| | |
|---|---|
| **CS381 Fall 2000** | **Practice Set 2, for Quiz 2** |
| **Nov.13, 2000** | |

**Example 1.** Find a Chomsky Normal Form of CFG $S \longrightarrow aXbY$, $X \longrightarrow aX \mid \epsilon$, $Y \longrightarrow bY \mid \epsilon$. Apply an algorithm from HO19.

Step 1: getting rid of $\epsilon$.
$S \longrightarrow aXbY \mid abY \mid aXb \mid ab, \quad X \longrightarrow aX \mid a, \quad Y \longrightarrow bY \mid b.$
Step 2. Replacing terminals in nontrivial productions.
$S \longrightarrow AXBY \mid ABY \mid AXB \mid AB, \quad X \longrightarrow AX \mid a, \quad Y \longrightarrow BY \mid b, \quad A \longrightarrow a, \quad B \longrightarrow b.$
Step 3. Shortening long productions by introducing extra nonterminals.
$S \longrightarrow AU \mid AV \mid AW \mid AB, \quad U \longrightarrow XV, \quad V \longrightarrow BY, \quad W \longrightarrow XB, \quad X \rightarrow AX \mid a,$ $Y \rightarrow BY \mid b, \quad A \longrightarrow a, \quad B \longrightarrow b.$

**Example 2.** Find a Greibach Normal Form of CFG $S \longrightarrow XbY, \ X \longrightarrow aX \mid a, \ Y \longrightarrow bY \mid a.$
The general algorithm converting CFG into GNF is not practical. Use common sense and some tricks instead. The first steps are easy: get rid of $\epsilon$ and terminals other then the first ones in their productions:
$S \longrightarrow XBY, \quad X \longrightarrow aX \mid a, \quad Y \longrightarrow bY \mid a, \quad B \longrightarrow b.$
Trace the leftmost substitutions on nonterminals occurring first in productions until a terminal comes first:
$S \stackrel{1}{\longrightarrow} XBY \stackrel{1}{\longrightarrow} aBY \stackrel{1}{\longrightarrow} \dots$
$S \stackrel{1}{\longrightarrow} XBY \stackrel{1}{\longrightarrow} aXBY \stackrel{1}{\longrightarrow} \dots.$
Replace old productions with first nonterminals by all possible productions obtained above:
$S \longrightarrow aBY \mid aXBY$, $X \longrightarrow aX \mid a$, $Y \longrightarrow bY \mid a$, $B \longrightarrow b$.

**Example 3.** Find a Greibach Normal Form of PAREN$-\{\epsilon\}$ (an old example in a new light). We start with the usual CFG $S \longrightarrow [S] \mid SS \mid [\ ]$. Getting rid of non-first terminals gives
$S \longrightarrow [SR \mid SS \mid [R, \quad R \longrightarrow ]\ .$
Tracing the leftmost substitutions in $S \longrightarrow SS$:
$S \stackrel{1}{\longrightarrow} SS \stackrel{1}{\longrightarrow} [SRS \dots$
$S \stackrel{1}{\longrightarrow} SS \stackrel{1}{\longrightarrow} [RS \dots$
Replacing the old production by the results of the above tracing:
$S \longrightarrow [R \mid [SR \mid [RS \mid [SRS, \quad R \longrightarrow ].$

**Example 4.** Prove that $A = \{a^n b^n a^n \mid n \geq 0\}$ is not a CFL (cf. Kozen, p.154, Ex. 22.3). Suppose $A$ is a CFL. By the Pumping Lemma there should be an integer $k \geq 1$ such that any $z \in A$ can be broken into $z = uvwxy$, $|vx| \geq 1$, $|vwx| \leq k$ such that $uv^iwx^iy \in A$ for any $i \geq 0$. Take $n > k$ and consider all possible partitions $uvwxy$ of the string $a^n b^n a^n$.

Case 1. Each of $v$, $x$ is inside some block of letter. Note, that one of the blocks of letters remains $v, x$-free. Then $uv^iwx^iy$ makes the blocks containing $v, x$ larger, then the third block, therefore, $uv^iwx^iy \notin A$.

Case 2. At least one of $v$, $x$ has intersections with two different blocks. Then either $v$ or $x$ contains both $a$'s and $b$'s and $uv^iwx^iy$ for $i \geq 2$ is not of the form $a^m b^m a^m$ and, therefore, is not in $A$.

**Example 5.** Prove that $B = \{a^n b^n c^n \mid n \geq 0\}$ is not a CFL. $A = h(B)$ where $A$ in from Example 4, and $h$ is homomorphism $h(c) = a$. Since CFLs are closed under homomorphisms, if $B$ were context free, then $A$ should be too, which is not the case.

