Normal Forms for Context-Free Grammars
CSE 322: Introduction to Formal Models in Computer Science
February 11, 2002

1. Putting a Context-Free Grammar in Normal Form

Definition: A context-free grammar $G = (V, \Sigma, R, S)$ is in normal form if and only if $R$ contains no rules of the form

1. $A \to \varepsilon$, for any $A \in V$, or
2. $A \to B$, for any $A, B \in V$.

Here is a procedure for converting a grammar $G$ into a grammar $G'$ such that $G'$ is in normal form, and $L(G') = L(G) - \{\varepsilon\}$. Throughout the procedure, $A$ and $B$ are arbitrary elements of $V$, and $u$ and $v$ are arbitrary strings in $(V \cup \Sigma)^*$.

1. (a) For every pair of rules $A \to \varepsilon$ and $B \to uAv$, add a new rule $B \to uv$. Continue doing this until no new rule can be added by this procedure.
   (b) Remove all rules $A \to \varepsilon$. ¹

2. (a) For every pair of rules $A \to B$ and $B \to u$, add a new rule $A \to u$. Continue doing this until no new rule can be added by this procedure.
   (b) Remove all rules $A \to B$.

2. Example

Put $G = (V, \Sigma, R, S)$ in normal form, where

\[
V = \{S, A, B\},
\]

\[
\Sigma = \{a, b\}, \text{ and}
\]

\[
R = \{S \to A, A \to SB, A \to B, B \to aAbB, B \to \varepsilon\}.
\]

(Since $\varepsilon \in L(G)$, the resulting normal form grammar will generate $L(G) - \{\varepsilon\}$.)

1. (a) Add $A \to S$, $A \to \varepsilon$, $B \to aAb$.
   Add $S \to \varepsilon$, $B \to abB$, $B \to ab$.

¹If $S \to \varepsilon$ is removed in this step, then $L(G) - L(G') = \{\varepsilon\}$; otherwise, $L(G) = L(G')$.  

1
(b) Remove $A \to \varepsilon$, $B \to \varepsilon$, $S \to \varepsilon$.

At this point, the set of rules is

$$\{S \to A, \\
A \to SB, A \to S, A \to B, \\
B \to aAbB, B \to aAb, B \to abB, B \to ab\}.$$ 

2. (a) Add $S \to SB$, $S \to S$, $S \to B$, $A \to A$, $A \to aAbB$, $A \to aAb$, $A \to abB$, $A \to ab$.

Add $S \to aAbB$, $S \to aAb$, $S \to abB$, $S \to ab$.

(b) Remove $S \to S$, $S \to A$, $S \to B$, $A \to S$, $A \to A$, $A \to B$.

The final set of rules is

$$\{S \to SB, S \to aAbB, S \to aAb, S \to abB, S \to ab, \\
A \to SB, A \to aAbB, A \to aAb, A \to abB, A \to ab, \\
B \to aAbB, B \to aAb, B \to abB, B \to ab\}.$$ 

As an example of how the equivalence works, consider the following derivation in the original grammar $G$: $S \Rightarrow A \Rightarrow B \Rightarrow aAbB \Rightarrow aSbbB \Rightarrow aABbb \Rightarrow aAbBBbb \Rightarrow aBbBBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB \Rightarrow aabBbBbB.$

This is simulated in the normal form grammar by the following derivation of the same terminal string: $S \Rightarrow aAbB \Rightarrow aSbBb \Rightarrow aabBbB \Rightarrow aabBbB \Rightarrow aabBbB \Rightarrow aabBbB.$

3. **Putting a Context-Free Grammar in Chomsky Normal Form**

**Definition:** A context-free grammar $G = (V, \Sigma, R, S)$ is in *Chomsky normal form* if and only if every rule in $R$ is of one of the following forms:

1. $A \to a$, for $A \in V$ and $a \in \Sigma$, or

2. $A \to BC$, for $A, B, C \in V$.

Here is a procedure for putting a normal form grammar in Chomsky normal form, without changing the language generated by the grammar. Throughout the procedure, $A$ and $B_1, B_2, \ldots, B_m$ are variables, and $X_1, X_2, \ldots X_m$ are arbitrary elements in $V \cup \Sigma$.

1. For each terminal symbol $a$, add a new variable $C_a$ and a new rule $C_a \to a$.

2. Let $A \to X_1X_2 \cdots X_m$ be a rule, with $m \geq 2$. For each $1 \leq i \leq m$, if $X_i$ is a terminal symbol $a$, replace $X_i$ in the right hand side of the original rule by $C_a$.

3. Let $A \to B_1B_2 \cdots B_m$ be a rule, with $m \geq 3$. Add new variables $D_1, D_2, \ldots, D_{m-2}$, and replace the rule $A \to B_1B_2 \cdots B_m$ by the rules

$$A \to B_1D_1, \ D_1 \to B_2D_2, \ldots, D_{m-3} \to B_{m-2}D_{m-2}, \ D_{m-2} \to B_{m-1}B_m.$$
4. Example

Put $G = (V, \Sigma, R, S)$ in Chomsky normal form, where

\[
\begin{align*}
V &= \{S, A\}, \\
\Sigma &= \{a, b\}, \text{ and} \\
R &= \{S \to aAb, A \to aAbS, A \to b\}.
\end{align*}
\]

Notice that $G$ is already in normal form.

The result of steps 1 and 2 is $G' = (V', \Sigma', R', S)$, where

\[
\begin{align*}
V' &= \{S, A, C_a, C_b\}, \\
\Sigma &= \{a, b\}, \text{ and} \\
R' &= \{S \to C_aAC_b, A \to C_aAC_bS, A \to b, C_a \to a, C_b \to b\}.
\end{align*}
\]

The result of step 3 is $G'' = (V'', \Sigma'', R'', S)$, where

\[
\begin{align*}
V'' &= \{S, A, C_a, C_b, D_1, E_1, E_2\}, \\
\Sigma &= \{a, b\}, \text{ and} \\
R'' &= \{S \to C_aD_1, D_1 \to AC_b, A \to C_aE_1, E_1 \to AE_2, E_2 \to C_bS, A \to b, C_a \to a, C_b \to b\}.
\end{align*}
\]

$G''$ is in Chomsky normal form.