CSCI 341 Workshop 4

Grammars

October 1, 2025

1 Some Pumping and Grammar Warmup

Problem 1 (Balanced Parentheses)**.** A string of parentheses, i.e., $)$ and $($, is called *balanced* if every left-parenthese $($ is eventually followed by a unique *matching* right-parenthese $)$. For example, the following strings of parentheses are not balanced:

$$(, \qquad))(), \qquad ((())() \tag{*}$$

but the following strings of parentheses are:

$$\varepsilon, \qquad (), \qquad (())(), \qquad ((())())() \tag{**}$$

Let $A = \{(,)\}$ and

$$L = \{w \in A^* \mid w \text{ is balanced}\}$$

(1) Show that L is not regular.

(2) Design a context-free grammar $\mathcal{G}$ with a variable x that derives L.

(3) In your grammar, derive or draw parse trees for the words in (**)

(4) Explain why the words in (*) are not derivable.

Problem 2 (Palindromes)**.** Let $A = \{a, c, e, r\}$ and recall that for any word $w = a_1 a_2 \cdots a_n$, we define $w^{\mathrm{op}} = a_n a_{n-1} \cdots a_2 a_1$. Consider the language below:

$$L_{pal} = \{w \in A^* \mid w = w^{\mathrm{op}}\}$$

The words in L_{pal} are precisely the *palindromes*.

(1) Show that L_{pal} is not regular.

(2) Design a grammar with a variable that derives L_{pal}.

(3) Draw a parse tree for $racecar$.

2 Some Normal Form Problems

Definition 2.1 (Unit Production)**.** Let $\mathcal{G} = (X, A, R)$ be a grammar. A *unit production* is a rewrite rule of the form $x \to y$, where both $x, y \in X$.

Problem 3 (Killing εs with a dagger?)**.** Consider the grammar (taken from *Sipser*'s book) $\mathcal{G}$ below.

$$\begin{aligned} x &\to a \mid b \mid xa \mid xb \mid x0 \mid x1 \\ y &\to x \mid (u) \\ z &\to y \mid z * y \\ u &\to z \mid u + z \end{aligned}$$

Observe that the rewrite rules $y \to x$, $z \to y$, and $u \to z$ are all unit productions.

(1) Find a derivation and parse tree that yields $a * b0 + b * a$.

(2) Find a grammar $\mathcal{G}'$ with a variable x' such that $\mathcal{G}'$ **has no unit productions at all** and yet $\mathcal{L}(\mathcal{G}', x') = \mathcal{L}(\mathcal{G}, x)$. In other words, *eliminate the unit productions in* $\mathcal{G}$.

Definition 2.2 (Usefulness)**.** Let $\mathcal{G}$ be a grammar with variables x, y. We say that y is *reachable from* x if there is a sequence of reqrites $x \Rightarrow \mu_1 \Rightarrow \cdots \Rightarrow \mu_n$ such that y is a variable that appears in μ_n. We say that y is *useful for* x if y is reachable from x and $\mathcal{L}(\mathcal{G}, y)$ is not empty (there is at least one derivation possible starting from y).

Problem 4 (Cutting the fat)**.** Consider the grammar $\mathcal{G}$ below.

$$
\begin{aligned}
x &\to yz \mid ux \\
y &\to 0 \\
z &\to zu \mid xy \\
u &\to 0z \mid 1
\end{aligned}
$$

We are going to find a grammar without useless symbols with a state that is equivalent to x.

(1) Which variables are reachable from x?

(2) Does every variable derive a nonempty language?

(3) Which variables are useless for x?

(4) Find a grammar $\mathcal{G}'$ with a variable x' such that

  (a) $\mathcal{G}'$ has no useless symbols for x',
  (b) $\mathcal{G}'$ has no unit productions, and
  (c) $\mathcal{L}(\mathcal{G}', x') = \mathcal{L}(\mathcal{G}, x)$.

Definition 2.3 (Chomsky Normal Form)**.** Let $\mathcal{G} = (X, A, R)$ be a grammar with a variable $x \in X$. We say that x is *in Chomsky Normal Form* if

(1) Every variable in $\mathcal{G}$ is useful for x.

(2) If $y \in X$ has a rewrite rule $y \to \varepsilon$, then $y = x$ (although this rewrite rule is not required to exist at all).

(3) All other rewrite rules (i.e., not $x \to \varepsilon$) in $\mathcal{G}$ are of one of the following two forms:

   (a) $y \to zu$ where $y, z, u \in X$

   (b) $y \to a$ where $y \in X$ and $a \in A$

Problem 5 (Manufacturing Chomsky Normal Forms)**.** Consider the grammar $\mathcal{G}$ below:

$$\begin{aligned} x &\to yxz \mid \varepsilon \\ y &\to 0yx \mid 1 \\ z &\to x1x \mid y \mid 11 \end{aligned}$$

Find a grammar $\mathcal{G}'$ with a variable x' such that x' is in Chomsky Normal Form and $\mathcal{L}(\mathcal{G}', x') = \mathcal{L}(\mathcal{G}, x)$.

Definition 2.4 (Greibach Normal Form)**.** Let $\mathcal{G} = (X, A, R)$ be a grammar with a variable x. We say that x is in *Greibach normal form* if no more than the variable x has a rewrite rule $x \to \varepsilon$, and if every other rewrite rule is of the form

$$y \to azu$$

for some $a \in A$ and some $y, z, u \in X$.

It is not an easy theorem, but it is known that every context-free grammar can be turned into one in Greibach Normal Form! This has significant consequences, which we might talk about next week.

Problem 6 (Challenging!!)**.** Consider the grammar $\mathcal{G}$ below:

$$\begin{aligned} x &\to yxz \mid \varepsilon \\ y &\to 0yx \mid 1 \\ z &\to x1x \mid y \mid 11 \end{aligned}$$

Find a grammar $\mathcal{G}'$ with a variable x' such that x' is in Greibach Normal Form and $\mathcal{L}(\mathcal{G}', x') = \mathcal{L}(\mathcal{G}, x)$.