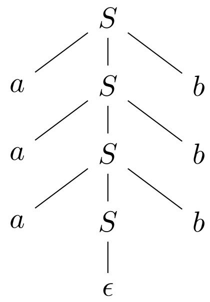
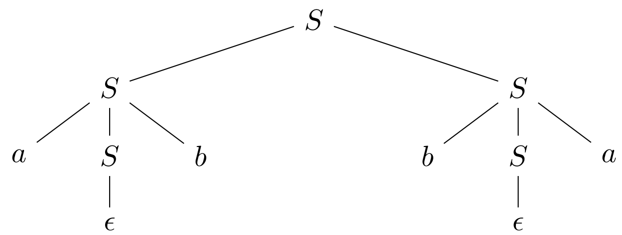
Sample solution to HW 3

(1) (i) a^3b^3 is in $L(\mathcal{G})$ with derivation tree:

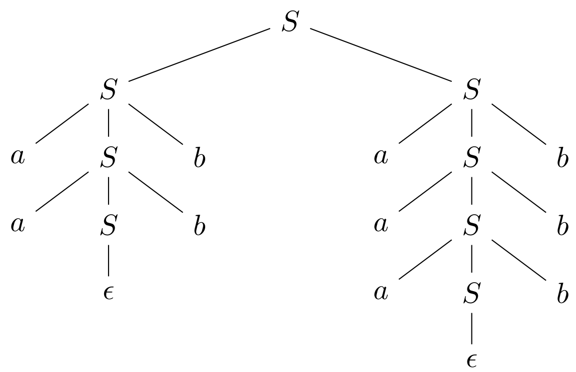


(ii) a^2b^3 is *not* in $L(\mathcal{G})$.
Note that each rule in $\mathcal{G}$ yields exactly one a and one b, thus, the number of a's and b's in a word generated by $\mathcal{G}$ must be the same.

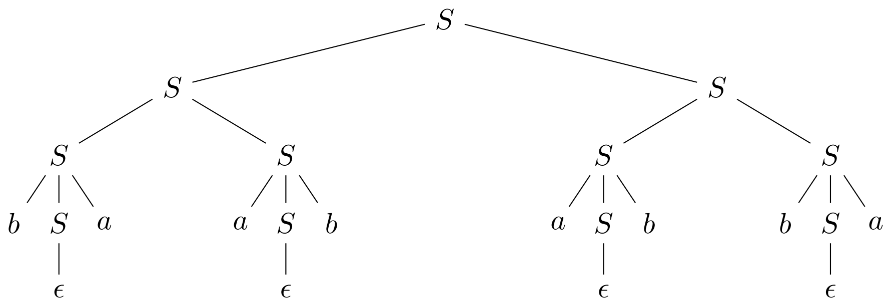
(iii) $abba$ is in $L(\mathcal{G})$ with derivation tree:



(iv) $a^2b^2a^3b^3$ is in $L(\mathcal{G})$ with derivation tree:



(v) $baababba$ is in $L(\mathcal{G})$ with derivation tree:



(2) (i) $L_1 = \{a^m b^n \mid m > n\}$ can be generated by the CFG $\mathcal{G}$ with the set of variables $V = \{S, T\}$, S is the start variable, and R contains the following rules:

$$\begin{array}{rcl} S & \to & aS \mid aT \\ T & \to & aTb \mid \epsilon \end{array}$$

(ii) $L_2 = \{a^m b^n \mid n > m\}$ can be generated by the CFG $\mathcal{G}$ with the set of variables $V = \{S, T\}$, S is the start variable, and R contains the following rules:

$$\begin{array}{rcl} S & \to & Sb \mid Tb \\ T & \to & aTb \mid \epsilon \end{array}$$

(iii) $L_3 = \{a^{2n} b^n \mid n \geqslant 0\}$ can be generated by the CFG $\mathcal{G}$ with the set of variables $V = \{S\}$, S is the start variable, and R contains the following rules:

$$S \to aaSb \mid \epsilon$$

(iv) $L_4 = \{w\$w^{\mathcal{R}} \mid w \in \{a,b\}^*\}$ can be generated by the CFG $\mathcal{G}$ with the set of variables $V = \{S\}$, S is the start variable, and R contains the following rules:

$$S \to aSa \mid bSb \mid \$$$

(v) L_5 is the complement of the language $\{a^n b^n \mid n \geqslant 0\}$ over the alphabet $\{a, b\}$. More formally, $L_5 = \Sigma^* - \{a^n b^n \mid n \geqslant 0\}$, where $\Sigma = \{a, b\}$.

