

3. (Identifying A -derivable variables, for each A) The S -derivable variables obviously include T , U , and V , and they also include W because of the production $V \rightarrow W$. The V -derivable variable is W .
4. (Eliminating unit productions) We add the productions

$$S \rightarrow aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b \quad V \rightarrow bW \mid b$$

before eliminating unit productions. At this stage, we have

$$S \rightarrow TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$$

$$T \rightarrow aTb \mid ab$$

$$U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid ac \mid bW \mid b$$

$$W \rightarrow bW \mid b$$

5. (Converting to Chomsky normal form) We replace a , b , and c by X_a , X_b , and X_c , respectively, in productions whose right sides are not single terminals, obtaining

$$S \rightarrow TU \mid X_aTX_b \mid X_aX_b \mid X_cU \mid c \mid X_aVX_c \mid X_aX_c \mid X_bW \mid b$$

$$T \rightarrow X_aTX_b \mid X_aX_b$$

$$U \rightarrow X_cU \mid c$$

$$V \rightarrow X_aVX_c \mid X_aX_c \mid X_bW \mid b$$

$$W \rightarrow X_bW \mid b$$

This grammar fails to be in Chomsky normal form only because of the productions $S \rightarrow X_aTX_b$, $S \rightarrow X_aVX_c$, $T \rightarrow X_aTX_b$, and $V \rightarrow X_aVX_c$. When we take care of these as described above, we obtain the final CFG G_1 with productions

$$S \rightarrow TU \mid X_aY_1 \mid X_aX_b \mid X_cU \mid c \mid X_aY_2 \mid X_aX_c \mid X_bW \mid b$$

$$Y_1 \rightarrow TX_b$$

$$Y_2 \rightarrow VX_c$$

$$T \rightarrow X_aY_3 \mid X_aX_b$$

$$Y_3 \rightarrow TX_b$$

$$U \rightarrow X_cU \mid c$$

$$V \rightarrow X_aY_4 \mid X_aX_c \mid X_bW \mid b$$

$$Y_4 \rightarrow VX_c$$

$$W \rightarrow X_bW \mid b$$

(We obviously don't need both Y_1 and Y_3 , and we don't need both Y_2 and Y_4 , so we could simplify G_1 slightly.)

EXERCISES

- 4.1. In each case below, say what language (a subset of $\{a, b\}^*$) is generated by the context-free grammar with the indicated productions.