# Languages and Computation (COMP 2049) Lab 06

## Proving Non-Regularity

(1) Consider the alphabet $\Sigma = \{a, b\}$ and the following language over $\Sigma$:

$$L_1 := \{a^n b^k \mid n, k \in \mathbb{N} \wedge n > k\}.$$

(a) Using the pumping lemma for regular languages, prove that $L_1$ is not regular.

(b) Demonstrate that $L_1$ is context-free by presenting the productions of a grammar $G$ for which $L_1 = L(G)$.

Solution

(a) Assume that $L_1$ is regular. Then, it must satisfy the pumping property. Let $n$ be the integer in the pumping lemma, and consider the string $x = a^{n+1}b^n \in L_1$. As $|x| = 2n + 1 \geq n$, there are strings $uvw$ such that $x = uvw$, and:

(1) $|uv| \leq n$

(2) $|v| > 0$

(3) $\forall i \in \mathbb{N} : uv^iw \in L_1$

From the first two properties, we infer that $v$ is of the form $a^k$, for some $k \geq 1$, which implies that $n + 1 - k \leq n$. If we take $i = 0$ in (3), we must have $a^{n+1-k}b^n \in L_1$, which is a contradiction. Hence, $L_1$ is not regular.

(b) The following grammar generates $L_1$:

$$\begin{aligned} S &\rightarrow aT \\ T &\rightarrow aT \mid aTb \mid \lambda \end{aligned}$$

(2) Using the pumping lemma for regular languages, prove that the following language is not regular:

$$L_2 := \{xx \mid x \in \{a, b\}^*\}.$$

Remark: The language $L_2$ is sometimes refered to as the copy language, denoted as $XX$ or $WW$. Not only is this language not regular, it is not even context-free.

Solution

Assume that $L_2$ is regular. Then, it must satisfy the pumping property. Let $n$ be the integer in the pumping lemma, and consider the string $x = a^nb^na^nb^n \in L_2$. As $|x| = 4n \geq n$, there are strings $uvw$ such that $x = uvw$, and:

(1) $|uv| \leq n$

(2) $|v| > 0$

(3) $\forall i \in \mathbb{N} : uv^iw \in L_2$

From the first two properties, we infer that $v$ is of the form $a^k$, for some $k \geq 1$. If we take $i = 0$ in (3), we must have $a^{n-k}b^na^nb^n \in L_2$, which is a contradiction. Hence, $L_2$ is not regular.

(3) Consider the alphabet $\Sigma = \{a, b, c\}$ and the following language over $\Sigma$:

$$L_3 := \{a^j b^k c^k \mid j \geq 1 \wedge k \geq 0\} \cup \{b^j c^k \mid j \geq 0 \wedge k \geq 0\}.$$

In simple terms, for every string $x \in L_3$

- If there is at least one $a$ in the string $x$, then the number of $b$'s and $c$'s must be the same.
- If there is no $a$ in the string $x$, then the number of $b$'s and $c$'s can be anything.

(a) Prove that the language $L_3$ satisfies the conclusions of the pumping lemma for regular languages.

(b) Using the pigeonhole principle, prove that the language $L_3$ is not regular.

Remark: This exercise demonstrates the following points:

- Although every regular language must have the pumping property, the converse is not true, i. e., there are languages that do have the pumping property, but are not regular.
- The pigeonhole principle is stronger than the pumping lemma.

Solution

(a) The languages $L_3$ indeed satisfies the pumping property. Take $n = 1$, and for every string $x \in L_3$ with $|x| \geq 1$, partition $x$ as $x = uvw$ in which:

(a) $u = \lambda$

(b) $|v| = 1$, i. e., $v$ consists of only the first symbol in $x$;

(c) $w$ is the tail of the string $x$.

There are three cases to consider:

- If the string $x$ begins with an '$a$', then it must be of the form $x = a^j b^k c^k$, with $j \geq 1$ and $k \geq 0$. From (a) and (b) we infer that $v = a$. Thus, for any $i \geq 0$, the string $x' = uv^i w$ is of the form $x' = a^{j-1+i} b^k c^k$ which is in $L_3$.
- If the string $x$ begins with a '$b$', then it must be of the form $x = b^j c^k$, with $j \geq 1$ and $k \geq 0$. From (a) and (b) we infer that $v = b$. Thus, for any $i \geq 0$, the string $x' = uv^i w$ is of the form $x' = b^{j-1+i} c^k$ which is in $L_3$.
- If the string $x$ begins with a '$c$', then it must be of the form $x = c^k$, with $k \geq 1$. From (a) and (b) we infer that $v = c$. Thus, for any $i \geq 0$, the string $x' = uv^i w$ is of the form $x' = c^{k-1+i}$ which is in $L_3$.

(b) As the language $L_3$ satisfies the pumping property, we cannot use the pumping lemma to prove that it is not regular. Instead, we directly use the pigeonhole principle.

Assume that $L_3$ is regular. By definition, there must be a deterministic finite automaton (DFA) $M = (Q, \Sigma, \delta, q_0, F)$ such that $L_3 = L(M)$. Let us assume that $Q = \{q_0, q_1, \ldots, q_{n-1}\}$, i. e., the automaton $M$ has $n$ states. Consider the following $n + 1$ strings in $L_3$:

$$\begin{cases} x_1 = ab \\ x_2 = ab^2 \\ \quad\vdots \\ x_{n+1} = ab^{n+1} \end{cases}$$

For each $i \in \{1, \ldots, n + 1\}$, we must have $\delta^*(q_0, x_i) \in Q$. As there are only $n$ states in $Q$, by the pigeonhole principle, for some $j \neq k$ we must have:

$$\delta^*(q_0, x_j) = \delta^*(q_o, x_k). \tag{1}$$

We know that $ab^jc^j \in L_3$. Thus, $\delta^*(\delta^*(q_0, x_j), c^j) \in F$. By (1), this implies that $\delta^*(\delta^*(q_0, x_k), c^j) \in F$, i. e., $ab^kc^j \in L_3$, which is a contradiction as $j \neq k$.
Therefore, $L_3$ cannot be regular.