A word $w \in \Sigma^*$ is *not* in $\{a^n b^n \mid n \geqslant 0\}$, if it satisfies one of the following conditions.

- In w some a appears after b, and such a word can be generated by the following rules:

$$\begin{array}{rcl} A & \to & ZbZaZ \\ Z & \to & aZ \mid bZ \mid \epsilon \end{array}$$

  Here the purpose of the variable Z is to generate arbitrary word.
- w is of the form: $a^m b^n$, where $m > n$, i.e., $w \in L_1$ defined in (i) above. Renaming the variables, we get the following rules to generate L_1:

$$\begin{array}{rcl} B & \to & aB \mid aC \\ C & \to & aCb \mid \epsilon \end{array}$$

- w is of the form: $a^m b^n$, where $m < n$, i.e., $w \in L_2$ defined in (ii) above. Renaming the variables, we get the following rules to generate L_2:

$$\begin{array}{rcl} D & \to & Db \mid Eb \\ E & \to & aEb \mid \epsilon \end{array}$$

We can combine the all the rules above to get the following grammar that generates the complement of $\{a^n b^n \mid n \geqslant 0\}$:

- $\Sigma = \{a, b\}$.
- $V = \{S, A, B, C, D, E, Z\}$.
- S is the start variable.

- R consists of all the rules above, as well as the rule:

$$
\begin{array}{rcl}
S & \to & A \mid B \mid D \\
A & \to & ZbZaZ \\
Z & \to & aZ \mid bZ \mid \epsilon \\
B & \to & aB \mid aC \\
C & \to & aCb \mid \epsilon \\
D & \to & Db \mid Eb \\
E & \to & aEb \mid \epsilon
\end{array}
$$

(3) This is quite straightforward. Simply follow the procedure described in Section 1 "From CFG to PDA" in Lesson 7.

(4) Show that the following languages are not CFL.

(i) $L_1 = \{a^k b^m c^n \mid k \leqslant m \leqslant n\}$ is not CFL.
The proof is via pumping lemma. Suppose to the contrary that L_1 is CFL. Let $\mathcal{G} = \langle \Sigma, V, R, S \rangle$ be its CFG.
Consider the word $w = a^k b^k c^k$, where $k \geqslant M^{|R|} + 1$ and M is the maximal length of the rule in R. By pumping lemma, there is a partition $w = sxyzt$ such that $|x| + |z| \geqslant 1$ and for each $i \geqslant 0$, $v\ sx^i yz^i t\ w \in L(\mathcal{G})$. There are a few cases.

(a) If either x or z consists of more than two symbols, then by pumping lemma, either some a's will appear after some b's or c's, or some c's will appear after some b's or a's. This violates the criteria to be in L_1.
(b) If x consists of a's and z consists of b's and both $x, z \neq \epsilon$, then sx^2yz^2t will contain more b's than c's. Again, this violates the criteria to be in L_1.
(c) If x consists of a's and z consists of c's and both $x, z \neq \epsilon$, then sx^2yz^2t will contain more a's than b's. Again, this violates the criteria to be in L_1.
(d) If x consists of b's and z consists of c's and both $x, z \neq \epsilon$, then $sx^0yz^0t = syt$ contains more a's than b's. Again, this violates the criteria to be in L_1.

The analysis is similar when one of x or z is ϵ. Therefore, we conclude that L_1 is not CFL.

(ii) $L_2 = \{a^m b^{2m} c^{3m} \mid m \geqslant 0\}$.
The proof is similar as above.