**Example 6.** Assume that $C = \{a^m b^n a^m b^n \mid m, n \geq 0\}$ is not a CFL (HW9, cf. Kozen, p.154, Example 22.4). Prove that $D = \{ww \mid w \in \{a, b\}^*\}$ is not a CFL. Use the theorem that intersection of a CFL with a regular language is a CFL. Note, that $C = D \cap a^*b^*a^*b^*$. If $D$ were CFL, then $C$ would be CFL too.

**Example 7.** Another example on the same lemma that CFL$\cap$ REG=CFL. We establish that the set $E$ of all strings over $\{a, b, c\}$ containing equal numbers of $a, b$ and $c$ is not a CFL. Note that $B = E \cap a^*b^*c^*$. Since the set $a^*b^*c^*$ is regular, if $E$ were CFL, then $B$ would be also a CFL, which contradicts Example 5.

**Example 8.** Convert an NPDA into a CFG. Let an NPDA $M$ have the transition function $\delta(s, a, \bot) = (s, XX)$, $\delta(s, a, X) = (s, X)$, $\delta(s, b, X) = (t, \epsilon)$, $\delta(t, a, X) = (t, \epsilon)$.
We first have to convert $M$ into a one-state NPDA $M'$. Use the algorithm from HO24. The corresponding $\delta'$ is

$$\begin{array}{ll}
\delta'(*, a, <s\bot s>) = (*, <sXs><sXs>), & \delta'(*, a, <s\bot s>) = (*, <sXt><tXs>) \\
\delta'(*, a, <s\bot t>) = (*, <sXs><sXt>), & \delta'(*, a, <s\bot t>) = (*, <sXt><tXt>), \\
\delta'(*, a, <sXs>) = (*, <sXs>), & \delta'(*, a, <sXt>) = (*, <sXt>), \\
\delta'(*, b, <sXt>) = (*, \epsilon), & \delta'(*, a, <tXt>) = (*, \epsilon).
\end{array}$$

Tracing for $M$ on the input $aaba$:
$(s, aaba, \bot) \xrightarrow{1} (s, aba, XX) \xrightarrow{1} (s, ba, XX) \xrightarrow{1} (t, a, X) \xrightarrow{1} (t, \epsilon, \epsilon)$

The corresponding tracing for $M'$ is:
$(*, aaba, <s\bot t> \xrightarrow{1} (*, aba, <sXt><tXt>) \xrightarrow{1} (*, ba, <sXt><tXt>) \xrightarrow{1}$
$\xrightarrow{1} (*, a, <tXt>) \xrightarrow{1} (*, \epsilon, \epsilon)$

From $M'$ we read the grammar $G$

$$\begin{array}{ll}
<s\bot s> \longrightarrow a <sXs><sXs>, & <s\bot s> \longrightarrow a <sXt><tXs> \\
<s\bot t> \longrightarrow a <sXs><sXt>, & <s\bot t> \longrightarrow a <sXt><tXt>, \\
<sXs> \longrightarrow a <sXs>), & <sXt> \longrightarrow a <sXt>, \\
<sXt> \longrightarrow b, & <tXt> \longrightarrow a.
\end{array}$$

Here is how $G$ generates the string $aaba$ from the initial nonterminal $<s\bot t>$:
$<s\bot t> \xrightarrow{1} a <sXt><tXt> \xrightarrow{1} aa <sXt><tXt> \xrightarrow{1} aab <tXt> \xrightarrow{1} aaba.$

Note that $M'$ contains many redundancies. After pruning transitions that never apply in accepting computations, we get much shorter NPDA:

$$\begin{array}{ll}
\delta'(*, a, <s\bot t>) = (*, <sXt><tXt>), & \delta'(*, a, <sXt>) = (*, <sXt>), \\
\delta'(*, b, <sXt>) = (*, \epsilon), & \delta'(*, a, <tXt>) = (*, \epsilon).
\end{array}$$

The same holds for the grammar. Here is a short equivalent of $G$:

$$<s\bot t> \longrightarrow a <sXt><tXt>, \quad <sXt> \longrightarrow a <sXt>, \quad <sXt> \longrightarrow b, \quad <tXt> \longrightarrow a.$$

Replace awkward notation of nonterminals by the usual upper case Latin letters:
$S \longrightarrow aVW, \quad V \longrightarrow aV \mid b, \quad W \longrightarrow a.$

Abandon Greibach Normal Forms and get a shorter CFG:
$S \longrightarrow aVa, \quad V \longrightarrow aV \mid b.$

A direct analisys of derivations allows us to come with yet shorter version:
$S \longrightarrow aS \mid aba.$