrow B$
- Remove all $A \rightarrow \varepsilon$ rules
- If $S_0 \in \mathcal{E}$ then add $S_0 \rightarrow \varepsilon$

$S_0 \rightarrow S$
$S \rightarrow AT \mid UB \mid \mathbf{T} \mid \mathbf{U}$
$A \rightarrow B \mid S$
$B \rightarrow b$
$U \rightarrow a$
$T \rightarrow SA \mid \mathbf{S}$

$\mathcal{E}=\{B,A\}$

Step 5



- Call rules of form A→B unit rules
- Call all other rules interesting ones
- For each A compute the set D(A) of all other variables reachable from A via unit rules
- Remove all unit rules and add all interesting rules on the RHS of vars in D(A) to the RHS for A

$S_0 \rightarrow S$

$S \rightarrow \underline{AT} \mid \underline{UB} \mid T \mid U$

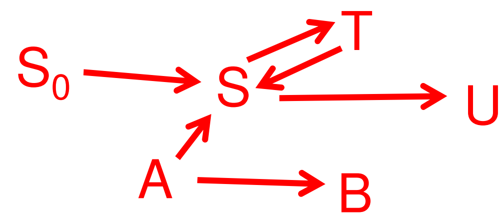
$A \rightarrow B \mid S$

$B \rightarrow \underline{b}$

$U \rightarrow \underline{a}$

$T \rightarrow \underline{SA} \mid S$

Step 5



- Call rules of form $A \rightarrow B$ unit rules
- Call all other rules interesting ones
- For each A compute the set D(A) of all other variables reachable from A via unit rules
- Remove all unit rules and add all interesting rules on the RHS of vars in D(A) to the RHS for A

$S_0 \rightarrow S$

$S \rightarrow \underline{AT} \mid \underline{UB} \mid T \mid U$

$A \rightarrow B \mid S$

$B \rightarrow \underline{b}$

$U \rightarrow \underline{a}$

$T \rightarrow \underline{SA} \mid S$

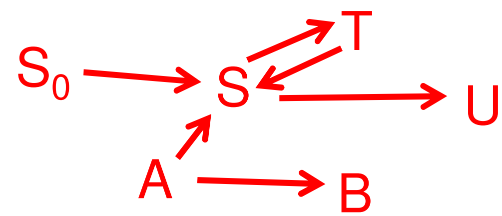
D(B)={B} D(U)={U}

D(T)=D(S)={S,T,U}

$D(S_0)=\{S_0,S,T,U\}$

D(A)={A,B,S,T,U}

Step 5



- Call rules of form A→B unit rules
- Call all other rules interesting ones
- For each A compute the set D(A) of all other variables reachable from A via unit rules
- Remove all unit rules and add all interesting rules on the RHS of vars in D(A) to the RHS for A

$S_0 \rightarrow$

$S \rightarrow$ AT | UB

$A \rightarrow$

$B \rightarrow$ b

$U \rightarrow$ a

$T \rightarrow$ SA

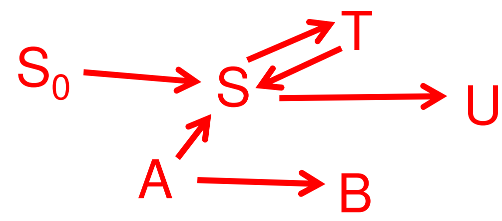
D(B)={B} D(U)={U}

D(T)=D(S)={S,T,U}

D(S_0)={S_0,S,T,U}

D(A)={A,B,S,T,U}

Step 5



- Call rules of form A→B **unit rules**
- Call all other rules **interesting** ones
- For each A compute the set D(A) of all other variables reachable from A via unit rules
- Remove all unit rules and add all interesting rules on the RHS of vars in D(A) to the RHS for A

$S_0 \rightarrow$ **AT | UB | a | SA**

$S \rightarrow$ AT | UB | **a | SA**

$A \rightarrow$ **AT | UB | a | SA | b**

$B \rightarrow$ b

$U \rightarrow$ a

$T \rightarrow$ **AT | UB | a |** SA

D(B)={B} D(U)={U}

D(T)=D(S)={S,T,U}

$D(S_0)=\{S_0,S,T,U\}$

D(A)={A,B,S,T,U}