<Lecture 16>

Chomsky Normal Form

A CFG is in Chomsky Normal Form if every rule is of the form

$A \to BC$

$A \to a$

where a is a terminal and A, B, C are variables (B and C are not the start variable)

In addition, we allow

$S \to \varepsilon$

where S is the start variable

Thm 2.9.

Any context-free language is generated by a context-free grammar in Chomsky Normal Form.

(1) add a new start variable S_0 and $S_0 \to S$

(2) remove ε-rules $A \to \varepsilon$, where A is not the start variable :

$R \to uAv$: add $R \to uv$

$R \to uAvAw$: add $R \to uvAw \mid uAvw \mid uvw$

$R \to A$: add $R \to \varepsilon$ unless $R \to \varepsilon$ was removed before

(3) remove unit rules $A \to B$

$B \to u$: add $A \to u$ unless this was a unit rule removed before

(4) Replace: $A \to u_1 u_2 \cdots u_k$ $(k \geq 3)$

with $A \to u_1 A_1$, $A_1 \to u_2 A_2$, $\cdots$ $A_{k-3} \to u_{k-2} A_{k-2}$, $A_{k-2} \to u_{k-1} u_k$

Replace $A \to u_1 u_2$ (u_1, u_2 terminals)

$A \to U_1 U_2$, $U_1 \to u_1$, $U_2 \to u_2$

- Reducing the variables on RHS

$A \to BCDEF$

$A \to BK_1$

$K_1 \to CK_2$

$K_2 \to DK_3$

$K_3 \to EF$

Removing unit rules

$$S \to Aa \mid B$$
$$B \to A \mid bb$$
$$A \to a \mid bc \mid B$$

$$\begin{array}{l} S \to XY \\ X \to A \\ A \to B \mid a \\ B \to b \end{array} \Rightarrow \begin{array}{l} S \to XY \\ X \to a \mid b \\ A \to a \mid b \\ B \to b \end{array}$$

Removing ε-rules

$$S \to ABaC$$
$$A \to BC$$
$$B \to b \mid \varepsilon$$
$$C \to D \mid \varepsilon$$
$$D \to d$$

$$\begin{array}{l} S \to aACa \\ A \to B \mid a \\ B \to C \mid c \\ C \to cC \mid \varepsilon \end{array} \Rightarrow \begin{array}{l} S \to aACa \mid aAa \mid aCa \mid aa \\ A \to B \mid a \\ B \to C \mid c \\ C \to cC \mid c \end{array}$$

Example 2.10

You should be able to do it without referring to the book.