Parsing Beyond CFG
Homework 1: CFG

Laura Kallmeyer, Tatiana Bladier

Summer 2018

Question 1

Give a CFG G_1 for the following language L_1:

$$L_1 = \{w \in \{a, b\}^* \mid w = a^n b^m a^n \text{ where } n, m \in N, n \geq 2, m \geq 1\}$$

Hint:

- *What is the minimal string which belongs to this language?*
- *Provide some example strings belonging to language L_1*
- *If G_1 is a tuple $\langle N, T, P, S\rangle$, provide the sets of $N, T, P,$ and S*
- *Provide some example productions of your resulting CFG*

Solution:

- The minimal string is: *aabaa*
- Possible strings: *aaabbaaa, aaaaaabbbbbaaaaaa*, ...
- $G_1 = \langle N, T, P, S\rangle$, where $N = \{S, B, C\}, T = \{a, b\}$,
  $P = \{S \rightarrow aBa, B \rightarrow aBa, B \rightarrow aCa, C \rightarrow bC, C \rightarrow b\}$
- Example derivations:
  $S \rightarrow aBa \rightarrow aaCaa \rightarrow aabaa$
  $S \rightarrow aBa \rightarrow aaaBaaa \rightarrow aaaaCaaaa \rightarrow aaaabCaaaa \rightarrow aaaabbaaaa$

Question 2

Consider the following CFG:

$$G = \langle \{S, A, B\}, \{a, b, c\}, \{S \rightarrow AB, A \rightarrow aAb \mid \epsilon, B \rightarrow cB \mid \epsilon\}, S\rangle$$

Which language L_2 does this CFG generate?

Solution:

$$L_2 = \{a^m b^m c^n | m, n \geq 0\}$$

Question 3

1. Show that the following language is not context-free:

   $$L_{a,b,c,d} = \{a^n b^m c^n d^m \mid n, m \geq 0\}$$

   Hint: Assume that $L_{a,b,c,d}$ is context-free. Then it must satisfy the pumping lemma. Show that this is not the case. Consequently, the initial assumption does not hold.

2. Show that the following language is not context-free:

   $$L = \{e^k (ag)^n e b^m f c^n f d^m \mid k, n, m \geq 0\}$$

   Hint: You can make use of the fact that $L_{a,b,c,d}$ (from 1.) is not context-free. And you can use the fact that CFLs are closed under homomorphisms.

Solution:

1. We assume that $L_{a,b,c,d}$ is context-free. Then it satisfies the pumping lemma with a certain constant $k \geq 1$.

   We consider the word $a^k b^k c^k d^k$. According to the pumping lemma this word must have a form xv_1yv_2z with $|v_1yv_2| \leq k$ such that v_1 and v_2 can be iterated. Because of $|v_1yv_2| \leq k$, v_1v_2 cannot contain all three terminals. With this and with $v_1v_2 \neq \varepsilon$, we have that after the first pumping, we obtain a word $xv_1^2yv_2^2z$ for which does not hold that the number of as is equal cs and that the number of bs is equal ds and that is therefore not in $L_{a,b,c,d}$. Contradiction to the pumping lemma. Consequently, $L_{a,b,c,d}$ does not satisfy the pumping lemma and is therefore not a CFL.

2. We assume that L is context-free. Then its image under the homomorphism ϕ with $\phi(e) = \phi(g) = \phi(f) = \varepsilon$, $\phi(a) = a$, $\phi(b) = b$, $\phi(c) = c$ and $\phi(d) = d$ must also be a CFL. But we have $\phi(L) = L_{a,b,c,d}$ which is not context-free. Contradiction, therefore L is not context-free either.