CSE 322: Introduction to Formal Models in Computer Science

Converting to Chomsky Normal Form

Paul Beame

Chomsky Normal Form

- Grammar rules allowed

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

$A \rightarrow a$ where $a \in \Sigma$

$S \rightarrow \varepsilon$

Step 1

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

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

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 $U_a \rightarrow a$

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

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 $U_a \rightarrow a$

$S_0 \rightarrow S$
$S \rightarrow ASA \mid UB$
$A \rightarrow B \mid S$
$B \rightarrow b \mid \varepsilon$
$U \rightarrow a$

Step 3

- For each rule of size >2 of the form $A \rightarrow B_1B_2 \ldots B_k$ add new variables $T_2, \ldots, T_{k-1}$ and rules
$A \rightarrow B_1T_2$
$T_2 \rightarrow B_2T_3$
…
$T_{k-2} \rightarrow B_{k-2}T_{k-1}$
$T_{k-1} \rightarrow B_{k-1}B_k$

$S_0 \rightarrow S$
$S \rightarrow ASA \mid UB$
$A \rightarrow B \mid S$
$B \rightarrow b \mid \varepsilon$
$U \rightarrow a$

Step 3

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

$S_0 \to S$
$S \to$ **AT** | UB
$A \to B \mid S$
$B \to b \mid \varepsilon$
$U \to a$
T → SA

Step 4

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

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

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

Step 4'

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

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

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

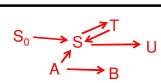
Step 4'

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

$S_0 \to S$
$S \to AT \mid UB \mid$ **T** | **U**
$A \to B \mid S$
$B \to b$
$U \to a$
$T \to SA \mid$ **S**

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

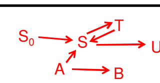
Step 5



- Call rules of form $A \to B$ 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

$S_0 \to S$
$S \to$ <u>AT</u> | <u>UB</u> | T | U
$A \to B \mid S$
$B \to$ <u>b</u>
$U \to$ <u>a</u>
$T \to$ <u>SA</u> | S

Step 5

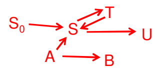


- Call rules of form $A \to B$ unit rules
- Call all other rules interesting ones
- For each A compute the set D(A) of all other variables reachable from A via unit rules
- Remove all unit rules and add all interesting rules on the RHS of vars in D(A) to the RHS for A

$S_0 \to S$
$S \to$ <u>AT</u> | <u>UB</u> | T | U
$A \to B \mid S$
$B \to$ <u>b</u>
$U \to$ <u>a</u>
$T \to$ <u>SA</u> | 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 → AT | UB

A →

B → b

U → a

T → SA

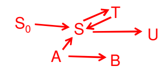
D(B)={B} D(U)={U}

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 → AT | UB | **a** | **SA**

A → **AT** | **UB** | **a** | **SA** | **b**

B → b

U → a

T → **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}