(iii) $L_3 = \{a^n \mid n \text{ is a prime number}\}$.
Again, the proof is via pumping lemma. Suppose to the contrary that L_3 is CFL. Let $\mathcal{G} = \langle \Sigma, V, R, S \rangle$ be its CFG.
Consider the word a^m, where $M^{|R|} + 1 \leqslant m \leqslant n$ and M is the maximal length of the rule in R. By pumping lemma, there is a partition $sxyzt$ such that $|x| + |z| \geqslant 1$ and for each $i \geqslant 0$, $sx^iyz^it \in L(\mathcal{G})$. Now, $|sx^iyz^it| = |s| + |y| + |t| + i(|x| + |z|)$.
If $|s| + |y| + |t| = 0$, the length $|v\ sx^iyz^it|$ is $|x^iz^i| = i(|x| + |z|)$, which is not a prime number. So, suppose $|s| + |y| + |t| \neq 0$, in which case, if we take $i = |s| + |y| + |t|$, the length of the word $v\ sx^iyz^it$ is $(|x| + |z| + 1)(|s| + |y| + |t|)$, which again is not a prime number. Thus, it contradicts the fact that $v\ sx^iyz^it\ w \in L(\mathcal{G})$, for each $i \geqslant 0$, and therefore, L_3 is not CFL.

(5) **(bonus point)** Consider the grammar defined in (1). Prove that $w \in L(\mathcal{G})$ if and only if w contains the same number of a's and b's.

Proof: For the "only if" direction, note that every rule in $\mathcal{G}$ generate the same number of a's and b's. Thus, every word generated by $\mathcal{G}$ have the same number of a's and b's.

Now, we prove the "if" direction. That is, we will show that if w contains the same number of a's and b's, then $w \in L(\mathcal{G})$. The proof is by induction on the length of w. The base case is when $w = \epsilon$, which is trivial.

For the induction hypothesis, we assume that it holds for every word of length $\leqslant m-1$. The induction step is as follows. Let w be a word of length m with the same number of a's and b's. There are a few cases.

- Case 1: $w = avb$, for some v. That is, w starts with a and ends with b.
  Thus, $|v| = m-2$. By induction hypothesis, $v \in L(\mathcal{G})$. That is, $S \Rightarrow^* v$. Now, due to the rule $S \to aSb$, we have $S \Rightarrow aSb$. Thus, $S \Rightarrow aSb \Rightarrow^* avb = w$. Therefore, $w \in L(\mathcal{G})$.
- Case 2: $w = bva$, for some v. That is, w starts with b and ends with a.
  This is similar to case 1. We have $S \Rightarrow bSa \Rightarrow^* bva = w$.
- Case 3: $w = ava$, for some v. That is, w starts with a and ends with a.
  Let $|w| = n$ and $w = d_1 \cdots d_n$, where each $d_i \in \{a, b\}$. Define the following function $f_w : \{1, \ldots, |w|\} \to \{1, \ldots, |w|\}$:

  $$f_w(i) \quad = \quad (\text{the number of } a\text{'s in } d_1 \cdots d_i) \quad - \quad (\text{the number of } b\text{'s in } d_1 \cdots d_i)$$

  Note that $|f_w(i+1) - f_w(i)| = 1$, that is, the value of $f_w(i+1)$ can only increase/decrease by 1 from $f_w(i)$ in the following sense.
  - If $f_w(i+1) = f_w(i) + 1$, then $d_{i+1} = a$.
  - If $f_w(i+1) = f_w(i) - 1$, then $d_{i+1} = b$.

  We have $f_w(1) = 1$, because w starts with a. Moreover, $f_w(n) = 0$, because w contains the same number of a's and b's. Hence, $f_w(n-1) = -1$, because w ends with a.
  Since $f_w(1) > 0$ and $f_w(n-1) < 0$, there is j such that

  $$f_w(j) = 0,$$

  thus, the number of a's and b's in $d_1 \cdots d_j$ is the same. Likewise, the number of the number of a's and b's in $d_{j+1} \cdots d_n$ is also the same. By induction hypothesis, $S \Rightarrow^* d_1 \cdots d_j$ and $S \Rightarrow^* d_{j+1} \cdots d_n$. Therefore, we have $S \Rightarrow SS \Rightarrow^* d_1 \cdots d_j S \Rightarrow^* d_1 \cdots d_j d_{j+1} \cdots d_n = w$. This proves that $w \in L(\mathcal{G})$.
- Case 4: $w = bvb$, for some v. That is, w starts with b and ends with b.
  This case is similar to Case 3.