Section 12.1 Context-Free Languages

We know that the language $\{a^n b^n \mid n \in \mathbf{N}\}$ is not regular, but it certainly has a non-regular grammar, such as

$$S \rightarrow aSb \mid \Lambda.$$

A *context-free grammar* has productions of the form

$$N \rightarrow w$$

Where N is a nonterminal and w is any string containing terminals and/or nonterminals.

A *context-free language* is the set of strings derived from a context-free grammar.

Example. $\{a^n b^n \mid n \in \mathbf{N}\}$ is a C-F language derived from the C-F grammar $S \rightarrow aSb \mid \Lambda$.

Example. Any regular grammar is context-free. So regular languages are C-F languages.

Quiz. Find a grammar for $\{a^n b^{n+2} \mid n \in \mathbf{N}\}$.

Answer. $S \rightarrow aSb \mid bb$.

Quiz. Find a grammar for$\{ww^R \mid w \in \{a, b\}^*\}$, where w^R is the reverse of w.

Answer. $S \rightarrow aSa \mid bSb \mid \Lambda$.

Techniques for Constucting Grammars:

Let L and M be two C-F grammars with disjoint sets of nonterminals and with start symbols A and B, respectively. Then

- $L \cup M$ has grammar $S \rightarrow A \mid B$.
- LM has grammar $S \rightarrow AB$.
- L^* has grammar $S \rightarrow AS \mid \Lambda$.

Example. Let L be the language of strings over $\{a, b\}$ with the same number of a's and b's. Does L have the following grammar?

$$S \to aSbS \mid bSaS \mid \Lambda.$$

It's easy to see that the language of the grammar is a subset of L.
What about the other way?
Assume that $w \in L$ and show w is derived by the grammar.
If $w = \Lambda$, then $S \Rightarrow \Lambda$.
Let $w \neq \Lambda$ and assume that if $s \in L$ and $|s| < |w|$, then $S \Rightarrow^+ s$.
Show that $S \Rightarrow^+ w$. Consider the four cases:

1. $w = asb$ for some string s. In this case, $s \in L$ and $|s| < |w|$. So by induction we have $S \Rightarrow^+ s$. Therefore, we have $S \Rightarrow aSbS \Rightarrow aSb \Rightarrow^+ asb = w$.
2. $w = bsa$ for some string s. Similar to case 1.
3. $w = axa$ for some string x. In this case, x has two more b's than a's. So $x \notin L$.
   What do we do now?
   Notice, for example, if $|w| = 4$, then $w = a\boldsymbol{bb}a$. If $|w| = 6$, then w has one of the forms

   $$a\boldsymbol{abbb}a,\ a\boldsymbol{babb}a,\ a\boldsymbol{bbab}a\ a\boldsymbol{bbba}a.$$

   We claim that x can be written in the form $x = ubbv$ where $u, v \in L$. (Can you prove it?)
   So by induction we have derivations $S \Rightarrow^+ u$ and $S \Rightarrow^+ v$. Therefore, we have

   $$S \Rightarrow aSbS \Rightarrow^+ aubS \Rightarrow aubbSaS \Rightarrow^+ aubbvaS \Rightarrow aubbva = axa = w.$$
4. $w = bxb$ for some string x. Similar to case 3.
   QED.

Examples/Quizzes. For a string x and letter a let $n_a(x)$ be the number of a's in x. Let $L = \{x \in \{a, b\}^* \mid n_a(x) = n_b(x)\}$. A grammar for L with start symbol E can be written as:

$$E \to aEbE \mid bEaE \mid \Lambda.$$

Use this information to find grammars for the following languages.

1. $\{x \in \{a, b\}^* \mid n_a(x) = 1 + n_b(x)\}$.
   Solution: $S \to EaE$.
2. $\{x \in \{a, b\}^* \mid n_a(x) = 2 + n_b(x)\}$.
   Solution: $S \to EaEaE$.
3. $\{x \in \{a, b\}^* \mid n_a(x) > n_b(x)\}$.
   Solution: $S \to EaET$
   $T \to aET \mid \Lambda$.
4. $\{x \in \{a, b\}^* \mid n_a(x) < n_b(x)\}$.
   Solution: $S \to EbET$
   $T \to bET \mid \Lambda$.
5. $\{x \in \{a, b\}^* \mid n_a(x) \neq n_b(x)\}$.
   Solution: This language is the union of the languages in (3) and (4). Rename the nonterminals in the grammars for (3) and (4) as follows:

   (3) $A \to EaET$
   $T \to aET \mid \Lambda$.
   (4) $B \to EbEU$
   $U \to bEU \mid \Lambda$.

   Then $S \to A \mid B$ is the desired grammar.