

CS 456: Quiz 6

Due on April 12, 2024

Nancy LaTorrette Section 1001

Christopher Howe

Problem

Convert the following CFG into Chomsky Normal form using the method demonstrated in the lecture.

$G = (\{S, A, B, C, D\}, \{a, b, c\}, S, P)$ where P is the following

$S \rightarrow ABDa|aa$

$A \rightarrow BC|a$

$B \rightarrow bB|b|\lambda$

$C \rightarrow CD|\lambda$

$D \rightarrow c$

Solution

The process demonstrated in class follows 5 main steps. First, ensure that the start variable is not produced by any other rules; if it is, add another start symbol. Second, remove any rule that produces the empty string (λ). Third, simplify/"shortcut" any unit rules. These are productions that follow the form $A \rightarrow B$. Fourth, break apart any productions that are more than two symbols and replace terminals in production rules with two symbols. Fifth, remove any unreachable or redundant rules.

Ensure the start variable is not the product of any other rules

This step can be skipped since no rules produce the start symbol.

Remove any rule that produces the empty string

Two variables that can produce the empty string, B and C. We will remove them by adding a rule to every production that has B to use the empty string instead. However, after removing the λ productions, a new one is produced for A which must also be removed.

$S \rightarrow ABDa|ADa|aa$

$A \rightarrow BC|C|a$

$B \rightarrow bB|b$

$C \rightarrow CD|\lambda$

$D \rightarrow c$

$S \rightarrow ABDa|ADa|aa$

$A \rightarrow BC|B|C|\lambda|a$

$B \rightarrow bB|b$

$C \rightarrow CD$

$D \rightarrow c$

$S \rightarrow ABDa|BDa|ADa|Da|aa$

$A \rightarrow BC|B|C|a$

$B \rightarrow bB|b$

$C \rightarrow CD$

$D \rightarrow c$

Productions after $B \rightarrow \lambda$ is removed

Productions after $C \rightarrow \lambda$ is removed

Productions after $A \rightarrow \lambda$ is removed

Remove Unit Rules

After removing all λ transitions, all unit rules can be removed and dropped since they serve no purpose. This can be done by determining what can be produced by each unit rule. In this grammar, there are only two unit rules: $A \rightarrow B$ and $A \rightarrow C$. To remove these, the LHS of the production can add the production rules of the RHS. The resulting productions of A are as follows.

$$A \rightarrow BC|a|bB|b|CD$$

Rule Splitting and terminal substitution

After removing all the unit rules and productions resulting in the empty string, the productions must be mutated to only contain rules adhering to Chomsky normal form. The productions for S , A , and B do not adhere to this form. To change them, create new variables to wrap illegal forms.

Fixed productions for S

$$S \rightarrow Y_1, Y_2|BY_2|AY_2|DU_a|U_aU_a$$

$$Y_1 \rightarrow AB$$

$$Y_2 \rightarrow DU_a$$

$$U_a \rightarrow a$$

Fixed productions for A

$$A \rightarrow BC|a|U_bB|b|CD$$

$$U_b \rightarrow b$$

Fixed productions for B

$$B \rightarrow U_bB|b$$

Grammar in Chomsky Normal Form

The resulting grammar is in Chomsky normal form. See below.

$G_{\text{Chomsky}} = (\{S, A, B, C, D, Y_1, Y_2, U_a, U_b\}, \{a, b, c\}, S, P_1)$ where P_1 is the following

$$S \rightarrow Y_1, Y_2|BY_2|AY_2|DU_a|U_aU_a$$

$$A \rightarrow BC|a|U_bB|b|CD$$

$$B \rightarrow U_bB|b$$

$$C \rightarrow CD$$

$$D \rightarrow c$$

$$Y_1 \rightarrow AB$$

$$Y_2 \rightarrow DU_a$$

$$U_a \rightarrow a$$

$$U_b \rightarrow b$$